

# Molecular glue degraders of cyclin K

## **Inauguraldissertation**

zur

Erlangung der Würde eines Doktors der Philosophie  
vorgelegt der Philosophisch-Naturwissenschaftlichen Fakultät  
der Universität Basel

von

**Zuzanna Kozicka**

Basel, 2023

Originaldokument gespeichert auf dem Dokumentenserver der Universität Basel  
[edoc.unibas.ch](http://edoc.unibas.ch)

Genehmigt von der Philosophisch-Naturwissenschaftlichen Fakultät

auf Antrag von

Dr Nicolas Thomä

Prof. Dirk Schübeler

Prof. Angela Koehler

Basel, den 15. November 2022

Prof. Marcel Mayor

The Dean of Faculty

## Abstract

Molecular glue degraders are small, drug-like compounds that induce interactions between an E3 ubiquitin ligase and a target, leading to target ubiquitination and subsequent degradation by the proteasome. Unlike traditional enzyme inhibitors, degraders can act sub-stoichiometrically to catalyse the rapid depletion of previously inaccessible targets. Despite the clinical success of this modality, only two molecular glue degrader classes, thalidomide analogues and aryl sulphonamides, have been well-described thus far and molecular glue discovery remains largely serendipitous. Hence, novel molecular glue degrader examples, as well as methods for their prospective discovery and rational design are highly sought-after.

In this work, a systematic search for molecular glue degraders was performed through bioinformatic mining of drug cytotoxicity data for correlations with E3 ligase component expression levels across hundreds of human cancer cell lines. This led to the identification of CR8, a preclinical cyclin-dependent kinase (CDK) inhibitor, as a molecular glue degrader of cyclin K. CR8 was found to facilitate the formation of a complex between CDK12–cyclin K and the CUL4 adaptor protein DDB1, thereby bypassing a canonical substrate receptor and directly presenting cyclin K for ubiquitination and degradation. A solvent-exposed pyridyl moiety of the compound, one that is absent in related inhibitors, protrudes from the kinase pocket and engages DDB1, which suggests more broadly that chemical modification of surface-exposed moieties can confer gain-of-function glue properties to an inhibitor. Notably, several structurally distinct cyclin K degraders were reported shortly after, including the dCeMM2-4 compounds, whose discovery through differential compound toxicity profiling in hyponeddylated and wild-type cellular models is described herein.

To examine this unexpected chemical diversity, the cyclin K degrader structure-activity relationship (SAR) was further investigated through the evaluation of over 90 putative degraders, including novel scaffolds and published inhibitors. The crystallographic dissection of 29 ternary complexes revealed that cyclin K degraders share a common mechanism, whereby kinase-binding scaffolds acquire gain-of-function properties through the engagement of interfacial DDB1 residues. This work demonstrated that degrader SAR differs considerably from that of inhibitors, offering conceptual learnings about compound-induced interfaces applicable to other molecular glue degrader classes. The cellular evaluation of cyclin K degrader activity identified compounds more potent than CR8, and established that cyclin K degradation versus CDK12 inhibition give rise to distinct transcriptional signatures. These findings therefore not only define a novel, diverse class of molecular glue degraders but also yield next-generation, interface-leveraging scaffolds for the selective inactivation of the emerging therapeutic targets CDK12 and cyclin K.

# Table of Contents

<b>ABSTRACT .....</b>	<b>3</b>
<b>TABLE OF CONTENTS .....</b>	<b>4</b>
<b>DISSERTATION OUTLINE.....</b>	<b>10</b>
<b>CHAPTER 1 : INTRODUCTION .....</b>	<b>12</b>
<b>1.1 TARGETED PROTEIN DEGRADATION .....</b>	<b>12</b>
1.1.1 THE UBIQUITIN PROTEASOME SYSTEM.....	12
1.1.2 CULLIN-RING LIGASES .....	14
1.1.3 CRL4 CLADE.....	15
1.1.4 THERAPEUTIC TARGETING OF THE UBIQUITIN-PROTEASOME SYSTEM.....	18
1.1.5 HIJACKING OF THE UBIQUITIN-PROTEASOME SYSTEM.....	20
1.1.6 CHEMICAL INDUCERS OF PROXIMITY.....	20
1.1.7 PLANT HORMONES AS INDUCERS OF PROTEIN DEGRADATION.....	22
1.1.8 INDUCIBLE DEGRADATION AND BIFUNCTIONAL DEGRADERS .....	23
1.1.9 MOLECULAR GLUE DEGRADERS .....	25
1.1.10 MOLECULAR GLUES FOR CRL4 <sup>CRBN</sup> .....	26
1.1.11 MOLECULAR GLUES FOR CRL4 <sup>DCAF15</sup> .....	32
1.1.12 BCL6 DEGRADERS .....	33
1.1.13 OTHER MOLECULAR GLUE DEGRADERS .....	34
<b>1.2 CYCLIN-DEPENDENT KINASES .....</b>	<b>35</b>
1.2.1 THE CDK FAMILY .....	35
1.2.2 THE ROLES OF CDK12/13-CYCLIN K.....	38
1.2.3 CDK12-CYCLIN K AS A BIOMARKER AND THERAPEUTIC TARGET .....	39
1.2.4 STRUCTURAL INSIGHTS INTO CDK12/13-CYCLIN K.....	41
1.2.5 THERAPEUTIC TARGETING OF CDK12/13-CYCLIN K.....	44
1.2.6 COVALENT INHIBITORS .....	46
1.2.7 ATP-COMPETITIVE INHIBITORS.....	47
1.2.8 PROTACs .....	50
1.2.9 SYNTHETIC LETHALITY AND COMBINATION TREATMENTS .....	51
<b>1.3 AIMS OF THE THESIS .....</b>	<b>54</b>
<b>CHAPTER 2 : DISCOVERY OF CYCLIN K DEGRADERS.....</b>	<b>55</b>
<b>2.1 THE CDK INHIBITOR CR8 ACTS AS A MOLECULAR GLUE DEGRADER THAT DEPLETES CYCLIN K</b> <b>(PUBLISHED MANUSCRIPT) .....</b>	<b>56</b>

2.1.1	ABSTRACT .....	56
2.1.2	INTRODUCTION .....	56
2.1.2	RESULTS .....	57
2.1.2.1	CR8 INDUCES PROTEASOMAL DEGRADATION OF CYCLIN K.....	57
2.1.2.2	CR8 DIRECTS CDK12 TO CRL4 CORE COMPONENT.....	64
2.1.2.3	CDK12-DDB1 INTERFACE IMPARTS SELECTIVITY.....	71
2.1.2.4	CR8 PHENYLPYRIDINE CONFERS GLUE ACTIVITY .....	73
2.1.2.5	CYCLIN K DEGRADATION ADDS TO CR8 TOXICITY.....	76
2.1.3	DISCUSSION .....	78
<b>2.2</b>	<b>RATIONAL DISCOVERY OF CYCLIN K MOLECULAR GLUE DEGRADERS VIA SCALABLE CHEMICAL PROFILING (SUMMARY OF A PUBLISHED MANUSCRIPT) .....</b>	<b>78</b>
2.2.1	COMPOUND TOXICITY PROFILING IN HYPONEDDYLATED VERSUS WILD-TYPE CELLS IDENTIFIES NOVEL DEGRADER COMPOUNDS .....	79
2.2.2	dCEMM2-4 ARE CYCLIN K DEGRADERS.....	81
2.2.3	SIGNIFICANCE .....	87

## **CHAPTER 3 : MOLECULAR GLUE DEGRADER STRUCTURE-ACTIVITY RELATIONSHIP AND DESIGN PRINCIPLES: A CASE STUDY FOR CYCLIN K DEGRADERS..... 88**

<b>3.1</b>	<b>ABSTRACT .....</b>	<b>88</b>
<b>3.2</b>	<b>INTRODUCTION .....</b>	<b>88</b>
<b>3.3</b>	<b>RESULTS .....</b>	<b>89</b>
3.3.1	CR8 RETAINS GLUE ACTIVITY DESPITE DIVERSE MODIFICATIONS .....	89
3.3.2	PI-CATION INTERACTIONS ARE IMPORTANT FOR ROBUST COMPLEX FORMATION.....	94
3.3.3	HYDROGEN BONDING INTERACTIONS OF THE GLUING MOIETY MORE WEAKLY MODULATE AFFINITY .....	95
3.3.4	DIVERSE R <sub>2</sub> MODIFICATIONS ARE TOLERATED FOR DDB1-CDK12 COMPLEX FORMATION.....	98
3.3.5	CDK12-SELECTIVE INHIBITORS CAN BE CYCLIN K DEGRADERS .....	100
3.3.6	HYDROPHILIC MOIETIES CAN ENGAGE DDB1 TO DEGRADE CYCLIN K .....	105
3.3.7	SPECIFIC FINGERPRINT DEFINES SMALL, FRAGMENT-LIKE CYCLIN K DEGRADERS.....	107
3.3.8	CONFORMATIONAL CHANGE TRIGGERED BY dCEMM4 SHOWS PPI PLASTICITY.....	110
3.3.9	<i>IN VITRO</i> VERSUS CELLULAR ACTIVITY .....	110
3.3.10	CYCLIN K DEGRADER DIVERSITY .....	111
3.3.11	KINASE INHIBITORY SELECTIVITY CAN BE DECOUPLED FROM GLUE ACTIVITY .....	116
3.3.12	CYCLIN K DEGRADERS AND CDK12 INHIBITORS SHOW DISTINCT TRANSCRIPTIONAL SIGNATURES .....	118
<b>3.4</b>	<b>DISCUSSION .....</b>	<b>120</b>

## **CHAPTER 4 : FLUORESCENCE-BASED BINDING STUDIES..... 122**

<b>4.1</b>	<b><i>IN VITRO</i> TR-FRET OPTIMISATION FOR MOLECULAR GLUE DEGRADER SCREENING AND EVALUATION</b>	<b>123</b>
4.1.1	THE PRINCIPLES OF TR-FRET.....	123
4.1.2	ORIGINAL ASSAY SETUP .....	125

4.1.3	TAG POSITION .....	127
4.1.4	TITRATION MODE .....	130
4.1.5	COMPOUND DISPENSING METHODS .....	132
4.1.6	FRET PAIR.....	134
4.1.7	PLATE TYPE.....	138
4.1.8	BUFFER COMPONENTS .....	138
4.1.9	ASSAY QUALITY CONTROL AND MINIATURISATION .....	141
<b>4.2</b>	<b>FLUORESCENCE-BASED METHODS TO STUDY PROTEIN-DNA INTERACTIONS .....</b>	<b>143</b>
4.2.1	FLUORESCENCE POLARIZATION .....	143
4.2.2	MECHANISMS OF OCT4-SOX2 MOTIF READOUT ON NUCLEOSOMES .....	144
4.2.2.1	INTRODUCTION.....	144
4.2.2.2	SeEN-SEQ REVEALS THE END-BINDING PREFERENCE OF OCT4-SOX2.....	145
4.2.2.3	STRUCTURAL STUDIES OF OCT4-SOX2 ON THE NUCLEOSOME .....	146
4.2.2.4	FLUORESCENCE POLARISATION MEASUREMENTS TO STUDY TF-DNA AND TF-NUCLEOSOME INTERACTIONS .....	149
4.2.2.5	THE OCT4-SOX2 PARTIAL MOTIF IS SUFFICIENT FOR CHROMATIN OPENING IN CELLS .....	153
4.2.3	STRUCTURAL MECHANISM OF cGAS INHIBITION BY THE NUCLEOSOME .....	154
4.2.3.1	INTRODUCTION.....	154
4.2.3.2	CRYO-EM STUDIES OF cGAS-NCP COMPLEXES .....	154
4.2.3.3	FLUORESCENCE POLARISATION STUDIES TO DISSECT THE NCP LINKER CONTRIBUTION TO cGAS BINDING .....	158
4.2.3.4	FLUORESCENCE POLARISATION STUDIES TO ELUCIDATE THE ROLE OF A ZINC FINGER MOTIF .....	159
4.2.3.5	DISRUPTION OF THE cGAS-NCP INTERACTION TRIGGERS IMMUNE ACTIVATION IN CELLS .....	160

## **CHAPTER 5 : DISCUSSION AND PERSPECTIVES .....** **163**

<b>5.1</b>	<b>CYCLIN K DEGRADERS .....</b>	<b>163</b>
<b>5.2</b>	<b>THERAPEUTIC UTILITY OF CYCLIN K DEGRADERS.....</b>	<b>165</b>
<b>5.3</b>	<b>GLUE-INDUCED INTERFACES.....</b>	<b>166</b>
<b>5.4</b>	<b>DEGRADER THEMES.....</b>	<b>167</b>
<b>5.5</b>	<b>SCREENING FOR MOLECULAR GLUE DEGRADERS .....</b>	<b>171</b>
<b>5.6</b>	<b>RATIONAL DESIGN FOR GLUING WEAK INTERACTORS .....</b>	<b>171</b>
<b>5.7</b>	<b>THE “PROTEIN PAIR PROBLEM” .....</b>	<b>174</b>
<b>5.8</b>	<b>COMPOUND-INDUCED STICKINESS.....</b>	<b>175</b>
<b>5.9</b>	<b>COMPOUND-INDUCED INTERFACES .....</b>	<b>177</b>
<b>5.10</b>	<b>ENDOGENOUS MOLECULAR GLUES.....</b>	<b>178</b>
<b>5.11</b>	<b>SIGNIFICANCE .....</b>	<b>180</b>

## **CHAPTER 6 : MATERIALS AND METHODS.....** **182**

<b>6.1</b>	<b>GENERAL MATERIALS AND METHODS .....</b>	<b>182</b>
------------	--	------------

6.1.1	MATERIALS .....	182
6.1.2	INSECT CELL CULTURE .....	183
6.1.3	PROTEIN EXPRESSION AND PURIFICATION .....	184
6.1.4	PROTEIN ANALYSIS .....	185
6.1.4.1	SDS POLYACRYLAMIDE-GEL ELECTROPHORESIS .....	185
6.1.4.2	DETERMINATION OF PROTEIN CONCENTRATION .....	185
<b>6.2</b>	<b>DISCOVERY OF CYCLIN K DEGRADERS (CHAPTER 2).....</b>	<b>185</b>
6.2.1	MAMMALIAN CELL CULTURE.....	185
6.2.2	COMPOUNDS .....	186
6.2.3	ANTIBODIES .....	186
6.2.4	REPORTER VECTORS .....	186
6.2.5	OLIGONUCLEOTIDES .....	186
6.2.6	BIOINFORMATIC SCREEN .....	186
6.2.7	CLONING AND LENTIVIRAL PACKAGING OF SGRNAs TARGETING 95 E3 LIGASES .....	187
6.2.8	VALIDATION OF DRUG-E3 LIGASE PAIRS FROM THE BIOINFORMATIC SCREEN .....	187
6.2.9	VALIDATION OF THE DDB1 RESISTANCE PHENOTYPE .....	188
6.2.10	WHOLE PROTEOME QUANTIFICATION USING TANDEM MASS TAG MASS SPECTROMETRY .....	188
6.2.11	QUANTITATIVE PCR .....	188
6.2.12	IMMUNOBLOTS FOR WHOLE PROTEIN LYSATE .....	188
6.2.13	CYCLIN K STABILITY REPORTER ANALYSIS.....	189
6.2.14	GENOME-WIDE CRISPR – CR8 RESISTANCE SCREEN .....	189
6.2.15	BISON CRISPR – CR8 RESISTANCE SCREEN .....	189
6.2.16	GENOME-WIDE CRISPR – CYCLIN K STABILITY REPORTER SCREEN .....	190
6.2.17	POOLED CRISPR SCREEN DATA ANALYSIS.....	190
6.2.18	DCAF ARRAYED SCREEN .....	190
6.2.19	PROTEIN PURIFICATION .....	190
6.2.20	CO-IMMUNOPRECIPITATION ASSAY.....	191
6.2.21	CRYSTALLISATION AND DATA COLLECTION.....	191
6.2.22	STRUCTURE DETERMINATION AND MODEL BUILDING.....	192
6.2.23	BIOTINYLATION OF DDB1 .....	192
6.2.24	TIME-RESOLVED FLUORESCENCE RESONANCE ENERGY TRANSFER (TR-FRET).....	192
6.2.25	DDB1-CUL4-RBX1 RECONSTITUTION AND <i>IN VITRO</i> CUL4 NEDDYLATION .....	193
6.2.26	IN-VITRO UBIQUITINATION ASSAYS.....	193
6.2.27	ISOTHERMAL TITRATION CALORIMETRY (ITC) .....	193
6.2.28	BIOLUMINESCENCE RESONANCE ENERGY TRANSFER (BRET) ANALYSES .....	194
6.2.29	DRUG SENSITIVITY ASSAYS .....	194
6.2.30	CYCLIN K STABILITY REPORTER ANALYSIS WITH CRBN OVEREXPRESSION .....	194
<b>6.3</b>	<b>CYCLIN K DEGRADER SAR (CHAPTER 3) .....</b>	<b>194</b>
6.3.1	COMPOUNDS .....	194
6.3.2	PROTEIN EXPRESSION AND PURIFICATION .....	195
6.3.2.1	EXPRESSION AND PURIFICATION OF DDB1 CONSTRUCTS .....	195
6.3.2.2	EXPRESSION AND PURIFICATION OF CDK12-CYCLIN K FOR BIOPHYSICAL ASSAYS.....	196
6.3.2.3	EXPRESSION AND PURIFICATION OF CDK12-CYCLIN K FOR CRYSTALLOGRAPHY.....	196
6.3.3	PROTEIN LABELLING .....	197

6.3.3.1 LABELLING WITH FLUOROPHORE-CONJUGATED MALEIMIDES .....	197
6.3.3.2 BIOTINYLATION .....	197
6.3.4 TR-FRET ASSAY .....	198
6.3.5 DDB1-CUL4-RBX1 RECONSTITUTION AND <i>IN VITRO</i> CUL4 NEDDYLATION .....	198
6.3.6 CRYSTALLOGRAPHIC METHODS .....	199
6.3.6.1 CRYSTALLISATION AND DATA COLLECTION .....	199
6.3.6.2 STRUCTURE DETERMINATION AND MODEL BUILDING .....	202
6.3.7 LIGAND DOCKING .....	203
6.3.8 LANTHASCREEN KINASE BINDING ASSAY .....	203
6.3.9 MAMMALIAN CELL CULTURE.....	204
6.3.10 REPORTER VECTORS .....	204
6.3.11 CYCLIN K REPORTER STABILITY ANALYSIS .....	204
6.3.12 DRUG SENSITIVITY ASSAYS .....	204
6.3.13 LABEL-FREE QUANTITATIVE MASS SPECTROMETRY .....	205
6.3.13.1 SAMPLE PREPARATION .....	205
6.3.13.2 LC-MS DATA ANALYSIS .....	206
6.3.14 RNA SEQUENCING .....	206
6.3.14.1 LIBRARY PREPARATION.....	206
6.3.14.2 RNA-SEQ ANALYSIS .....	206
<b>6.4 FLUORESCENCE-BASED METHODS (CHAPTER 4) .....</b>	<b>207</b>
6.4.1 TR-FRET .....	207
6.4.1.1 COMPOUND TITRATIONS.....	207
6.4.1.2 FORWARD TITRATIONS .....	207
6.4.1.3 COUNTER TITRATIONS.....	208
6.4.2 INTERACTIONS OF OCT4-SOX2 WITH DNA AND NUCLEOSOMES .....	208
6.4.2.1 HUMAN OCTAMER HISTONES EXPRESSION, PURIFICATION, AND RECONSTITUTION .....	208
6.4.2.2 DNA PREPARATION .....	208
6.4.2.3 NUCLEOSOME ASSEMBLY .....	208
6.4.2.4 PROTEIN EXPRESSION AND PURIFICATION OF OCT4 AND SOX2 .....	209
6.4.2.5 DETERMINATION OF OCT4 AND/OR SOX2 BINDING AFFINITY FOR DNA BY FP FORWARD TITRATIONS.....	209
6.4.2.6 DETERMINATION OF OCT4 AND/OR SOX2 BINDING AFFINITY FOR NUCLEOSOMES BY FP COUNTER TITRATIONS .....	209
6.4.3 INTERACTIONS OF cGAS WITH DNA AND NUCLEOSOMES .....	210
6.4.3.1 DETERMINATION OF cGAS BINDING AFFINITY FOR DNA BY FLUORESCENCE POLARISATION FORWARD TITRATIONS .....	210
6.4.3.2 DETERMINATION OF cGAS BINDING AFFINITY FOR NUCLEOSOMES BY FLUORESCENCE POLARISATION COUNTER-TITRATIONS .....	210
<b><u>REFERENCES .....</u></b>	<b><u>211</u></b>
<b><u>ABBREVIATIONS.....</u></b>	<b><u>237</u></b>
<b><u>ACKNOWLEDGEMENTS .....</u></b>	<b><u>240</u></b>
<b><u>SUPPLEMENTARY INFORMATION .....</u></b>	<b><u>241</u></b>





## Dissertation outline

This doctoral thesis consists of six chapters. Collaborations and co-authorships are indicated at the beginning of each chapter. The thesis is based on both published and unpublished work performed at the Friedrich Miescher Institute for Biomedical Research, Basel, Switzerland, under the supervision of Dr Nicolas H. Thomä.

**Chapter 1** is an introduction to the ubiquitin-proteasome system and targeted protein degradation, with a focus on molecular glue degraders. Several sections and figures have been adapted or reproduced from the following manuscript:

*Haven't got a glue: Protein surface variation for the design of molecular glue degraders.*

**Z. Kozička**, N. Thomä

*Cell Chemical Biology* (2021), <https://doi.org/10.1016/j.chembiol.2021.04.009>

Later sections of this chapter introduce cyclin-dependent kinases and discuss CDK12 and cyclin K as emerging therapeutic targets.

**Chapter 2**, **Chapter 3**, and **Chapter 4** are experimental chapters.

⇒ **Chapter 2** describes the discovery of cyclin K degraders. It contains published data reproduced or adapted from the following manuscripts:

*The CDK inhibitor CR8 acts as a molecular glue degrader that depletes cyclin K*

M. Ślabicki\*, **Z. Kozička**\*, G. Petzold\*, Y.-D. Li, M. Manojkumar, R. Bunker, K. Donovan, Q. Sievers, J. Koepfel, D. Suchyta, A. Sperling, E. Fink, J. Gasser, L. Wang, S. Corsello, R. Sellar, M. Jan, D. Gillingham, C. Scholl, S. Fröhling, T. Golub, E. Fischer, N. Thomä<sup>†</sup>, B. Ebert<sup>†</sup>.

*Nature* (2020), <https://doi.org/10.1038/s41586-020-2374-x> (\*equal contribution)

*Rational discovery of molecular glue degraders via scalable chemical profiling*

C. Mayor-Ruiz, S. Bauer, M. Brand, **Z. Kozička**, M. Siklos, H. Imrichova, I. Kaltheuner, E. Hahn, K. Seiler, A. Koren, G. Petzold, M. Fellner, C. Bock, A. Muller, J. Zuber, M. Geyer, N. Thomä, S. Kubicek, G. Winter.

*Nature Chemical Biology* (2020), <https://doi.org/10.1038/s41589-020-0594-x>

⇒ **Chapter 3** explores the structure-activity relationships underlying the gain-of-function glue activity of CR8 and describes the discovery of new cyclin K degrader scaffolds. This chapter is unpublished.

⇒ **Chapter 4** focuses on fluorescence-based methods to study molecular glue-induced proximity as well as the interactions of DNA-binding proteins with DNA or nucleosomes. It contains unpublished data as well as published data based on the following manuscripts:

*Mechanisms of OCT4-SOX2 motif readout on nucleosomes*

A. Michael\*, R. Grant\*, L. Isbel\*, S. Cavadini, **Z. Kozicka**, R. Bunker, A. Schenk, A. Graff-Meyer, G. Pathare, J. Weiss, S. Matsumoto, L. Burger, D. Schübeler<sup>†</sup>, N. Thomä<sup>†</sup>  
*Science* (2020); <https://doi.org/10.1126/science.abb0074>

*Structural mechanism of cGAS inhibition by the nucleosome*

G. Pathare, A. Decout, S. Glück, S. Cavadini, K. Makasheva, R. Hovius, G. Kempf, J. Weiss, **Z. Kozicka**, B. Guey, P. Melenec, B. Fierz, N. Thomä<sup>†</sup>, A. Ablasser<sup>†</sup>  
*Nature* (2020), <https://doi.org/10.1038/s41586-020-2750-6>

⇒ **Chapter 5** is a discussion of the results described in chapters 2 and 3. It integrates the findings into our current understanding of molecular glue degraders and considers strategies to rationalise degrader design and discovery. Several sections and figures have been adapted from the manuscript:

*Haven't got a glue: Protein surface variation for the design of molecular glue degraders.*

**Z. Kozicka**, N. Thomä

*Cell Chemical Biology* (2021), <https://doi.org/10.1016/j.chembiol.2021.04.009>

⇒ **Chapter 6** describes the materials and methods pertaining to Chapters 2, 3 and 4

⇒ All references, the list of abbreviations, the acknowledgments, and my curriculum vitae can be found after Chapter 6

⇒ **Supplementary Information** contains additional data related to Chapter 3

⇒ The **Appendix** features all published manuscripts listed above, as well as additional publications that do not constitute a part of this thesis.

# Chapter 1 : Introduction

In recent decades, small molecules able to remodel the protein-protein interaction landscape of their targets have been identified. Molecular glue degrader drugs that cause target protein degradation via induced proximity with a ubiquitin ligase have attracted particular attention and have already yielded clinically approved compounds, with lenalidomide (*Revlimid*) regarded as one of the most successful anticancer drugs of our time<sup>1</sup>. Molecular glue degraders, and more specifically a novel class of small molecules that degrade cyclin K that was discovered through the work described herein, will be the main focus of this dissertation.

The following chapter will provide an overview of the ubiquitin-proteasome system and targeted protein degradation, with a focus on molecular glue degraders. In later sections, cyclin-dependent kinases are introduced. In particular, the CDK12-cyclin K complex, the rationale for its therapeutic targeting and current efforts aimed at its pharmacological modulation are discussed.

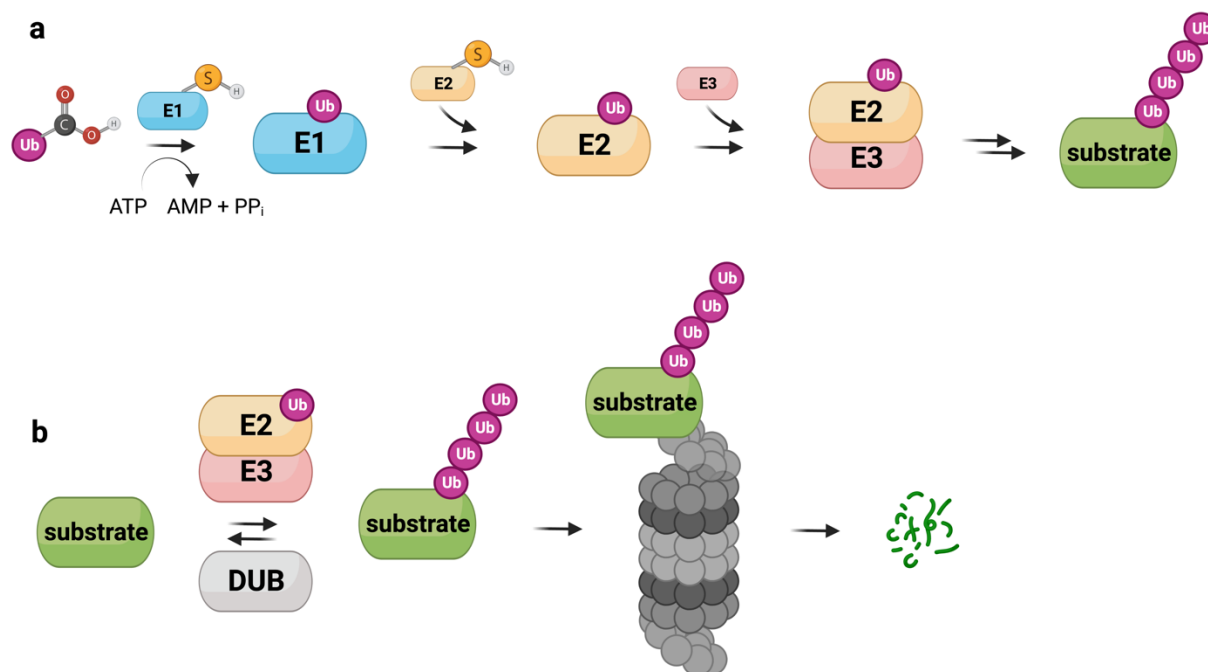
## 1.1 Targeted protein degradation

### 1.1.1 The ubiquitin proteasome system

Cellular homeostasis requires a fine balance between protein synthesis and degradation. The landmark discoveries of Aaron Ciechanover, Avram Hershko, and others begun the elucidation of the complex biochemical mechanisms of intracellular proteolysis<sup>2</sup>. Their work revealed that the 76-amino acid protein initially termed APF-1 (ATP-dependent Proteolysis Factor 1) is required for ATP-dependent protein degradation and that APF-1 becomes covalently tethered to proteolytic substrates via an isopeptide bond to a substrate lysine residue<sup>3,4</sup>. Subsequent studies revealed that a protein earlier identified to covalently attach to histones through an isopeptide bond and APF-1 are in fact the same protein, which was then renamed *ubiquitin* due to its high conservation and ubiquitous expression in eukaryotes<sup>5,6</sup>.

These early studies recognised that ubiquitinated substrate proteins are directed for proteasomal degradation and, in the following decades, the work of many laboratories elucidated the molecular details of the tightly regulated proteolytic cycle<sup>7</sup>. Substrate ubiquitination occurs via a sequential reaction catalysed by three enzymes, namely ubiquitin-activating (E1), ubiquitin-conjugating (E2), and ubiquitin ligase (E3) proteins<sup>8</sup>. The canonical ubiquitination cascade is shown in **Figure 1.1a**. Firstly, ubiquitin is activated by the E1 in an ATP-dependent manner, with ubiquitin tethered to the catalytic cysteine of the enzyme via a high-energy thioester bond to its C-terminus<sup>9</sup>. Ubiquitin is then transferred onto the catalytic cysteine of the E2 enzyme through a transthioesterification reaction<sup>8,10,11</sup>. The E2-ubiquitin conjugate subsequently associates with the E3 enzyme, which catalyses the final ligation step that attaches ubiquitin onto an amino group in the E3-bound substrate<sup>8</sup>. This results in an isopeptide bond between the C-terminus of ubiquitin and the substrate<sup>5,8</sup>. The cycle can repeat several times to yield a polymeric chain, with ubiquitin-ubiquitin isopeptide bonds forming. As ubiquitin possesses seven lysine residues (K6, K11, K27, K29, K33, K48, K63), both branched and linear chains with various linkages can form, and their nature governs the fate of the ubiquitinated substrate<sup>12</sup>. K48-linked polyubiquitin chains serve as a mark for degradation by the 26S proteasome (**Figure 1.1b**), while individual

ubiquitin marks, K63-linked or branched chains are associated with other signalling roles or are only partly understood<sup>12</sup>. The reverse reaction of ubiquitin removal is catalysed by deubiquitinating enzymes (DUBs), a family of *circa* 100 hydrolases that frequently show both substrate and linkage specificity<sup>8,13</sup>.



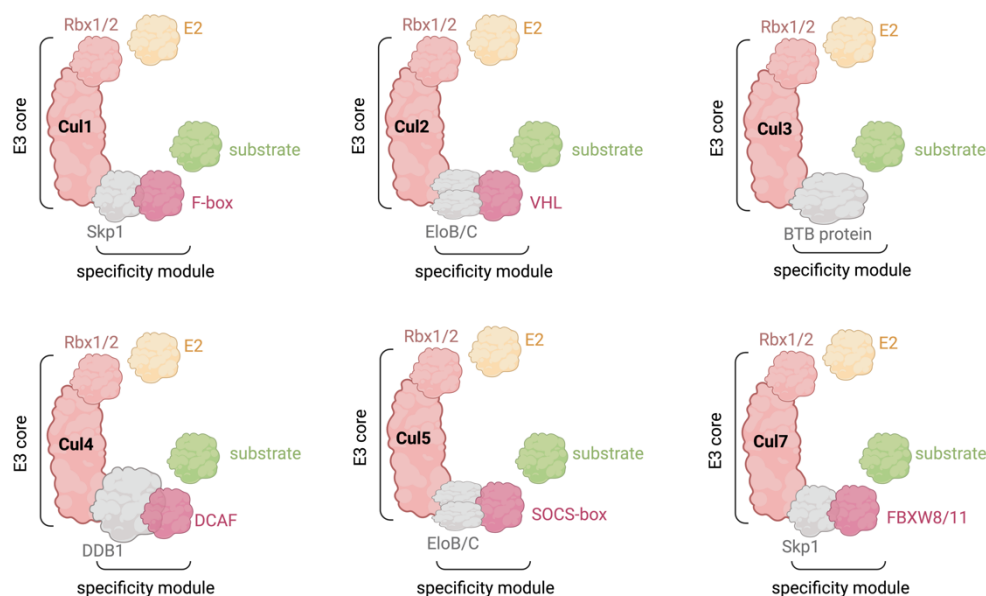
**Figure 1.1 Overview of the ubiquitin-proteasome system.** **a**, The ubiquitination cascade begins with activation of ubiquitin, an ATP-dependent process catalysed by the E1 (ubiquitin-activating) enzyme. The activated ubiquitin is then transferred to the E2 (ubiquitin-conjugating) enzyme and, with the aid of an E3 ligase, covalently attached to a lysine residue on the substrate. **b**, Ubiquitin can be transferred directly from the E2-E3 complex onto the substrate (RING ligases) or first conjugated to the E3 ligase (HECT and RBR ligases, *not shown*). Deubiquitinating enzymes (DUBs) can cleave ubiquitin chains. K48-linked polyubiquitin chains mark substrates for degradation by the 26S proteasome. Created with BioRender.com.

The extraordinary specificity of the ubiquitin-proteasome system stems from its hierarchical structure. Whereas only two E1 enzymes (UBA1 and UBA6) exist in humans, there are *circa* 40 E2 enzymes and over 600 E3 enzymes<sup>14,15</sup>. The ubiquitin system therefore gains specificity as the cascade progresses. While E1 and E2 enzymes have largely conserved structures, E3 ligases are much more diverse and often contain additional domains. The two major families of E3 ubiquitin ligases are RING (Really Interesting New Gene)-type ligases and HECT (Homologues to E6AP C-Terminus) ligases. RING ligases contain a RING domain that coordinates two  $Zn^{2+}$  ions and usually have binding affinity for their targets, allowing for direct ubiquitin transfer from the E2 to the E3-bound substrate (**Figure 1.1**)<sup>16</sup>. In contrast, substrate ubiquitination by HECT-type ligases proceeds via an additional step, whereby ubiquitin is first transferred to the catalytic cysteine of the HECT domain and only then onto the substrate<sup>17</sup>. RING ligases make up the bulk of human E3 ligases and among them, cullin-RING ligases (CRLs) are a particularly prominent class.

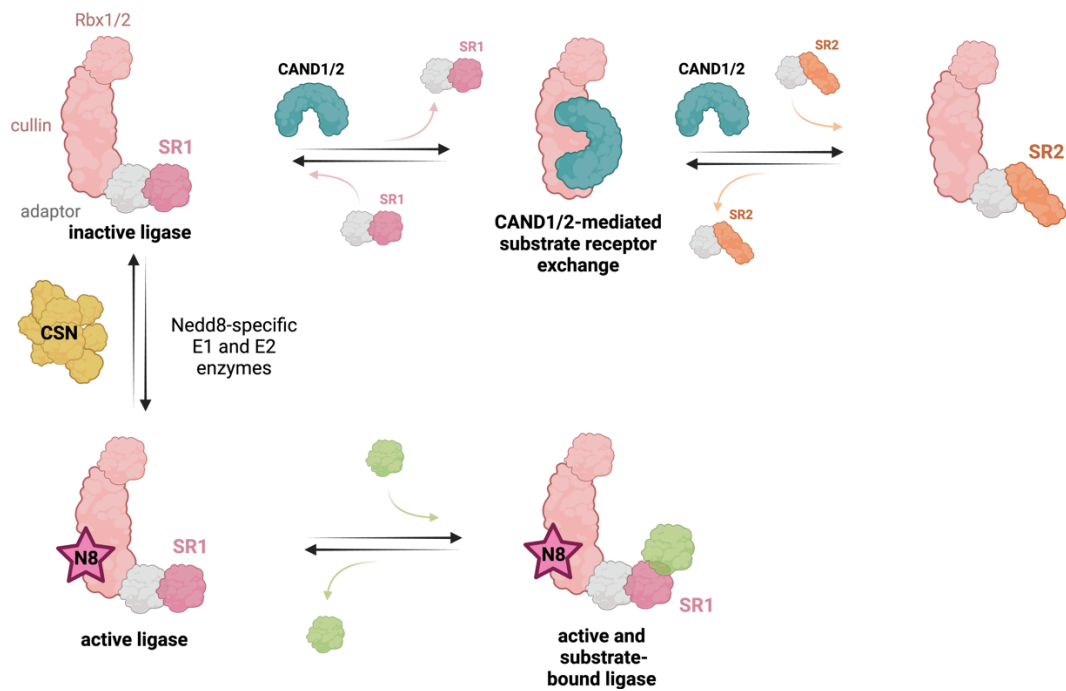
### 1.1.2 Cullin-RING ligases

The cullin-RING ligase (CRL) family is comprised of the CRL 1-5 and 7 clades, which together with ~200 substrate receptors are responsible for ~20% of proteasome-mediated protein degradation in the cell<sup>18</sup> (**Figure 1.2**). CRLs are multiprotein complexes containing a dedicated substrate receptor module that confers ubiquitination specificity through direct protein-protein interactions with the substrate<sup>19,20</sup>. These receptors connect the substrate directly, or with the help of an adaptor protein, to the N-terminus of the cullin ligase arm<sup>20,21</sup>. Through its C-terminus, the cullin engages the Rbx1/2 subunit, which serves as the recruitment site for the ubiquitin-loaded E2, allowing for direct ubiquitin transfer onto the substrate.

CRLs are dynamic assemblies governed by multiple regulatory mechanisms (**Figure 1.3**). These E3 ligases only become fully active upon the conjugation of Nedd8, a ubiquitin-like protein, to a conserved lysine residue on the cullin, and a subsequent conformational change<sup>22-25</sup>. Neddylation is enacted by Nedd8-specific enzymes, the heterodimeric E1 NAE1-Uba3 and two E2s, UBE2M and UBE2F<sup>26</sup>. The COP9 signalosome (CSN), a multiprotein isopeptidase complex, can remove the Nedd8 modification from idle (*i.e.* not substrate-bound) cullins using a steric gating mechanism and thereby acts to inhibit CRLs<sup>27</sup>. Nonetheless, CSN was shown to be required for CRL activity as it enables the exchange of ligase substrate receptors in the de-neddylated state<sup>28,29</sup>. CAND1/2 (Cullin-Associated and Neddylated-Dissociated 1/2) only bind to cullins that are not neddylated or associated with adaptor proteins or substrate receptors and were initially thought to sequester cullins in their inactive state<sup>30</sup>. However, recent studies elegantly demonstrated that CAND1/2 in fact function as exchange factors for CRL substrate receptors and are required for efficient substrate degradation<sup>31</sup>.



**Figure 1.2. Modular architecture of cullin-RING ligases.** The CRL core comprises a scaffolding protein (cullin; Cul1-5, 7) and a small RING finger protein (Rbx1/2) that contains the E2 binding site. Through its N-terminus, the cullin binds to an adaptor protein (Skp1, EloB/C, or DDB1), which in turn is bound to a substrate receptor (F-box, VHL, DCAF or SOCS-box) that bestows substrate specificity. In Cul3 ligases, the cullin instead binds to a BTB protein that directly interacts with its cognate substrates. Created with BioRender.com.

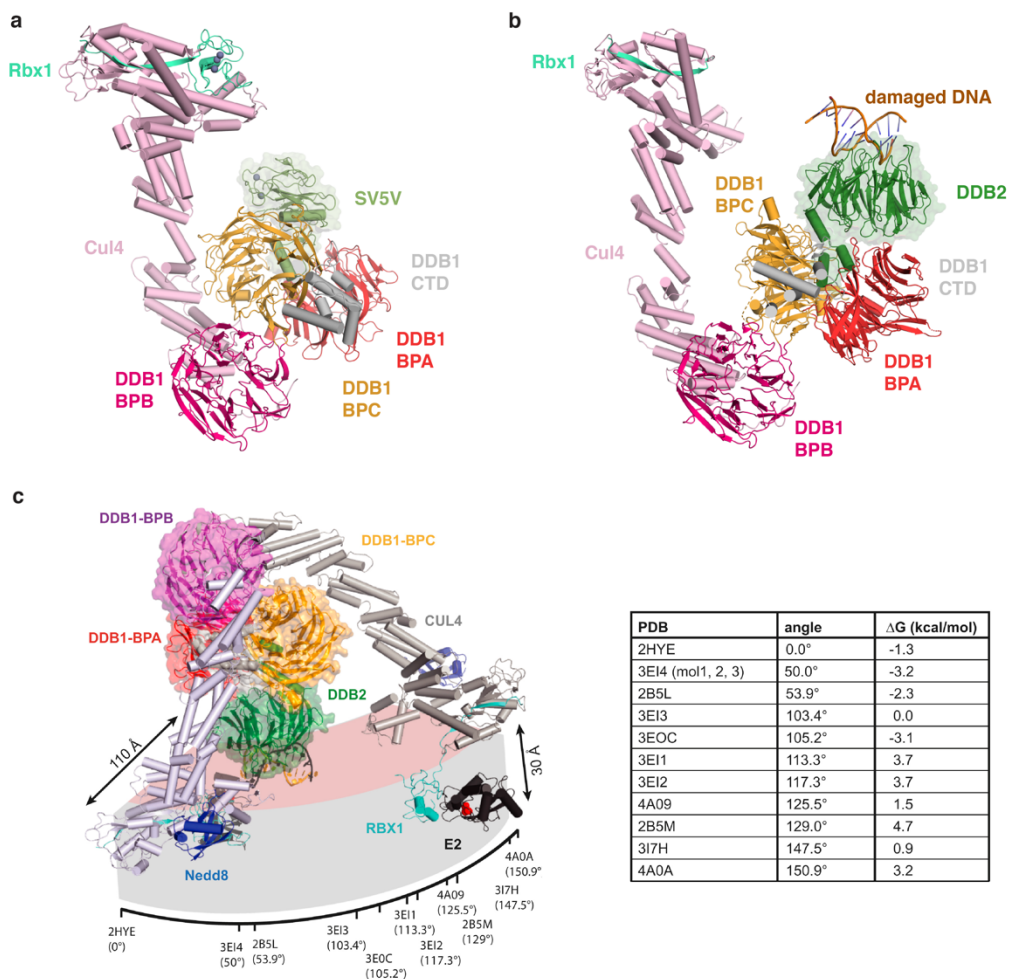


**Figure 1.3. Cullin-RING ligase regulation.** Nedd8 conjugation activates CRLs and de-neddylation of idle ligases by the CSN inhibits their activity. The CAND1-mediated exchange of substrate receptors occurs in the de-neddylated state and is essential for CRL activity. N8: Nedd8, SR: substrate receptor. Created with BioRender.com.

### 1.1.3 CRL4 clade

CRL4 ligases are multiprotein complexes consisting of Cul4, RBX1, the adaptor protein DDB1 (Damage DNA Binding protein 1), and a DCAF (DDB1-Cul4-Associated Factor), which usually functions as a substrate receptor<sup>32</sup>. Cul4 is the only cullin that in humans has two isoforms, Cul4a and Cul4b, with distinct subcellular localisations and only partially overlapping functions<sup>33</sup>. While other CRLs have structurally-related, BTB domain-containing adaptors, the CRL4-specific DDB1 is unusually large (127 kDa) and has a distinct fold with three WD40  $\beta$ -propellers (BPA, BPB, BPC) arranged in a trilateral fashion and stabilised by a helical C-terminal domain (**Figure 1.4**). Structural and biochemical studies revealed that DDB1 interacts with the N-terminus of Cul4 through its BPB domain, and that the overall architecture of the CRL4 complex resembles that of a previously described CRL1 assembly<sup>30,32</sup>.

Seminal work of the Zheng laboratory identified putative CRL4 substrate receptors through DDB1 pulldown experiments and coined the term DCAF<sup>32</sup>. It confirmed known DDB1 interactors such as DDB2 and revealed that DCAFs have a common WD40  $\beta$ -propeller domain architecture, while possessing little sequence similarity otherwise. One exception is a small DxR motif located on the bottom surface of the DCAF, which Angers *et al.* showed to be essential for DDB1 binding through mutational analysis<sup>32</sup>. Subsequent studies identified multiple additional DCAFs<sup>34-36</sup> and recent work by Reichermeier *et al.* found that DDB1-substrate receptor modules differ up to 200-fold in abundance despite comparable DCAF expression levels, with substrate availability additionally shaping the composition of the CRL4 ligase network<sup>31</sup>.



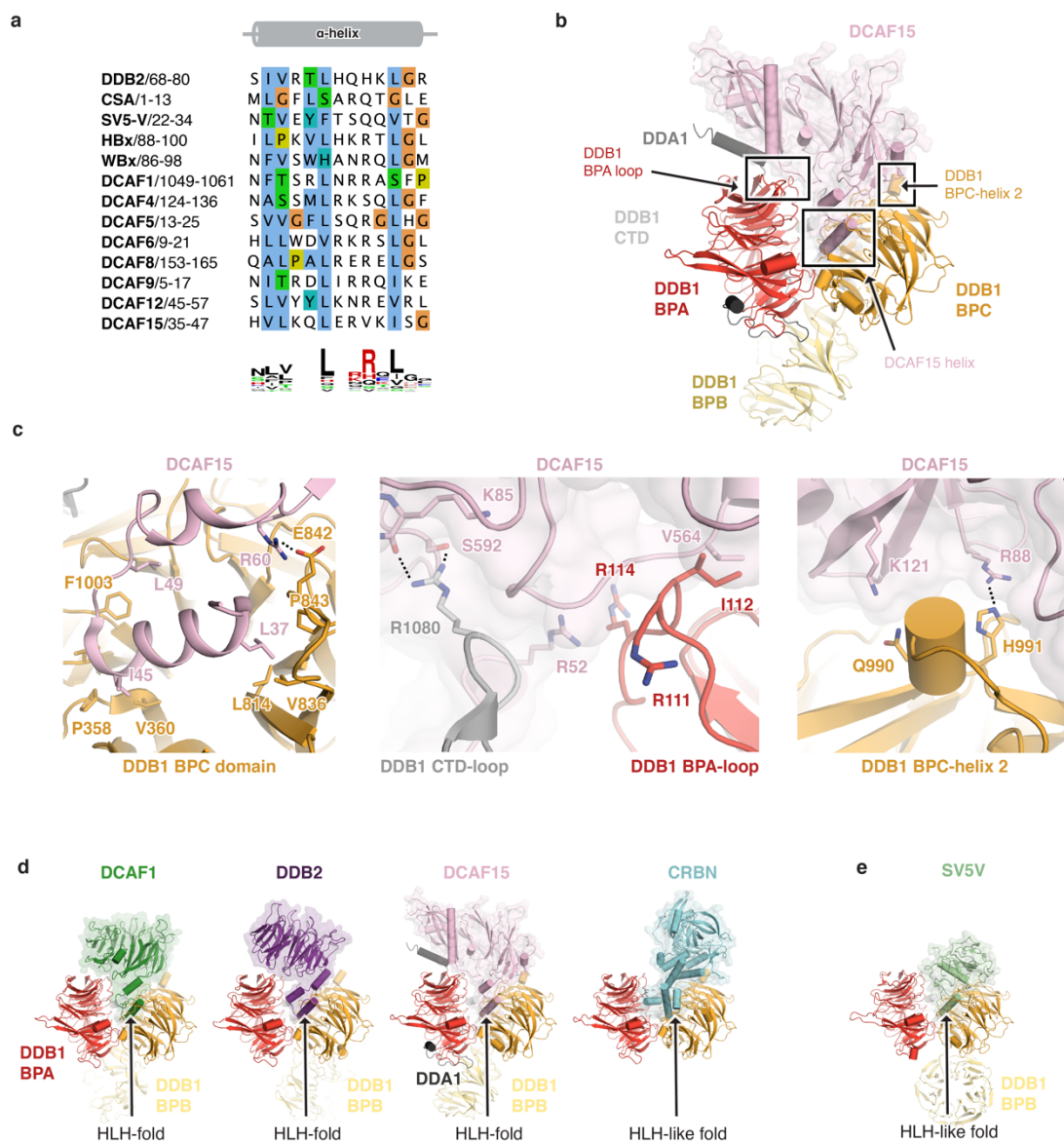
**Figure 1.4. Structural features of CRL4 enzymes.** **a**, Structure of the SV5V-DDB1-Cul4-Rbx1 complex (PDB 2HYE)<sup>37</sup>. **b**, Structure of the DDB2-DDB1-Cul4-Rbx1 complex engaged with a substrate, an basic-site containing DNA duplex (PDB 4A0K)<sup>38</sup>. **c**, Illustration of the rotational mobility of the BPB domain of DDB1 and the resulting orientations of Cul4-Rbx1. This flexibility gives CRL4s a large search radius and a ubiquitination zone of about 340 Å × 110 Å × 30 Å from the receptor<sup>38,39</sup>. Panel (c) was adapted from<sup>39</sup>.

Interestingly, even before CRL4 DCAFs were described, DDB1 was shown to interact with the paramyxovirus V protein SV5V (**Figure 1.4a**)<sup>37</sup>. The viral protein effectively hijacks the CRL4 function and recruits the host transcription factor STAT1 to DDB1 to cause STAT1 ubiquitination and degradation, thereby acting as a viral, exogenous substrate receptor. Its binding to DDB1 was shown to be mutually exclusive with endogenous DCAFs, and later structural studies demonstrated that viral and endogenous substrate receptors utilise the same interface on DDB1<sup>37,40</sup> (**Figure 1.4b**).

Multiple structures of DDB1-DCAF complexes have been solved. In the first structure, Scrima *et al.* visualised CRL4<sup>DDB2</sup>, a complex involved in nucleotide excision repair, to reveal that the DCAF, DDB2, utilises a helix-loop-helix (HLH) fold to engage the cavity between BPA and BPC domains of DDB1 (**Figure 1.4b**)<sup>19</sup>. A later study by Li *et al.* characterised the binding mode of nine endogenous and viral DCAFs to DDB1 through co-crystallisation of DDB1 with the relevant 13-mer HLH peptides<sup>40</sup>. The study defined an H-box motif spanning the key DDB1-interacting helix, which bears some sequence similarity across all DCAFs, as an



essential feature for DDB1 engagement (**Figure 1.5a**). Further structures of DDB1 with DCAF1<sup>41</sup>, DCAF15<sup>42</sup>, and a somewhat structurally atypical DCAF CRBN<sup>39</sup> showed that while the HLH-motif interaction is universally present, full-length DCAFs engage multiple additional DDB1 interfaces (**Figure 1.5b-e**).



**Figure 1.5. DDB1-DCAF interactions.** **a**, Protein sequence alignment of the key helix-loop-helix (HLH)-domain helix from various DCAFs and viral proteins. **b**, Structure of the DCAF15-DDB1 complex (PDB 6PAI<sup>42</sup>). DCAF15 binds a cleft between the BPA and BPC domains of DDB1. Protein-protein interaction hotspots are highlighted with boxes. **c**, An HLH motif of DCAF15 engages DDB1 (left). Additional protein-protein interaction hotspots (middle, right). **d**, Diverse DCAFs, including DCAF1 (PDB 5JK7<sup>43</sup>), DDB2 (PDB 4A0B<sup>38</sup>), and CRBN (PDB 5FQD<sup>44</sup>) bind the cleft between the BPA and BPC domains of DDB1 through HLH- or HLH-like folds. **e**, Viral proteins such as SV5V (PDB 2HYE<sup>32</sup>) can bind in a DCAF-like manner to hijack the activity of DDB1.

The CRL4 clade has several properties that make it particularly amenable to reprogramming by viruses and, as detailed in subsequent sections, by small molecules. One important factor is that CRLs ubiquitinate diverse substrate proteins when brought into direct proximity to the ligase: CRL4<sup>DDB2</sup>, for example, is recruited to sites of DNA damage, where it ubiquitinates not the DNA lesion itself but various proteins in the immediate vicinity of the damage site<sup>45</sup>. Moreover, the rotational flexibility of the DDB1 BPB domain and the resulting mobility of the Cul4 arm spanning over 150° establish a ubiquitination zone with dimensions as large as 340 Å × 110 Å × 30 Å (**Figure 1.4c**)<sup>38,39</sup>. The CRL4 ligases are promiscuous, targeting any lysine residues that cross the ubiquitination zone, often including sites on the DCAF itself<sup>39</sup>. This flexible attachment and thermal motion-driven rotation give rise to a large search radius, facilitating the search for lysine residues in non-native substrates, for example upon viral exploitation of CRL4s.

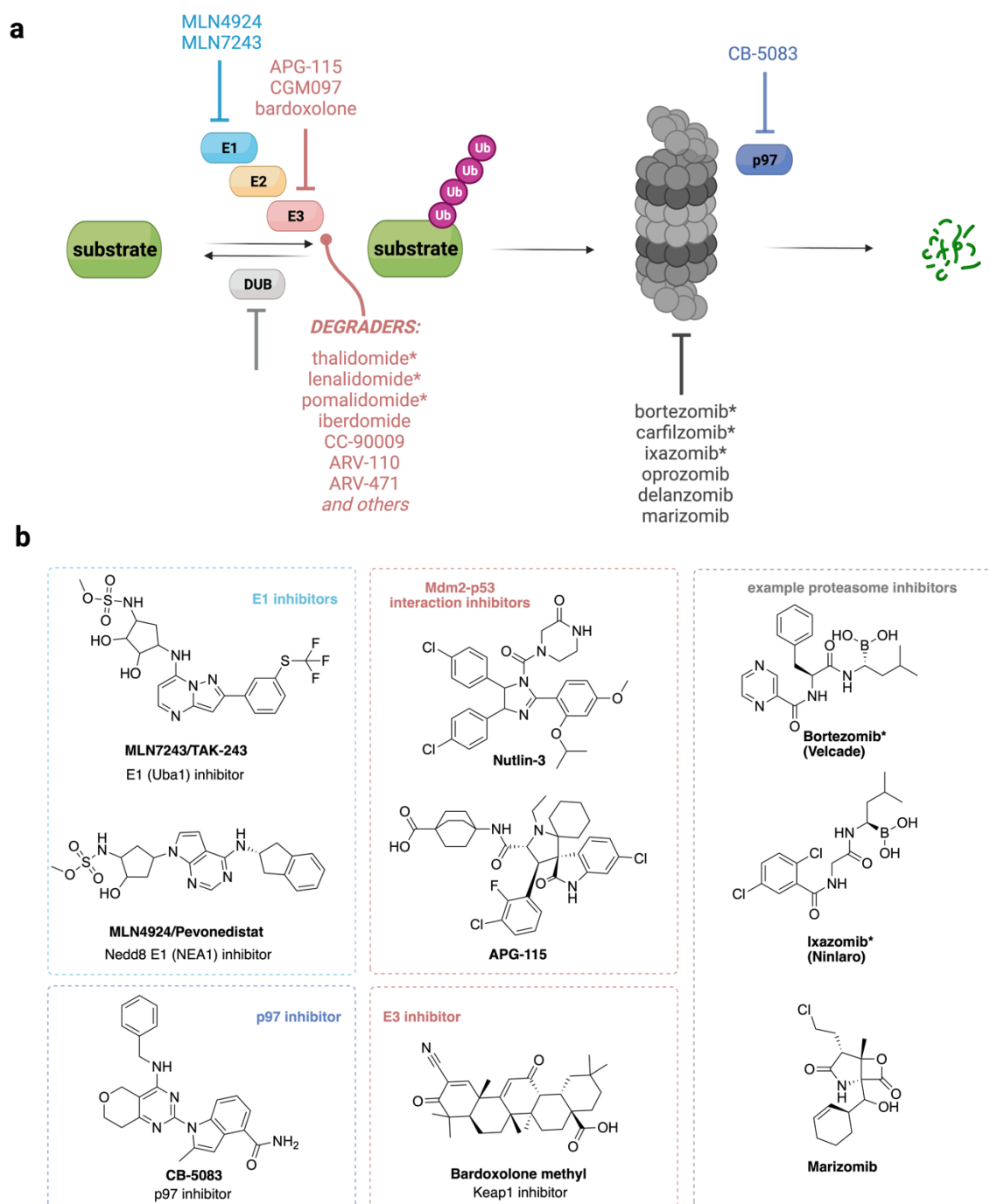
#### 1.1.4 Therapeutic targeting of the ubiquitin-proteasome system

Tightly controlled protein turnover is essential for homeostasis. In addition, the ubiquitin-proteasome system (UPS) regulates a multitude of vital cellular processes, including gene expression and the cell cycle, and the roles of protein ubiquitination extend far beyond the regulation of proteolysis<sup>7,12</sup>. As aberrations within ubiquitin-dependent pathways are frequently associated with disease states, the UPS constitutes an attractive target for therapeutic intervention and multiple small molecules that modulate it have been developed<sup>46</sup>.

The selectivity of UPS-directed therapeutics depends greatly on where in the pathway does the drug intervene (**Figure 1.6a**). For instance, inhibition of the proteasome will have pleiotropic effects. Yet, the first UPS-targeting drug to receive FDA approval was the proteasome inhibitor bortezomib (*Velcade*), a dipeptide boronic acid analogue developed by Millennium Pharmaceuticals (**Figure 1.6b**)<sup>47</sup>. Despite the lack of any substrate specificity, bortezomib demonstrated clinical efficacy in refractory or relapsed multiple myeloma with manageable toxicities<sup>48</sup>. The therapeutic window is thought to be afforded by an increased dependence of the rapidly dividing tumour cells on the UPS<sup>49</sup>. Multiple analogues followed, including an FDA-approved orally available inhibitor ixazomib (*Ninlaro*, developed by Takeda), and a structurally distinct β-lactone marizomib currently in clinical trials (**Figure 1.6**)<sup>50</sup>.

While the direct targeting of the E1-3 enzymes remains challenging due to the lack of classically druggable sites on their surface, considerable progress has been made in recent years<sup>51</sup>. MLN7243/TAK-243 (Millennium Pharmaceuticals) is an inhibitor of the E1 enzyme Uba1 and is currently in clinical studies for solid tumours<sup>52</sup>. The compound achieves nanomolar inhibition of Uba1 through the formation of a drug-ubiquitin adduct that cannot be processed by the enzyme<sup>52</sup>. However, such E1 inhibition has largely nonspecific effects.

Another E1 inhibitor in clinical trials, though one that in fact inhibits the CRL class of E3 ubiquitin ligases, is MLN4924 (Millennium Pharmaceuticals). MLN4924 targets the Nedd8-specific E1 enzyme NAE1, similarly forming an adduct with Nedd8 and thereby disabling the enzyme<sup>53</sup>. Inhibition of NAE1 prevents cullin neddylation and thus widely disrupts CRL-mediated ubiquitination, effectively inhibiting this class of E3 ubiquitin ligases.



**Figure 1.6. Therapeutic targeting of the ubiquitin-proteasome system.** **a**, Drug targets within the UPS. Clinical candidates and approved drugs (marked with an asterisk) are listed. Figure panel partly based on <sup>51</sup> and created with Biorender.com. **b**, Chemical structures of selected UPS-targeting inhibitors. Degradation compounds are depicted and discussed in detail in the following sections.

The direct targeting of individual E3 ligases or E3-substrate interactions constitutes the most specific intervention in the UPS pathway. Covalent inhibitors of the CRL3 adaptor protein Keap1, such as the triterpenoid bardoxolone methyl (**Figure 1.6b**), prevent the degradation of its target, the transcription factor Nrf2<sup>54</sup>. This potentially activates the Nrf2 pathway and could offer therapeutic benefits in cancer and chronic

kidney disease, with clinical trials ongoing<sup>54</sup>. However, most E3 ligases have proven to be challenging targets as they lack conventional small molecule-binding pockets, and the efforts have extended beyond blocking their catalytic activity and into protein-protein interaction inhibition. Nutlins, a class of small molecules that block the interaction between the RING-family ligase Mdm2 (Mouse double minute 2) and its substrate p53 are a seminal example of direct pharmacological targeting of E3 ligases<sup>55</sup>. These cis-imidazoline derivatives bind to an extended cavity on the Mdm2 surface and prevent the binding of p53, resulting in the stabilisation of the tumour suppressor. Multiple generations of Mdm2-p53 inhibitors have since been developed and nine are currently in clinical trials (**Figure 1.6b**)<sup>56</sup>.

Small-molecule binders of the von Hippel-Lindau tumour suppressor (VHL) E3 ligase such as VH298 have been developed based on peptidomimetics (inspired by its endogenous substrate HIF1 $\alpha$ ) and structure-guided evolution of fragment-derived scaffolds<sup>57</sup>. Similarly, ligands of the E3 ligase IAP (Inhibitor of Apoptosis Protein) have been obtained, and several scaffolds are currently in clinical development for oncology indications<sup>58,59</sup>. Binders of CRBN, a CUL4 DCAF substrate receptor, such as thalidomide (*Thalomid*, *Contergan*) and lenalidomide (*Revlimid*), have been retrospectively identified but many display an unexpected gain-of-function mechanism of action (see **section 1.1.10**). As described in the following sections, all E3 ligase ligands, even if not inhibitory in nature, have become very highly sought after with the advent of targeted protein degradation.

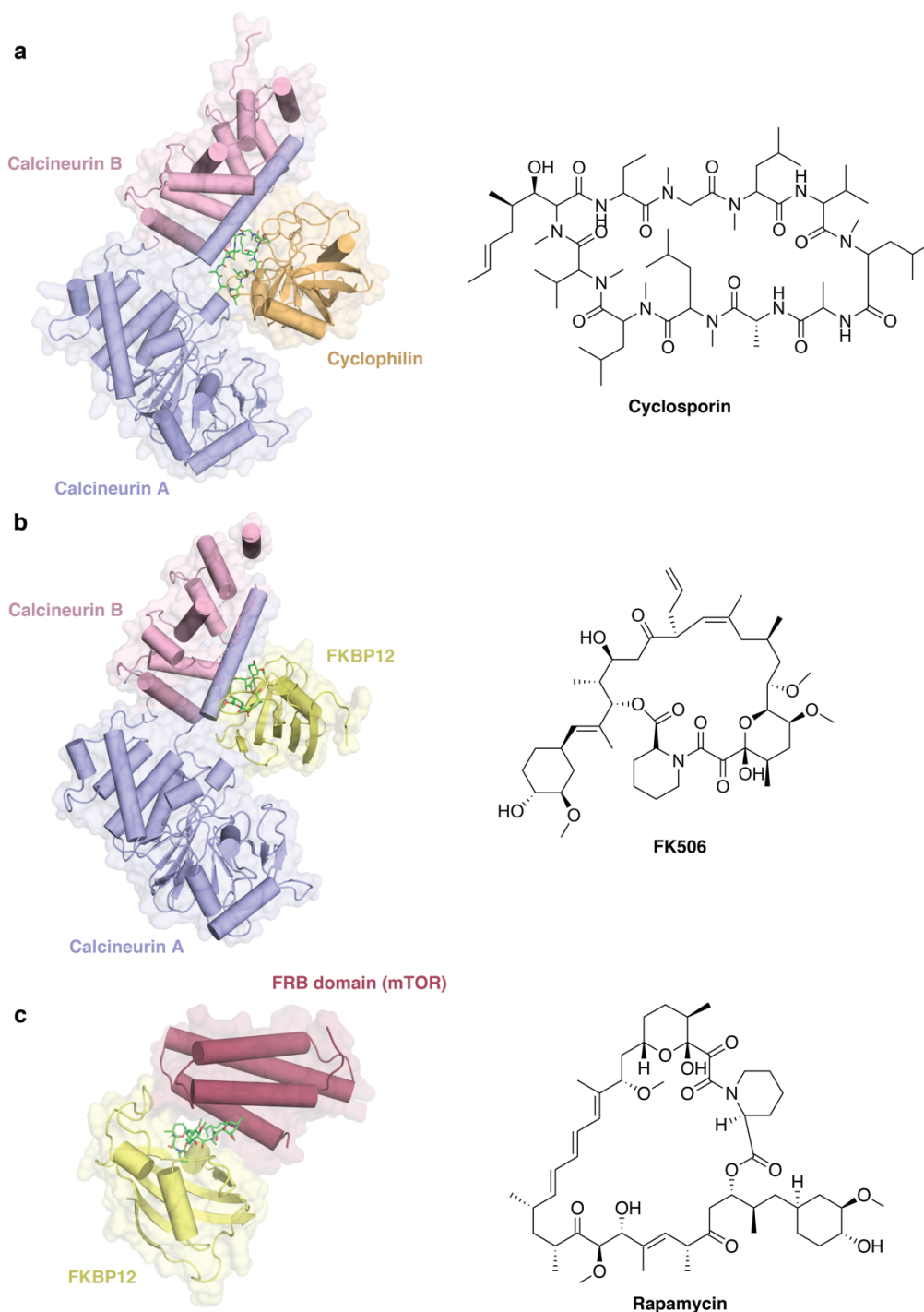
Preclinical small molecules that target other components of the pathway, *e.g.*, DUBs (dual USP14/UCHL5 inhibitor VLX1750<sup>60</sup>, multiple USP7 inhibitors<sup>61,62</sup>), the CSN5 subunit of the COP9 signalosome (CSN5i-3<sup>63</sup>), or p97, an ATPase involved in the extraction of ubiquitinated substrates (CB-5083<sup>64</sup>), have also been described (**Figure 1.6**).

### 1.1.5 Hijacking of the ubiquitin-proteasome system

#### 1.1.6 Chemical inducers of proximity

Molecular recognition events between two proteins, or between a protein and a small molecule, are at the heart of every biological process. In drug discovery efforts the predominant, yet technically challenging, approach of modulating protein-protein interactions has long been to develop interaction inhibitors<sup>65</sup>. Inducing *de novo* protein-protein interactions with a compound, on the other hand, has been considered infeasible<sup>66</sup> but this perception shifted over the last years with the discovery of molecular glues. The elucidation of the proximity-inducing properties of macrolides cyclosporin, FK506, and rapamycin (**Figure 1.7**)<sup>67-70</sup> paved the way for the discovery of numerous compounds that elicit their effects through prompting new protein-protein contacts<sup>71</sup>. The three macrolides all showed immunosuppressive effects – the natural product cyclosporin (*Sandimmune*) was approved by the FDA in 1983 for the prevention of organ transplant rejection, followed by the approvals of FK506 (*Tacrolimus*) and rapamycin (*Sirolimus*) in the coming years<sup>72,73</sup>. Seminal work from the Schreiber and Crabtree laboratories uncovered that these compounds trigger the formation of calcineurin-cyclosporin-cyclophilin and calcineurin-FK506-FKBP12 ternary complexes<sup>74</sup>, and later work from Schreiber and Snyder

laboratories concurrently uncovered that rapamycin induces the interaction of FKBP12 with the kinase mTOR<sup>75,76</sup>. The structural studies that followed revealed that all three compounds are effectively buried at the protein-protein interface, triggering extensive interactions between the proteins they bring in proximity (**Figure 1.7**)<sup>77-79</sup>.

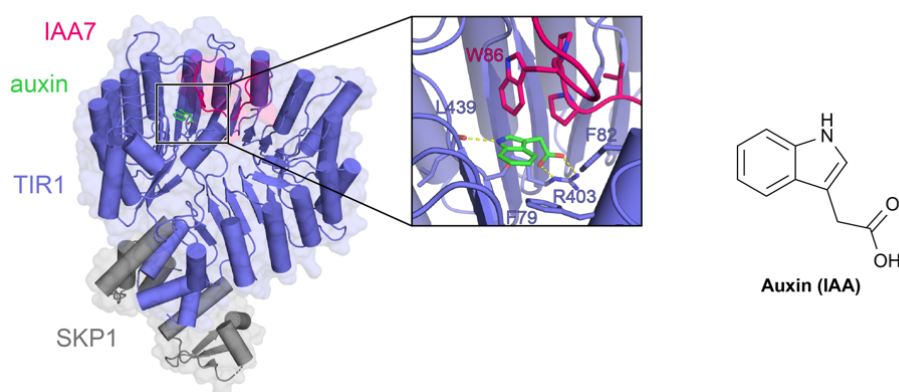


**Figure 1.7. Macrolide molecular glues.** **a**, Cyclosporin induced the interaction of cyclophilin with calcineurin (PDB 1MF8)<sup>79</sup>. **b**, FK506 induces the interaction of FKBP12 with calcineurin (PDB 6TZ7)<sup>78</sup>. **c**, Rapamycin triggers complex formation between FKBP12 and the FRB domain of mTOR (PDB 1FAP)<sup>77</sup>.

Small molecule-induced proximity is particularly consequential for E3 ubiquitin ligases. As described above, the CRL4 class of ligases is especially amenable to ubiquitinating diverse substrate proteins when brought into direct proximity to the ligase: CRL4<sup>DDB2</sup>, for example, is recruited to sites of DNA damage, where it ubiquitinates not the DNA lesion itself but various proteins in its proximity<sup>45</sup>. This property is exploited by several virus families, with viral proteins binding and thereby redirecting ligases to ubiquitinate neosubstrates to benefit viral replication. Examples include HBx, which directly binds the CRL4 adaptor DDB1 and leads to the degradation of the HBV restriction factors SMC5/6<sup>40</sup>, and S5V which brings about degradation of host transcription factors STAT in an analogous manner<sup>80</sup>. A similar mechanism is observed for substrate receptor hijacking, with the HIV-1 protein Vpr binding the CRL4 substrate receptor DCAF1 and directing the ligase to ubiquitinate several host proteins including UNG2<sup>81</sup> and the adenovirus early gene e1a hijacking CRL<sup>DCAF10</sup> to degrade RUVBL1/2<sup>82</sup>. Ubiquitination by proximity was also observed for other ubiquitin ligases such as the Anaphase Promoting Complex (APC), where grafting the destruction box of cyclin B onto non-substrate proteins rendered them subject to APC-mediated ubiquitination and degradation<sup>83,84</sup>.

### 1.1.7 Plant hormones as inducers of protein degradation

The first biological indication that a small molecule can induce protein degradation came from plant hormones<sup>85</sup>. The phytohormones auxin and methyl jasmonate were found to bind the plant F-box CRL receptors (TIR1 and COI1, respectively) and facilitate their interaction with target proteins, two transcription factors, resulting in enhanced degradation of these targets<sup>85,86</sup>. Seminal crystallographic studies from the Zheng laboratory revealed that the compact small-molecule hormones can facilitate nanomolar target-ligase interactions through a small protein-ligand interface (**Figure 1.8**). The molecular mechanism of these compounds differs substantially from that of large macrolides (**Figure 1.7**), in that the phytohormones rely predominantly on large and complementary protein-protein interfaces, with the relatively small compound/neosubstrate interface conferring sufficient additional binding energy for tight complex formation<sup>87</sup>.

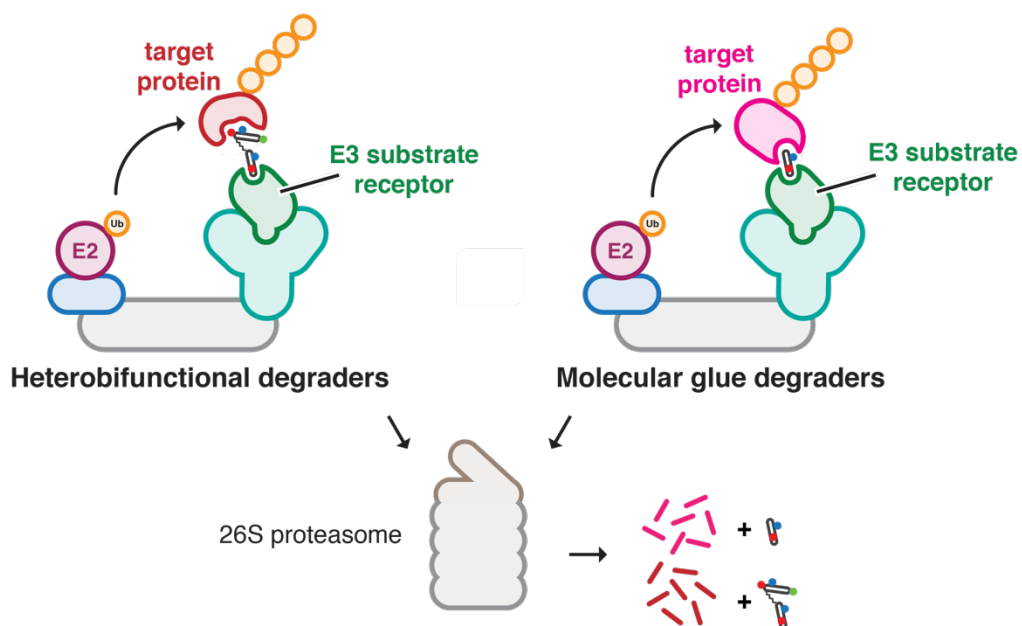


**Figure 1.8.** SKP1-TIR1-auxin-IAA7 complex (PDB 2P1Q)<sup>85</sup> and the chemical structure of auxin. Important interface residues are shown as sticks. Key interactions are marked with dashed yellow lines. The protein-protein interface spans 657 Å<sup>2</sup>, while the auxin-IAA7 interface is only 65 Å<sup>2</sup>.

### 1.1.8 Inducible degradation and bifunctional degraders

Following these discoveries, the auxin-inducible degron (AID) system was developed for controlled cellular degradation of non-physiological substrate proteins fused to an IAA7 degron<sup>88</sup>. However, the fact that viruses and plants can reprogram ubiquitin ligases through induced proximity suggested that targeted protein degradation that does not require protein engineering could be achieved with appropriately designed compounds. The man-made proof of concept came from broad successes with rationally-designed heterobifunctional degrader molecules (PROTACs, Proteolysis Targeting Chimeras)<sup>89-92</sup>. These compounds, comprising a ligase-binding ligand and a target-engaging warhead connected by a linker, position the target in proximity to the ubiquitin ligase, leading to target ubiquitination and its proteasomal degradation (**Figure 1.9**, **Figure 1.10**). The approach of targeted protein degradation has attracted keen interest from both academia and industry as it offers multiple advantages over traditional inhibition<sup>93</sup>. Firstly, the degradation mode of action allows for the targeting of proteins with no enzymatic activity or for which no effective inhibitors could be developed, as merely a ligand binding anywhere on the target is required for PROTAC design<sup>94</sup>. Moreover, target degradation eliminates all functions of the protein of interest, including *e.g.* scaffolding roles, and, while reversible, it phenotypically resembles a genetic knockout, which can lead to superior efficacy<sup>95-97</sup>. Compound-induced degradation proceeds via an event-driven mode of action, whereby one molecule of the degrader facilitates iterative rounds of target ubiquitination and degradation and can therefore be considered catalytic<sup>92</sup>. This contrasts with occupancy-based inhibitors and allows degraders to be used at sub-stoichiometric concentrations. Finally, as degrader action necessitates the formation of a stable ligase-compound-target complex, this dependence on complementary protein-protein interfaces affords degraders additional selectivity, even in the case of PROTAC ligands deriving from promiscuous inhibitors<sup>98-100</sup>.

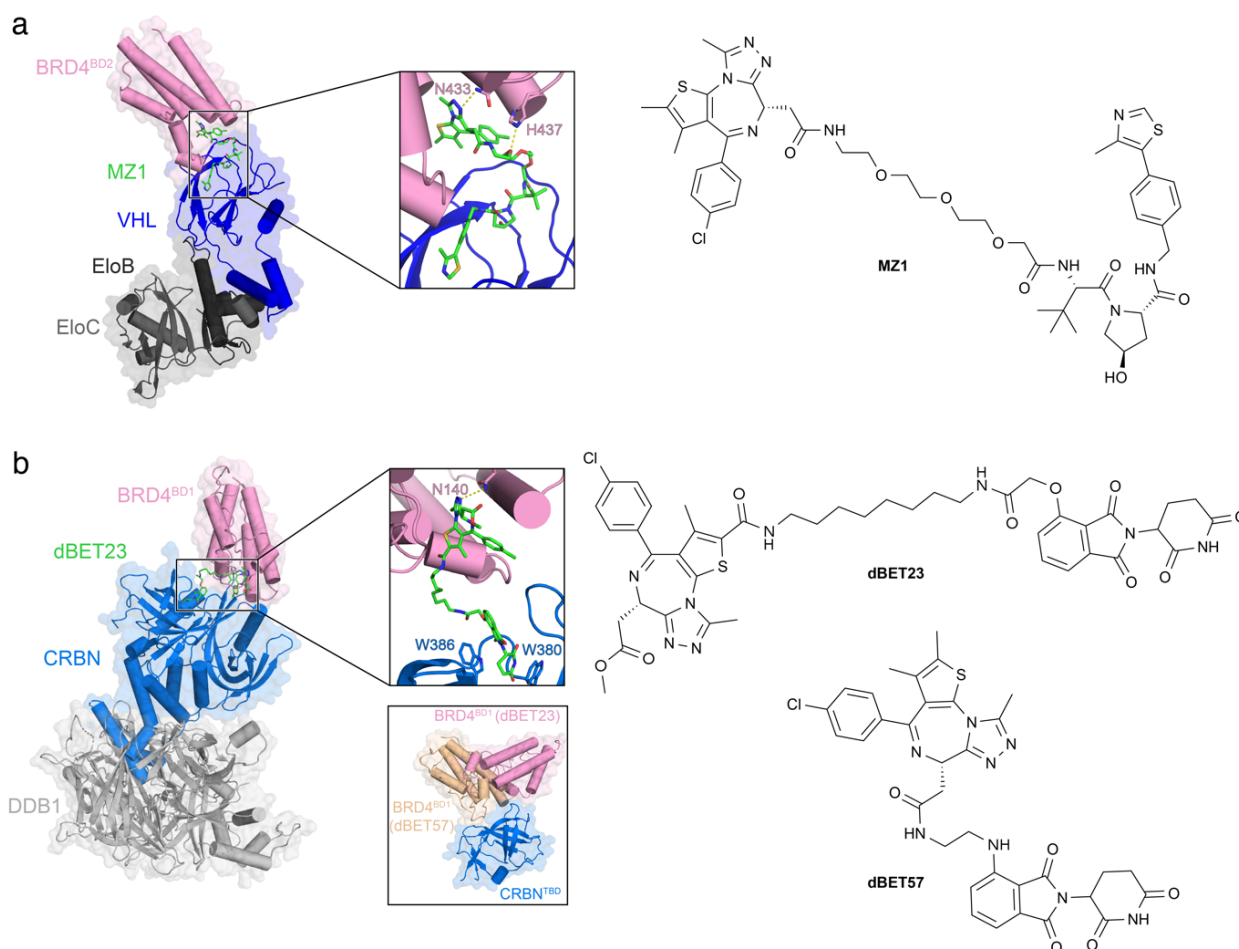
A key aspect of PROTAC's appeal is the fact that bifunctionality allows the exchange of the target-binding warhead in a plug-and-play fashion to inactivate other proteins. This modularity, at the same time, also highlights their key shortcomings. Firstly, it underscores the fact that having a target- and a ligase-binding compound is a prerequisite for designing a bifunctional degrader, therefore effectively limiting the target space to ligandable (and often otherwise druggable) targets. The ligases employed include primarily CRBN and VHL, due to the limited availability of chemical matter for targeting other ligases (see **section 1.1.4**). The described heterobifunctionality also renders the compounds comparatively large (**Figure 1.10**), making it more challenging to progress these molecules into the clinic. Nevertheless, several PROTACs are currently in clinical trials, with the CRBN-based ARV-110 (androgen receptor degrader, phase II) and ARV-471 (oestrogen receptor degrader, phase II) developed by Arvinas being the most advanced<sup>101</sup>.



Rational design	Serendipitous discovery
Dependent on ligand availability	Less reliant on ligands/pockets
High molecular weight	Small and drug-like
At least twelve currently in clinical trials <sup>101</sup>	Three approved drugs ( <i>Thalomid</i> , <i>Revlimid</i> , <i>Pomalyst</i> ) and at least six in clinical trials <sup>101</sup>

**Figure 1.9.** Schematic representation of heterobifunctional degraders (PROTACs) and molecular glue degraders, and a table with their key characteristics. These compounds promote ternary complex formation between the target protein and the ubiquitin ligase complex, which leads to target ubiquitination and subsequent degradation by the proteasome. The small molecules do not get degraded and can facilitate iterative cycles of target degradation.





**Figure 1.10. Examples of PROTAC-mediated ternary complexes.** **a**, EloB/C-VHL-MZ1-BRD4<sup>BD2</sup> complex (PDB 5T35)<sup>90</sup>. **b**, DDB1-CRBN-dBET23-BRD4<sup>BD1</sup> complex (PDB 6BN7)<sup>102</sup>. Bottom panel: overlay of dBET23 and dBET57-induced conformations of BRD4<sup>BD1</sup> (PDB 6BNB) aligned on CRBN thalidomide-binding domain (TBD). The two PROTACs engage different BD1 interfaces depending on the linker and exit vector used. The BPB domain of DDB1 was omitted for clarity. Important interface residues are shown as sticks. Key interactions are marked with dashed yellow lines. **(a, b)** The protein-protein interfaces span 348-518 Å<sup>2</sup>, while the PROTAC-target interfaces are 505-508 Å<sup>2</sup>.

### 1.1.9 Molecular glue degraders

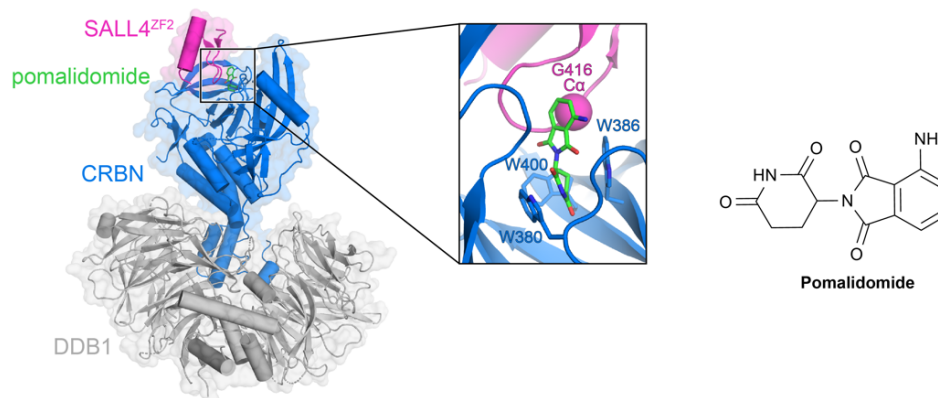
Molecular glue degraders are a second class of targeted protein degraders. Akin to PROTACs, they also facilitate dimerization of a target protein and an E3 ubiquitin ligase, but they are single, linker-less scaffolds that induce proximity by relying on extensive and direct protein-protein interactions (**Figure 1.9**). These drugs alter the interactome of the ligase or the target, strengthening existing or enabling *de novo* protein-protein contacts, which in the context of a ubiquitin ligase can result in degradation. As molecular glues rely on cooperative protein-protein interactions, they are much less dependent on independently ligandable pockets and known molecular glues predominantly target previously undruggable proteins (see **sections 1.1.10-1.1.12**). In some regards, the difference between molecular glues and bifunctional degraders is one in degree, not kind, as PROTACs can also induce direct protein-protein interactions<sup>90,102,103</sup> (**Figure 1.10**). Yet, differences in size and mono- versus bivalency remain to distinguish the classes. A significant disadvantage of molecular glues is that they are difficult to find, and all have thus far been discovered serendipitously for a given target, with one

notable exception<sup>104</sup>. In the following sections, the known molecular glue degrader compounds are reviewed with a focus on examples that have been structurally characterised to date.

### 1.1.10 Molecular glues for CRL4<sup>CRBN</sup>

Thalidomide was first introduced in West Germany in 1957 as a sedative and antiemetic, and was found to be highly teratogenic, affecting more than 10,000 patients before being withdrawn from the market in 1962<sup>105</sup>. Through a series of unexpected discoveries, thalidomide and its analogues re-emerged as powerful therapeutics in cancer and haematological malignancies. The long-sought after understanding of their mode of action came with the identification of CRBN as the efficacy target of thalidomide by the Handa laboratory<sup>106</sup> and subsequent structural studies showing that CRBN plays the role of the substrate receptor of the CRL4<sup>CRBN</sup> E3 ubiquitin ligase<sup>39,107</sup>. Thalidomide and derivatives (termed IMiDs for IMmunomodulatory Drugs) were eventually identified to be molecular glue degraders that bind a highly conserved tri-tryptophan pocket on CRBN, with the composite CRBN-ligand interface recruiting several different neosubstrates for ubiquitination. Degradation of transcription factors Ikaros/Aiolos (IKZF1/3) has been linked to the clinical efficacy of thalidomide analogues in multiple myeloma, while Casein kinase 1 $\alpha$  (CK1 $\alpha$ ) is the likely efficacy target of lenalidomide in 5q-myelodysplastic syndrome<sup>39,106–111</sup>.

Structural studies of glue-induced ternary complexes of CK1 $\alpha$ <sup>44</sup>, GSPT1<sup>112</sup>, Ikaros<sup>113</sup>, ZNF692<sup>113</sup> and SALL4<sup>114,115</sup> all revealed a structural degron in the neosubstrate comprising a characteristic  $\beta$ -hairpin loop and a key glycine residue that engages the otherwise solvent-exposed phthalimide moiety of IMiDs and contacts residues lining the IMiD pocket on CRBN (**Figure 1.11**)<sup>44,113</sup>. With the help of a relatively small protein-protein interface ( $\sim 350$ - $700$  Å<sup>2</sup>), the compound-bound CRBN complex provides a nanomolar binding platform for neosubstrates.

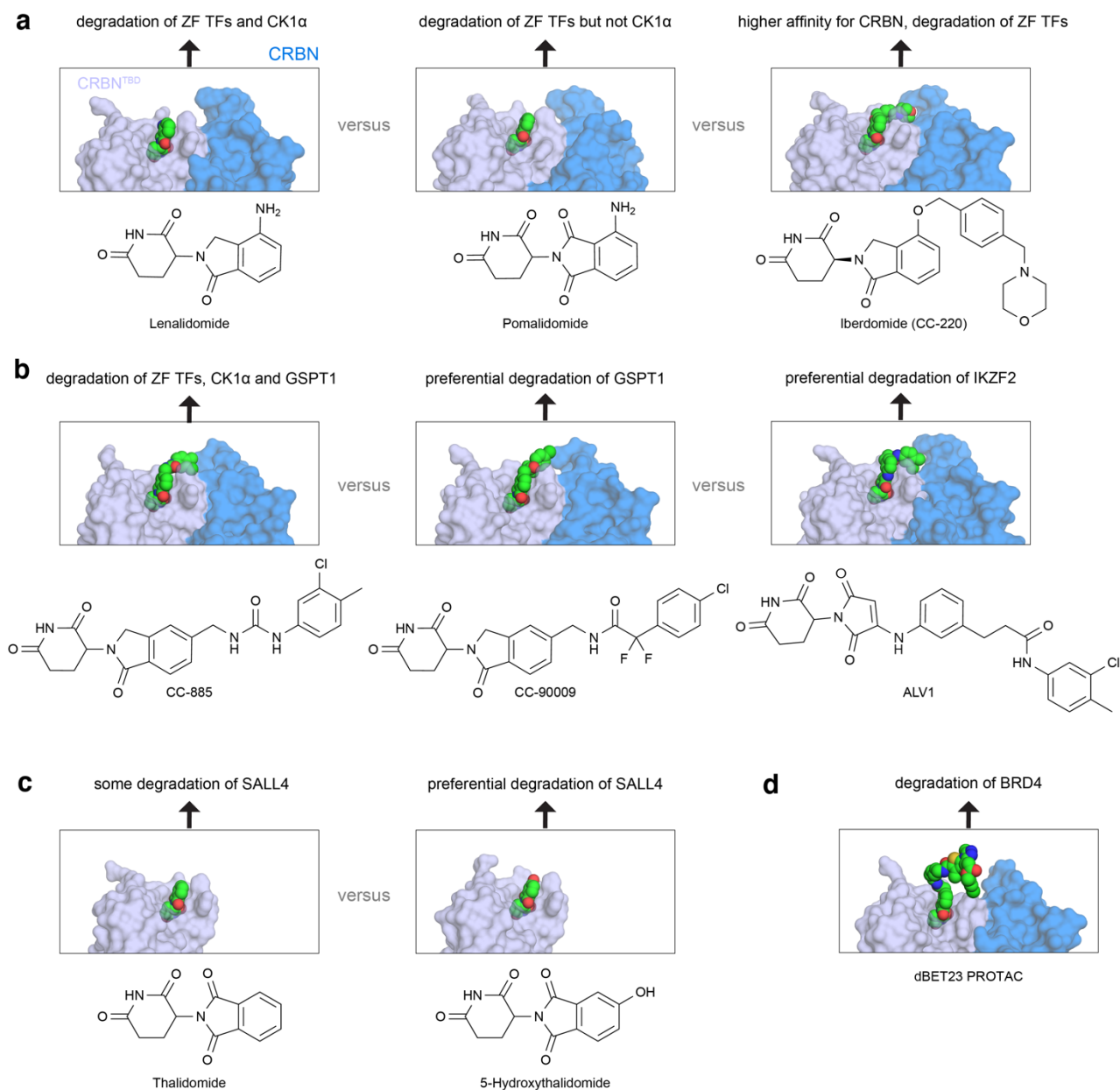


**Figure 1.11.** DDB1-CRBN-pomalidomide-SALL4<sup>ZF2</sup> complex (PDB 6UML)<sup>115</sup>. The C $\alpha$  of G416 in the  $\beta$ -hairpin loop of SALL4<sup>ZF2</sup> is shown as a sphere. The BPB domain of DDB1 was omitted for clarity. Important interface residues are shown as sticks. Key interactions are marked with dashed yellow lines. The protein-protein interface is 354 Å<sup>2</sup>, while the pomalidomide-SALL4 interface spans 89 Å<sup>2</sup>.

Arguably the biggest structural class of neosubstrates for CRBN are zinc-finger transcription factors (ZF TFs). Structures of thalidomide derivatives with IKZF1, ZNF692 and SALL4 showed that binding of the

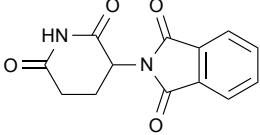
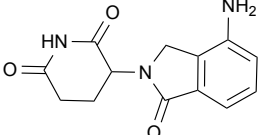
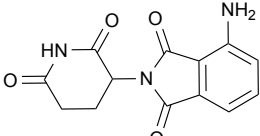
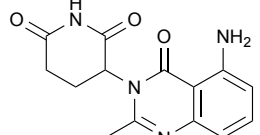
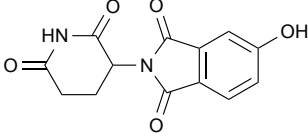
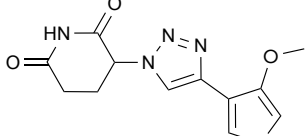
small and distinct zinc-finger fold is similarly driven by a  $\beta$ -hairpin-containing ZF, while the surrounding ZFs appear to facilitate binding *in vitro*. In cells, only those ZF-containing proteins that carry an additional ZF close by were proficient in degradation of the corresponding the full-length protein<sup>113,116</sup>, but the role of the nearby ZFs in specificity and degradation proficiency requires further investigation. Several other Cys<sub>2</sub>His<sub>2</sub> (C2H2) ZF TFs have been shown to be degraded by thalidomide and derivatives (**Table 1.1**). For example, candidate acute promyelocytic leukaemia therapeutics CC-647 and CC-3060 have recently been reported to promote ZBTB16 degradation but each engage different ZF domains<sup>117</sup>. Multiple additional ZF proteins were found to bind CRBN *in vitro*, with the engagement being insufficient to drive cellular degradation<sup>113,118</sup>.

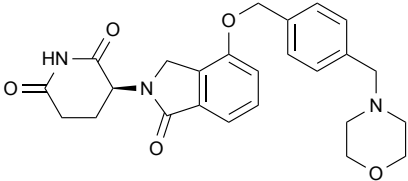
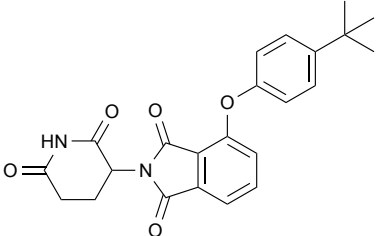
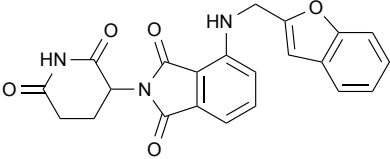
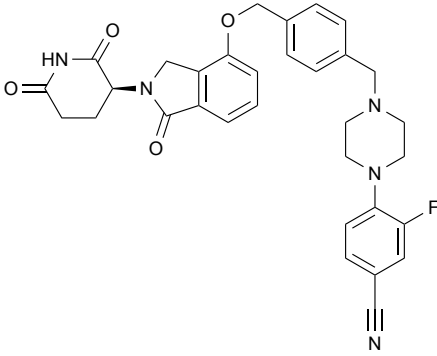
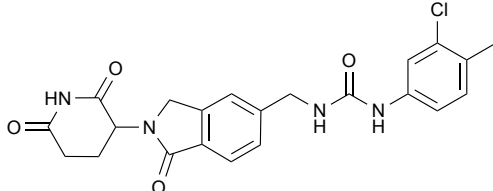
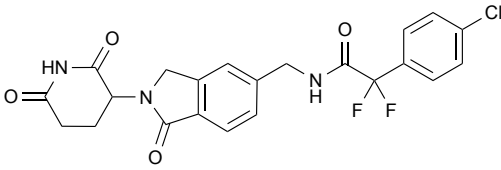
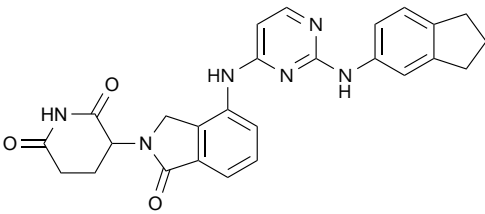
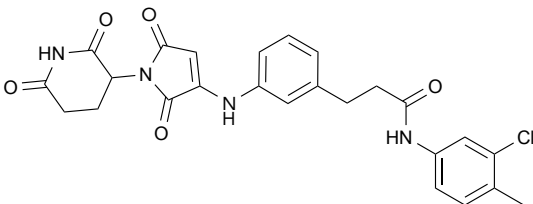
Many new thalidomide derivatives have been developed in recent years, and several compounds with improved efficacy have entered clinical trials<sup>119–121</sup> (see **Table 1.1** for a complete overview). Following the clinical success of early thalidomide analogues, IMiDs have also been successfully modified to recruit novel substrates (**Figure 1.12**). Still, even if apparent specificity is observed in given cell types, many of these compounds appear to be intrinsically polypharmacological<sup>113</sup>. For instance, growing out the isoindolinone scaffold of lenalidomide led to the pleiotropic ZF/GSPT1 degrader CC-885<sup>112</sup>, and later yielded the more GSPT1-selective compounds CC-90009<sup>122</sup> and ZXH-1-161 (**Figure 1.12b**)<sup>123</sup>. Small changes in compound structure strongly alter neosubstrate degradation specificity, with a recent report on 5-hydroxythalidomide, a thalidomide metabolite, illustrating how the presence of a single hydroxyl leads to a striking change in degradation selectivity from IKZF1 to SALL4, possibly explaining the teratogenicity of thalidomide (**Figure 1.12c**)<sup>114,116</sup>. Notably, a recent study reported the elegant, structure-guided modification of CC-885, which shows only weak affinity for IKZF2, into ALV1 and ALV2, compounds that robustly degrade IKZF2<sup>124</sup> (**Table 1.1, Figure 1.12b**). Nevertheless, it remains highly challenging to rationally tune IMiD selectivity or to predict which compound will trigger degradation of which subset of  $\beta$ -hairpin containing proteins, and hence to rationally decode the relationship between IMiD used and target degraded. Future efforts are required to map those  $\beta$ -hairpin-containing proteins in the genome that can be recruited to CRL4<sup>CRBN</sup> and investigate whether the spectrum of structural CRL4<sup>CRBN</sup> degrons can be further extended, possibly to also target non-glycine containing protein loops.

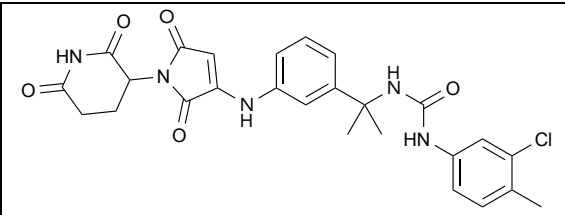
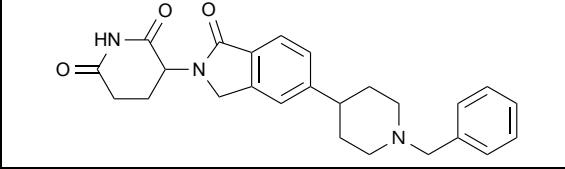
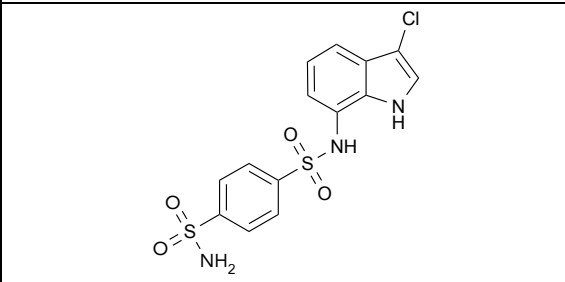
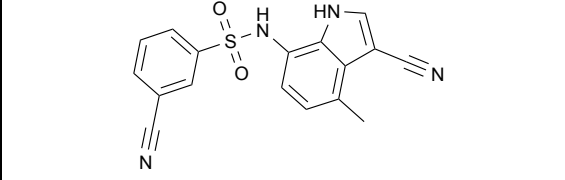
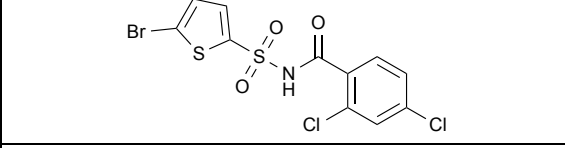
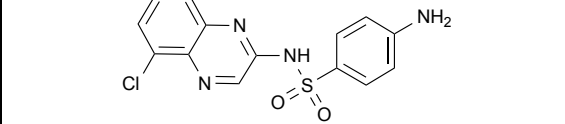


**Figure 1.12. Modifications of the CRBN surface effected by bound compounds trigger recruitment of various neosubstrates.** **a**, CRBN-bound lenalidomide (PDB 5FQD)<sup>44</sup>, pomalidomide (PDB 6UML)<sup>115</sup>, and iberdomide (PDB 5V30)<sup>125</sup>. **b**, CRBN-bound CC-885 (PDB 5HXB)<sup>112</sup>, CC-90009 (PDB 6XK9)<sup>122</sup>, and ALV1 (PDB 7LPS)<sup>124</sup>. **c**, CRBN<sup>TBD</sup>-bound thalidomide (PDB 7BQU) and 5-hydroxythalidomide (PDB 7BQV)<sup>114</sup>. **d**, CRBN-bound dBET23 PROTAC (PDB 6BN7)<sup>102</sup>. **a-c**, CRBN is shown in a surface representation with 40 % transparency. The compounds are shown as spheres (scale 1.0) and their chemical structures are given under each panel (for dBET23 the chemical structure can be found in **Figure 1.10b**). Neosubstrate degradation selectivity is indicated for each compound. TBD: thalidomide-binding domain.

**Table 1.1.** Chemical diversity and neosubstrate selectivity among known molecular glue degrader classes.

Compound	Chemical structure	Reported degradation substrates	References
<b>Molecular glues for CRL4<sup>CRBN</sup></b>			
Thalidomide		<u>IKZF1</u> , <u>IKZF3</u> , <u>ZNF692</u> , <u>ZNF276</u> , <u>SALL4</u> , <u>RNF166</u> , <u>ZBTB16</u> , FAM83F, p63	113,116,126,127
Lenalidomide		<u>IKZF1</u> , <u>IKZF3</u> , <u>ZFP91</u> , <u>ZFP692</u> , <u>ZNF276</u> , <u>ZNF653</u> , <u>ZNF827</u> , <u>SALL4</u> , <u>RNF166</u> , <u>WIZ1</u> , CK1α, FAM83F, RAB28	108,111,113,116,128,129
Pomalidomide		<u>IKZF1</u> , <u>IKZF3</u> , <u>ZFP91</u> , <u>ZFP692</u> , <u>ZNF276</u> , <u>ZNF653</u> , <u>ZNF827</u> , <u>SALL4</u> , <u>RNF166</u> , <u>GZF1</u> , <u>ZBTB39</u> , <u>ZNF98</u> , <u>WIZ1</u> , <u>ZBTB16</u> , FAM83F, RAB28, DTWD1	108,113,116–118,128,129
Avadomide (CC-122)		<u>IKZF1</u> , <u>IKZF3</u> , <u>ZFP91</u>	130
5- hydroxythalidomide		<u>SALL4</u> , <u>ZBTB16</u>	114,127
FPFT-2216		<u>IKZF1</u> , CK1α	131

Iberdomide (CC-220)		<u>IKZF1, IKZF3,</u> <u>ZFP91, ZNF98</u>	116,125
CC-647		<u>ZBTB16</u>	117
CC-3060		<u>ZBTB16,</u> <u>IKZF1, ZFP91,</u> <u>ZNF276</u>	117
CC-92480		<u>IKZF1, IKZF3</u>	121
CC-885		<u>IKZF1, IKZF3,</u> GSPT1, CK1α, PLK1, HBS1L	112,122,132
CC-90009		GSPT1	122
ZXH-1-161		GSPT1, GSPT2	123
ALV1		<u>IKZF2, IKZF1,</u> <u>IKZF3,</u> <u>ZNF653,</u> <u>ZNF692</u>	124

ALV2		<u>IKZF2</u> , <u>IKZF1</u> , <u>IKZF3</u> , <u>ZNF692</u> , <u>ZFN324</u>	124
DKY709		<u>IKZF2</u>	133
<b>Molecular glues for CRL4<sup>DCAF15</sup></b>			
Indisulam		RBM39, RBM23	134–136
E7820		RBM39, RBM23	135–137
Tasisulam		RBM39, RBM23	134,135
CQS (chloroquinoxaline sulfonamide)		RBM39, RBM23	134

Cys<sup>2</sup>-His<sup>2</sup> (C2H2) zinc finger-containing proteins are underlined. Only targets that have been shown to be degraded in the context of the full-length protein are reported.

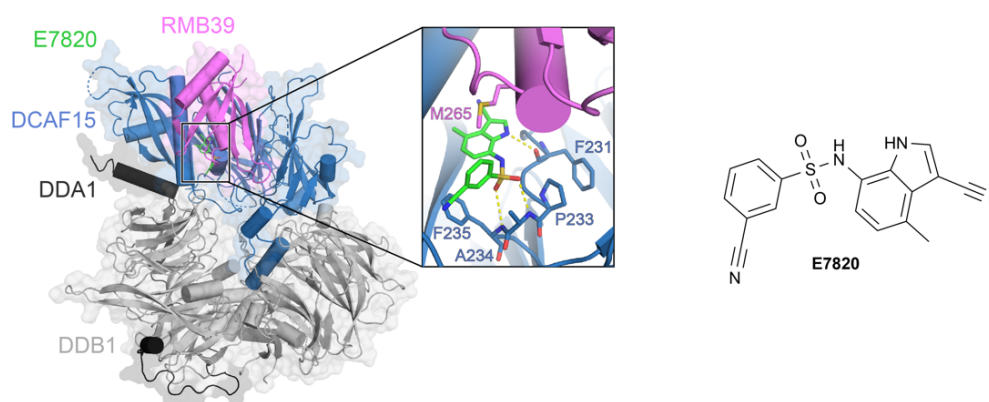
The structure activity relationship (SAR) for IMiD-induced complexes is somewhat unusual in that mutating the CRBN/neosubstrate interface or using less efficient molecular glue compounds that give rise to two- to five-fold weaker neosubstrate recruitment *in vitro* can trigger steep transitions between degradation *versus* non-degradation *in vivo* in what appears to be an all-or-nothing manner<sup>113</sup>. Relatively small changes in the binding mode appear to convert already engaged substrates into degraded targets, while at the same time rendering other tightly bound neosubstrates less stably engaged and therefore not degraded<sup>113</sup>.

As described in **section 1.1.2**, CRL activity is regulated by the CSN in a manner whereby ligases that are engaged with substrates remain neddylylated, while idle CRLs are de-neddylylated<sup>138</sup>. Thalidomide-induced substrate complexes have been shown to be sufficiently substrate-like to maintain the ligase in a neddylylated and active state<sup>138</sup>. All glue-induced neosubstrates for CRLs must pass this substrate criterion, and it is unclear to

what extent very small domains or linear peptides are sufficiently good mimics to support the ubiquitination process<sup>38,138,139</sup>. Analogously, other components of the ubiquitin-proteasome system such as p97, an ATPase involved in ubiquitinated substrate extraction<sup>140</sup>, and the proteasome also require the neosubstrate to be recognized as sufficiently substrate-like to enable efficient degradation<sup>100</sup>.

### 1.1.11 Molecular glues for CRL4<sup>DCAF15</sup>

Following the discovery of thalidomide analogues and their mechanism of action, aryl sulphonamides such as indisulam and E7820 were found to drive degradation of the essential RNA-binding protein RBM39 (RNA Binding Motif Protein 39) and RBM23 by the CRL4<sup>DCAF15</sup> E3 ubiquitin ligase<sup>134,136,141</sup>. Structural studies of DCAF15-DDB1-DDA1 in complex with RBM39 and indisulam or E7820 found that aryl sulphonamides functionalise a shallow, non-conserved cavity on DCAF15 to selectively recruit the splicing factors through an  $\alpha$ -helical degron motif on the RRM2 domain<sup>42,137,142</sup> (**Figure 1.13**). Both  $\alpha$ -helices and  $\beta$ -hairpins, the two elemental units of protein folds, are thus principally readable by protein-ligand interfaces<sup>143</sup>.



**Figure 1.13.** DDB1-DDA1-DCAF15-E7820-RBM39 complex (PDB 6PAI)<sup>42</sup>. Important interface residues are shown as sticks. Key interactions are marked with dashed yellow lines. The protein-protein interface spans 995 Å<sup>2</sup>, while the gluc-RBM39 interface spans 146 Å<sup>2</sup>.

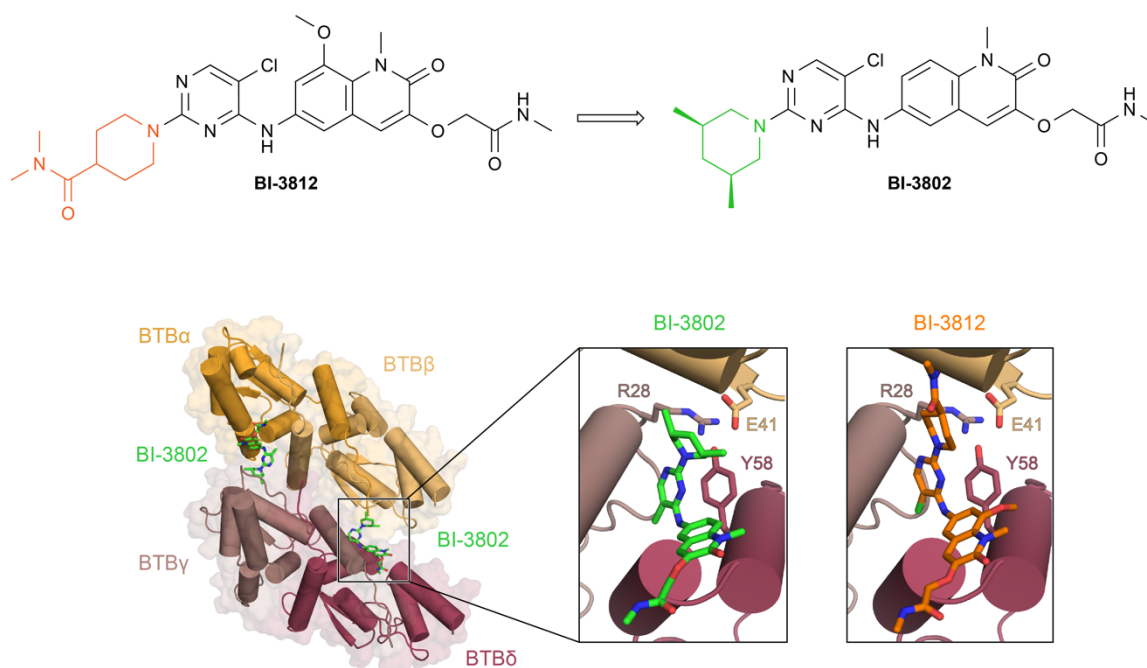
All RBM39 degraders reported thus far contain the aryl sulphonamide moiety, with the sulfonyl oxygens hydrogen-bonding with backbone nitrogen atoms of DCAF15 and the sulfonyl nitrogen interacting with RBM39 side chains via water-mediated hydrogen bonds (**Table 1.1**). In contrast to IMiDs being high-affinity CRBN binders, aryl sulphonamides have only a low affinity for the isolated DCAF15 receptor ( $\sim 3 \mu\text{M}$  for E7820,  $>50 \mu\text{M}$  for indisulam)<sup>137</sup>. These findings demonstrate that molecular glue degraders do not critically depend on high-affinity interactions, and this conceptually paves the way for pursuing traditionally non-ligandable ligases for targeted protein degradation. The low binary affinity of aryl sulphonamides to DCAF15 is mitigated by an extensive network of hydrophobic DCAF15-RBM39 interactions, yielding a high-affinity ternary complex. The DCAF15 interface with RBM39/23 is not highly conserved, suggesting a drug-induced neomorphic binding event<sup>137</sup>.



In the case of IMiDs, a portion of the CRBN-bound glue molecule protrudes into the solvent allowing for recruitment of diverse proteins possessing the  $\beta$ -hairpin structural degron. In contrast, sulphonamides are buried within a shallow groove of DCAF15 and appear to only recruit neosubstrates that exhibit a particular geometry. Thus far, the buried nature of the compound and the extensive protein-protein interaction interface have made it difficult to tune the degradation selectivity by modifying the compound to direct the CRL<sup>DCAF15</sup> ligase to other RRM2 proteins or beyond. The enclosed nature of the cavity further makes it challenging to turn aryl sulphonamides into PROTAC warheads, although a PROTAC thought to engage DCAF15 was recently reported<sup>144</sup>.

### 1.1.12 BCL6 degraders

BI-3802 is a small molecule that binds the BTB domain of the oncogenic TF BCL6 and results in its proteasomal degradation<sup>145</sup>. A recent report, concomitant with the work presented herein, revealed that the compound induces polymerization, sequestration into cellular foci, and subsequent degradation of BCL6<sup>146</sup>. The TF was reported to be a substrate for the E3 ubiquitin ligase SIAH1, with the BI-3802-induced formation of BCL6 filaments strongly facilitating BCL6 ubiquitination by SIAH1 and its subsequent degradation. Cryo-EM studies showed that the solvent-exposed dimethylpiperidine moiety of BI-3802 drives polymerisation through direct hydrophobic interactions with distal amino acids of an adjacent BCL6 dimer (**Figure 1.14**). Further modelling explained the lack of polymerisation with BI-3812, a closely related *bona fide* BCL6 inhibitor, by revealing a steric clash between the extended carboxamide group and residues of the adjacent BTB dimer. BI-3802 is thus an example of a target-binding compound that, upon modification of its solvent-facing moiety, shows gain-of-function molecular glue activity leading to enhanced target inactivation. It will be important to investigate whether other BTB domains as well as other protein folds share a small molecule-induced polymerisation-dependent mechanism of degradation and to uncover what governs these reprogramming processes.



**Figure 1.14.** Structure of two BTB dimers brought together by BI-3802<sup>146</sup>. Left close-up panel: BI-3812 at the dimer interface. BCL6 residues R28 (BTB $\gamma$ ), E41 (BTB $\beta$ ), Y58 (BTB $\delta$ ) are shown as sticks. Right close-up panel: model of BI-3812 at the dimer interface. The inhibitor was docked to the crystal structure of BCL6 BTB (PDB 5MW2) and was modelled at the interface by superposition of BTB domains. The solvent-exposed moiety of BI-3812 clashes with a helix of BTB $\beta$ .

### 1.1.13 Other molecular glue degraders

The preceding sections focused on well-studied molecular glue degrader examples, for which structural information exists. However, multiple other compounds have been proposed to function through a similar mechanism, and a selected few will be briefly discussed below.

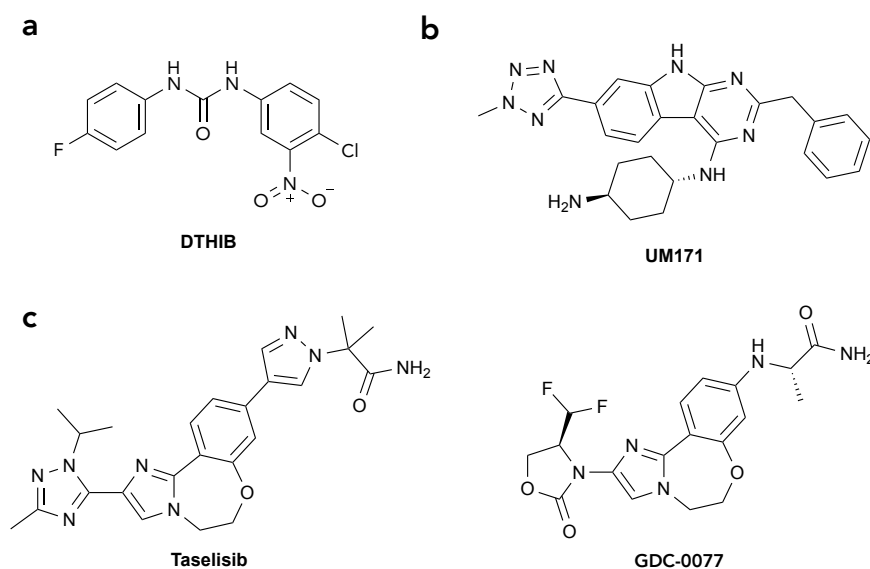
A compound postulated to degrade the transcription factor heat shock factor 1 (HSF1) has recently been described<sup>147</sup>. While DTHIB (**Figure 1.15a**) was originally developed as an inhibitor, the study found that it selectively degraded the nuclear fraction of HSF1 in a proteasome-dependent manner, likely by strengthening the interaction with its endogenous ubiquitin ligase FBXW7.

UM171 (**Figure 1.15b**) is a compound promoting the expansion of haematopoietic stem cells (HSCs) through a hitherto undefined mechanism, and UM171-expanded HSCs are currently in clinical trials for haematological disorders<sup>148</sup>. Recently, it has been proposed that UM171 causes the polyubiquitination and proteasomal degradation of RCOR1, LSD1, and HDAC2, all of which are components of the chromatin remodelling complex CoREST<sup>149,150</sup>. A subsequent study postulated that this degradation is dependent on the CUL3<sup>KBTBD4</sup> ligase and showed that through CoREST degradation, UM171 can re-establish H3K4me2 and H3K27ac marks in HSCs *ex vivo*, leading to prolonged cell self-renewal<sup>149</sup>.

GDC-0077 and tasisib (**Figure 1.15c**) are inhibitors of the PI3K catalytic subunit p110 $\alpha$ . A recent study from Genentech demonstrated that these compounds cause selective degradation of mutant p110 $\alpha$  while

sparing the wild-type protein<sup>151</sup>. The degradation was dependent on the activity of the proteasome and E1 enzyme, as well as on the receptor tyrosine kinase-mediated recruitment of mutant p110 $\alpha$  to the membrane<sup>151</sup>.

For all the putative molecular glue degraders described above, more research will be needed to ascertain their precise mode of action. Biochemical reconstitutions and, ultimately, structural studies, will serve to verify whether they function through promoting direct target-ligase association.



**Figure 1.15.** Examples of compounds postulated to work as molecular glue degraders.

Additionally, a compound (structure and name pending disclosure) that promotes complex formation between VHL and the cysteine dioxygenase CDO1 has recently been identified by Novartis using protein microarrays<sup>152</sup>. It is unclear whether degradation of CDO1 can confer any therapeutic benefit, but this discovery illustrates that molecular glue degraders are likely to exist for many E3 ligases beyond the most widely exploited example, CRBN.

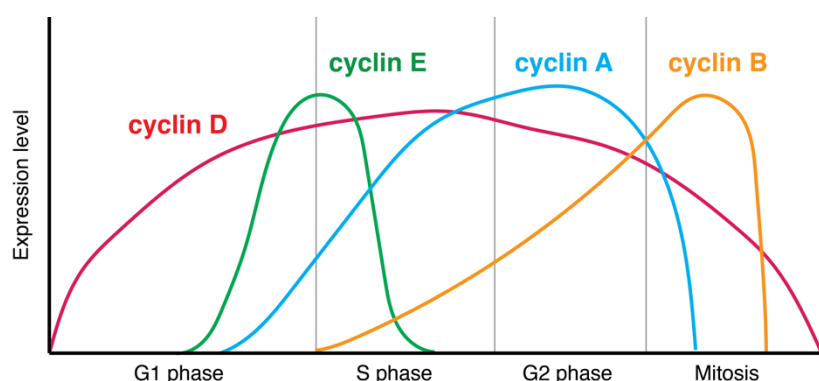
## 1.2 Cyclin-dependent kinases

### 1.2.1 The CDK family

Cyclin-dependent kinases (CDKs) are a family of evolutionarily conserved serine-threonine kinases involved in regulating many essential cellular processes. The human genome encodes 21 CDKs, which can be broadly divided into kinases that regulate cell cycle progression (*e.g.* CDK1-3, CDK4/6), and transcriptional regulators (CDK7-13, CDK19-20)<sup>153,154</sup>.

All CDKs are dependent on association with regulatory subunits, typically their cognate cyclin, for activity. The expression of cyclins that heterodimerise with cell-cycle CDKs (mainly cyclins A, B, D, E) markedly varies across the cell cycle (**Figure 1.16**), and these changes constitute a key driver of cell cycle progression.

Upon mitogenic signals, CDK4 and CDK6 promote entry into the cell cycle<sup>155</sup>. The CDK4/6-cyclin D complexes orchestrate the phosphorylation of multiple substrates, most notably of the tumour suppressor retinoblastoma protein (Rb), which unleashes gene expression governed by the E2F TFs, thereby initiating the transcription of cyclins A and E<sup>156,157</sup>. Cyclin E subsequently activates CDK2, contributing to the initiation of DNA replication<sup>158</sup>. Then, CDK1 associates with cyclin A to enact target phosphorylation in G2 phase, and the CDK1-cyclin B complex facilitates the progression into mitosis<sup>158</sup>. Degradation of cyclin A is finally required for anaphase progression and the generation of two daughter cells<sup>159</sup>. Various checkpoints can inhibit cell cycle CDK activity and thus halt the cycle progression, for instance through increased expression of CDK inhibitors such as p27 or Ink4<sup>158</sup>.

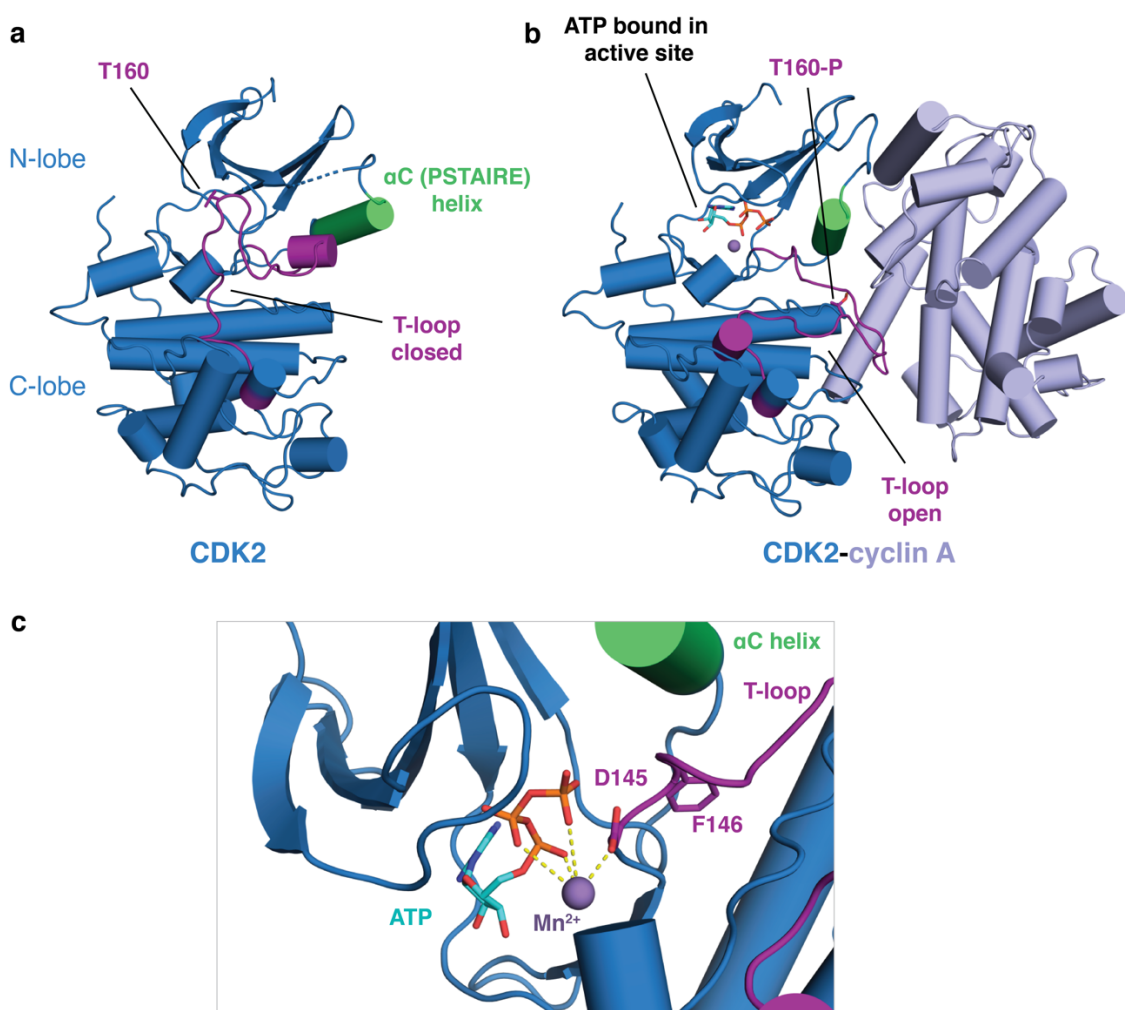


**Figure 1.16.** Expression levels of cyclins associated with CDK1, CDK2, and CDK4/6 vary across the cell cycle. The figure is based on <sup>160</sup>.

Beyond its paramount role in cell cycle progression, the CDK family is involved in transcriptional regulation, primarily through the phosphorylation of the C-terminal domain (CTD) of RNA polymerase II (RNAPII)<sup>161</sup>. RNAPII activity across the distinct stages of transcription (initiation, elongation, and termination) is dependent on the phosphorylation state of serine residues in the heptad repeats (YSPTSPS) in its CTD<sup>162</sup>. Phosphorylation of Ser5 (serine at position 5) by CDK7 is required for the initiation of transcription, while Ser2 phosphorylation, enacted by CDK9, CDK12, and CDK13, supports elongation<sup>163,164</sup>. The expression of cyclins that associate with transcriptional CDKs is constant throughout the cell cycle, but the name persists due to their structural similarity to cyclins A, B, E, and D.

CDKs assume a bilobal structure common to all kinases. The kinase domain comprises an N-lobe (containing a  $\beta$ -sheet and an  $\alpha$ C helix), a C-lobe (consisting of multiple  $\alpha$ -helices), and a hinge region (**Figure 1.17a**)<sup>165</sup>. The hinge, located between the two lobes, contains several conserved residues essential for ATP binding and catalytic activity. In the absence of a cyclin, the CDK remains in an inactive conformation, whereby the T-loop (also known as the activation loop) occludes the active site (**Figure 1.17a**)<sup>166</sup>. Upon cyclin association, the T-loop conformation changes from closed to open, allowing for substrate binding and ATP hydrolysis (**Figure 1.17b**)<sup>167</sup>. While for most kinases such a conformational change is associated with a so-called DFG flip, this has not been observed for most CDKs so far. The conserved DFG (Asp-Phe-Gly) motif, a

principal regulatory structure that modulates kinase activity, is involved in ATP binding, with the aspartic acid residue (D of the DFG) able to interact with  $Mg^{2+}$  ions coordinated to ATP only in the DFG-in (aspartate pointing into the active site) conformation (**Figure 1.17c**). As the DFG motif is located at the N-terminal end of the T-loop, the movement of the loop typically corresponds to a change from a DFG-out (inactive) to a DFG-in (active) conformation. However, for CDKs, the DFG-out conformation was postulated to be inaccessible, with a crystal structure of monomeric CDK2 revealing a DFG-in conformation with the T-loop in a closed conformation (**Figure 1.17a**)<sup>165,168</sup>. For full activity, CDKs additionally require phosphorylation of a threonine residue in their T-loop, which can be accomplished by the CAK complex (CDK7-MAT1-cyclin H) or through autophosphorylation in the case of some CDKs. The T-loop phosphorylation stabilises the CDK-cyclin complex and can also stabilise a CDK-bound substrate (**Figure 1.17b**)<sup>169</sup>.

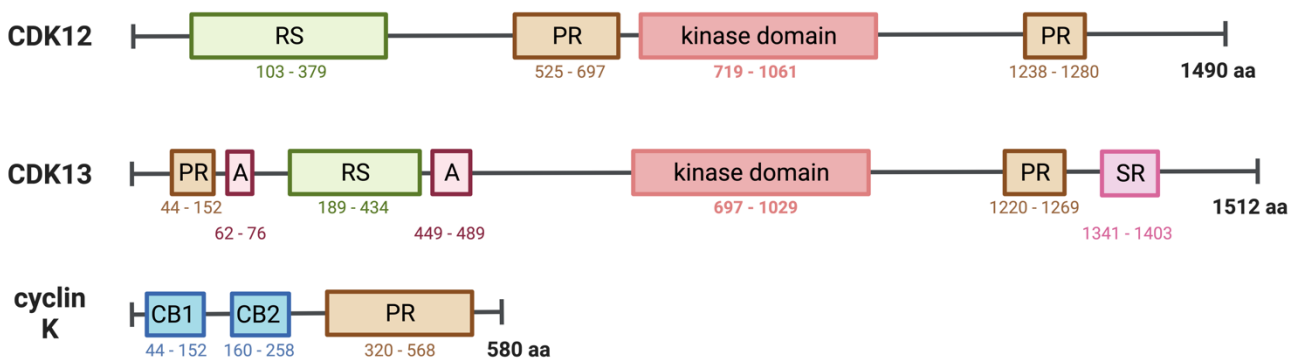


**Figure 1.17. CDK structure and activation by cyclin binding.** **a**, Crystal structure of inactive, monomeric CDK2 (PDB 1HCL)<sup>168</sup>. In the cyclin-free form, the  $\alpha$ C (PSTAIRE) helix and the T-loop occlude the active site. **b**, Active CDK2-cyclin A structure (PDB 1JST)<sup>169</sup>. Upon cyclin binding, the T-loop flips and the  $\alpha$ C helix forms a part of the CDK-cyclin interface, rendering the CDK active site accessible. This conformation is stabilised by T-loop phosphorylation (T160-P). **c**, Zoomed-in view on the active site of the structure presented in (b). This illustrates the DFG-in conformation, and the aspartic acid and phenylalanine residues of the DFG motif are shown as sticks.

## 1.2.2 The roles of CDK12/13-cyclin K

CDK12 (also known as CRK7 and CrkRS) and CDK13 (CDC2L5, CHED) are paralogues that both pair with cyclin K to phosphorylate Ser2 in the CTD of RNAPII and therefore regulate transcription<sup>170</sup>. CDK12 and CDK13 are unusually large for CDKs (1490 and 1512 aa, respectively) and, in addition to a central kinase domain, they feature additional arginine/serine (RS), proline-rich (PR), serine-rich (SR; CDK13 only) and alanine-rich (A; CDK13 only) motifs thought to function as protein-protein interaction domains (**Figure 1.18**)<sup>171</sup>. CDK12 and CDK13 display 43% sequence identity overall, with the kinase domain being 93% identical<sup>172</sup>. While CDK12 is widely and highly expressed<sup>173</sup>, the data for CDK13 is less clear<sup>174</sup>. The two kinases have partially overlapping roles, and this redundancy likely explains why individually CDK12/13 are not pan-essential genes (essential in 605/1086 (CDK12) and 66/1086 (CDK13) cancer cell lines according to the DepMap database)<sup>175</sup>.

Cyclin K is the obligate cognate cyclin of CDK12/13. However, it may also exert additional independent functions as discussed below. CCNK (cyclin K) is a common essential gene, with the knockout shown to be lethal in 1085 out of the 1086 tested cancer cell lines<sup>175</sup>. Cyclin K knockout was also demonstrated to have an embryonic lethal phenotype in mice<sup>176</sup>. The level of cyclin K expression is low in non-proliferative tissues and correlates with proliferation<sup>174,177</sup>, with one study showing that it is expressed in mammalian testes in a developmentally regulated manner<sup>177</sup>. Another report found cyclin K to be essential for embryonic stem cell maintenance, with its knockout inducing differentiation<sup>178</sup>. Cyclin K was also shown to have a long half-life and, unlike other cyclins, its abundance is not regulated by proteasomal degradation (*c.f.* cyclin A and APC/C or cyclin D and APC/C, AMBRA1, Fbxo4, Fbxo31,  $\beta$ -TrCP)<sup>174,179</sup>.



**Figure 1.18. Schematic diagram of the domain organisation of CDK12, CDK13 and cyclin K.** RS: arginine/serine-rich motif, PR: proline-rich motif, A: alanine-rich motif, SR: serine-rich motif, CB: cyclin box. RS are frequently found in proteins that regulate mRNA splicing<sup>180</sup>. PRs commonly exist in proteins involved in signal transduction<sup>181</sup>.

While in invertebrates only CDK12 is present, vertebrates possess the paralogues CDK12 and CDK13, which have both distinct and overlapping roles<sup>171</sup>. Beyond RNAPII phosphorylation, CDK12 and CDK13 are involved in controlling co-transcriptional processes, such as mRNA splicing or 3'-end RNA processing, and these roles are briefly discussed below.

Genetic depletion of CDK12 or CDK13 showed that both CDKs regulate the expression of *ca.* 1000 overlapping genes (including those implicated in RNA processing), whilst each kinase additionally regulates a distinct subset<sup>176,182</sup>. CDK13 depletion, but not CDK12 depletion, results in decreased expression of genes encoding proteins that regulate protein translation, while CDK12 loss but not CDK13 loss is associated with the decreased expression of DNA damage response (DDR) genes<sup>182</sup>. It has been shown that upon CDK12 loss or inhibition, transcription shows length-dependent elongation defects, defaulting to premature cleavage or polyadenylation<sup>183</sup>. This is thought to principally affect the transcription of DDR genes, which are typically long and contain many exons<sup>183</sup>.

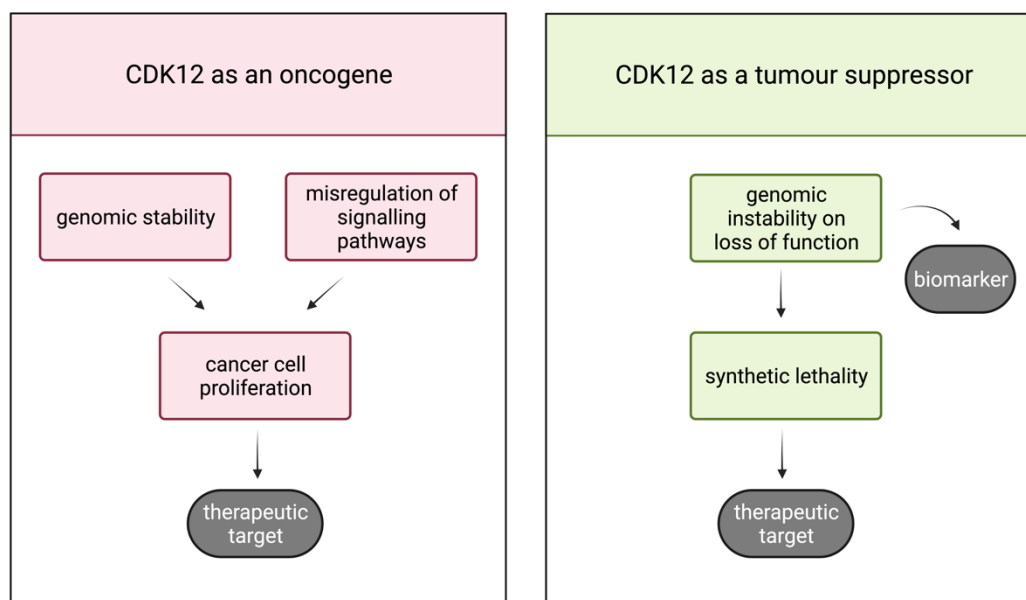
CDK12/13 have additional important roles in RNA processing. On one hand, these kinases act indirectly through phosphorylation of RNAPII, which in turn recruits processing factors (*e.g.* CsfF-64 and CstF-77)<sup>184,185</sup>. Alternatively, CDK12/13 can modulate RNA processing via physical interactions with other factors through the RS domain<sup>182,184-186</sup>. CDK12 was found to co-localise with the splicing factor SC35 (also known as SRSF2<sup>187</sup>), and to bind in exon junction complexes with other splicing factors<sup>182,186</sup>. Moreover, CDK12 also regulates the translation of several mRNAs<sup>188</sup>. Accordingly, CDK12 loss, but also the truncation of its N-terminal RS domain, was shown to cause mRNA splicing and 3'-end processing defects<sup>185</sup>.

Interestingly, a recent study demonstrated that CDK12-cyclin K mediates phosphorylation of cyclin E to block its interaction with CDK2 during pre-replication complex assembly<sup>174</sup>. CDK12-cyclin K thereby indirectly control cell cycle progression by interfering with the formation of the CDK2-cyclin E complex. Moreover, another study found that CDK12 regulates the ability of RNAPII to synthesise core DNA replication genes (such as CDT1, CDC6 or TOPBP1) and therefore controls G1/S progression<sup>189</sup>. Hence, the catalytic activity of CDK12-cyclin K provides a distinct link between the regulation of transcription and cell cycle progression.

### 1.2.3 CDK12-cyclin K as a biomarker and therapeutic target

In cancer, CDK12 has the highest alteration frequency among all CDK<sup>190</sup>. Reported abnormalities include mutations, rearrangements, deletion, amplification, and fusions and have been found in multiple malignancies, *e.g.*, ovarian carcinoma, metastatic lung adenocarcinoma, gastric, breast, cervical, prostate, and colon cancer<sup>170,191-199</sup>. Interestingly, CDK12 can exhibit both tumour-suppressive and oncogenic properties (**Figure 1.19**). Inactivation of CDK12 (*e.g.* through loss-of-function mutations) leads to genomic instability as downregulation of DDR genes causes impairment of homologous recombination (HR) and appearance of focal tandem duplications (FTDs)<sup>200</sup>. Conversely, CDK12 overexpression results in increased expression of oncogenes, given that CDK12 is an important regulator of multiple signalling pathways that govern cancer cell proliferation<sup>201</sup>. For example, the kinase has been shown to regulate the expression of the important cancer driver c-Myc<sup>201</sup> as well as to modulate NF- $\kappa$ B activity via controlling the expression of NIK (MAP3K14) and p100 (NF $\kappa$ B2), key activators of the non-canonical NF- $\kappa$ B pathway<sup>202</sup>. CDK12 was also reported to activate the ErbB-PI3K-AKT and Wnt/ $\beta$ -catenin signalling pathways via RNAPII phosphorylation, allowing breast cancer

cells to maintain stemness and develop resistance to anti-HER2 therapy<sup>196</sup>. Activation of the Wnt pathway could be ascribed to CDK12 promoting the synthesis of Wnt ligands, with subsequent pathway activation causing upregulation of  $\beta$ -catenin, c-Myc, and cyclin D1, as well as to CDK12 inactivating DNAJB6, a protein involved in  $\beta$ -catenin degradation, through alternative last exon splicing<sup>196,203</sup>. Additionally, CDK12 was postulated to directly interact with and phosphorylate PAK2, an oncogene in gastric cancer, and thereby activate the MAPK pathway<sup>199</sup>.



**Figure 1.19.** CDK12 has both oncogenic and tumour-suppressive properties in cancer. Created with BioRender.com

As CDK12 loss leads to DDR deficiency, CDK12 is a potential therapeutic target for cancers with HR pathway defects. However, it can also be exploited through various synthetic lethality approaches. For example, CDK12 inhibition sensitizes cells to PARP inhibition and cisplatin<sup>204,205</sup> and patients with CDK12 mutations were shown to respond better to platinum-based chemotherapy<sup>204</sup>. CDK12 mutations are also a predictor of sensitivity to CHK1 inhibitors<sup>206</sup>. Recent studies show that the presence of FTDs and the resulting gene fusions increases neoantigen load in ovarian and prostate cancers, making CDK12 a possible biomarker of sensitivity to immune checkpoint inhibitors, such as programmed cell death-1 (PD-1) inhibitors, in these malignancies<sup>200,207</sup>. Inhibition of transcriptional CDKs was also postulated to be beneficial for cancers that are strongly dependent on transcriptional programs for oncogenesis, *e.g.* tumours that rely on aberrant transcription factor expression. Accordingly, CDK12 inhibition showed effectiveness in MYCN-dependent neuroblastoma<sup>208</sup> and knockouts of CDK12 or cyclin K showed synthetic lethality with c-Myc overexpression<sup>209</sup>. Moreover, CDK12-targeting compounds effectively inhibited EWS/FLI-dependent DDR in Ewing sarcoma, demonstrating synthetic lethal effects<sup>210</sup>. Several of these approaches are currently being explored in preclinical and clinical studies, as described in **section 1.2.9**.

Beyond oncology, CDK12 was identified as a therapeutic target in myotonic dystrophy DM1, a disease characterised by accumulation of the wrongly transcribed dystrophin myotonic protein kinase<sup>211</sup>. A recent report



described a significant reduction in nuclear foci formation upon a pan-CDK inhibitor treatment<sup>212</sup>. The study identified CDK12 as the relevant target, with CDK12 levels also found to be elevated in muscle biopsies from DM1 patients<sup>212</sup>.

Cyclin K has been reported to be frequently overexpressed in cancer cells<sup>174,213</sup>. While the cyclin is not considered to be a classically druggable protein due to the lack of ligand-binding cavities, it can be indirectly targeted via the inhibition of its cognate kinases. However, beyond its roles alongside CDK12 and CDK13, cyclin K may exert additional functions independently or through interactions with other kinases such as CDK9<sup>214</sup>. Several examples of the latest discoveries on the roles and interactions of cyclin K are listed below.

Recently, CCNK emerged as a gene implicated in a novel neurodevelopmental disorder<sup>215</sup>. The syndrome, involving facial dysmorphism and intellectual and motor impairment, was attributed to heterozygous deletions or mutations in CCNK. In one patient, the mutated residue (K111E) was located at the cyclin K-CDK interaction interface, indicating that the syndrome could result at least partly from the impairment of CDK12/13 activity<sup>215</sup>. Moreover, a study by Hoshii *et al.* described the interaction of cyclin K with SETD1A, a histone methyltransferase, and SET1D was postulated to recruit cyclin K to chromatin to effect the expression of DNA damage response genes<sup>216</sup>. Notably, loss-of-function mutations in SETD1A and CCNK deficiency have overlapping phenotypes characterised by intellectual disability<sup>217</sup>.

A recent publication also described a novel role for cyclin K in lung cancer and reported that cyclin K is often overexpressed and relates to poor prognosis in patients<sup>213</sup>. The study identified a direct protein-protein interaction between cyclin K and  $\beta$ -catenin that leads to  $\beta$ -catenin stabilisation and subsequent induction of cyclin D1 expression, which in turn facilitates tumourigenesis<sup>213</sup>. The authors postulate this as the molecular mechanism of radiotherapy resistance in lung cancer, highlighting cyclin K as a potential biomarker in lung cancer therapy.

The complex roles of CDK12-cyclin K, partly overlapping and partly distinct from other CDKs, highlight the importance of the development of selective inhibitors for their effective modulation. It also remains to be determined whether the direct therapeutic targeting of cyclin K over CDK12/13 inhibition, should it become possible, offers any added therapeutic benefit.

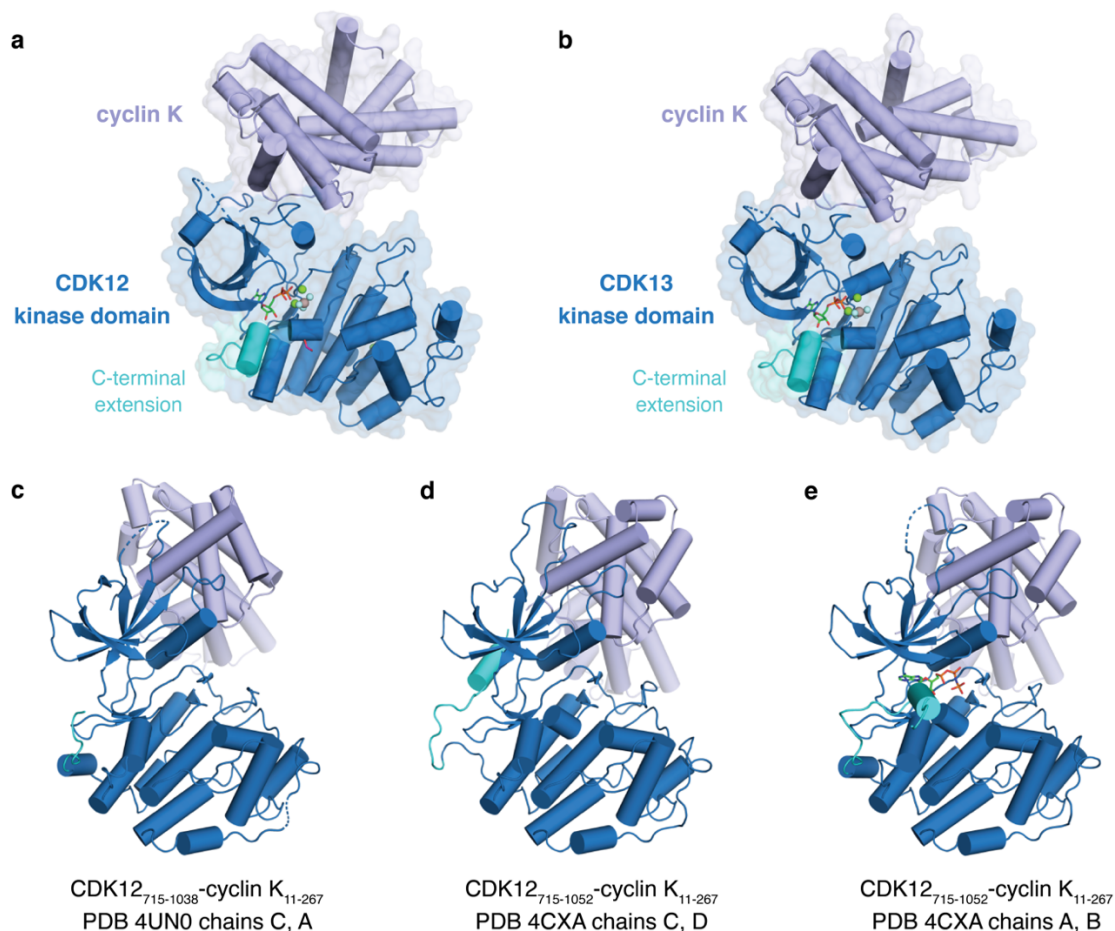
#### 1.2.4 Structural insights into CDK12/13-cyclin K

To date, several structures of CDK12-cyclin K and CDK13-cyclin K in their apo or ligand-bound form have been solved by X-ray crystallography (Table 1.2; Figure 1.20a, b). Structural studies, all utilising solely the kinase domain of the CDKs, revealed that they assume the same overall heterodimeric arrangement as other CDK-cyclin complexes (*c.f.* Figure 1.17b). One distinctive feature of CDK12/13 is an extension of the kinase domain towards the C-terminus (amino acids (aa) 1027-1052 in CDK12, referred to as the *C-terminal tail* or *C-terminal extension* in the literature and herein, although it is an internal fragment when considered in terms of the *ca.* 1500 aa full-length protein). Importantly, in CDK12 this C-terminal tail displays significant conformational variability among the solved structures, ranging from highly mobile and therefore absent from

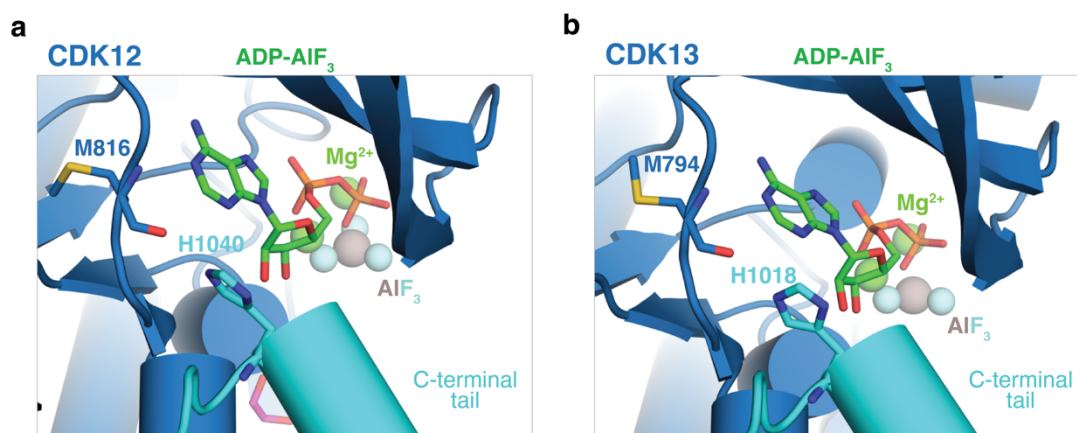
the crystal structure (**Figure 1.20c**), to wrapping around the kinase domain (**Figure 1.20d**), to enclosing the active site and directly contacting the bound ligand (**Figure 1.20e**)<sup>218,219</sup>. In the lattermost case, a histidine (H1040) residue in the CDK12 C-terminal extension was found to interact with the nucleotide cofactor through hydrogen bonding interactions, suggesting the tail could function to stabilise the bound ligand (**Figure 1.21a, b, e**)<sup>219</sup>. Accordingly, truncation of the C-terminal extension led to the loss of kinase activity, with truncation from 1082 to 1063 only minimally affecting catalytic activity, while restricting the construct further to 1044 led to a four-fold loss of activity<sup>218,219</sup>.

**Table 1.2.** Published structures of CDK12, CDK13, or cyclin K.

Proteins	Ligand	PDB code and reference	Figure panel
CDK12-cyclin K	-	4UN0 <sup>219</sup>	Figure 1.20c
	AMP-PNP	4CXA <sup>219</sup>	Figure 1.20d, e
	ADP-AIF <sub>3</sub>	4NST <sup>218</sup>	Figure 1.20a, Figure 1.21a
	THZ531	5ACB <sup>220</sup>	Figure 1.23b
	“Compound 5”	6B3E <sup>221</sup>	Figure 1.24b
	“Compound 29”	6CKX <sup>222</sup>	Figure 1.24d
	BSJ-01-175	7NXX <sup>223</sup>	Figure 1.23d
CDK13-cyclin K	ADP-AIF <sub>3</sub>	5EFQ <sup>172</sup>	Figure 1.20b, Figure 1.21b
	THZ531	7NXJ <sup>223</sup>	Figure 1.23c
Cyclin K	-	2I53 <sup>224</sup>	-



**Figure 1.20. Overview of CDK12/13-cyclin K structures and the conformational flexibility of the CDK12 C-terminal extension.** **a**, Crystal structure of the human CDK12-cyclin K complex bound to ADP-AIF<sub>3</sub> (PDB 4NST)<sup>218</sup>. **b**, Crystal structure of human CDK13-cyclin K complex bound to ADP-AIF<sub>3</sub> (PDB 5EFQ)<sup>172</sup>. **c**, Crystal structure of the human apo CDK12-cyclin K complex with the C-terminal extension not observed (PDB 4UN0)<sup>219</sup>. **d**, CDK12-cyclin K complex expected to be bound to AMP-PNP but crystallised in the apo state, with the C-terminal tail wrapping around the kinase domain (PDB 4CXA, chains C and D)<sup>219</sup>. **e**, CDK12-cyclin K complex bound to AMP-PNP, with the C-terminal tail enclosing the active site and contacting the ligand (PDB 4CXA, chains A and B)<sup>219</sup>.

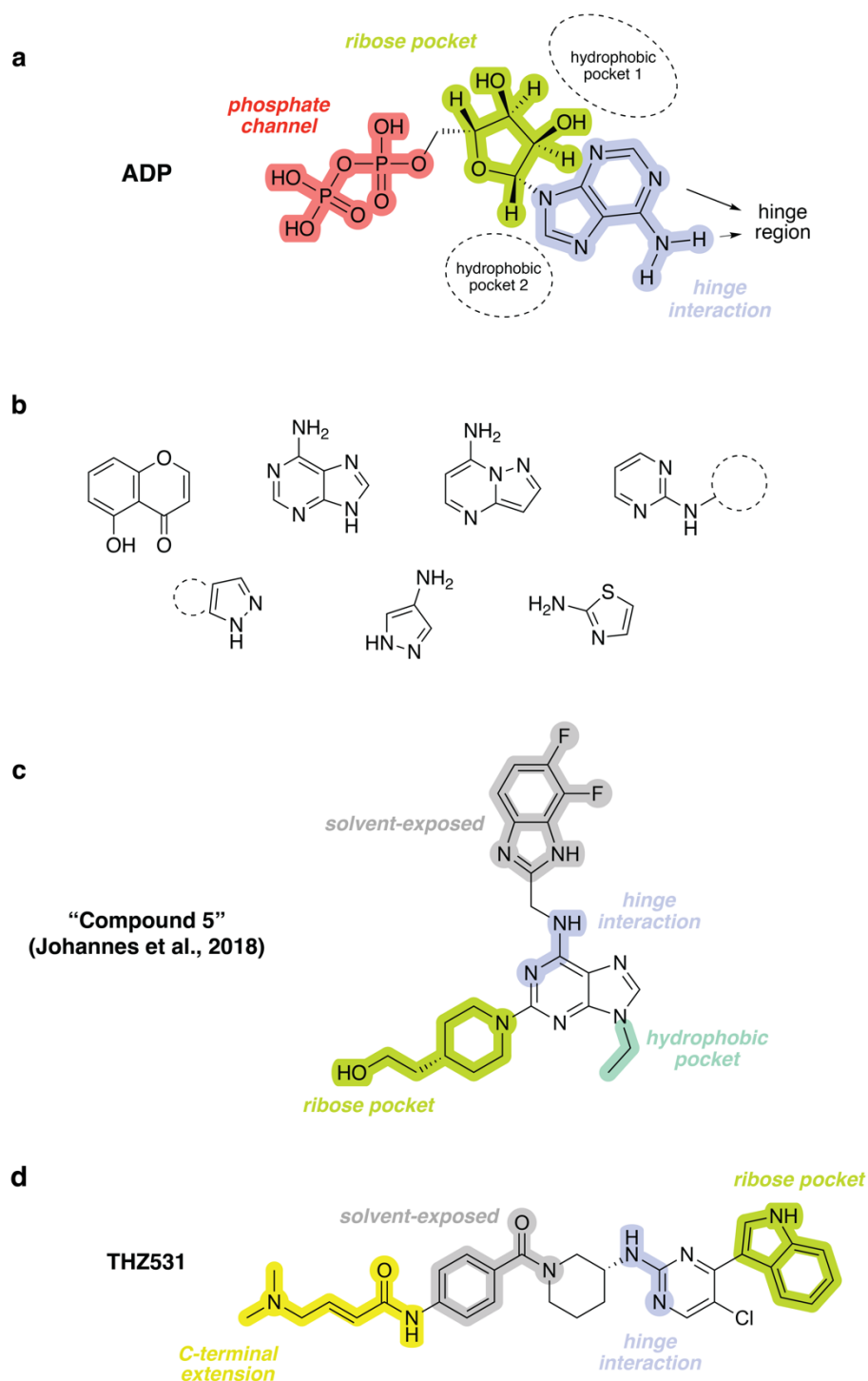


**Figure 1.21. Binding mode of endogenous ligands to CDK12/13.** **a**, Crystal structure of human CDK12-cyclin K complex bound to ADP-AIF<sub>3</sub> (PDB 4NST)<sup>218</sup>. **b**, Crystal structure of human CDK13-cyclin K complex (PDB 5EFQ)<sup>172</sup>.

### 1.2.5 Therapeutic targeting of CDK12/13-cyclin K

Therapeutic targeting of kinases has been hindered by the fact that the kinase domain fold and the geometry of the active site are highly similar among this large protein superfamily<sup>225</sup>. Within the CDK family itself, the development of selective inhibitors has been particularly challenging due to the high conservation of key residues and severe off-target effects associated with spurious inactivation of cell-cycle CDKs. The careful fine-tuning of all substituents is required in each case to impart selectivity towards a particular CDK, a process aided considerably by the increasing availability of structural data<sup>226</sup>. Thus far, the only FDA-approved CDK-targeting drugs are four CDK4/6 inhibitors: palbociclib (*Ibrance*, Pfizer), ribociclib (*Kisqali*, Novartis), abemaciclib (*Verzenio*, Eli Lilly) and trilaciclib (*Cosela*, G1 Therapeutics). Their clinical progression as cancer therapeutics was facilitated by the clear understanding of the paramount role of CDK4/6 in the G1 cell cycle arrest and by the fact that dual selective inhibitors could be obtained. Multiple other CDK inhibitors, many of which are multi- or pan-selective, are currently in clinical trials<sup>226,227</sup>.

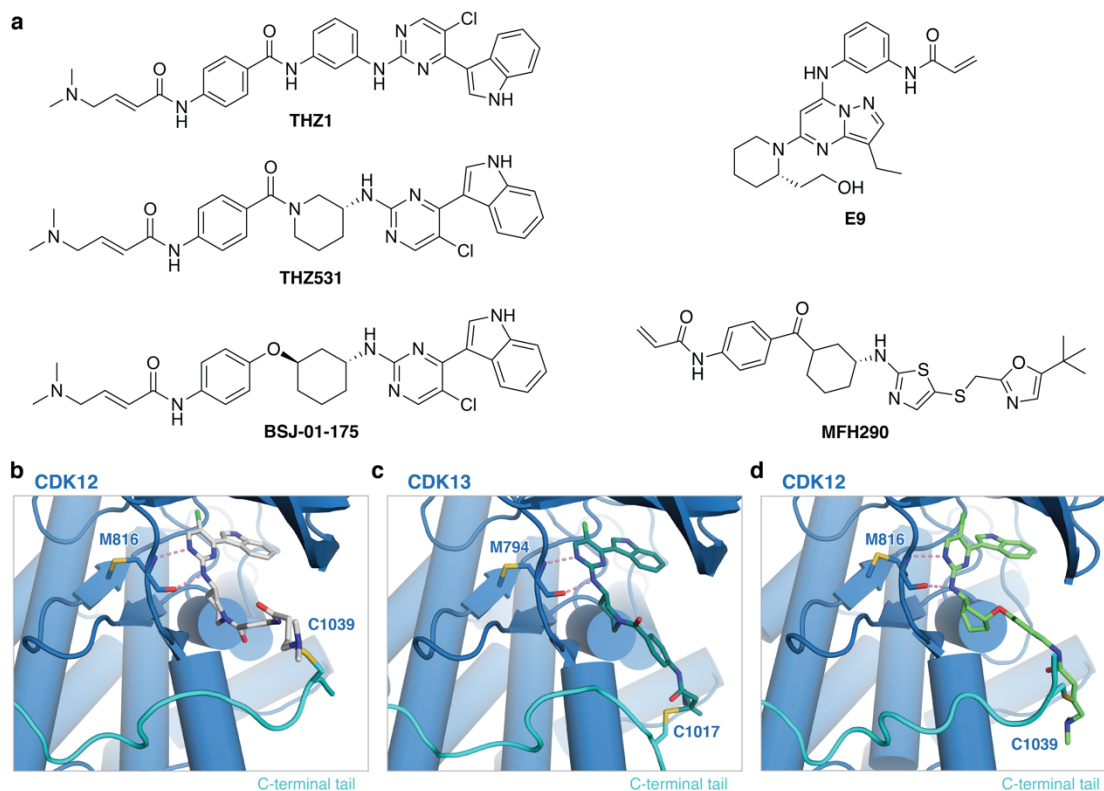
The majority of CDK inhibitors are the so-called type I, ATP-competitive inhibitors. Such compounds bear moieties able to form two critical hydrogen bonding interactions with the hinge region of the kinase akin to endogenous ligands (**Figure 1.22a**), such as *e.g.* pyrimidin-2-amine, 1*H*-pyrazole, or pyrazolo[1,5-*a*]pyrimidin-7-amine (**Figure 1.22b**). Inhibitors based on these scaffolds partly mimic the binding mode of nucleotide cofactors (**Figure 1.22c, d**). Compounds that act via a covalent binding mechanism have also been developed to enhance inhibition selectivity (**Figure 1.22d**), as the biggest challenge in the therapeutic targeting of CDK12 has been achieving selectivity over other related CDKs such as CDK9 or CDK2. The following sections briefly review CDK12/13 inhibitors with a focus on structurally characterised examples.



**Figure 1.22. CDK inhibitors and their binding mode.** **a**, Annotated chemical structure of ADP with the kinase interaction sites indicated in different colours. **b**, Typical kinase inhibitory scaffolds among CDK-targeting drugs under clinical investigation. All scaffolds exhibit a hydrogen bond donor-hydrogen bond acceptor pattern that enables productive interactions with the hinge region of the kinase<sup>227</sup>. **c**, Example of an ATP-competitive CDK12 inhibitor with chemical moieties assigned to the CDK active site regions they occupy or contact based on the co-crystal structure. **d**, As in (c) but for a covalent CDK12 inhibitor THZ531.

## 1.2.6 Covalent inhibitors

Five covalent inhibitors that target CDK12/13 have been described to date (**Figure 1.23a**) and all take advantage of the C-terminal tail of the kinase bearing a single cysteine residue (Cys1039 in CDK12 and Cys1017 in CDK13) being proximal to the active site. This feature differentiates CDK12 and CDK13 from other kinases in the family, offering a path to the selective targeting of CDK12/13.



**Figure 1.23. Covalent inhibitors of CDK12/13.** **a**, Chemical structures of five known covalent inhibitors of CDK12/13. **b**, Crystal structure of the CDK12-cyclin K complex bound to THZ531 (PDB 5ACB)<sup>220</sup>. Interactions with the hinge methionine residue are shown in pink dashes. The C-terminal tail of the kinase is shown in cyan with the covalently modified cysteine residue depicted in a sticks representation. **c**, as in (**b**) but CDK13-cyclin K complex bound to THZ531 (PDB 7NXJ)<sup>223</sup>. **d**, as in (**b**) but CDK12-cyclin K bound to BSJ-01-175 (PDB 7NXK)<sup>223</sup>.

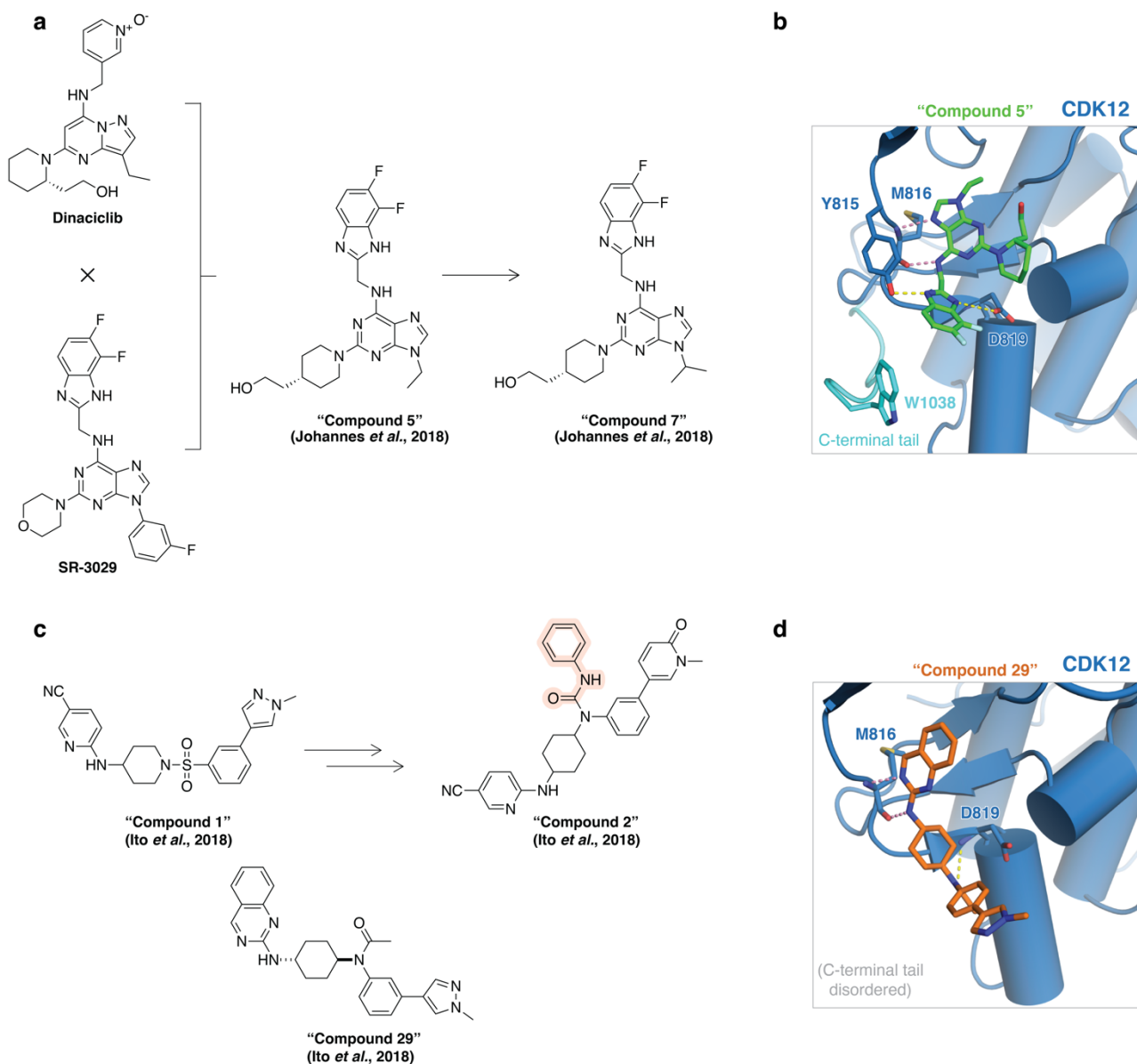
First, a covalent inhibitor developed for CDK7, THZ1, was found to also target CDK12/13<sup>228</sup>. Substituting the phenylbenzamide moiety in THZ1 with a phenyl(1-piperidinyl)methanone yielded a potent and CDK12/13-selective covalent inhibitor THZ531<sup>220</sup>. As the pharmacokinetic properties of THZ531 precluded animal studies<sup>223,229</sup>, the scaffold was further evolved to BSJ-01-175, CDK12/13-selective covalent inhibitor with demonstrable efficacy *in vivo*<sup>223</sup>. A distinct inhibitor E9 was then developed based on a pleiotropic non-covalent CDK inhibitor dinaciclib, with the aim of designing a compound that is not subject to ABC transporter-mediated efflux, a phenomenon observed to give rise to resistance to the THZ compound series<sup>230</sup>. The final covalent inhibitor, MFH290, was developed based on THZ531 and a pan-CDK inhibitor SNS032<sup>231</sup>. However, the compound again showed poor pharmacokinetics and lacked stability *in vivo*<sup>231</sup>. The above compounds bind the active site of the kinase, primarily via two hydrogen bonding interactions with the hinge region (Met816 in

CDK12 and Met794 in CDK13), and also extend out of the pocket, forming a covalent bond to the C-terminal tail's cysteine residue (**Figure 1.23b-d**).

### 1.2.7 ATP-competitive inhibitors

Several non-covalent, ATP-competitive inhibitors for CDK12/13 have also been reported. However, most of these compounds are pleiotropic and target multiple CDKs, while none can effectively differentiate between CDK12 and CDK13<sup>232</sup>. All are type I inhibitors, meaning that they bind the active, DFG-in conformation of the kinase. Type II inhibitors, which bind to an inactive, DFG-out kinase conformation are rarely observed for CDKs, likely due to the bulky nature of their gatekeeper residue (phenylalanine in almost all CDKs). This is thought to limit the compound's access to the hydrophobic back-pocket and disfavour the inactive, DFG-out conformation of the kinase in the presence of ligands (the sorafenib-CDK8-cyclin C structure is a notable exception in the CDK family<sup>233</sup>). Interestingly, as CDK12/13 were discovered several years after most other CDK family members, many CDK inhibitors were not tested against these CDKs during selectivity profiling and only later found to possess cryptic CDK12 inhibitory activity.

In 2016, a pan-CDK inhibitor dinaciclib (**Figure 1.24a**) was reported to be a potent inhibitor of CDK12<sup>234</sup>. SR-3029 (**Figure 1.24a**), a dual CK1 $\delta$ /CK1 $\epsilon$  inhibitor, was also found to target CDK12, yet with little potency at physiological ATP concentration<sup>221,235</sup>. In a 2018 study, AstraZeneca took an interesting approach of hybridising dinaciclib with SR-3029 to achieve both high potency and family selectivity (**Figure 1.24a**)<sup>221</sup>. The process, followed by some optimisation, yielded "Compound 5", which was subsequently crystallised with CDK12-cyclin K. The structure revealed the expected hinge interactions of the purine core and showed the 2-(2-piperidinyl)ethanol moiety occupying the ribose pocket (**Figure 1.24b**). The difluorobenzimidazole substituent, postulated to be a determinant of family selectivity, is largely solvent-exposed, forming hydrogen bonds with Tyr815 and Asp819, as well as possibly contacting Trp1038 of the C-terminal tail (**Figure 1.24b**)<sup>221</sup>. The ethyl substituent was later replaced with an isopropyl group ("Compound 7"), leading to improved selectivity but decreased potency<sup>221</sup>.



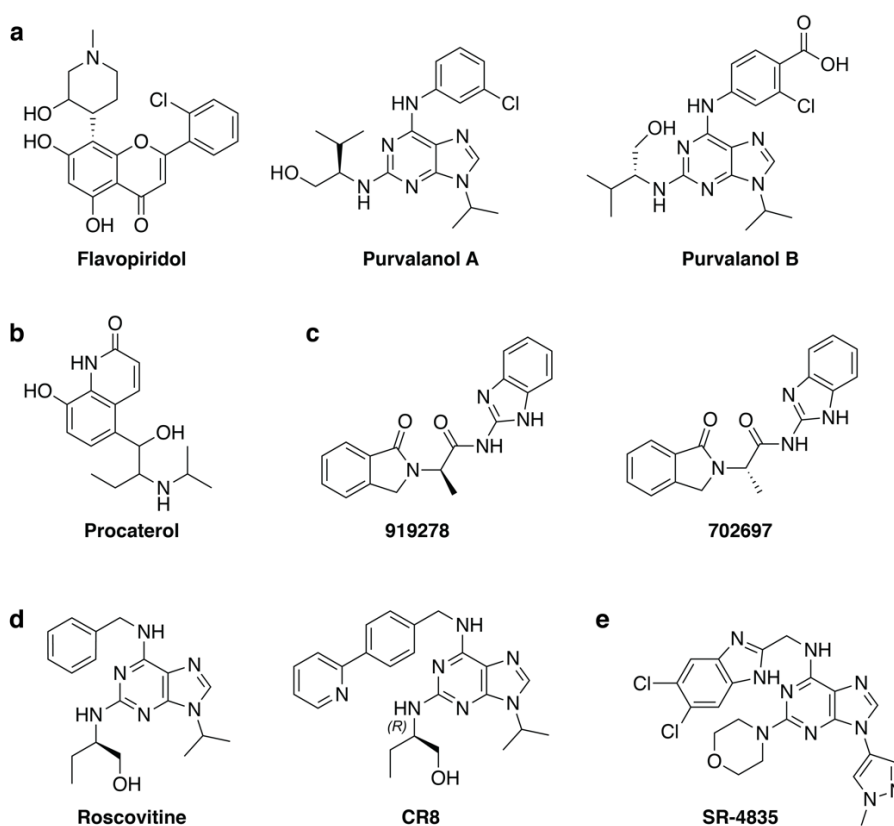
**Figure 1.24. ATP-competitive CDK12 inhibitor optimisation.** **a**, Chemical structures of dinaciclib, SR-3029, and two compounds resulting from their hybridization. **b**, Binding mode of "Compound 5" to CDK12-cyclin K (PDB 6B3E)<sup>221</sup>. **c**, Chemical structures of the initial hit ("Compound 1"), optimised inhibitor ("Compound 2"), and an intermediate derivative ("Compound 29"). The benzyl urea moiety (orange highlight) of "Compound 2" contributes to CDK12 versus CDK2 selectivity. **d**, Binding mode of "Compound 29" to CDK12-cyclin K (PDB 6CKX)<sup>222</sup>.

Later in 2018, a series of trisubstituted urea derivatives was reported as novel CDK12 inhibitors<sup>222</sup>. The initial hit ("Compound 1") (Figure 1.24c) was identified through high-throughput screening (HTS), yet it also showed strong CDK2 inhibition ( $IC_{50}(\text{CDK12}) = 0.36 \mu\text{M}$  versus  $IC_{50}(\text{CDK2}) = 0.01 \mu\text{M}$ ). Hit optimisation yielded "Compound 2" (CDK12  $IC_{50} = 0.052 \mu\text{M}$ , CDK13  $IC_{50} = 0.010 \mu\text{M}$ , CDK2  $IC_{50} > 100 \mu\text{M}$ ), with CDK12 versus CDK2 selectivity achieved through replacing the sulphonyl group with a benzyl urea moiety (Figure 1.24c). An intermediate derivative, "Compound 29", was crystallised in complex with CDK12-cyclin K to reveal the expected hinge contacts, as well as a hydrogen bonding interaction of the oxygen atom of the acetyl group with the backbone of Asp819 (Figure 1.24d).



Multiple studies additionally tested panels of published CDK inhibitors against CDK12-cyclin K. Bosken *et al.* found that flavopiridol, an ATP-competitive kinase inhibitor derived from a natural alkaloid Rohitkine predominantly active on CDK9, also targets CDK12 (**Figure 1.25a**)<sup>218</sup>. Other compounds, such as purvalanol A, purvalanol B, and procaterol, a  $\beta$ 2 adrenergic receptor agonist, were also shown to inhibit CDK12 (**Figure 1.25a, b**)<sup>218</sup>.

In a recent study conducted by Biogen, a HTS for inhibitors of the non-canonical NF- $\kappa$ B pathway was undertaken<sup>202</sup>. It yielded 919278 (**Figure 1.25c**), a small molecule later identified to be a CDK12 inhibitor, and reported that 919278 selectively inhibits non-canonical pathways while preserving canonical NF- $\kappa$ B signalling<sup>202</sup>. Notably, the S enantiomer of the compound (702697) showed very limited activity, highlighting the importance of stereochemistry for its inhibition of CDK12 (**Figure 1.25c**)<sup>202</sup>.



**Figure 1.25. Chemical structures of other ATP-competitive CDK12 inhibitors.** **a**, Examples of compounds identified by Bosken *et al.* to inhibit CDK12, including flavopiridol, purvalanol A, and purvalanol B. **b**, Procaterol, a  $\beta$ 2 adrenergic receptor agonist. **c**, CDK12/13 inhibitor 919278 and its inactive enantiomer 702697<sup>202</sup>. **d**, A pan-CDK inhibitor roscovitine and its potent derivative CR8. **e**, SR-4835, a CDK12/13-selective inhibitor.

One compound that will be discussed extensively in **Chapter 2** is CR8, a CDK inhibitor discovered as a much more potent analogue of a pan-CDK inhibitor roscovitine (**Figure 1.25d**). Despite extensive research, the reason for its increased potency could not be identified<sup>236-241</sup>. CR8 was later discovered to also modestly inhibit CDK12<sup>218</sup>, yet this finding still did not account for its unusually high cytotoxicity.

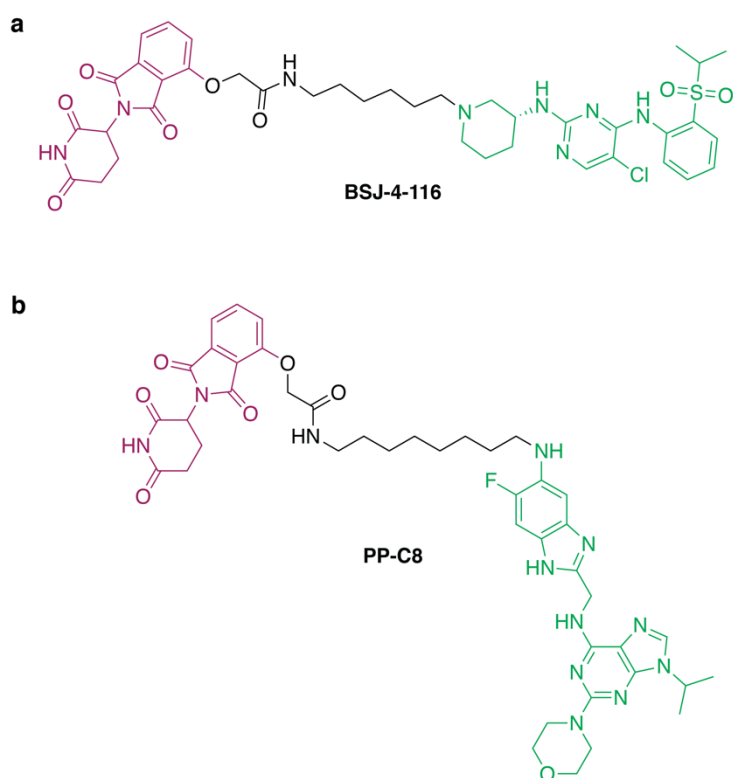
In 2019, the Duckett laboratory identified SR-4835 (**Figure 1.25e**) as a potent dual inhibitor of CDK12/13 ( $K_{d(\text{CDK12})} = 98 \text{ nM}$ ,  $K_{d(\text{CDK13})} = 5 \text{ nM}$ )<sup>242</sup>. Studies in a triple-negative breast cancer (TNBC) patient-derived xenograft models showed the compound displays suitable pharmacokinetics and efficacy both as a single agent and in combination with other drugs<sup>242</sup>. These results provided a robust preclinical proof-of-concept for the use of CDK12/13 inhibitors in TNBC.

### 1.2.8 PROTACs

With the recent emergence of CDK12 as a therapeutic target, two PROTACs targeting the CDK12-cyclin K complex have been described (**Figure 1.26**). There are two main advantages of targeting CDK12 using degraders. Firstly, CDK12 degradation abrogates both kinase-dependent and kinase-independent functions of the protein. Secondly, due to the reliance on protein-protein interaction between the ligase and the target, degraders can offer increased family selectivity and therefore could enable the targeting of CDK12 with some selectivity over CDK13 and other CDKs.

In 2021, a CDK12 PROTAC BSJ-4-116 (**Figure 1.26a**) was reported<sup>243</sup>. The CDK12-targeting warhead of BSJ-4-116 is based on the covalent CDK12/13 inhibitor THZ531 and pomalidomide is used for CRBN recruitment. BSJ-4-116 induces complementary protein-protein interactions, and sequence differences between CDK12 and CDK13 at CRBN contact sites partly explain the selectivity for CDK12 over CDK13 degradation, with nanoBRET assays further demonstrating no appreciable CDK13-PROTAC-CRBN complex formation. BSJ-4-116 significantly inhibited the phosphorylation of RNAPII Ser2 and Thr4 without inhibiting p-Ser5 and p-Ser7, illustrating its CDK family selectivity beyond CDK12/13. Interestingly, the PROTAC did not cause any destabilization of cyclin K and only depletion of CDK12 was observed<sup>244</sup>. Notably, the authors identified two point mutations in CDK12 that confer resistance to the degrader, demonstrating that cancer cells can acquire resistance to PROTACs via mutations in the target rather than the UPS components as previously shown<sup>244</sup>.

Another bifunctional degrader of CDK12, PP-C8 (**Figure 1.26b**), was recently published<sup>245</sup>. The CDK12-targeting warhead is based on an amalgamation of SR-4835 and inhibitors reported by Johannes *et al.*<sup>221</sup>, while CRBN is recruited via pomalidomide. It also displayed considerable selectivity for CDK12 over CDK13 but, unlike BSJ-4-116, PP-C8 caused the degradation of both CDK12 and cyclin K<sup>245</sup>.



**Figure 1.26. CDK12 PROTAC structures.** **a**, BSJ-4-116 based on a covalent CDK12/13 inhibitor THZ531 (green) and CRBN binder pomalidomide (purple). **b**, PP-C8 with the CDK12-targeting warhead based on an amalgamation of SR-4835 and inhibitors reported by Johannes *et al.*<sup>221</sup> (green) and CRBN recruited through pomalidomide (purple).

### 1.2.9 Synthetic lethality and combination treatments

CDK12-targeting compounds have been evaluated both as single agents and in combination with other drugs. In particular, synthetic lethality upon concomitant CDK12 and PARP inhibition has been extensively assessed. In isolation, PARP inhibitors are an effective treatment for BRCA-mutated or homologous recombination (HR)-deficient cancers through a synthetic lethality mechanism<sup>246</sup>. However, PARP inhibitors lack broad applicability due to insufficient activity in BRCA-WT or HR-proficient cancers, as well as resistance appearing upon HR restoration or BRCA mutation<sup>246</sup>. Combinations of such inhibitors with drugs able to impair HR have been explored as a way of sensitizing all cancer cells to PARP inhibition. Given the paramount role of CDK12 in controlling the expression of DDR genes, the effect of CDK12 silencing on PARP inhibitor sensitivity was evaluated. Dinaciclib, a pan-CDK inhibitor that potently inactivates CDK12, was shown to effectively sensitize BRCA-WT and mutant TNBC cells to PARP inhibition and clinical trials combining dinaciclib with veliparib are currently ongoing (NCT01434316)<sup>234</sup>. A CDK12/13-selective inhibitor SR-4835 was also found to sensitize TNBC cells to PARP inhibitors, as well as to DNA crosslinking agents, topoisomerase I inhibitors, and DNA replication targeting agents<sup>242</sup>. Both CDK12-targeting PROTACs were also shown to exhibit synthetic lethal effects in combinations with PARP inhibitors<sup>243,245</sup>. Further examples of synergistic drug combinations for

CDK12-targeting agents, such as those with CHK1 inhibitors<sup>206</sup> or anti-PD1 antibodies<sup>247,248</sup>, are detailed in **Table 1.3**.

**Table 1.3.** Synergistic drug combinations reported for CDK12-targeting agents (adapted from <sup>249,250</sup>).

CDK12 drug	Combination drug	Assay	Indication	References
THZ531	Olaparib (PARP inhibitor)	Mouse xenograft models	Ovarian cancer, Ewing sarcoma, myeloma	204,210,251
	KU-0060648 (DNA-PK inhibitor)	Cellular	Myeloma	251
	Doxorubicin (topoisomerase II inhibitor)	Cellular	Anaplastic thyroid carcinoma	252
	GSK3 or UNC1999 (EZH2 inhibitors)	Cellular	Aggressive B-cell lymphomas	253
	Sorafenib (pan-kinase inhibitor)	Cellular	Hepatocellular carcinoma	254
SR-4835	Olaparib (PARP inhibitor)	Cellular	TNBC	242
	Cisplatin (DNA damaging agent)	Mouse xenograft model	TNBC	242
	Irinotecan (topoisomerase inhibitor)	Mouse xenograft model	TNBC	242
	Anti-PD1 antibody (checkpoint inhibitor)	Mouse xenograft model	Breast cancer	247
BSJ-4-116 PROTAC	Olaparib (PARP inhibitor)	Cellular	T-cell acute lymphoblastic leukaemia	243
PP-C8 PROTAC	Olaparib (PARP inhibitor)	Cellular	TNBC	245
THZ1	Olaparib (PARP inhibition)	Mouse xenograft model	Ewing sarcoma	201
Dinaciclib	Olaparib or veliparib (PARP inhibitors)	Mouse xenograft model	TNBC, multiple myeloma	234,255
	Cisplatin (DNA damaging agent)	Mouse xenograft model	Ovarian cancer	256
	Trastuzumab (monoclonal antibody)	Mouse xenograft model	HER2-positive breast cancer	257
	Doxorubicin or etoposide (topoisomerase inhibitors)	Cellular	Neuroblastoma	258
	Vemurafenib (BRAF inhibitor)	Mouse xenograft model	Melanoma	259
	Anti-PD1 antibody (checkpoint inhibition)	Mouse xenograft model	PD1-dependent cancers	248

	Gemcitabine (DNA synthesis inhibitor)	Mouse xenograft model	Pancreatic cancer	260
	MK-2206 (AKT inhibitor)	Mouse xenograft model, phase I (no benefit)	Pancreatic cancer	261,262
CDK12 silencing (shRNA)	Sorafenib (pan-kinase inhibitor)	Mouse xenograft model	Hepatocellular carcinoma	254
	SCH900776, LY2603618 (CHK1 inhibitors)	Mouse xenograft model	Ovarian cancer	206
CDK12 silencing (siRNA)	Veliparib (PARP inhibitor)	Cellular	Ovarian cancer	205
	Cisplatin or melphalan (DNA damaging agents)	Cellular	Ovarian cancer	205

In summary, multiple CDK12-targeting compounds have been described thus far but most lack selectivity or, in the case of covalent inhibitors and PROTACs, suitable physicochemical properties. Although gaps exist in our understanding of the biological roles of CDK12 and cyclin K, the complex emerges as an important and actionable therapeutic target in oncology and beyond<sup>190,199,242,249,250,263</sup>. CDK12-cyclin K inhibitors are likely to find particular clinical utility in conjunction with other agents (**Table 1.3**), and phase I clinical trials with a combination of veliparib and dinaciclib are currently ongoing. Dinaciclib is the first (deliberately) CDK12-targeting compound in clinical trials and while it is potent, it remains a pan-CDK inhibitor. Given the clinical relevance of CDK12-cyclin K, the development of novel CDK12- or CDK12/13-selective compounds is likely to yield small molecules of an even higher therapeutic value.

The mechanism of action of several of the above-described compounds will be revisited in the experimental sections of this thesis (**Chapter 2** and **Chapter 3**).

### 1.3 Aims of the thesis

Molecular glue degraders are a highly promising novel therapeutic modality. As exemplified by thalidomide analogues, such degrader drugs can show remarkable clinical efficacy and, unlike for PROTACs, their drug-like properties will likely facilitate the transition of new examples into the clinic. However, with only two well-described compound classes, the molecular glue degrader discovery and design rules are poorly understood. Therefore, the overarching aim of this thesis was to dissect how such small molecules can reprogram protein-protein interfaces to lead to selective target degradation.

More specifically, the goals of this work were as follows:

- 1) The first aim of the present study was to identify novel molecular glue degrader compounds and to structurally characterise their binding mode to inform on how they induce target degradation. To address this goal, systematic searches for molecular glue degrader compounds were conducted in collaboration with the Ebert laboratory (DFCI, Boston) and the Winter laboratory (CeMM, Vienna). The resulting candidate compounds were studied through biochemical reconstitutions, as well as enzymatic and biophysical assays to confirm their mode of action as molecular glue degraders. In the case of CR8, the compound-induced ternary complex was crystallised to reveal how the molecular glue compound is positioned at the protein-protein interface. This work is described in **Chapter 2**.
- 2) The second aim of this thesis was to further the understanding of the molecular glue degrader structure-activity relationship in order to enable their rational modification and, ultimately, design. For this, a set of over 90 putative degraders of cyclin K was extensively studied using a suite of biochemical, biophysical, structural, and cellular methods. These results are presented in **Chapter 3**.
- 3) The final aim of the study was to develop and fine-tune biophysical assays that enable robust molecular glue degrader evaluation. To address this goal, a time-resolved Förster resonance energy transfer (TR-FRET) assay was developed and systematically optimised into a robust, HTS-appropriate assay setup. This work is featured in **Chapter 4**, and is presented alongside the optimisation of other fluorescence-based assays for adjacent projects.

## Chapter 2 : Discovery of cyclin K degraders

This chapter contains published data and additional unpublished data. I planned and performed experiments presented in **sections 2.1.2.2 – 2.1.2.4**, several in collaboration with Georg Petzold (Thomä laboratory, FMI, Basel). Crystallographic data analysis was performed primarily by Richard Bunker (Thomä laboratory, FMI, Basel), though with my and Georg's active participation. The bioinformatic correlation, CRISPR validation and screening (**section 2.1.2.1**) and cellular evaluation (**section 2.1.2.5 and section 2.1.2.4 Figure 2.14f and Figure 2.15b,c**) were performed by the members of the Ebert laboratory (Dana Farber Cancer Institute, Boston) led by Mikołaj Słabicki). **Section 2.2** describes an effort led by the Winter laboratory (Research Center for Molecular Medicine of the Austrian Academy of Sciences, Vienna; spearheaded by Cristina Mayor-Ruiz), to which I contributed through biochemical complex reconstitution and complex formation TR-FRET assays (**Figure 2.21j, Figure 2.22**).

Several sections and figures in this chapter are reproduced or adapted from the following manuscripts:

*The CDK inhibitor CR8 acts as a molecular glue degrader that depletes cyclin K*

M. Słabicki\*, **Z. Kozicka\***, G. Petzold\*, Y.-D. Li, M. Manojkumar, R. Bunker, K. Donovan, Q. Sievers, J. Koepfel, D. Suchyta, A. Sperling, E. Fink, J. Gasser, L. Wang, S. Corsello, R. Sellar, M. Jan, D. Gillingham, C. Scholl, S. Fröhling, T. Golub, E. Fischer, N. Thomä<sup>†</sup>, B. Ebert<sup>†</sup>.

*Nature* (2020), <https://doi.org/10.1038/s41586-020-2374-x> (\*equal contribution)

*Rational discovery of molecular glue degraders via scalable chemical profiling*

C. Mayor-Ruiz, S. Bauer, M. Brand, **Z. Kozicka**, M. Siklos, H. Imrichova, I. Kaltheuner, E. Hahn, K. Seiler, A. Koren, G. Petzold, M. Fellner, C. Bock, A. Muller, J. Zuber, M. Geyer, N. Thomä, S. Kubicek, G. Winter.

*Nature Chemical Biology* (2020), <https://doi.org/10.1038/s41589-020-0594-x>

The published manuscripts can be found in the **Appendix** of this thesis.

## 2.1 The CDK inhibitor CR8 acts as a molecular glue degrader that depletes cyclin K (published manuscript)

### 2.1.1 Abstract

Molecular glue compounds induce protein-protein interactions that, in the context of a ubiquitin ligase, lead to protein degradation. Unlike traditional enzyme inhibitors, such molecular glue degraders act sub-stoichiometrically to catalyse the rapid depletion of previously inaccessible targets. They are clinically effective and highly sought-after but have thus far only been discovered serendipitously. Through the systematic mining of databases for correlations between the cytotoxicity of 4,518 compounds and E3 ligase expression levels across hundreds of human cancer cell lines, we identified CR8, a CDK inhibitor, as a compound that acts as a molecular glue degrader. A solvent-exposed pyridyl moiety of CR8, in its CDK-bound form, induces CDK12-cyclin K complex formation with DDB1, the CUL4 adaptor protein, bypassing the requirement for a substrate receptor and presenting cyclin K for ubiquitination and degradation. These studies demonstrate that chemical alteration of surface-exposed moieties can confer gain-of-function glue properties to an inhibitor, and we propose this as a broader strategy to turn target binders into molecular glues.

### 2.1.2 Introduction

Molecular glues are a class of small molecule drugs that induce or stabilise protein-protein interactions<sup>264</sup>. In the context of a ubiquitin ligase, drug-induced interactions can lead to protein degradation, and this constitutes an emerging strategy for the inactivation of therapeutic targets intractable by conventional pharmacological means<sup>265,266</sup>. Known molecular glue degraders bind to substrate receptors of E3 ubiquitin ligases and recruit target proteins for their ubiquitination and subsequent degradation by the proteasome.

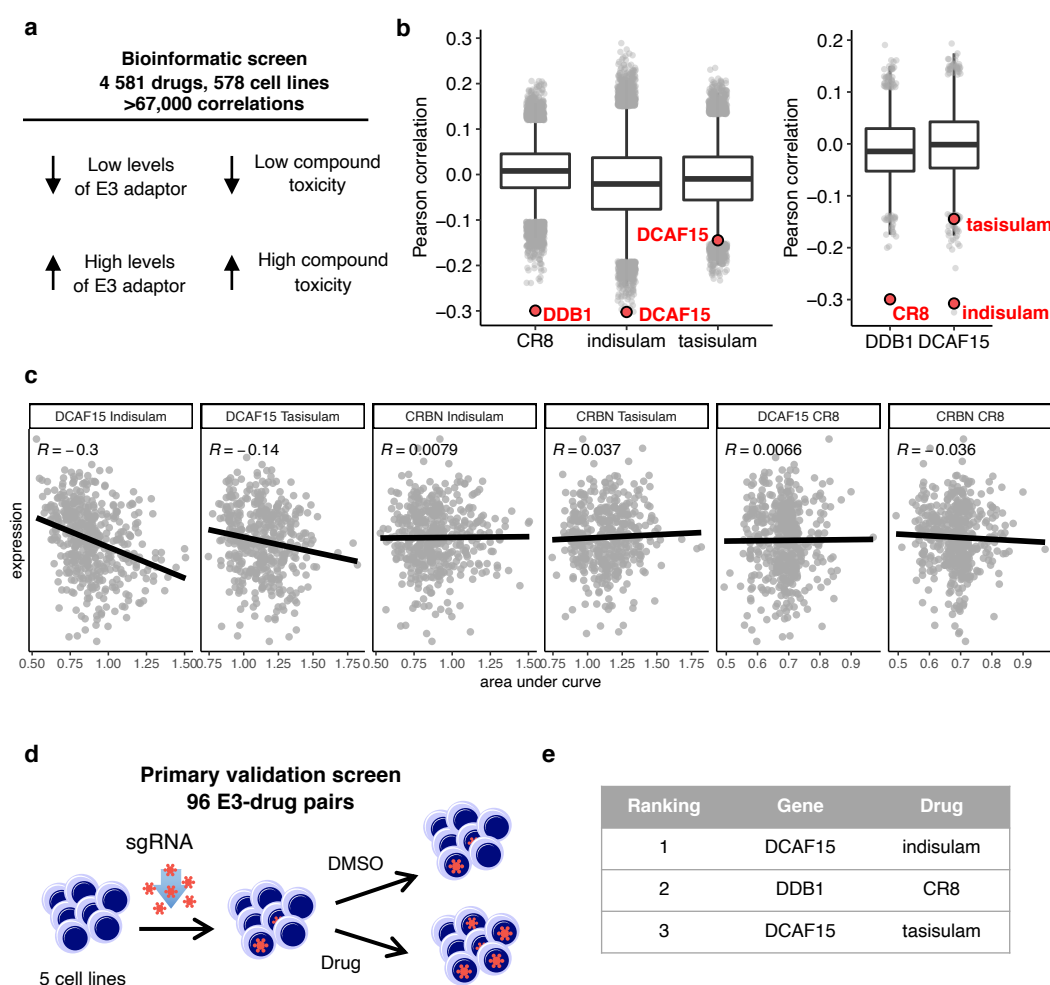
Thalidomide analogues and aryl sulphonamides are two classes of drugs that act as molecular glue degraders, as detailed in **section 1.1.9**. Widely used in the clinic, thalidomide analogues have proven to be an effective treatment for multiple myeloma, other B cell malignancies, and myelodysplastic syndrome with a deletion in chromosome 5q<sup>267</sup>. Thalidomide analogues recruit zinc-finger TFs and other targets to CRBN<sup>44,268-270</sup>, the substrate receptor of the cullin-RING E3 ubiquitin ligase CUL4A/B-RBX1-DDB1-CRBN (CRL4<sup>CRBN</sup>)<sup>39</sup>. Similarly, aryl sulphonamides degrade the essential RNA-binding protein RBM39 by engaging DCAF15, the substrate receptor of the CRL4<sup>DCAF15</sup> E3 ubiquitin ligase<sup>134,136,271</sup>. In these examples, the degraders are not dependent on a ligandable pocket on the target protein, but instead leverage complementary protein-protein interfaces between the receptor and the target. By reprogramming ubiquitin ligase selectivity, these molecules divert the ligase to drive multiple rounds of target ubiquitination in a catalytic manner<sup>92</sup>. Such compounds can therefore circumvent the limitations of classical inhibitors, expanding the repertoire of “druggable” proteins. Although highly sought-after, molecular glue degraders have been found predominantly serendipitously, and there are currently limited strategies available for identifying or designing compounds with such a mode of action.



## 2.1.2 Results

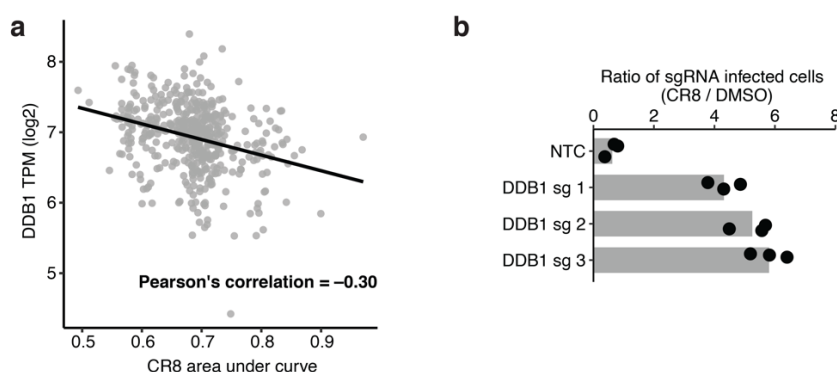
### 2.1.2.1 CR8 induces proteasomal degradation of cyclin K

To identify small molecules that mediate targeted protein degradation through an E3 ubiquitin ligase, we correlated drug sensitivity data for 4,518 clinical and pre-clinical drugs tested against 578 cancer cell lines<sup>272,273</sup> with the mRNA expression levels for 499 E3 ligase components<sup>274</sup> (**Figure 2.1a**). *DCAF15* gene expression correlated with indisulam and tasisulam toxicity, consistent with its known function as a degrader of the essential protein RBM39 by the CRL4<sup>DCAF15</sup> E3 ubiquitin ligase, thus demonstrating the potential of the approach (**Figure 2.1b, c**). We sought to validate the high-scoring ligase-drug correlations by examining whether CRISPR-mediated inactivation of the identified E3 ligase component would rescue the respective drug-induced toxicity (**Figure 2.1d**). These experiments confirmed that sgRNAs targeting *DCAF15* confer resistance to indisulam and tasisulam. In addition, we also observed a correlation between cytotoxicity of the CDK-inhibitor *R*-CR8 and mRNA expression levels of the CUL4 adaptor DDB1 (**Figure 2.2a and Figure 2.1e**). Consistently, sgRNAs targeting *DDB1* conferred resistance to *R*-CR8 (**Figure 2.2b**).



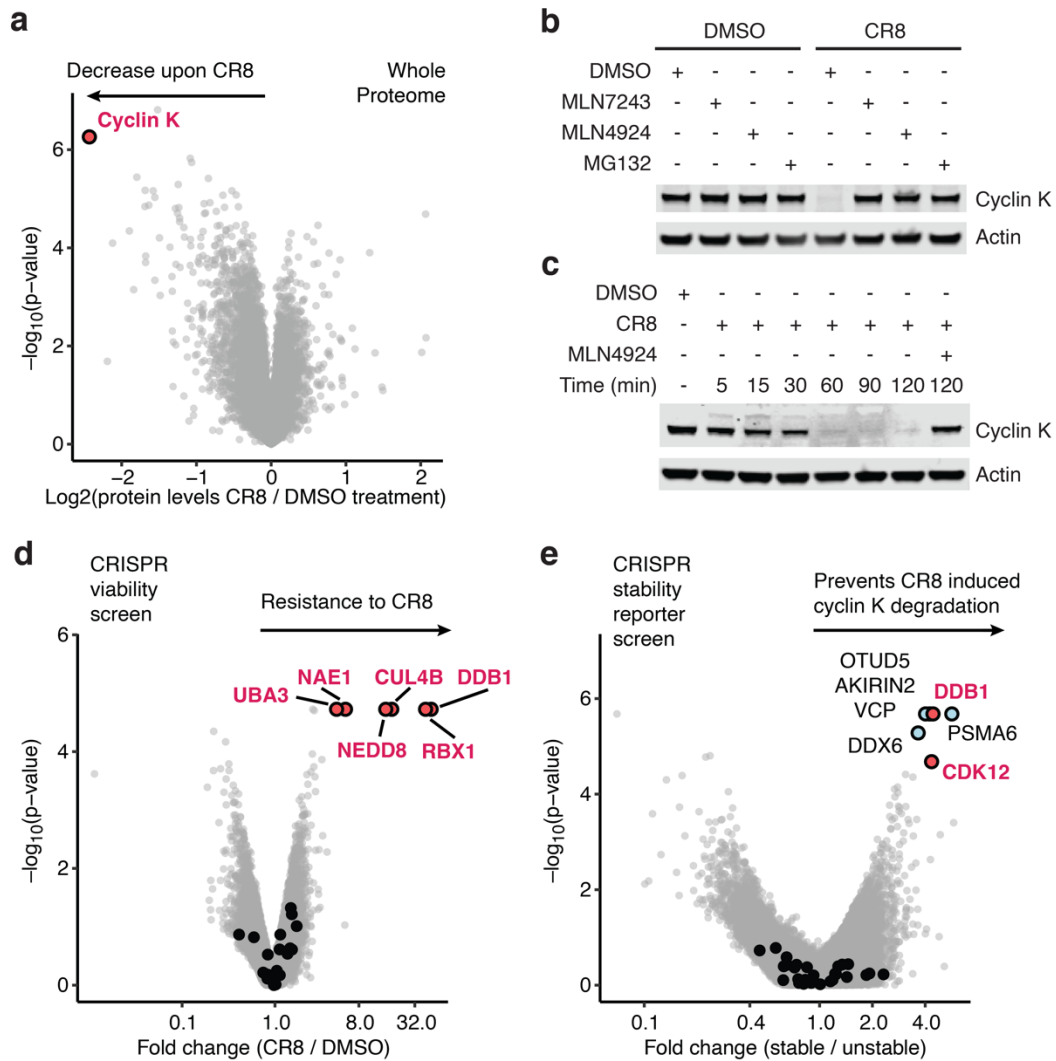
**Figure 2.1. CR8-induced degradation of cyclin K correlates with DDB1 expression.** **a**, Schematic of bioinformatic screen for identification of novel correlations between drug toxicity and E3 ligase mRNA expression. **b**, Box-and-whisker plot representing expression–sensitivity correlations for CR8, indisulam and tasisulam sensitivity as well as DDB1 and DCAF15 expression. **c**, Correlation of indisulam and tasisulam toxicity with mRNA expression of DCAF15. Selected

negative correlations are included for comparison. **d**, Schematic of flow-based primary validation screen. **e**, Top three hits from the primary validation screen in 5 cell lines, performed according to schematic outline in (**d**). *I have not contributed to these results.*

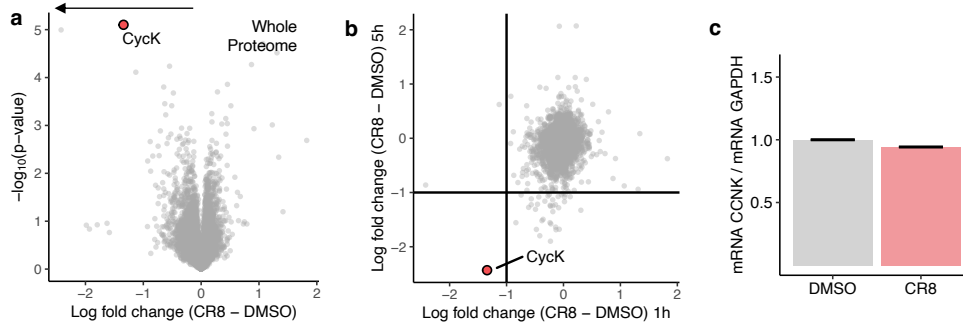


**Figure 2.2. Bioinformatic correlation shows CR8 toxicity correlates with DDB1 expression.** **a**, Correlation between CR8 toxicity and mRNA expression of DDB1 in 578 cancer cell lines. Lower area under the curve (AUC) corresponds to higher drug toxicity. TPM, transcripts per million. **b**, HEK293T<sub>Cas9</sub> cells were transduced with BFP (blue fluorescent protein) labelled sgRNAs, treated with DMSO or 1  $\mu$ M CR8 and analysed by flow cytometry (n=3). *I have not contributed to these results.*

As DDB1-dependent cytotoxicity of *R*-CR8 implicated ubiquitin ligase-mediated degradation of one or more essential proteins, we performed quantitative proteome-wide mass spectrometry to evaluate protein abundance following compound treatment. Of the >8,000 quantified proteins, cyclin K was the only protein that consistently showed decreased abundance following *R*-CR8 addition (**Figure 2.3a-c** and **Figure 2.4a, b**). As expected, *R*-CR8 did not alter the cyclin K mRNA levels (**Figure 2.4c**) and compound-induced cyclin K degradation could be rescued by inhibition of the E1 ubiquitin-activating enzyme (MLN7243), cullin neddylation (MLN4924) and the proteasome (MG132) (**Figure 2.3b**). Together, these results suggest that *R*-CR8 triggers rapid proteasomal degradation of cyclin K (**Figure 2.3c**) through the activity of a DDB1-containing cullin-RING ubiquitin ligase.



**Figure 2.3. CR8-induced degradation of cyclin K depends on DDB1 and CDK12.** **a**, Molt-4 cells were exposed to 1  $\mu\text{M}$  CR8 or DMSO for 5 hours followed by whole proteome quantification using tandem mass tag mass spectrometry (mean  $\log_2$  fold change, p value calculated by a moderated t-test,  $n=3$  (DMSO),  $n=1$  (CR8)). **b**, HEK293T<sub>Cas9</sub> cells were treated with DMSO, 0.5  $\mu\text{M}$  MLN7243 (ubiquitin activating enzyme inhibitor), 1  $\mu\text{M}$  MLN4924 (NEDD8-activating enzyme inhibitor), or 10  $\mu\text{M}$  MG132 (proteasome inhibitor) for 2 hours followed by exposure to 1  $\mu\text{M}$  CR8 for 2 hours. Lysates were immunoblotted for the indicated targets ( $n=3$ , representative image shown). **c**, HEK293T<sub>Cas9</sub> cells were treated with DMSO or 1  $\mu\text{M}$  MLN4924 for 2 hours followed by exposure to 1  $\mu\text{M}$  CR8 ( $n=3$ , representative image shown). **d**, Median fold change of read counts (CR8/DMSO treatment) and corresponding p values (empirical rank-sum test-statistic) for single guide RNAs (sgRNAs) targeting 19,112 human genes conferring resistance to CR8 treatment. DCAF substrate receptors are labelled with black dots ( $n=1$ ). **e**, Median fold change of read counts (cyclin K stable/unstable) and corresponding p values (empirical rank-sum test-statistic) for sgRNAs targeting 19,112 human genes preventing CR8 induced cyclin K<sub>GFP</sub> degradation. DCAF substrate receptors are labelled with black dots ( $n=3$ ). *I have not contributed to these results.*

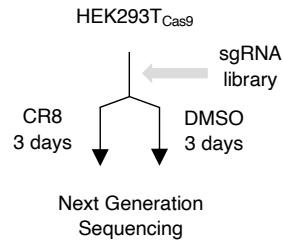
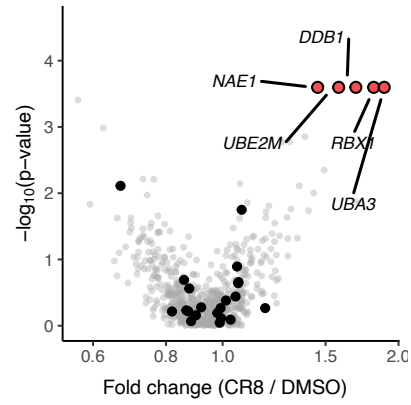


**Figure 2.4 CR8-induced degradation of cyclin K.** **a**, Molt-4 cells were exposed to 1  $\mu$ M CR8 or DMSO for 1 hour followed by whole proteome quantification using tandem mass tag mass spectrometry (mean log<sub>2</sub> fold change, p value calculated by a moderated t-test), n=3 (DMSO), n=1 (CR8). **b**, The log<sub>2</sub> fold changes in whole proteome quantification after 1 and 5 hours of exposure to CR8 are plotted against each other. **c**, HEK293T<sub>Cas9</sub> cells were treated with 1  $\mu$ M CR8 or DMSO for 2 hours and CCNK mRNA levels were measured by quantitative PCR. Bars represent the mean  $\pm$  s.d. (n=3). *I have not contributed to these results.*

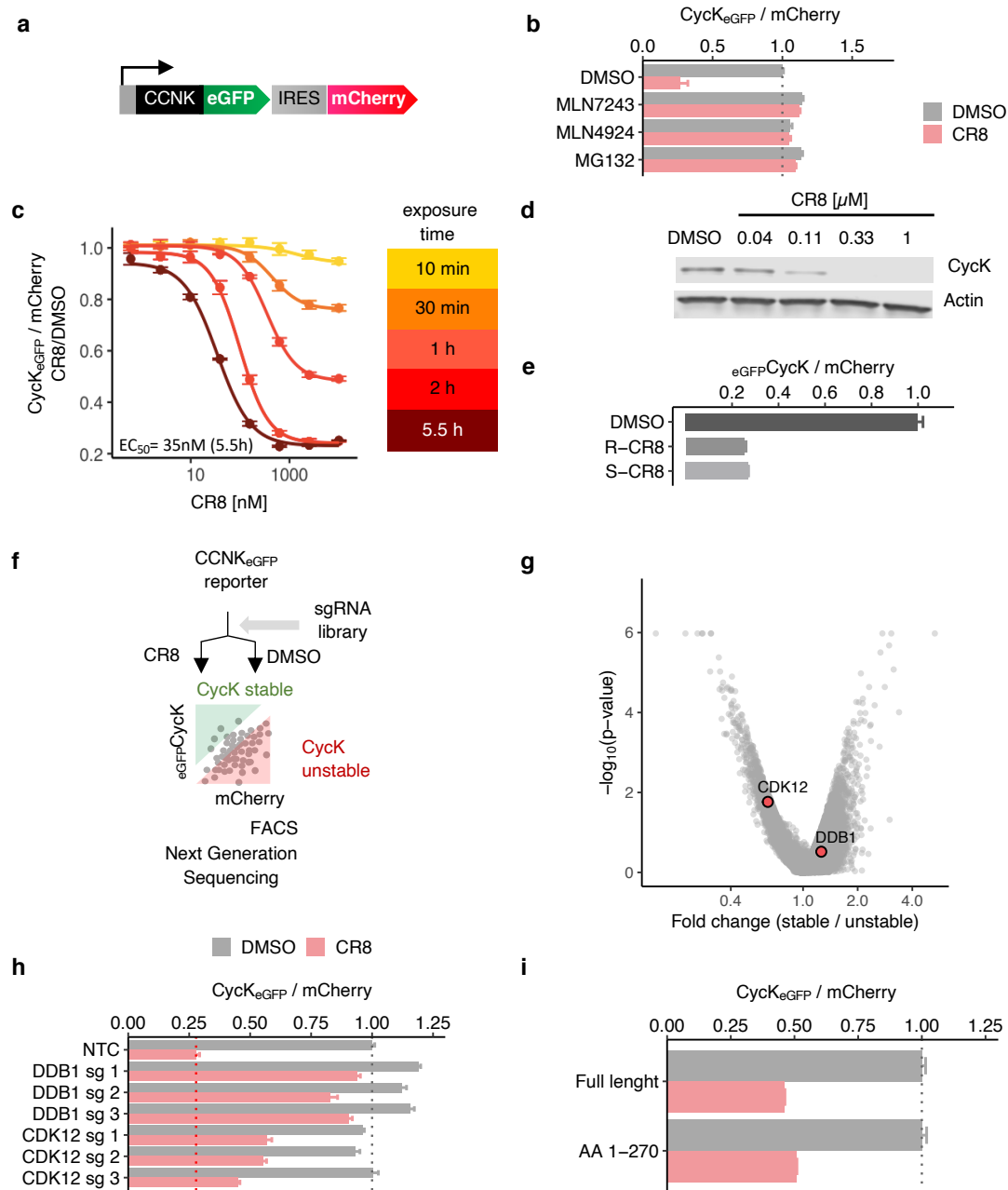
To dissect the molecular machinery required for *R*-CR8 toxicity, we performed genome-wide and E3 ubiquitin ligase-focused CRISPR-Cas9 resistance screens (**Figure 2.3d** and **Figure 2.5**). SgRNAs targeting *DDB1*, *CUL4B*, *RBX1*, the cullin-RING activator *NEDD8*, and the NEDD8-activating enzyme (*NAE1/UBA3*) were significantly enriched in the *R*-CR8-resistant cell population. These proteins are all required for CRL activity, and our results thus provide genetic evidence for the involvement of a functional CUL4-RBX1-DDB1 ubiquitin ligase complex in mediating *R*-CR8 cytotoxicity.

Thus far, all known cullin-RING ligases engage their substrates through specific substrate receptors, and DDB1 serves as an adaptor protein able to bind over 20 different substrate receptors (also known as DDB1-CUL4-associated-factors, DCAFs)<sup>275,276</sup> to recruit them to the CUL4-RBX1 E3 ubiquitin ligase core. As no DCAF was identified in our viability screens, we constructed a fluorescent reporter of cyclin K stability (**Figure 2.6c**), in which *R*-CR8-mediated degradation of endogenous cyclin K could be recapitulated with a cyclin K<sub>eGFP</sub> fusion protein (**Figure 2.3e** and **Figure 2.6a-d**). Using the stability reporter, in which the extent of degradation can be determined by measuring cyclin K<sub>eGFP</sub> levels normalised to mCherry expression, we found that *S*- and *R*-CR8 facilitated cyclin K<sub>eGFP</sub> degradation to the same extent (**Figure 2.6e**; henceforth, CR8 refers to *R*-CR8). We then performed a genome-wide CRISPR-Cas9 screen for genes involved in cyclin K reporter stability and validated the involvement of DDB1 in CR8-mediated cyclin K degradation (**Figure 2.3e**, **Figure 2.6f-h**), but not in compound-independent cyclin K degradation. In addition, we identified cyclin-dependent kinase 12 (CDK12), which is a known target of CR8<sup>218</sup> and whose activity depends on the interaction with cyclin K, as a crucial component for CR8-induced cyclin K<sub>eGFP</sub> destabilisation (**Figure 2.3e**, **Figure 2.6f-i**).

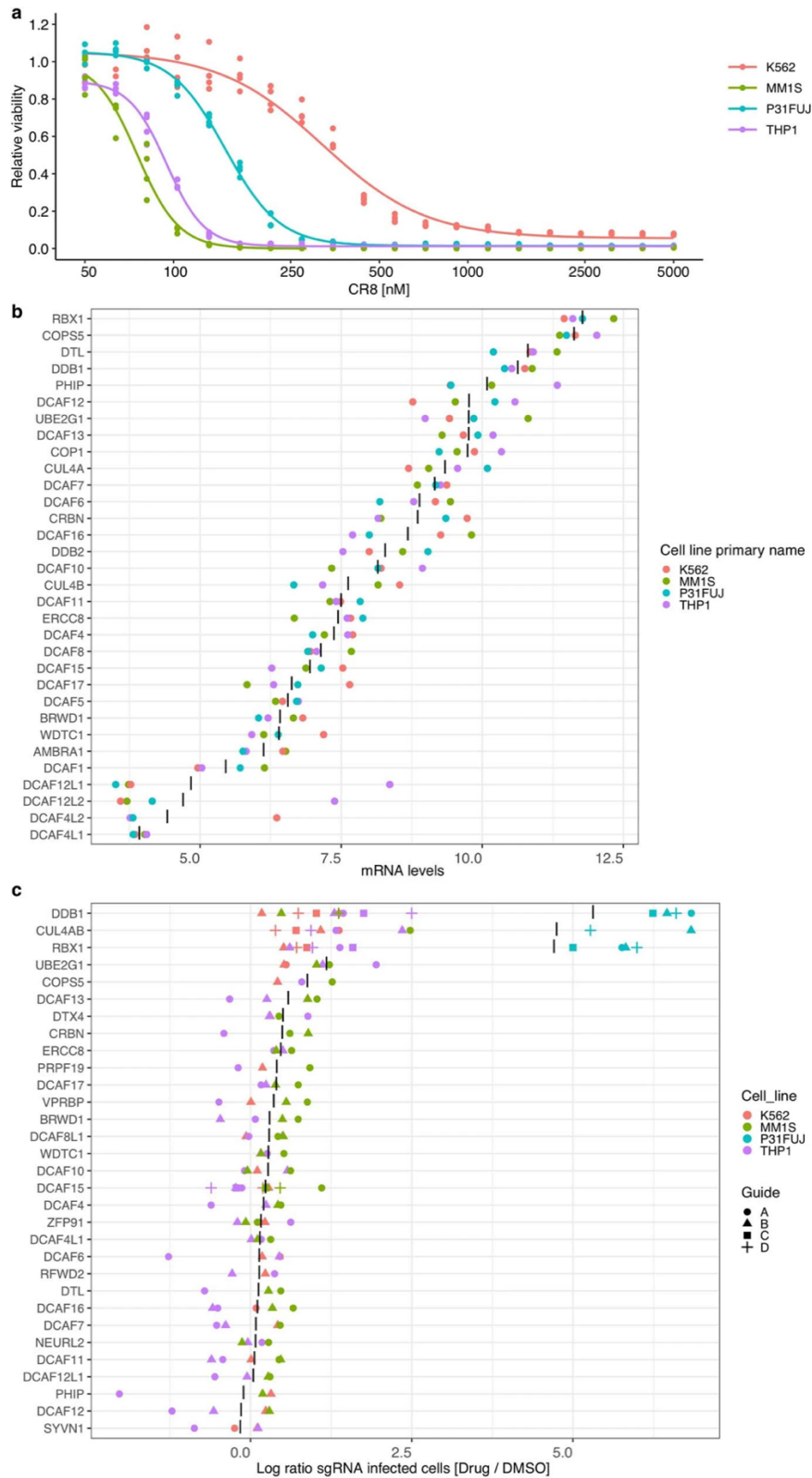
As neither the cyclin K<sub>eGFP</sub> stability reporter screen nor the CR8 resistance screen identified a substrate receptor, we performed additional CRISPR screens targeting 29 genes encoding known DCAFs or DCAF-like candidate proteins in four different cell lines. While sgRNAs targeting the previously identified components of the CUL4-RBX1-DDB1 complex consistently caused resistance to CR8, a DCAF substrate receptor could not be identified (**Figure 2.7**).

**a****b**

**Figure 2.5. Ligase-focused screen identifies CRL components as essential for CR8-induced cyclin K degradation. a,** Schematic of the genome-wide CRISPR-Cas9 resistance screen. **b,** Median fold change of read counts (CR8/DMSO treatment) and corresponding p values (empirical rank-sum test-statistic) for sgRNAs targeting 713 E1, E2, E3, DUB and control genes (BISON library) conferring resistance to CR8 treatment. DCAF substrate receptors are labelled with black dots (n=3). *I have not contributed to these results.*



**Figure 2.6 Reporter screen identifies CDK12 as required for CR8-induced cyclin K degradation.** **a**, Schematic of the cyclin K (CCNK) stability reporter. eGFP, enhanced green fluorescent protein, IRES, internal ribosome entry site. **b**, Cyclin K<sub>eGFP</sub> HEK293T<sub>Cas9</sub> cells were analysed by flow cytometry. Bars represent the mean  $\pm$  s.d. (n=3). **c**, Cyclin K<sub>eGFP</sub> HEK293T<sub>Cas9</sub> cells were exposed to various concentration of CR8 and analysed by flow cytometry. Data represent mean  $\pm$  s.d. Lines represent standard four-parameter log-logistic curves (n=3). **d**, HEK293T<sub>Cas9</sub> cells were treated with indicated concentration of CR8 for 2 hours and lysates were immunoblotted for the indicated targets. **e**, Cyclin K<sub>eGFP</sub> HEK293T<sub>Cas9</sub> cells were treated with 1  $\mu$ M of the indicated compound for 2 hours and analysed by flow cytometry. Data represent mean  $\pm$  s.d. Line represents standard four parameter log-logistic curves (n=3). **f**, Schematic of the genome-wide CRISPR-Cas9 reporter screen. **g**, Median fold change of read counts (cyclin K stable/unstable) and corresponding p values (empirical rank-sum test-statistic) in the absence of CR8 for sgRNAs targeting 19,112 human genes in HEK293T<sub>Cas9</sub> cells (n=2). **h**, Cyclin K<sub>eGFP</sub> HEK293T<sub>Cas9</sub> cells were transfected with blue fluorescent protein (BFP) and sgRNAs, treated with DMSO or 1  $\mu$ M CR8 and analysed by flow cytometry. Bars represent the mean  $\pm$  s.d. (n=3). **i**, Cyclin K<sub>eGFP</sub><sup>Full Length</sup> HEK293T<sub>Cas9</sub> or Cyclin K<sub>eGFP</sub><sup>AA1-270</sup> HEK293T<sub>Cas9</sub> were treated with DMSO or 1  $\mu$ M CR8 and analysed by flow cytometry. Bars represent the mean  $\pm$  s.d. (n=3). *I have not contributed to these results.*



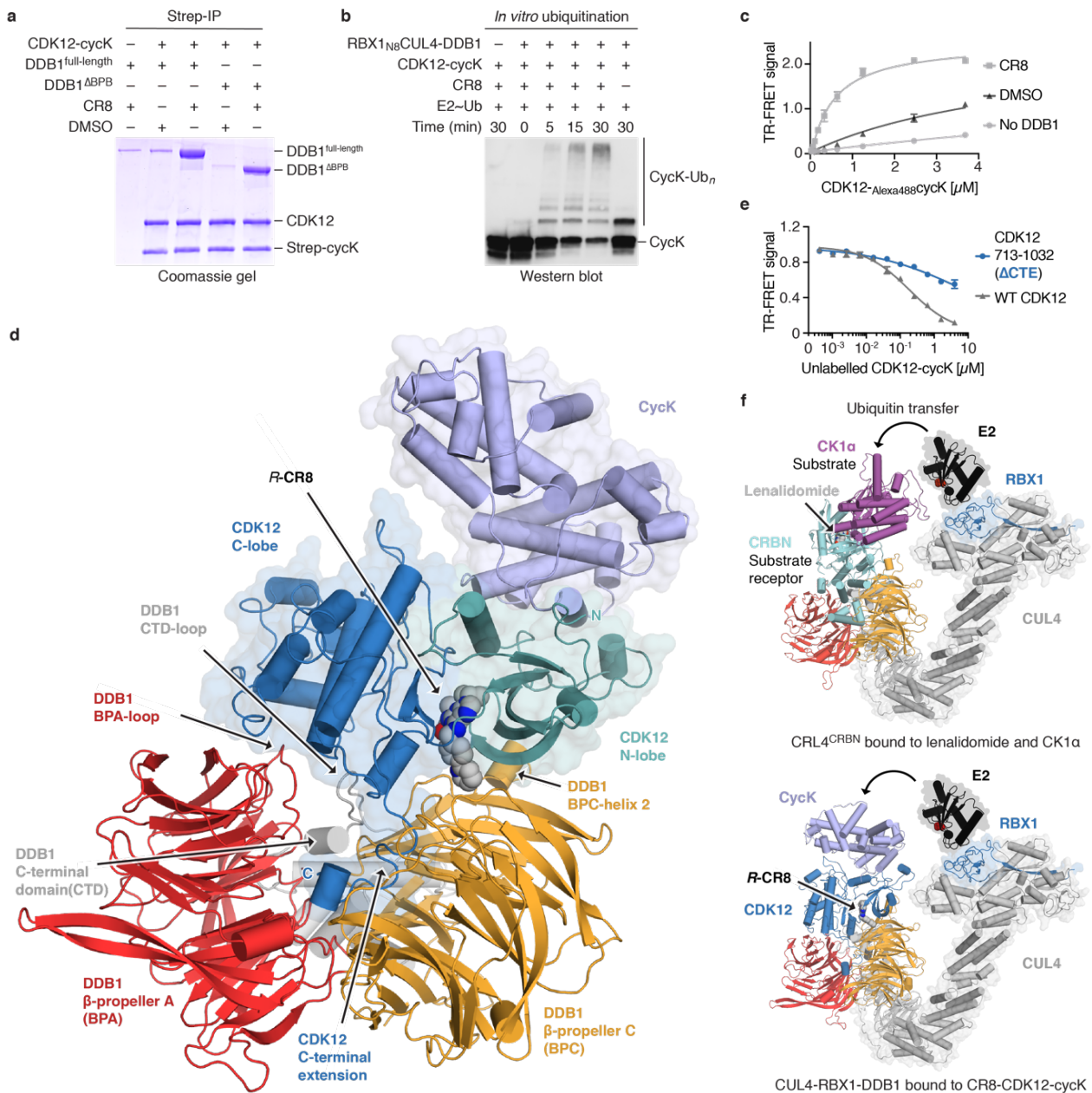
**Figure 2.7. CR8-induced cyclin K degradation is not dependent on any DCAF substrate receptor.** **a**, K562<sub>Cas9</sub>, P31FUJ<sub>Cas9</sub>, THP1<sub>Cas9</sub> and MM1S<sub>Cas9</sub> cells were exposed to the indicated concentrations of CR8 for 3 days. Data represent mean  $\pm$  s.d. Lines represent standard four-parameter log-logistic curves. **b**, mRNA expression levels for all genes included in DCAF library. **c**, K562<sub>Cas9</sub>, P31FUJ<sub>Cas9</sub>, THP1<sub>Cas9</sub> and MM1S<sub>Cas9</sub> were individually transduced with arrayed DCAF

library and treated with 1  $\mu$ M CR8 (K562<sub>Cas9</sub>, P31FUJ<sub>Cas9</sub>, THP1<sub>Cas9</sub>) or 0.1  $\mu$ M CR8 (MM1S<sub>Cas9</sub>) and ratio of transduced to untransduced cells was determined using flow cytometry (n=1). *I have not contributed to these results.*

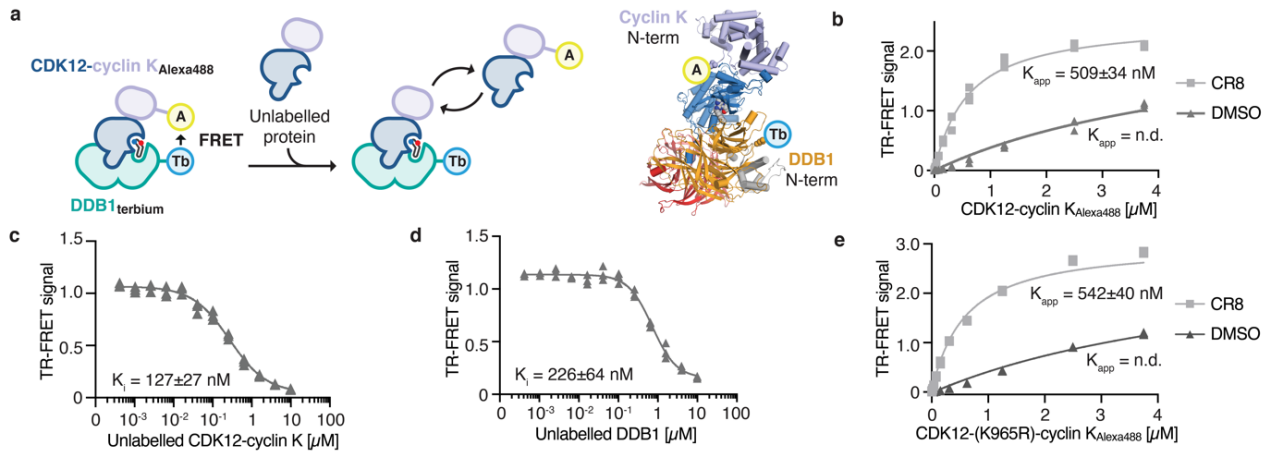
### 2.1.2.2 CR8 directs CDK12 to CRL4 core component

Since none of our genetic screens highlighted a DCAF required for cyclin K degradation, we tested whether CR8-engaged CDK12-cyclin K directly binds one of the CUL4-RBX1-DDB1 ligase components in the absence of a substrate receptor. We therefore performed *in vitro* co-immunoprecipitation experiments using recombinantly purified proteins. The kinase domain of CDK12 (CDK12<sup>713-1052</sup>) bound to cyclin K<sup>1-267</sup> did not markedly enrich DDB1 over the bead binding control in the absence of CR8, whereas equimolar amounts of the compound led to stoichiometric complex formation (**Figure 2.8a**). DDB1  $\beta$ -propeller domains A (BPA) and C (BPC)<sup>37</sup>, which are otherwise involved in DCAF binding, were sufficient for drug-induced CDK12-cyclin K recruitment. DDB1  $\beta$ -propeller B (BPB), which binds CUL4 and is not involved in DCAF binding, was dispensable for the drug-dependent interaction with CDK12-cyclin K (**Figure 2.8a**). *In vitro* ubiquitination assays confirmed that the CUL4A-RBX1-DDB1 ligase core alone is sufficient to drive robust cyclin K ubiquitination (**Figure 2.8b**). Quantification of the interaction showed that CR8 stimulated binding between CDK12-cyclin K and DDB1 in the range of 100-500 nM depending on the experimental setup (**Figure 2.8c**, **Figure 2.9**, and **Figure 2.10a**). While weak CDK12-cyclin K-DDB1 interaction was still detectable in the absence of the compound *in vitro*, CR8 strengthened complex formation 500- to 1000-fold as estimated by isothermal titration calorimetry (ITC) (**Figure 2.10a, b**). Thus, our data indicate that CR8-engaged CDK12-cyclin K is recruited to the CUL4-RBX1-DDB1 ligase core through DDB1, and the compound tightens the complex sufficiently to drive CR8-induced cyclin K degradation in the absence of a canonical DCAF substrate receptor.

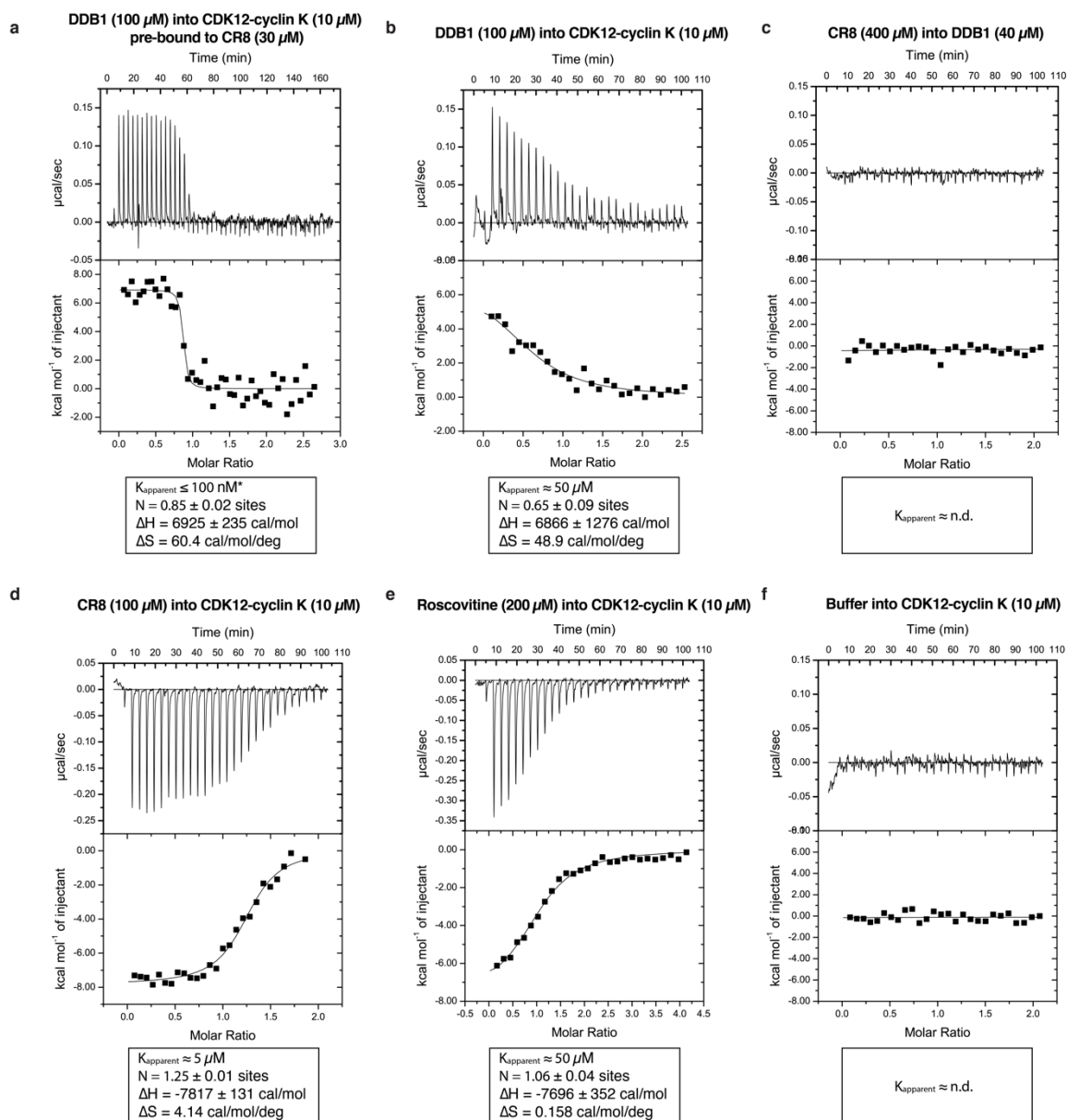




**Figure 2.8. CR8-bound CDK12 interacts with DDB1 in a DCAF-like manner.** **a**, Co-immunoprecipitation (IP) experiments with recombinant proteins. **b**, *In vitro* ubiquitination of cyclin K by the RBX1<sub>N8</sub>CUL4-DDB1 ubiquitin ligase core. **c**, TR-FRET signal for CDK12-Alexa488cyclin K titrated to <sup>Terbium</sup>DDB1 in DMSO or 10  $\mu$ M CR8. The no DDB1 control contains streptavidin-terbium and shows concentration-dependent fluorophore effects. Data represent the mean  $\pm$  s.d. (n=3). **d**, Cartoon representation of the DDB1<sup>ΔBPP</sup>-R-CR8-CDK12-cyclin K crystal structure. **e**, TR-FRET counter titration of unlabelled wild-type or mutant CDK12-cyclin K (0-10  $\mu$ M) into pre-assembled <sup>Terbium</sup>DDB1-CR8-CDK12-Alexa488cyclin K complex. Data represent the mean  $\pm$  s.d. (n=3). **f**, Structural models of CRL4<sup>CRBN</sup> bound to lenalidomide and CK1 $\alpha$  (top) and RBX1-CUL4-DDB1 (CRL4) bound to the R-CR8-CDK12-cyclin K complex (bottom). The active site cysteine of the E2 enzyme (red spheres) provides ubiquitin through a thioester bond.



**Figure 2.9. Characterization of the CR8-induced DDB1-CDK12-cyclin K interaction by TR-FRET.** **a**, Schematic of the TR-FRET setup. Positions of the FRET donor (Terbium-streptavidin (T)) and acceptor (Alexa488-labeled SpyCatcher (A)) are indicated in the structural model of DDB1<sup>ABPB</sup>CDK12-CR8-cyclin K on the right. **b**, TR-FRET titration of CDK12-Alexa488cyclin K (0-3.75  $\mu$ M) to 50 nM TerbiumDDB1 and 5  $\mu$ M CR8 or DMSO. Data represent the mean  $\pm$  s.d. (n=3). **c**, TR-FRET counter titration of unlabelled wild-type CDK12-cyclin K to 50 nM TerbiumDDB1, 500 nM CDK12-Alexa488cyclin K and 12.5  $\mu$ M CR8. Data represent the mean  $\pm$  s.d. (n = 3). **d**, TR-FRET counter titration of unlabelled wild-type DDB1 to 50 nM TerbiumDDB1, 500 nM CDK12-Alexa488cyclin K and 1  $\mu$ M CR8. Data represent the mean  $\pm$  s.d. (n=3). **e**, TR-FRET titration of CDK12(R965K)-Alexa488cyclin K (wild-type sequence of canonical isoform of CDK12; 0-3.75  $\mu$ M) to 50 nM TerbiumDDB1 and 5  $\mu$ M CR8 or DMSO. Data represent the mean  $\pm$  s.d. (n=3). The CDK12 K965R variant (residue distal from the interface with DDB1 and cyclin K), which was used throughout our *in vitro* studies (see methods), shows a binding affinity indistinguishable from that of wild-type CDK12.



**Figure 2.10. Characterization of the CR8-induced DDB1-CDK12-cyclin K interaction by isothermal titration calorimetry.** (a-f), ITC experiments ( $n = 1-3$ , additional replicates provided in the Supplementary Information of the online manuscript file). Specifications of the individual experiments are given in the panel. Asterisk marking the approximate  $K_{\text{apparent}}$  value in the first panel denotes that the binding affinity was too high to allow confident fitting of the binding curve.

We then crystallised CDK12<sup>713-1052</sup>-cyclin K<sup>1-267</sup> bound to CR8 and DDB1 <sup>$\Delta$ BPB</sup> and determined the 3.5 Å resolution structure of this complex (**Figure 2.8d**, **Table 2.1**). In the structure, CDK12 forms extensive protein-protein interactions ( $\sim 2000 \text{ \AA}^2$ ) with DDB1. CR8 binds the active site of CDK12 and bridges the CDK12-DDB1 interface, while cyclin K binds CDK12 on the opposite site and does not contact DDB1. The N- and C-lobes of CDK12 are proximal to DDB1 residues located in a loop of the BPA domain (amino acid (aa) 111-114), BPC-helix 2 (aa 986-990), and a loop in the C-terminal domain (aa 1078-1081) that are otherwise involved in DCAF binding (**Figure 2.11**). In addition, the C-terminal extension of CDK12 binds the cleft between the DDB1 domains BPA and BPC, a hallmark binding site of DDB1-DCAF interactions (**Figure 2.11a-**

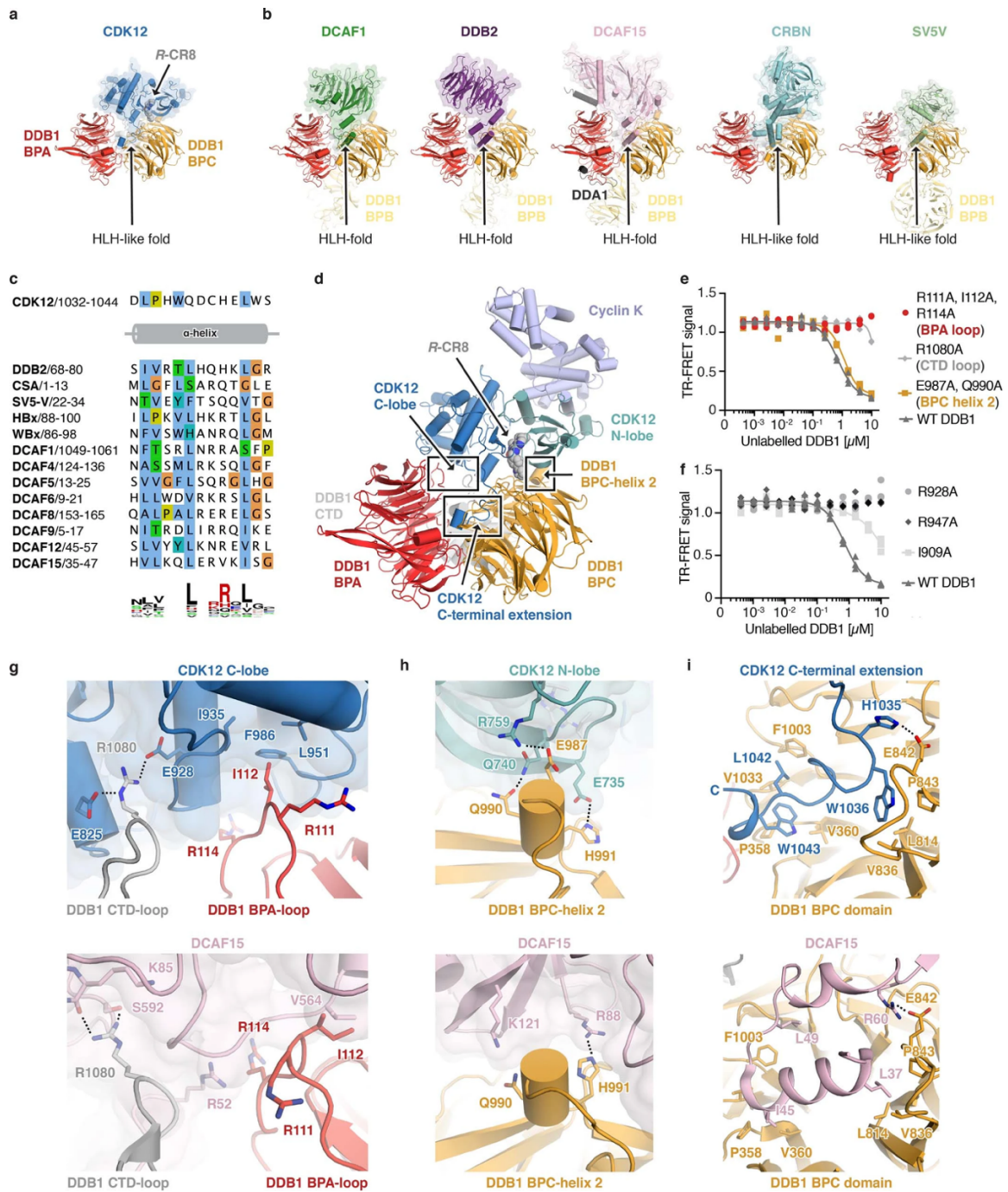
**d, h; c.f. section 1.1.3**). The density for this region could only be tentatively assigned, likely due to the presence of multiple conformations, but the CDK12 C-terminal tail clearly engages with DDB1 and assumes a conformation different from those seen in isolated CDK12-cyclin K structures (**Figure 2.12a, b, d; c.f. section 1.2.4**)<sup>218,219</sup>. Structure-guided mutational analyses combined with time-resolved fluorescence resonance energy transfer (TR-FRET) assays was used to assess the contribution of these interactions to CR8-dependent CDK12-DDB1 complex formation (**Figure 2.8e and Figure 2.11e**). CDK12 thus assumes the role of a glue-induced substrate receptor and places cyclin K in a position that is typically occupied by CRL4 substrates (**Figure 2.8f**). This renders CDK12-cyclin K binding to DDB1 mutually exclusive with that of DCAFs and provides a structural framework for why a canonical substrate receptor is dispensable for cyclin K ubiquitination.

**Table 2.1.** Data collection and refinement statistics for the CR8-induced complex.

<b>DDB1<sup>ABPB</sup>-CR8-CDK12<sup>713-1052</sup>-Cyclin K<sup>1-267</sup></b>	
<b>Data collection</b>	
Space group	<i>P</i> 3 <sub>1</sub> 21
Cell dimensions	
<i>a, b, c</i> (Å)	250.75, 250.75, 217.92
$\alpha, \beta, \gamma$ (°)	90, 90, 120
Resolution (Å)	54–3.46 (3.63–3.46)*
<i>R</i> <sub>meas</sub>	0.318 (>4.00)
<i>I</i> / $\sigma I$	7.2 (0.9)
Completeness (%)	95.1 (68.3) <sup>†</sup>
Redundancy	12.0 (11.6)
<b>Refinement</b>	
Resolution (Å)	54–3.46
No. reflections	89,183
<i>R</i> <sub>work</sub> / <i>R</i> <sub>free</sub>	0.1934 / 0.220
No. non-hydrogen atoms	
Protein	33,781
<i>R</i> -CR8	96
<i>B</i> -factors (Å <sup>2</sup> )	
Protein	59.9
<i>R</i> -CR8	39.6
R.m.s. deviations	
Bond lengths (Å)	0.009
Bond angles (°)	1.01

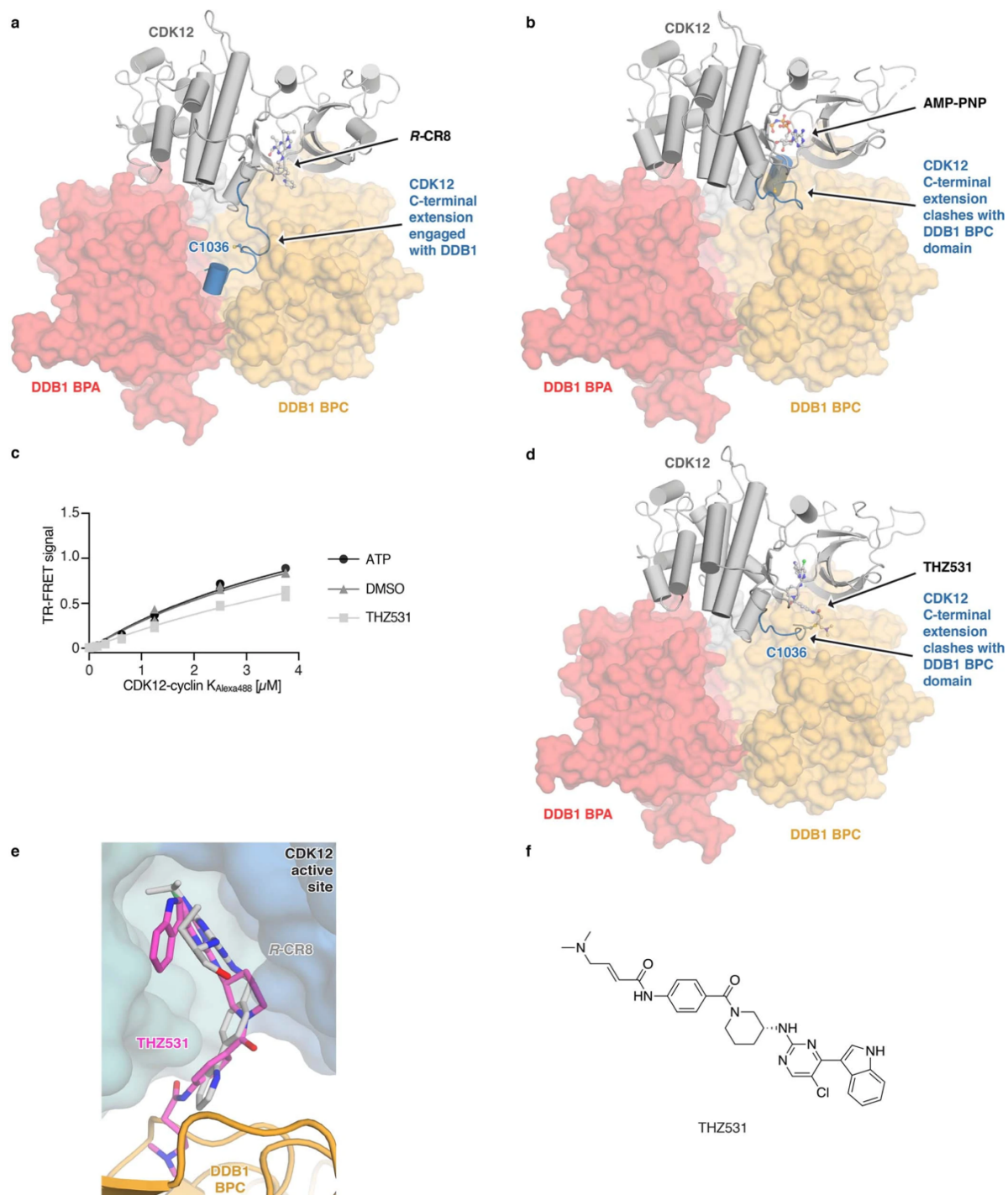
\*Values in parentheses are for the highest-resolution shell.

<sup>†</sup> From STARANISO<sup>277</sup> assuming a local weighted  $CC_{1/2}=0.3$  resolution cut-off



**Figure 2.11. CDK12 contacts sites on DDB1 otherwise implicated in DCAF binding.** **a**, Structure of the CDK12-R-CR8-DDB1<sup>ΔBPB</sup> complex. The CDK12 C-terminal domain binds a cleft between the BPA and BPC domains of DDB1 (arrow) and adopts an helix-loop-helix (HLH)-like fold. **b**, Diverse DDB1-CUL4 associated factors (DCAFs) bind DDB1 BPA and BPC domains through HLH- or HLH-like folds. **c**, Protein sequence alignment of identically positioned HLH-domain helices from different DCAFs. **d**, Protein-protein interaction hotspots highlighted in the structure of the DDB1<sup>ΔBPB</sup>CDK12-CR8-cyclin K complex. **e**, TR-FRET counter titration of unlabelled wild-type or mutant DDB1 (0-10 μM) into pre-assembled <sup>Terbium</sup>DDB1-CR8-CDK12-Alexa488cyclin K complex. Data represent the mean ± s.d. (n = 3). **f**, Close-up of the interface between the CDK12 C-lobe and DDB1 loop residues provided by the BPA and CTD domains (top). These DDB1 residues also form contacts with DCAF15 (bottom). **g**, Close-up of the interface between the CDK12 N-lobe and DDB1 BPC-helix 2 (top). Similar DDB1 residues contact DCAF15 (bottom). **h**, Close-up of the interface

between the CDK12 C-terminal extension and BPC domain of DDB1 (top). Similar DDB1 residues interact with DCAF15 (bottom).

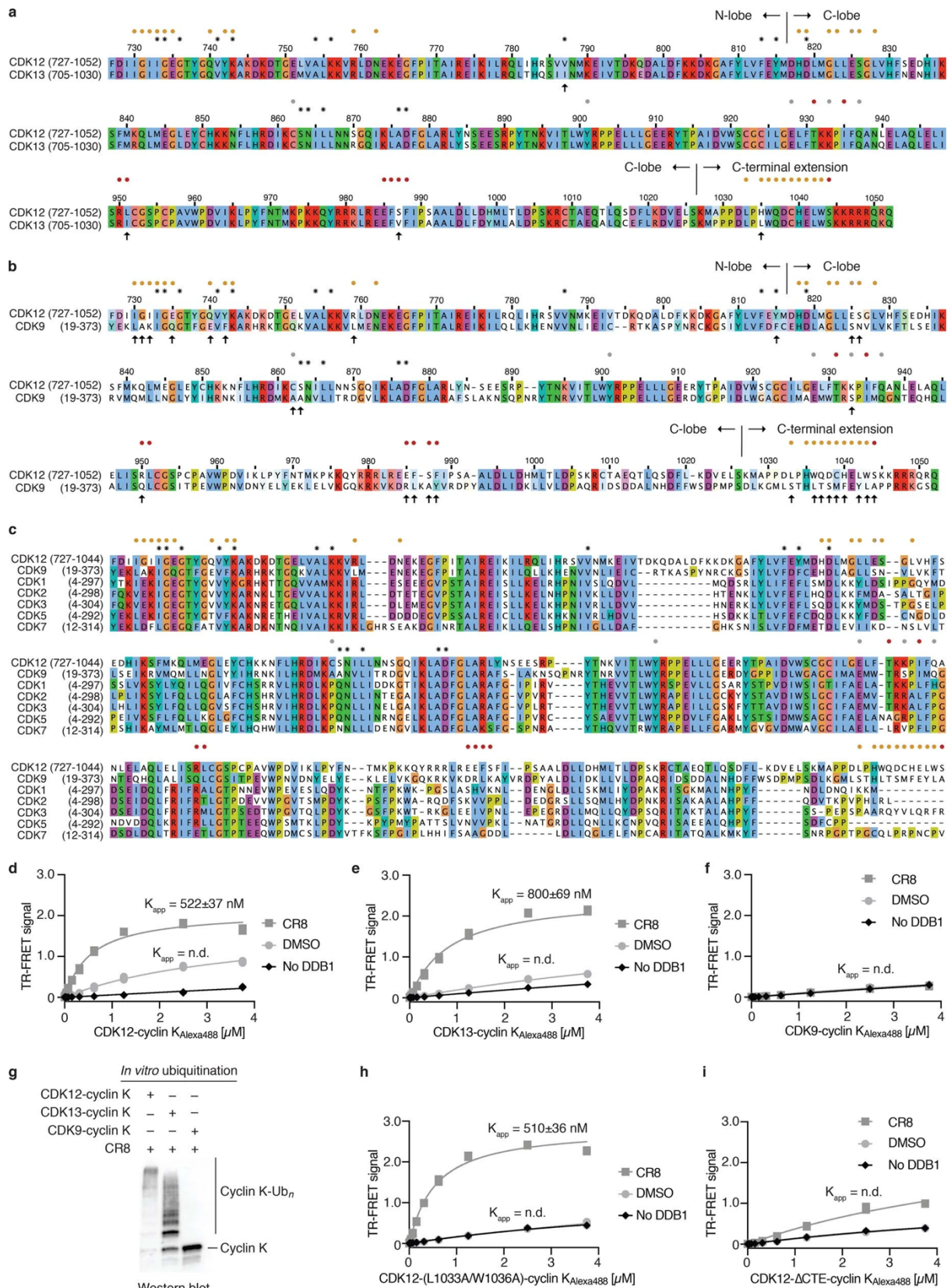


**Figure 2.12. C-terminal extension of CDK12 adopts different conformations.** **a**, Structure of the CDK12-CR8-DDB1<sup>ABPB</sup> complex highlighting position and conformation of the CDK12 C-terminal extension binding the cleft between the BPA and BPC domains of DDB1. **b**, Structure of CDK12 bound to AMP-PNP (PDB entry 4CXA) superimposed onto CDK12 in the CDK12-CR8-DDB1<sup>ABPB</sup> complex. In the AMP-PNP-bound form of CDK12, the C-terminal extension is in a conformation that allows contacts with the nucleotide analogue. This conformation of the CDK12 C-terminal extension is incompatible with DDB1 binding. **c**, TR-FRET titration of CDK12-Alexa488-cyclin K (0-3.75 μM) to 50 nM TerbiumDDB1 in the presence of 5 μM THZ531, ATP or DMSO. Data represent the mean ± s.d. (n=3). **d**, Structure of CDK12 bound to THZ531 (PDB entry 5ACB) superimposed onto CDK12 in the CDK12-CR8-DDB1<sup>ABPB</sup> complex. This conformation of

the CDK12 C-terminal extension is incompatible with DDB1 binding. **e**, THZ531 binding pose in the active site of CDK12 (PDB entry 5ACB) superimposed on the CR8-bound CDK12 in the DDB1<sup>ΔB<sup>BP</sup>B</sup>CR8-CDK12-cyclin K complex.

### 2.1.2.3 CDK12-DDB1 interface imparts selectivity

CR8 is a pleiotropic CDK inhibitor reported to bind to CDK1/2/3/5/7/9/12<sup>218,236</sup>, yet in cells we observed selective cyclin K destabilization in the presence of the drug. As cyclin K was reported to associate with CDK9, CDK12 and CDK13, we tested whether the other cyclin K-dependent kinases are also recruited to DDB1. The closely related CDK13 (90.8 % sequence identity in the kinase domain, **Figure 2.13a**), but not the more divergent CDK9 (45.5 % sequence identity in the kinase domain, **Figure 2.13b, c**), was recruited to DDB1 in the presence of CR8, albeit with a slightly lower binding affinity (**Figure 2.13d-f**). Analogously, less productive *in vitro* cyclin K ubiquitination was observed for CDK13 compared to CDK12 (**Figure 2.13g**). The main difference between CDK9 and CDK12/13 primary sequence lies in their C-terminal extension (**Figure 2.13a, b**), which in our structure nestles up against DDB1 BPA and BPC propellers (**Figure 2.8d and Figure 2.11d, h**). Mutations in, or truncation of, the CDK12 C-terminal extension (**Figure 2.11c**) abolished any compound-independent binding between CDK12 and DDB1, whereas complex formation could still be facilitated by CR8 to a varying extent (**Figure 2.13h, i**). Hence, our data show that the pan-selective CDK inhibitor CR8 induces specific protein-protein interactions between CDK12/13 and DDB1 and suggest that the C-terminal extension, while contributing to binding, is not essential for drug-dependent kinase recruitment.



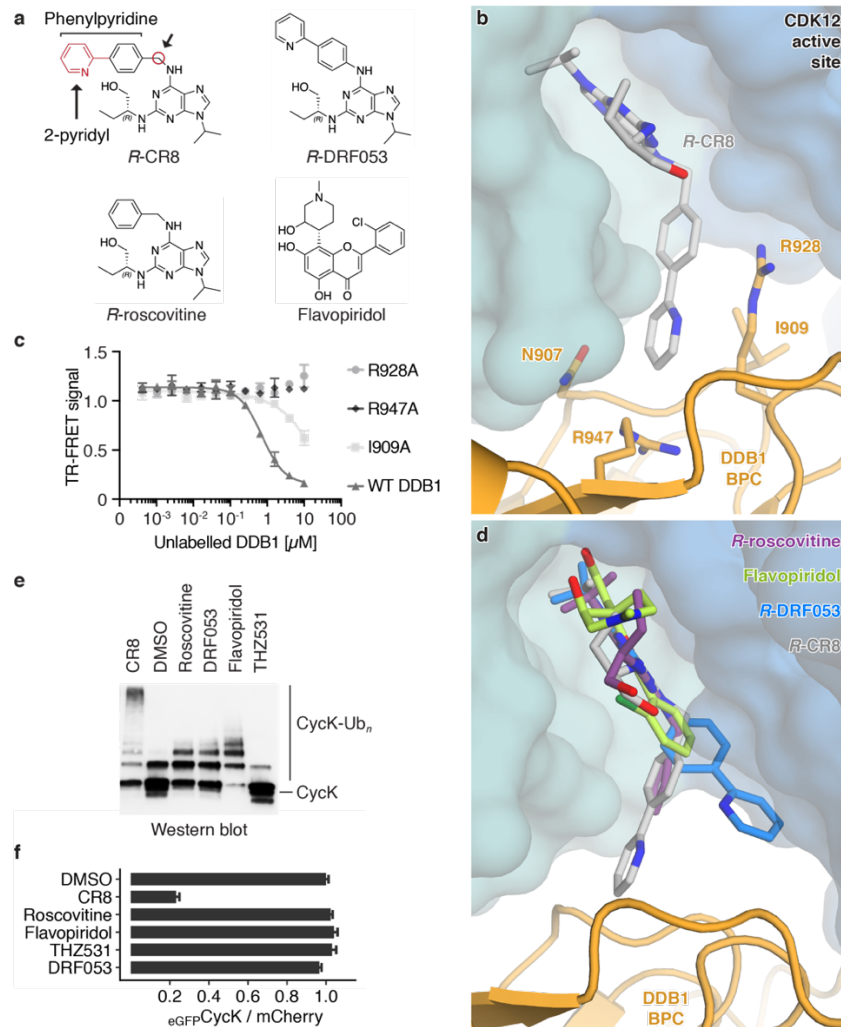
**Figure 2.13. Interface differences between CDK12 and other CDKs highlight residues involved in CR8-mediated recruitment to DDB1.** **a**, Protein sequence alignment of CDK12 (727-1052) and CDK13 (705-1030). In this and later panels residues that contact CR8 are marked by an asterisk, those that contact DDB1 by dots that are coloured according to the domain organisation of DDB1 (see **Figure 2.8**). Arrows mark differences at the DDB1-CR8-CDK interface. **b**, Protein sequence alignment of CDK12 (727-1052) and CDK9 (19-373). **c**, Protein sequence alignment of CDK12 (727-



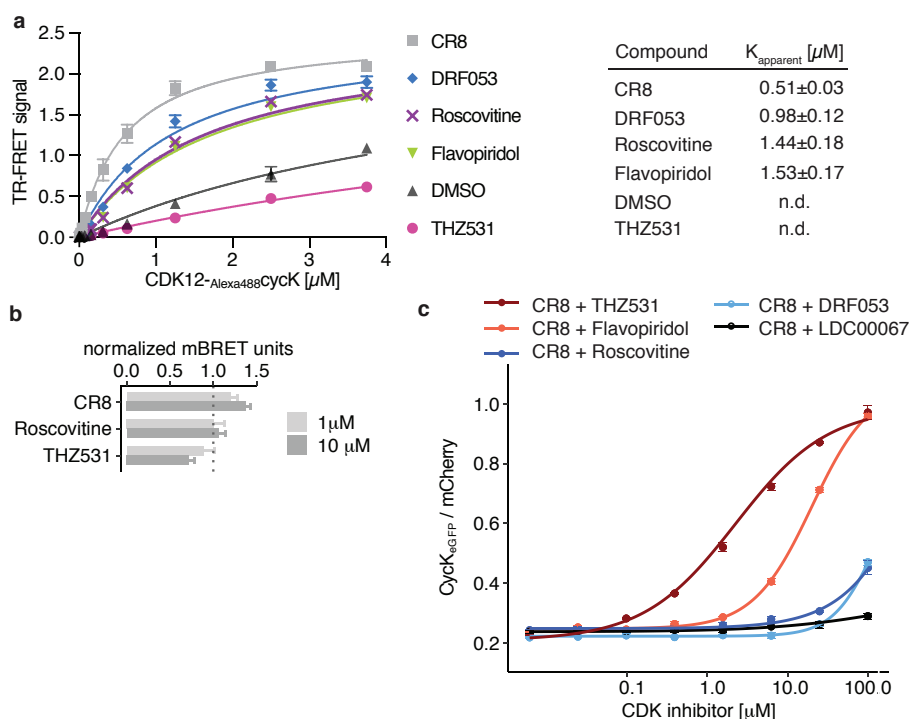
1052), CDK9 (19-373), CDK1 (4-297), CDK2 (4-298), CDK3 (4-304), CDK5 (4-292), CDK7 (12-314). **d**, TR-FRET titration of CDK12-Alexa488cyclin K (0-3.75  $\mu$ M) to 50 nM <sup>Terbium</sup>DDB1 and 5  $\mu$ M CR8 or DMSO. The no-DDB1 control in this and subsequent panels contained streptavidin-terbium and shows concentration-dependent fluorophore effects. Data represent the mean  $\pm$  s.d. (n=3). **e**, TR-FRET titration of CDK13-Alexa488cyclin K (0-3.75  $\mu$ M) to 50 nM <sup>Terbium</sup>DDB1 and 5  $\mu$ M CR8 or DMSO. Data represent the mean  $\pm$  s.d. (n=3). **f**, TR-FRET titration of CDK9-Alexa488cyclin K (0-3.75  $\mu$ M) to 50 nM <sup>Terbium</sup>DDB1 and 5  $\mu$ M CR8 or DMSO. Data represent the mean  $\pm$  s.d. (n=3). **g**, *In vitro* ubiquitination of cyclin K by the RBX1<sub>N8</sub>CUL4-DDB1 ubiquitin ligase core in the presence of CDK12, CDK13 or CDK9. **h**, TR-FRET titration of CDK12-Alexa488cyclin K (CDK12 mutant (L1033A, W1036A); 0-3.75  $\mu$ M) to 50 nM <sup>Terbium</sup>DDB1 and 5  $\mu$ M CR8 or DMSO. Data represent the mean  $\pm$  s.d. (n=3). **i**, TR-FRET titration of CDK12-Alexa488cyclin K (CDK12 tail truncation (713-1032); 0-3.75  $\mu$ M) to 50 nM <sup>Terbium</sup>DDB1 and 5  $\mu$ M CR8 or DMSO. Data represent the mean  $\pm$  s.d. (n=3).

#### 2.1.2.4 CR8 phenylpyridine confers glue activity

CR8 occupies the ATP binding pocket of CDK12 and forms discrete contacts with residues in the BPC domain of DDB1 ( $\sim$ 150  $\text{\AA}^2$ ) through its hydrophobic phenylpyridine ring system (**Figure 2.14a, b**). Mutation of the DDB1 residues Ile909, Arg928, and Arg947 each diminished drug-induced recruitment of the kinase (**Figure 2.14c**), highlighting the contribution of the phenylpyridine moiety to complex formation. To evaluate the structure-activity relationship underlying the gain-of-function activity of CR8, we probed other CDK inhibitors for their ability to drive complex formation between DDB1 and CDK12. DRF053<sup>239</sup>, a CR8-related inhibitor that carries a differently linked phenylpyridine ring system (**Figure 2.14a, d**), induced binding with two-fold lower affinity than CR8 (**Figure 2.15a**). Roscovitine<sup>238</sup>, the parent compound of CR8 that lacks the 2-pyridyl substituent but retains the phenyl ring proximal to Arg928 (**Figure 2.14a, d**), also facilitated complex formation, albeit with a three-fold lower apparent affinity (**Figure 2.15a**). The affinity rank-order observed in our TR-FRET assay correlated with the degree of cyclin K ubiquitination *in vitro*, in which DRF053 and roscovitine showed less processive ubiquitination (**Figure 2.14e**). As neither DRF053 nor roscovitine induced degradation of the cyclin K<sub>eGFP</sub> reporter in cells (**Figure 2.14f**), our results demonstrate that the presence and correct orientation of the 2-pyridyl on the surface of CDK12 confer the gain-of-function activity of CR8 leading to cyclin K degradation.



**Figure 2.14. Molecular glue degrader activity of CR8 is conferred by a surface-exposed 2-pyridyl moiety.** **a**, Chemical structures of R-CR8, R-DRF053, R-roscovitine and flavopiridol. Arrows indicate differences between R-CR8, R-DRF053 and R-roscovitine. **b**, Close-up of the CDK12-CR8-DDB1 interface. The phenylpyridine moiety of CR8 contacts DDB1 residues. **c**, Titration of unlabelled wild-type or mutant DDB1 (0-10  $\mu$ M) into pre-assembled <sup>Terbium</sup>DDB1-CR8-CDK12-Alexa488cyclin K complex. Data represent the mean  $\pm$  s.d. (n=3). **d**, R-roscovitine (PDB entry 2A4L), R-DRF053 and flavopiridol (PDB entry 3BLR) in the active site of CDK12 in the DDB1-CR8-CDK12-cyclin K complex through superposition of the respective kinase domains or the purine moiety (in case of DRF053). **e**, *In vitro* ubiquitination of CDK12-cyclin K complex by RBX1<sub>N8</sub>CUL4-DDB1 in the absence (DMSO) or presence of 2  $\mu$ M compound. **f**, Cyclin K<sub>eGFP</sub> HEK293T<sub>Cas9</sub> cells were treated with 1  $\mu$ M of the indicated compound for 2 hours and analysed by flow cytometry. Data represent the mean  $\pm$  s.d. (n=3).



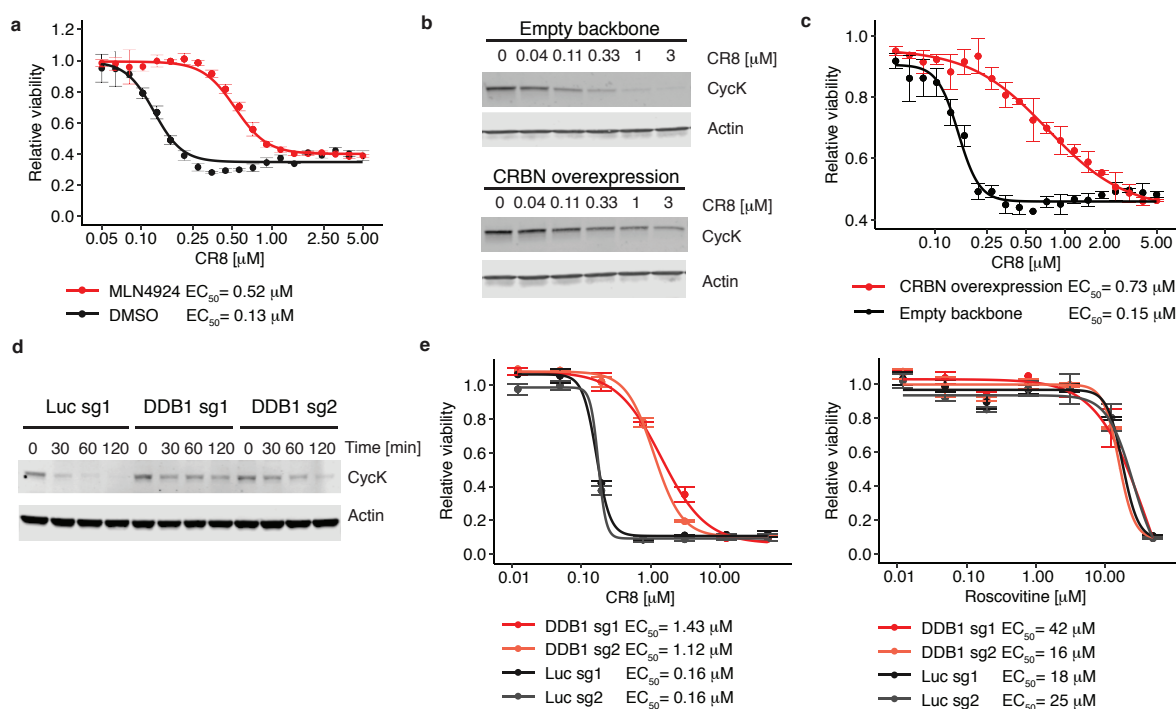
**Figure 2.15. CDK inhibitors block CR8-induced cyclin K degradation.** **a**, TR-FRET. CDK12-Alexa488Cyclin K titrated to TerbiuimDDB1 in the absence (DMSO) or presence of 10  $\mu\text{M}$  CR8, roscovitine, DRF053, flavopiridol or THZ531. Data represent the mean  $\pm$  s.d. (n=3). **b**, HEK293T cells transiently transfected with NanoLucCDK12<sup>713-1052</sup> and HaloTagDDB1<sup>ABPB</sup> constructs were treated with DMSO or inhibitors for 2 hours and the mBRET signal was measured. Data represent the mean  $\pm$  s.d. (n=3). **c**, Cyclin K<sub>eGFP</sub> HEK293T<sub>Cas9</sub> cells were treated with 1  $\mu\text{M}$  CR8 and varying concentrations of competitive CDK inhibitor and analysed by flow cytometry. Data represent the mean  $\pm$  s.d. (n=3). *I have not contributed to the results presented in panels b and c.*

To test whether any ligand could in principle drive the interaction of CDK12 with DDB1, we tested the endogenous CDK nucleotide cofactor ATP for its ability to promote complex formation. ATP neither facilitated nor abrogated the interaction over basal binding observed in the presence of DMSO (**Figure 2.12c**), suggesting that although the nucleotide-bound conformation of CDK12 seems incompatible with the approach of DDB1 (**Figure 2.12b**), the C-terminal extension is in fact free to adopt multiple conformations<sup>219</sup>. THZ531<sup>220</sup>, a bulky covalent CDK12/13 inhibitor predicted to clash with DDB1(**Figure 2.12d, e** and **section 1.2.6**), locks the CDK12 C-terminal tail in a conformation incompatible with DDB1 recruitment (**Figure 2.12d**). Consistently, THZ531 further decreased the TR-FRET signal and diminished cyclin K ubiquitination *in vitro* below DMSO control levels (**Figure 2.14e** and **Figure 2.12c**)<sup>220</sup>. Flavopiridol<sup>278</sup>, a natural product-derived inhibitor structurally distinct from CR8 (**Figure 2.14a, d**), also stimulated the binding of CDK12-cyclin K to DDB1 (**Figure 2.15a**). While flavopiridol gave rise to moderate cyclin K ubiquitination *in vitro* (**Figure 2.14e**), it failed to degrade cyclin K in cells (**Figure 2.14f**). Our results thus show that the interactions between the compound and DDB1 display a significant plasticity and that structurally diverse surface-exposed moieties in CR8, DRF053, roscovitine and flavopiridol can facilitate CDK12-cyclin K recruitment. Small differences in their ability to stabilise the DDB1-CDK12 complex translate, in an almost binary fashion, into cellular

degradation of cyclin K or lack thereof. This behaviour is reminiscent of CRL4<sup>CRBN</sup> and thalidomide analogues<sup>44,113</sup>, where an apparent affinity threshold must be overcome to drive drug-induced target degradation.

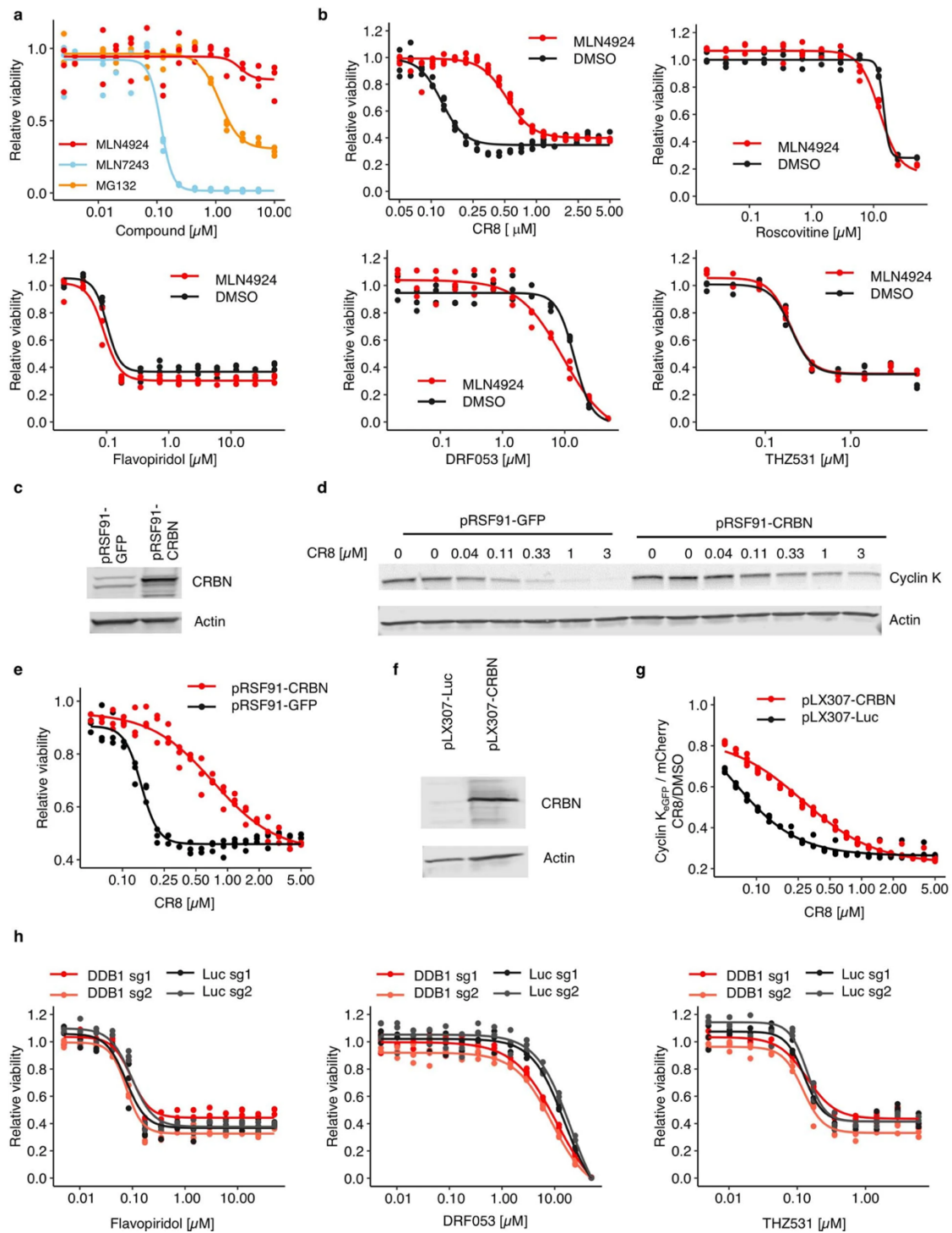
### 2.1.2.5 Cyclin K degradation adds to CR8 toxicity

Finally, to delineate the contribution of CRL4-mediated cyclin K degradation to the cellular cytotoxicity of CR8 over non-degradative CDK inhibition, we compared compound toxicity in wild-type HEK293T<sub>Cas9</sub> cells to cells that were pre-treated with MLN4924 (NEDD8-activating enzyme inhibitor), genetically depleted for DDB1, or subject to DCAF overexpression. Global inhibition of CRL activity by MLN4924 had only minor effects on cell viability (**Figure 2.17a**) but resulted in decreased sensitivity to CR8 (**Figure 2.16a**), showing that CRL neddylation significantly contributes to CR8 toxicity. Overexpression of the substrate receptor CRBN also affected sensitivity to CR8 and decreased cyclin K degradation (**Figure 2.16b, c** and **Figure 2.17c-e**), presumably by reducing the free pool of DDB1. As expected, CR8-induced endogenous cyclin K degradation was dependent on DDB1 (**Figure 2.16d**) and, consistently, we found that cytotoxicity of CR8, but not that of the other CDK inhibitors, was ten-fold lower in cells depleted for DDB1 (**Figure 2.16e** and **Figure 2.17f**). Together, the data demonstrate that the CLR4-dependent gain-of-function glue degrader activity of CR8 strongly contributes to its cellular potency and provides an additional layer of ortholog-specific CDK inactivation through cyclin K degradation.



**Figure 2.16. CR8-mediated cyclin K degradation contributes to its cellular potency.** **a**, HEK293T<sub>Cas9</sub> cells were exposed to DMSO or 100 nM MLN4924 at indicated concentrations of CR8 for 3 days. This dose of MLN4924 alone did not influence cell viability (**Figure 2.17a**). Data represent mean  $\pm$  s.d. Lines represent standard four-parameter log-logistic curves ( $n=3$ ). **b**, HEK293T<sub>Cas9</sub> cells were transiently transfected with control or CRBN overexpression vectors and after 48h lysates were immunoblotted for the indicated targets. **c**, HEK293T<sub>Cas9</sub> cells were transiently transfected with control or CRBN overexpression vectors and after 48h were exposed to the indicated concentration of CR8 for 3 days. Data

represent mean  $\pm$  s.d. Lines represent standard four-parameter log-logistic curves ( $n=3$ ). **d**, HEK293T<sub>Cas9</sub> cells were transfected with a fluorophore and sgRNAs targeting DDB1 or a non-targeting control (NTC) and exposed to CR8 for the indicated time. Protein lysates were immunoblotted for the indicated targets ( $n=2$ , representative image shown). **e**, HEK293T<sub>Cas9</sub> cells were transduced with sgRNAs targeting DDB1 or luciferase and exposed to the indicated concentration of CR8, roscovitine, or DMSO for 3 days. Data represent mean  $\pm$  s.d. Lines represent standard four-parameter log-logistic curves ( $n=3$ ). *I have not contributed to these results.*



**Figure 2.17. Cytotoxicity of CR8 analogues does not depend on CRL4 components.** **a**, HEK293T<sub>Cas9</sub> cells were exposed to indicated concentrations of MLN4924, MLN7243 or MG132 for 3 days. Data represent mean  $\pm$  s.d. Lines represent standard four-parameter log-logistic curves. **b**, HEK293T<sub>Cas9</sub> cells were exposed to DMSO or 100 nM MLN4924

in combination with indicated concentrations of roscovitine, flavopiridol, DRF053 or THZ531 for 3 days. Data represent mean  $\pm$  s.d. Lines represent standard four-parameter log-logistic curves. **c, d**, HEK293T<sub>Cas9</sub> cells were transiently transfected with control or CRBN overexpression vector and after 48h lysates were immunoblotted for the indicated targets. pRSF91-GFP and pRSF91-CRBN are denoted as empty backbone or CRBN overexpression respectively. **e**, Cyclin K<sub>eGFP</sub> HEK293T<sub>Cas9</sub> cells were transiently transfected with control or CRBN overexpression vector and after 48h were exposed to the indicated concentrations of CR8 for 2h and analysed by flow cytometry. Data represent mean  $\pm$  s.d. Lines represent standard four-parameter log-logistic curves. **f**, HEK293T<sub>Cas9</sub> cells were transduced with sgRNAs targeting DDB1 or luciferase and exposed to the indicated concentrations of CR8, roscovitine or DMSO for 3 days. Data represent mean  $\pm$  s.d. Lines represent standard four-parameter log-logistic curves. *I have not contributed to these results.*

### 2.1.3 Discussion

Kinase inhibitors have long been suspected to have a degradation component to their mode of action<sup>279,280</sup>, and our work provides the first characterization and structural dissection of how a kinase inhibitor scaffold acquires degrader properties. Small molecule glue degraders have thus far only been shown to target substrate-recruiting E3 ligase modules. CDK12 is not a constitutive E3 ligase component, but instead serves as a drug-induced substrate receptor, linking DDB1 to the ubiquitination target, bypassing the requirement for a canonical DCAF. While cyclin K is the primary ubiquitination target, CDK12 may become subject to autoubiquitination upon prolonged compound exposure akin to canonical DCAFs<sup>106,281</sup>. As CR8 binds the active site of CDK12 and does not require a ligandable pocket on the ligase, developing target-based glue degraders to essential ligase components such as DDB1 could greatly expand the repertoire of ubiquitin ligases accessible to targeted protein degradation. Furthermore, as kinase inhibitors often show poor target selectivity, small molecule-induced kinase inactivation that leverages specific protein-protein interactions could offer a path towards improved drug selectivity and may facilitate the pursuit of CDK12 as an emerging therapeutic target<sup>234</sup>.

The gain-of-function glue degrader activity of CR8 is attributed to a 2-pyridyl moiety exposed on the kinase surface. Surface-exposed single residue mutations have been shown to promote higher-order protein assemblies, as the haemoglobin Glu to Val mutation, for example, induces polymerization in sickle cell anaemia<sup>282</sup>. Accordingly, single residue mutations designed to increase surface hydrophobicity give rise to ordered protein aggregates<sup>283,284</sup>. Bound compounds, such as enzyme inhibitors, can in principle mimic such amino acid changes with dramatic effects on the protein interaction landscape, suggesting that compound-induced protein-protein interactions may be more common than previously recognised. Taken together, our results suggest that modifications of surface-exposed regions in target-bound small molecules offer a rational strategy to develop molecular glue degraders for a given protein target.

## 2.2 Rational discovery of cyclin K molecular glue degraders via scalable chemical profiling (*summary of a published manuscript*)

This section summarises a study led by the Winter laboratory (CeMM, Vienna), to which I contributed through *in vitro* reconstitution of the DDB1-CDK12-cyclin K ternary complex with molecular glue degraders identified

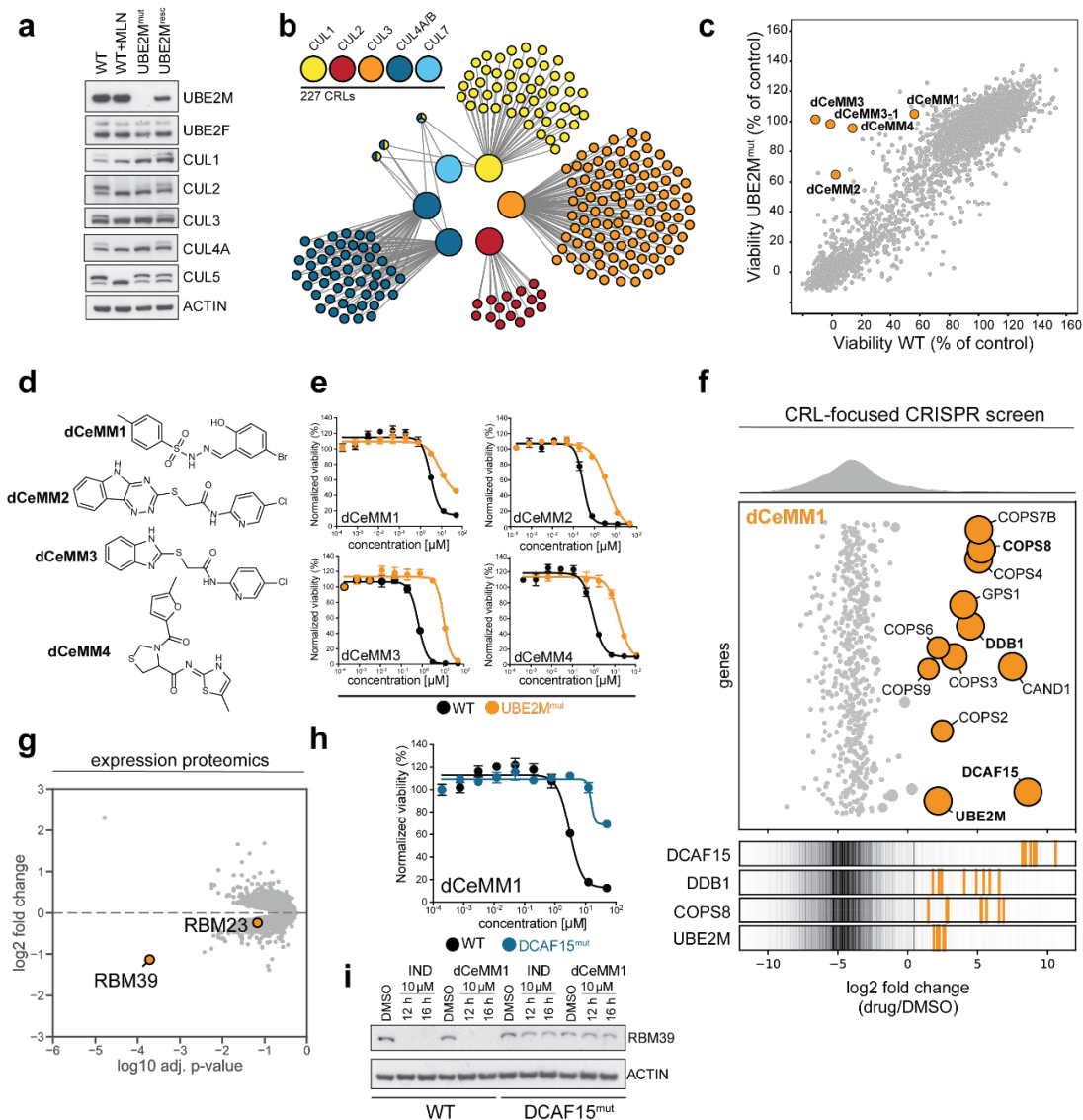
in this work, and the quantification of these interactions through TR-FRET assays. My contribution is displayed in **Figure 2.21j** and **Figure 2.22**.

### **2.2.1 Compound toxicity profiling in hyponeddylated versus wild-type cells identifies novel degrader compounds**

To discover novel molecular degraders in a systematic manner, we developed a scalable strategy towards glue degrader discovery that is based on chemical screening in hypo-neddylated cells. Almost all known molecular glue degraders function by hijacking CRLs, whose activity is dependent on the reversible attachment of NEDD8 (see **section 1.1.2**). CRISPR/Cas9-induced mutation of the NEDD8 E2 enzyme UBE2M was shown to give rise to resistance to degrader compounds<sup>244</sup>. We characterised the hypomorphic phenotype in the near-haploid KBM7 cell line (UBE2M<sup>mut</sup>) and found that CRL activity was inhibited in a manner comparable to drug-induced inhibition of the NEDD8 E1 enzyme with MLN4924, but with no associated proliferation defects (**Figure 2.18a, b**).

We hypothesised that comparative chemical profiling in UBE2M<sup>mut</sup> versus neddylation-proficient cells could constitute a viable strategy to screen for compounds whose mechanism of action depends on functional CRLs. We therefore screened a library of 2,000 small molecules both in wildtype and UBE2M<sup>mut</sup> KBM7 cells, and the cross-comparison between the two datasets led us to identify several compounds that appeared to be functionally dependent on CRL activity (**Figure 2.18c**). Through individual validation, we confirmed that dCeMM1, dCeMM2, dCeMM3, and dCeMM4 display UBE2M-dependent cytotoxicity profiles (**Figure 2.18d, e**), and prioritised these four compounds for target deconvolution.

Firstly, for dCeMM1, we performed a focused CRISPR/Cas9 screen using a custom-designed sgRNA library covering the known CRLs and their associated regulators, which confirmed a critical dependence of dCeMM1 efficacy on UBE2M, but also revealed its reliance on the CRL4<sup>DCAF15</sup> ligase (**Figure 2.18f**). This is in agreement with previous studies describing aryl sulphonamide molecular glue degrader compounds that direct DCAF15 to ubiquitinate RBM39 and RBM23<sup>42,134,135,137,142</sup>. Quantitative proteomics experiments showed that dCeMM1 selectively degrades RBM39 (**Figure 2.18g**), and CRISPR-induced frameshift-mutations in DCAF15 were found to desensitize cells to dCeMM1 toxicity and rescue RBM39 depletion (**Figure 2.18h, i**). These results therefore identify a novel aryl sulphonamide molecular glue degrader and offer a proof-of-concept for the degrader discovery pipeline.



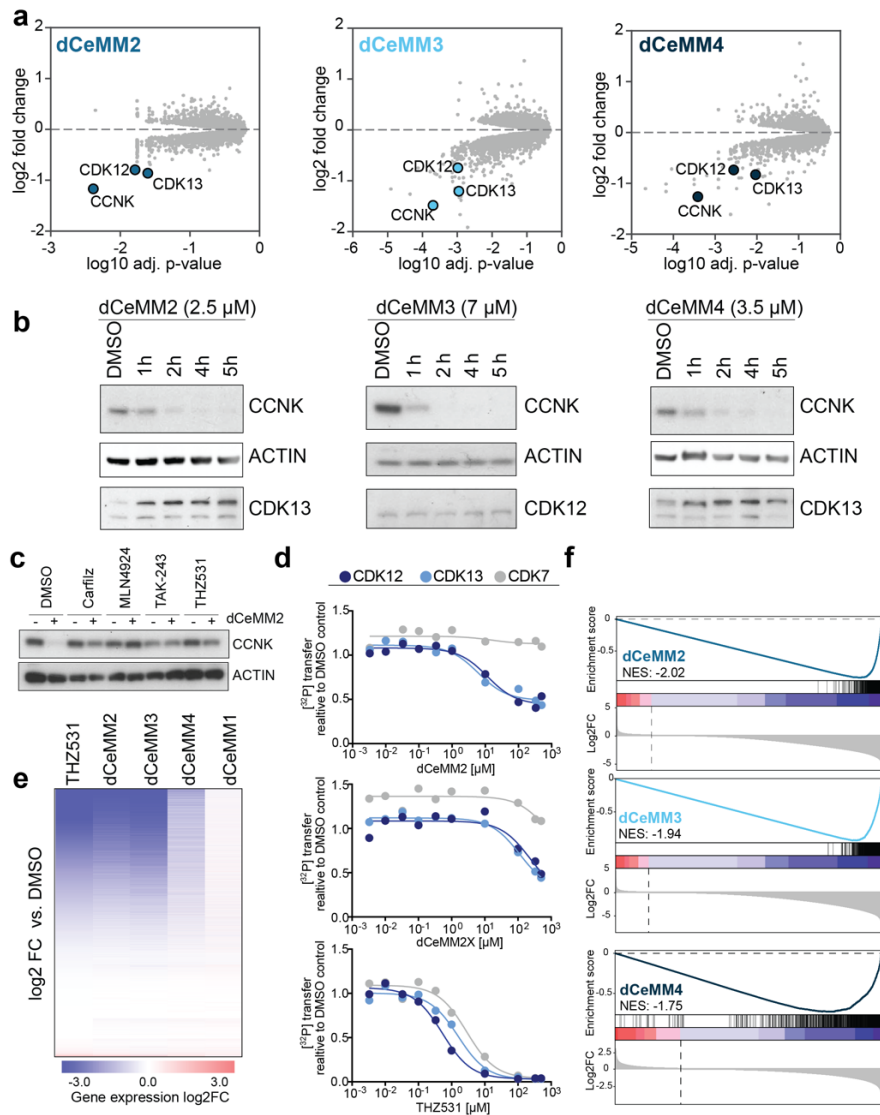
**Figure 2.18** **a**, Protein levels of UBE2M, UBE2F and neddylation status of different cullin backbones in WT, MLN4924-treated (1  $\mu$ M, 1 h), UBE2M-deficient and UBE2M-reconstituted KBM7 cells. **b**, SRs expressed in KBM7 cells that associate with hyponeeddylated cullin scaffolds. **c**, Primary screening data comparing DMSO-normalised viability of WT and UBE2M<sup>mut</sup> KBM7 cells treated with an (approximately) 2,000 cytotoxic/cytostatic small-molecule library for 3 d. Doses tested were 10  $\mu$ M and 500 nM. **d**, Chemical structure of the four prioritised chemical scaffolds (dCeMM1/2/3/4). dCeMM3-1 was structurally identical to dCeMM3 with a Br instead of the Cl. **e**, Dose-resolved, DMSO-normalised viability after 3-d dCeMM1/2/3/4 treatment in WT and UBE2M<sup>mut</sup> KBM7 cells. Mean  $\pm$  s.e.m.;  $n = 3$  independent treatments. EC<sub>50</sub>s dCeMM1/2/3/4 ( $\mu$ M) WT = 3, 0.3, 0.6 and 0.4 and UBE2M<sup>mut</sup> = 8, 4.2, 10.7 and 7. **f**, CRL-focused CRISPR resistance screen for dCeMM1. Top, bubble plot displaying median sgRNA enrichment over DMSO, bubble size indicates significance. Bottom, sgRNAs enrichment targeting indicated genes, background indicates distribution of all sgRNAs. Results shown are the median of two independent screens. **g**, Expression proteomics after dCeMM1 treatment (25  $\mu$ M, 12 h). **h**, DMSO-normalised viability after 3-d dCeMM1 treatment in WT and DCAF15<sup>mut</sup> KBM7 cells. Mean  $\pm$  s.e.m.;  $n = 3$  independent treatments. **i**, RBM39 levels in WT and DCAF15<sup>mut</sup> KBM7 cells after indisulam (IND) or dCeMM1 treatment. Reproduced from <sup>285</sup>. *I have not contributed to these results.*



### 2.2.2 dCeMM2-4 are cyclin K degraders

To identify the mode of action of compounds dCeMM2-4, we followed the same multi-omics target deconvolution approach as for dCeMM1. Expression proteomics revealed that all three compounds lead to robust destabilization of cyclin K, with milder depletion of its cognate kinases CDK12 and CDK13 (**Figure 2.19a**). Through immunoblotting, we showed that near-complete cyclin K degradation occurred within two hours (**Figure 2.19b**) and that this degradation was dependent on the proteasome (carfilzomib), E1 ubiquitin-activating enzyme (TAK-243), and NAE (MLN4924) (**Figure 2.19c**). Importantly, saturating the active site of CDK12/13 through treatment with THZ531, a CDK12/13 covalent inhibitor, rescued cyclin K degradation, indicating that engagement of the CDK12/13 ATP pocket is required for cyclin K degradation (**Figure 2.19c**). We therefore tested whether dCeMM2-4 inhibit CDK12/13 and observed a measurable activity, albeit ten-fold lower than that of THZ531 and hence unlikely to account for their cytotoxicity (**Figure 2.19d**).

To further validate CDK12/13-cyclin K as the target of dCeMM2-4, we performed mRNA-sequencing after drug exposure and observed that the resulting transcriptional profiles resembled those obtained upon selective CDK12/13 inhibition with THZ531, showing robust induction of apoptosis without a phase-specific cell-cycle arrest (**Figure 2.19e, f**). This functionally validated cyclin K as the phenotypically relevant target of dCeMM2-4.

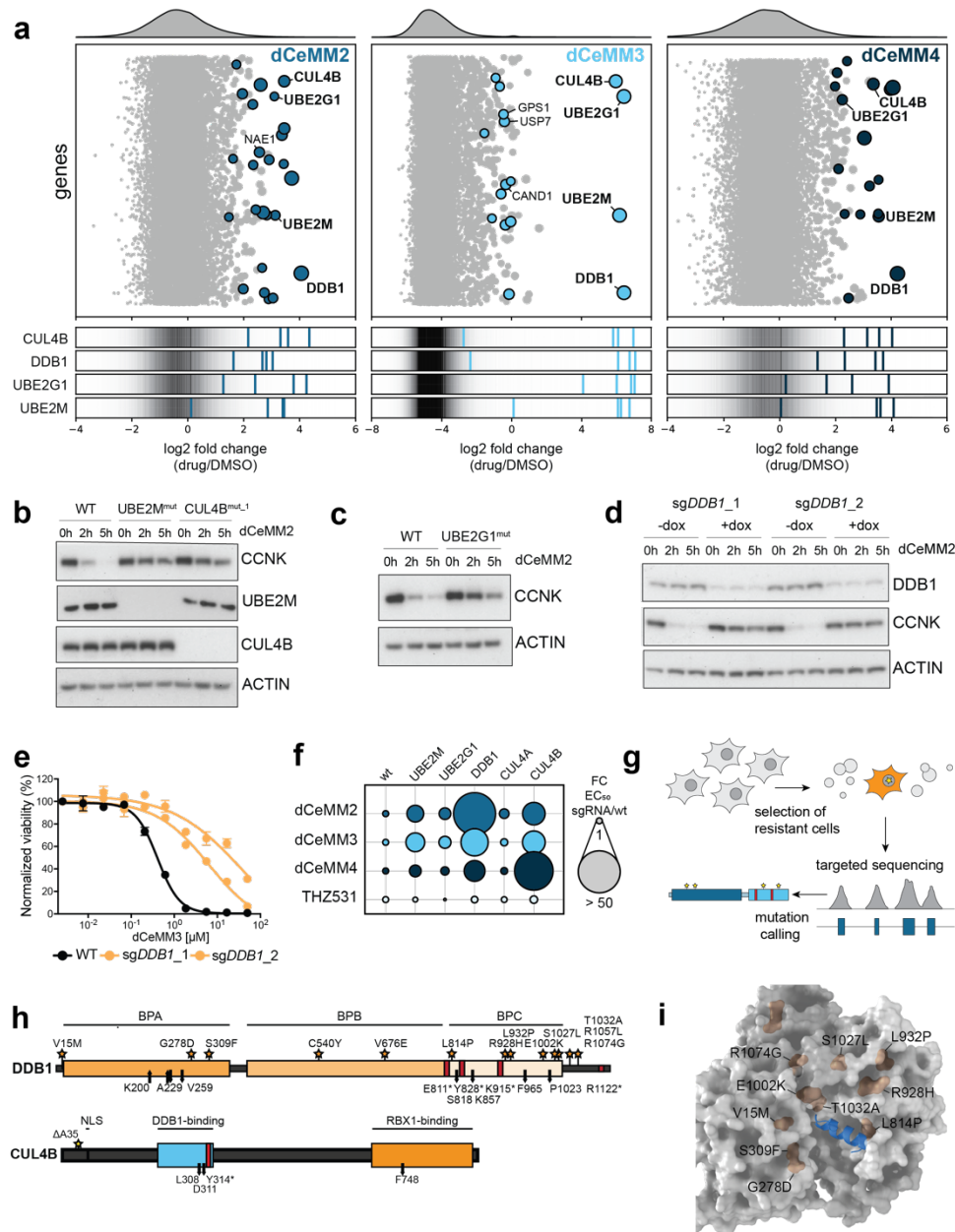


**Figure 2.19** **a**, DMSO-normalised expression proteomics after 5 h dCeMM2/3/4 treatment (2.5, 7 and 3.5  $\mu\text{M}$ ) in KBM7 cells. Limma statistical analysis was used. **b**, Cyclin K levels after dCeMM2/3/4 treatment in KBM7 cells. **c**, dCeMM2 (2.5  $\mu\text{M}$ , 5 h) destabilises cyclin K. 30 min pre-treatment with 1  $\mu\text{M}$  carfilzomib, 1  $\mu\text{M}$  MLN4924, 10  $\mu\text{M}$  TAK243 or 1  $\mu\text{M}$  THZ531 rescues cyclin K destabilization. **d**, Recombinant kinase assays of dCeMM2/2X and THZ531 inhibition on enzymatic activity of CDK12/13/7. Mean  $\pm$  s.d.,  $n = 2$ . **e**, dCeMM2/3/4 induce global transcriptional downregulation with phenotypic similarity to CDK12/13 inhibition by THZ531. Heatmap displays DMSO-normalised  $\log_2\text{FC}$  in gene expression for 27,051 transcripts, ranked by THZ531  $\log_2\text{FC}$ . **f**, GSEA of 984 genes ( $\log_2\text{FC} < -4$  and adj.  $P < 0.05$ ) significantly downregulated after THZ531 treatment in comparison to dCeMM2/3/4 (FDR  $< 0.001$ ). GSEA pre-ranked function was used (1,000 permutations). Reproduced from <sup>285</sup>. *I have not contributed to these results.*

To further study the mode of action of dCeMM2-4, we performed a CRISPR/Cas9 screen using a CRL-focused library and showed that all three drugs depend on CUL4B and the adaptor protein DDB1, alongside identifying UBE2Z and UBE2G1 as the relevant ubiquitin priming and extending E2 enzymes. As these screens did not identify a CRL substrate receptor, we turned to genome-scale CRISPR/Cas9 screens to exclude the possibility of an orphan substrate receptor mediating cyclin K degradation. However, we again only identified

the same subset of genes (DDB1, CUL4B, UBE2M and UBE2G1) (**Figure 2.20a**), which we also individually validated to abrogate cyclin K degradation (**Figure 2.20b, c, d**).

We sought to further validate these findings through isolation of clones that develop resistance following treatment of various cell types with excess concentrations of the compounds. We evaluated mutations in DDB1, CUL4A, CUL4B, UBE2M, UBE2G1, CDK12/13 and CCNK via a hybrid-capture based approach coupled to next-generation sequencing (**Figure 2.20g**). We observed a profound enrichment of missense mutations or focal deletions in DDB1, in particular in the cleft between BPA and BPC that has previously been identified as the interaction site of viral proteins that hijack CRL4 activity (**Figure 2.20h, i**, see also **section 1.1.3**). We also identified mutations in the DDB1-binding domain of CUL4B (**Figure 2.20h**), and in CDK12 and CDK13, albeit outside of their kinase domains. A hotspot of spontaneously arising resistance mutations in a DDB1 cavity previously employed by viral proteins for ligase reprogramming suggested that dCeMM2-4 could mediate cyclin K recruitment and degradation in a way that bypasses a canonical substrate receptor module.

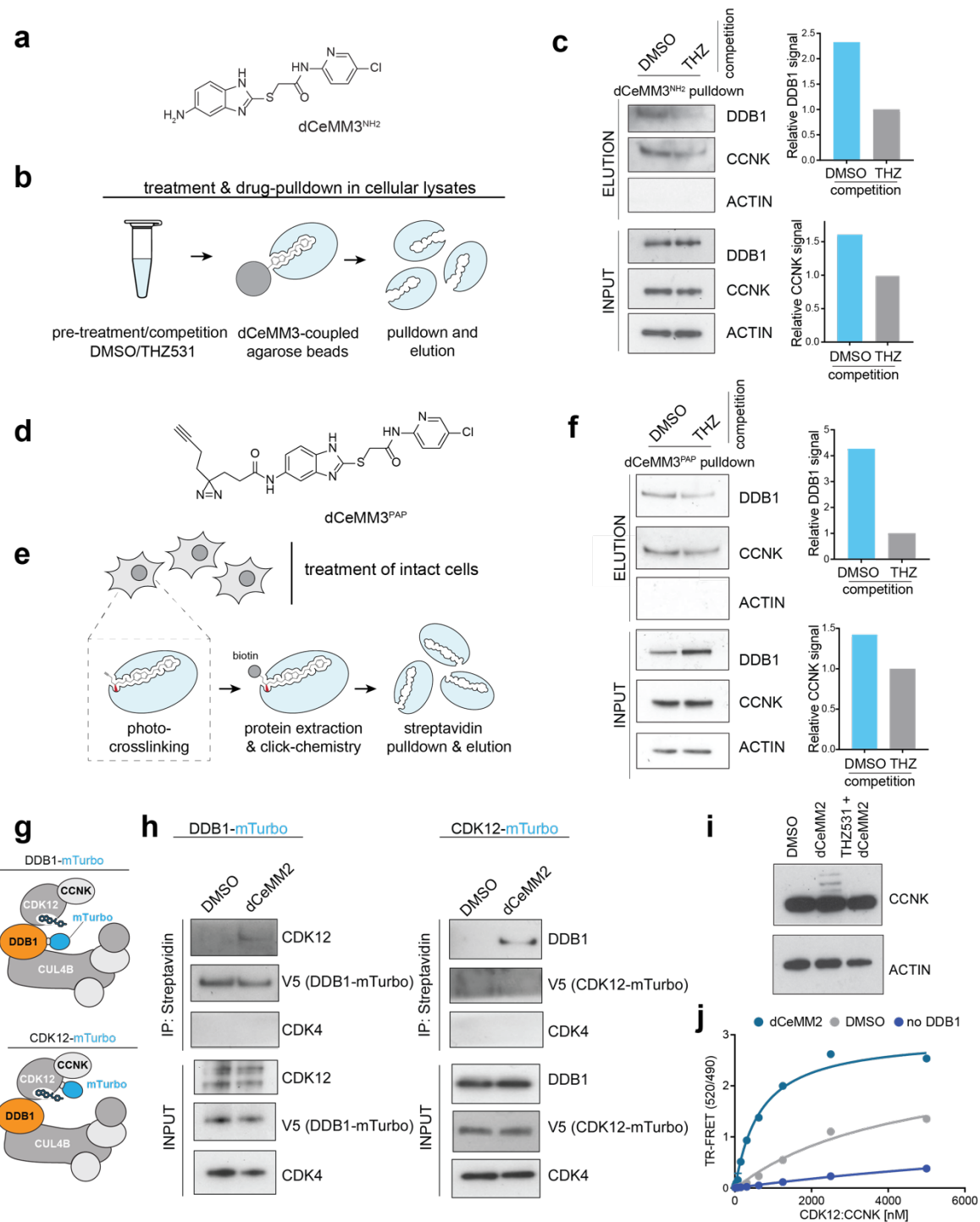


**Figure 2.20 a**, Genome-wide CRISPR dCeMM2/3/4 resistance screens. Top, bubble plot displaying median sgRNA enrichment over DMSO, bubble size indicates significance. Bottom, sgRNA enrichment targeting indicated genes, background indicates distribution of all sgRNAs. **b-d**, dCeMM2-induced cyclin K degradation (2.5 μM) is rescued in UBE2M-, CUL4B- (**b**), UBE2G1- (**c**) and DDB1-deficient cells (**d**). **e**, DMSO-normalised viability in WT and 3-d doxycycline (dox) pre-treated sgDDB1\_1 and sgDDB1\_2 dox-inducible Cas9 KBM7 cells after 3-d dCeMM3 treatment. Mean ± s.e.m.; *n* = 3 independent treatments. **f**, WT-normalised FC in EC<sub>50</sub> based after dose-resolved 3-d viability experiments on exposure to the indicated drugs in the indicated genetic backgrounds. **g**, Targeted hybrid-capture approach coupled to next-generation sequencing to identify mutations in spontaneously dCeMM2/3/4-resistant cells. **h**, Depiction of DDB1 and CUL4B mutations identified by hybrid-capture sequencing in drug-resistant cell pools. Stars indicate point mutations. Red bars indicate premature stop codons. Arrows indicate frameshift mutations. **i**, Structure of DDB1 (grey) in complex with SV5V peptide (amino acids 21–39, blue), PDB 4HYE. DDB1 point mutations located within 20 Å of the peptide are highlighted in orange. Reproduced from <sup>285</sup>. *I have not contributed to these results.*

We aimed to validate a drug-induced interaction between CUL4B-DDB1 and CDK12-cyclin K and set out to perform drug affinity chromatography using an amine-derivatised dCeMM3 analogue (dCeMM3<sup>NH2</sup>) immobilised on sepharose beads (**Figure 2.21a, b**). This indicated cyclin K and DDB1 as interacting proteins (**Figure 2.21c**), while the treatment of lysates with THZ531 prevented DDB1 and CCNK enrichment, which is in line with our earlier observations (**Figure 2.19c, d**). We also employed an alternative pulldown strategy, whereby cells were treated with dCeMM3 containing a photoactive diazirine moiety and an alkyne handle (dCeMM3<sup>PAP</sup>), UV-crosslinked, lysed, and the alkyne handle in CeMM3<sup>PAP</sup> was biotinylated for immobilization on streptavidin beads (**Figure 2.21d**). DDB1 and cyclin K appeared to again be enriched in the eluates, suggesting target engagement in intact cells (**Figure 2.21e, f**). These results therefore indicate simultaneous engagement of cyclin K and DDB1 in a CDK12/13-dependent manner by dCeMM3.

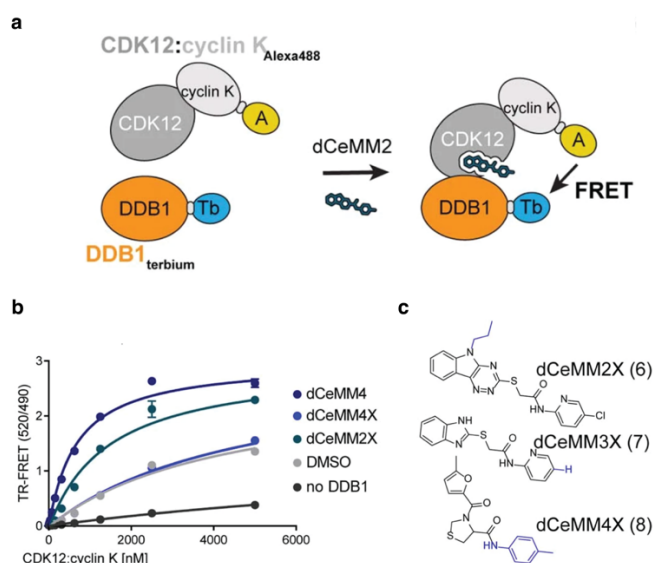
To corroborate a direct interaction between CDK12-cyclin K and CUL4B-DDB1 in the absence of a CRL substrate receptor we employed a proximity-based labelling strategy. For this, we transiently expressed C-terminally tagged DDB1 or CDK12 fusions with the biotin ligase miniTurbo (mTurbo)<sup>286</sup> in HEK293s, treated cells with dCeMM2 or DMSO, and evaluated the difference in biotinylated (and therefore bait-proximal) proteins through streptavidin pulldowns (**Figure 2.21g, h**). We found that CDK12 was only identified as a DDB1 interactor in the presence of dCeMM2, and vice versa (**Figure 2.21h**).

For a conclusive, final confirmation of direct, drug-induced binding of CDK12-cyclin K to DDB1, we reconstituted this interaction using recombinant proteins. Drug-induced proximity between CK12-cyclin K and DDB1 was detected using a TR-FRET assay (**Figure 2.22a**). While weak binding between DDB1 and CDK12-cyclin K was observed in the presence of vehicle control (DMSO), treatment with dCeMM2 facilitated an interaction of DDB1 and CDK12-cyclin K with a  $K_{\text{apparent}}$  of 628 nM (**Figure 2.21j**). A similar affinity was measured for the dCeMM4-induced interaction, while inactive analogues of both compounds (dCeMM2X, dCeMM4X) failed to strengthen the binding (**Figure 2.22b, c**).



**Figure 2.21** **a**, dCeMM3-NH<sub>2</sub> chemical structure (dCeMM3 tethered analogue). **b**, Drug-affinity chromatography strategy based on probe-coupled agarose beads pull-downs after DMSO or THZ531 (competition) pre-treatment in lysates. **c**, Cyclin K and DDB1 enrichment in dCeMM3-NH<sub>2</sub>-based pull-downs. For quantification, eluted protein was normalised to protein available (input panels). THZ531-competed (100 μM, 1 h) ratios were set to 1. **d**, Chemical structure of dCeMM3-PAP (PAP, photoaffinity probe). **e**, Drug-target enrichment strategy based on cellular dCeMM3-PAP cotreatment with DMSO or THZ531 (100 μM, competition) after Carfilzomib pre-treatment (10 μM, 30 min). **f**, Cyclin K and DDB1 enrichment in dCeMM3-PAP-based pull-downs. For quantification, eluted protein was normalised to protein available (input panels). THZ531-competed (100 μM, 1 h) ratios were set to 1. **g**, Proximity labelling strategy to assess drug-induced dimerization in intact cells based on the biotin ligase mTurbo. **h**, Biotin-labelled CDK12 enrichment following 1 h DMSO or dCeMM2 treatment in the presence of carfilzomib (10 μM) in human embryonic kidney cells transfected with DDB1-mTurbo fusion. **i**, Direct cyclin K immunoblotting of carfilzomib-pre-treated KBM7 cells after 2 h DMSO, dCeMM2 or THZ531 + dCeMM2 (10 μM) treatments. **j**, TR-FRET signal for CDK12-Alexa488/cyclin K (0–5 μM) titrated to terbiumDDB1

in DMSO or 10  $\mu\text{M}$  dCeMM2. ‘No DDB1’ only contains streptavidin–terbium. Data are means  $\pm$  s.d. ( $n = 3$ ).  $K_{\text{apparent}}$  (nM): DMSO: not determined, ‘no DDB1’: not determined, dCeMM2 = 628. Reproduced from <sup>285</sup>. *I have not contributed to the results shown in panels (a-i).*



**Figure 2.22.** **a**, Illustration of the TR-FRET assay setup. **b**, TR-FRET signal for CDK12-Alexa488cyclin K (0–5  $\mu\text{M}$ ) titrated to TerbiumDDB1 in DMSO or 10  $\mu\text{M}$  dCeMM4/4 $\times$ /2 $\times$ . ‘No DDB1’ only contains streptavidin–Terbium and is a control for fluorophore-mediated effects. Data are means  $\pm$  SD ( $n = 3$ ).  $K_{\text{apparent}}$  (nM): DMSO = n.d., no DDB1 = n.d., dCeMM4 = 651, dCeMM4X = n.d., dCeMM2X = n.d **c**, Chemical structures of inactive dCeMM2–4 analogues. Adapted from <sup>285</sup>.

### 2.2.3 Significance

Hence, this work established a scalable strategy towards glue degrader discovery by comparative chemical screening in hypo-neddylated versus wild-type cellular models. It also identified a novel DCAF15-dependent molecular glue degrader (dCeMM1), as well as three compounds (dCeMM2–4) that induce ubiquitination and degradation of cyclin K by prompting an interaction of CDK12–cyclin K with a partial CRL4 ligase complex.

The convergence of all three scaffolds on this mechanism of cyclin K degradation is unexpected given the structural differences among dCeMM2–4, as well as between these compounds and the other described molecular glue degrader CR8. Both in the case of dCeMM2–4 and of CR8, this interaction is independent of a dedicated substrate receptor, thus functionally separating this mechanism from previously described classes of thalidomide analogues and aryl sulphonamides.

## Chapter 3 : Molecular glue degrader structure-activity relationship and design principles: a case study for cyclin K degraders

This chapter contains unpublished data. I planned and performed the majority of experiments presented in the following sections. Chemical synthesis of over 60 derivatives was performed by Dakota Suchyta (Thomä laboratory, FMI). Vivian Focht, a Master student I supervised, performed many protein purifications, and assisted in TR-FRET experiments and crystallography. Crystallographic data analysis was performed with the help of Georg Kempf (Thomä laboratory, FMI). In the initial months of the project, Georg Petzold (Thomä laboratory, FMI) also actively participated in the biochemical and crystallographic experiments as well as in X-ray data processing. *In cellulo* results were collected by Mikołaj Słabicki, Charles Zou and Christina di Genua, who are (or were) all members of the Ebert laboratory (DFCI). Proteomics experiments were contributed by Katherine Donovan (Fischer laboratory, DFCI), while the analysis of RNA sequencing data was performed by Marius Jentsch (Schmidt-Burgk laboratory, University of Bonn).

### 3.1 Abstract

Molecular glue degraders are an effective therapeutic modality, but their design principles are not well understood. Recently, several unexpectedly diverse compounds were reported to deplete cyclin K by linking CDK12-cyclin K to the DDB1-CUL4-RBX1 E3 ligase. To investigate how chemically dissimilar small molecules trigger cyclin K degradation, we evaluate over 90 candidate degraders in structural, biophysical, and cellular studies and reveal all compounds acquire glue activity via simultaneous CDK12 binding and engagement of DDB1 interfacial residues, in particular Arg928. While we identify multiple published kinase inhibitors as cryptic degraders, we also show that these glues do not require inhibitory properties for activity and that the relative degree of CDK12 inhibition versus cyclin K degradation is tuneable. We further demonstrate cyclin K degraders have transcriptional signatures distinct from CDK12 inhibitors and therefore offer unique therapeutic opportunities. The systematic structure-activity relationship analysis presented herein provides a conceptual framework for rational molecular glue design.

### 3.2 Introduction

The modulation of protein-protein interactions has become an important avenue of therapeutic intervention<sup>65,66</sup>. Recent advances in targeted protein degradation illustrate that compound-induced proximity between a ubiquitin ligase and a target protein can lead to target ubiquitination and degradation<sup>101,287,288</sup>. Of particular interest are molecular glue degraders, which are drug-like compounds that leverage complementary protein-protein interfaces to induce cooperative ligase-target interactions leading to target depletion<sup>289</sup>. Despite the clinical success of thalidomide derivatives and several recent, largely serendipitous discoveries of other molecular glue degraders<sup>134,152,285,290–292</sup>, the rules that govern their discovery, design and rational optimisation remain poorly defined<sup>293</sup>.



We recently reported that CR8, a preclinical cyclin-dependent kinase (CDK) inhibitor, is a molecular glue degrader that binds CDK12-cyclin K and recruits the DDB1-CUL4-RBX1 E3 ligase core to ubiquitinate cyclin K<sup>294</sup>. Structural studies revealed that CR8 binds the ATP pocket of CDK12, leaving a phenylpyridine moiety exposed on the kinase surface to induce complex formation with the ligase adaptor DDB1, which orients the CDK12-associated cyclin K in a position normally adopted by CUL4 substrates. CR8 therefore hijacks CDK12 in a manner that bypasses the requirement of a canonical substrate receptor (DCAF) typical for CUL4-based E3 ligases. The extensive (~2100 Å<sup>2</sup>) DDB1-CDK12 interface is highly complementary and reveals a helical motif in the CDK12 C-terminal extension that engages DDB1 in a DCAF-like manner. While a basal affinity of ~50 μM was measured between DDB1 and CDK12-cyclin K in the absence of a compound, CR8 enhanced this affinity into the low nanomolar range, leading to robust ubiquitination and degradation of cyclin K<sup>294</sup>. CR8 therefore shows dual activity, promiscuous CDK inhibition and selective cyclin K degradation, which leads to robust inactivation of CDK12, an emerging therapeutic target in oncology and beyond<sup>190,212,242,249</sup>. Recently, chemically distinct compounds have been found to degrade cyclin K<sup>285,290,291,295,296</sup>, but as several lack any obvious chemical similarity to CR8<sup>285,290,291</sup>, their precise mode of action remained unknown.

The prospective development of molecular glue degraders would be greatly facilitated by systematic studies of large compound sets that evaluate ternary complex structures to inform on the structure-activity relationship (SAR). In comparison to other molecular glue degraders<sup>297,298</sup>, the chemical diversity across published cyclin K degraders is unusually broad, which presents a unique opportunity to delineate more generally how target-binding compounds can acquire gain-of-function molecular glue degrader activity. Herein, we perform a systematic dissection of the cyclin K degrader SAR by evaluating over 90 putative degraders, notably also identifying cyclin K degraders among unrelated scaffolds and published kinase inhibitors. Through crystallographic evaluation of 29 compound-induced ternary complexes, we find that while the small molecules are chemically diverse, they lead to the formation of an analogous ternary complex. We identify the key underlying interactions and show that the relative degree of kinase inhibition versus cyclin K degradation can be tuned with modifications and ultimately decoupled. We further demonstrate that cyclin K degraders have transcriptional signatures distinct from CDK12 inhibitors and therefore can offer unique therapeutic opportunities. Taken together, these findings provide a set of learnings for converting target binders into molecular glue degraders.

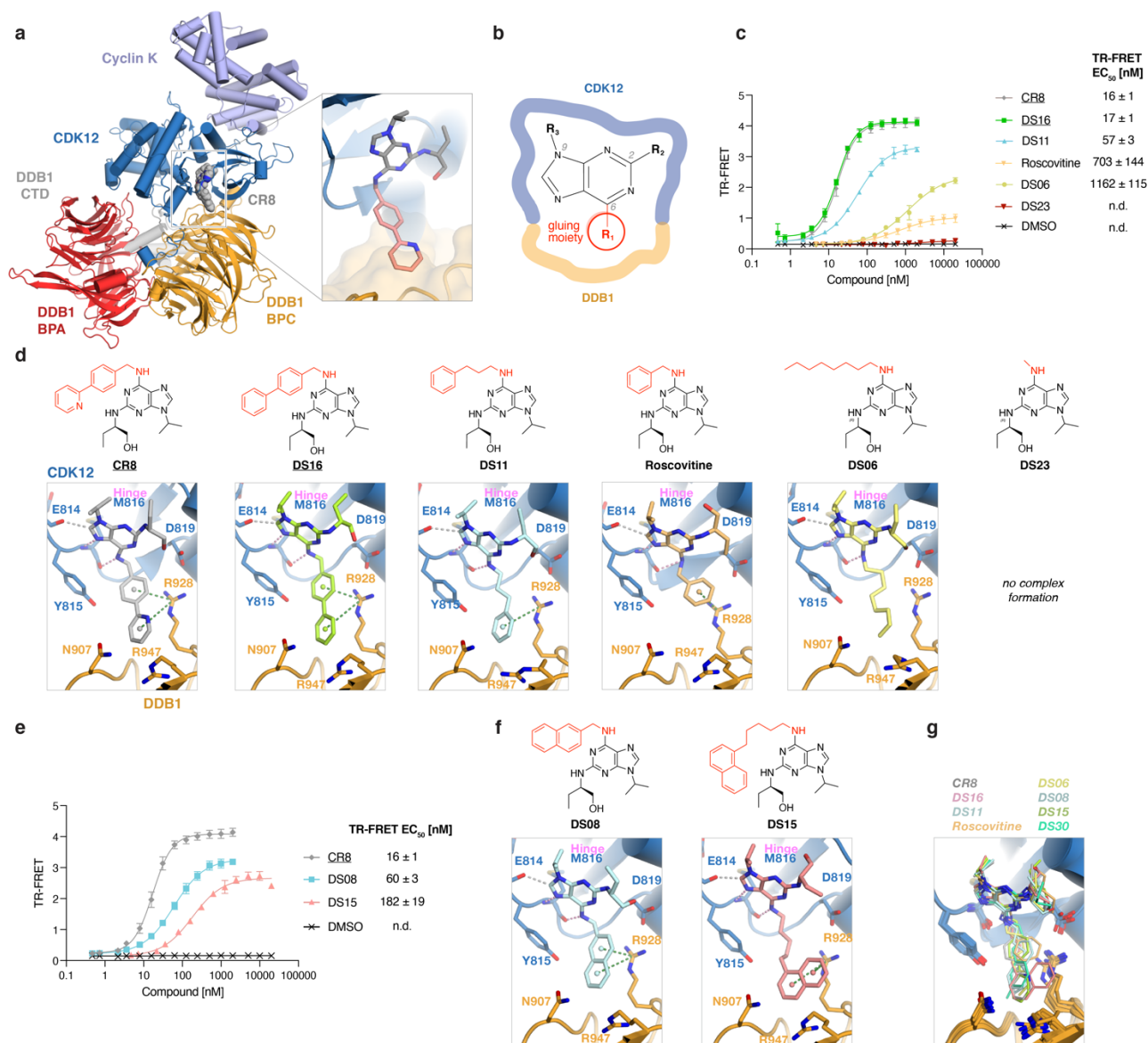
### 3.3 Results

#### 3.3.1 CR8 retains glue activity despite diverse modifications

To understand how chemically dissimilar compounds can commit cyclin K for degradation, we first focussed on the CR8 scaffold. The DDB1-CR8-CDK12-cyclin K complex crystal structure found the compound bound at the CDK12-DDB1 interface, with the phenylpyridine moiety of CR8 protruding towards DDB1 and bridging this interface (**Figure 3.1a, b**)<sup>294</sup>. We set out to dissect the CR8 SAR and explore whether moieties other than the phenylpyridine (substituents at position R<sub>1</sub>, hereafter referred to as *gluing moieties* (C6 position in the 2,6,9-

trisubstituted nomenclature; **Figure 3.1b**) could promote CDK12-DDB1 interactions. For this, we developed an optimised time-resolved fluorescence energy transfer (TR-FRET) assay that accurately measured *in vitro* complex formation between CDK12-cyclin K and DDB1 in the presence of closely related small molecules and confirmed that CR8 induces tight complex formation ( $EC_{50} = 16 \pm 1$  nM) (**Figure 3.1c, d** and **Figure 3.2a-c**; see also **Chapter 4** for a detailed account of TR-FRET assay optimisation). To identify essential features of the gluing moiety, we also visualised the binding geometry through crystallization of ternary complexes (**Tables 3.1-3.5**, **Figure 3.3**, and **Supplementary Figures S3.1** and **S3.2**) with systematically varied derivatives.

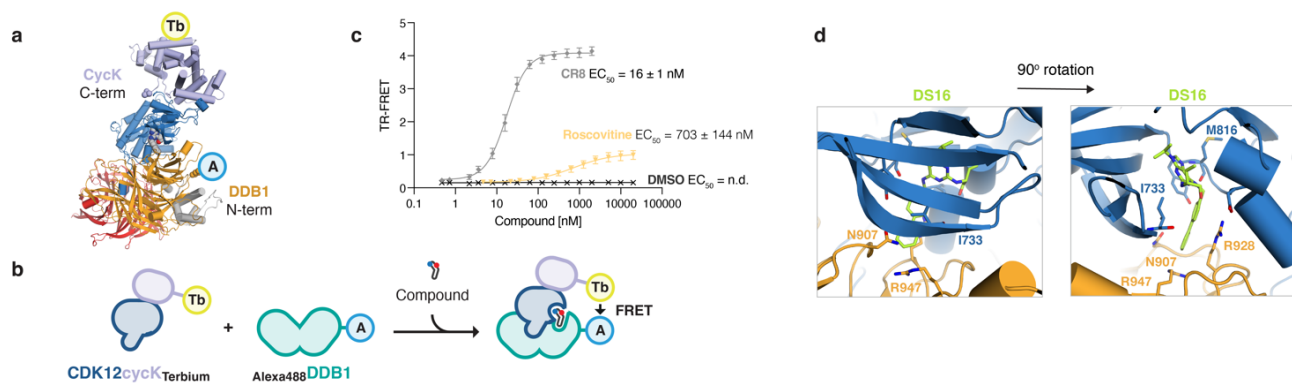
First, we set out to gauge the steepness of the CR8 SAR through the gradual simplification of its gluing moiety. A derivative bearing a biphenyl substituent (DS16) instead of the phenylpyridine displayed the same binding mode as CR8, with the purine core held in the CDK12 active site by two hydrogen bonds to the hinge (Met816), a region of the kinase, which connects the N- and C-terminal lobes of the catalytic domain (**Figure 3.1c, d**). The compound is enclosed by a CDK12 loop (aa 731-743, with Ile733 approaching the ligand (**Figure 3.2d**); omitted in most figure panels for clarity), and the gluing moiety engages in  $\pi$ -cation interactions with Arg928 of DDB1 (**Figure 3.1d**). Interestingly, DS16 showed activity equivalent to CR8, demonstrating that the hydrogen bond acceptor (HBA) in the ring is not required for robust complex formation *in vitro* (**Figure 3.1c, d**). While a phenyl ring on a propyl chain (DS11) supported robust complex formation ( $EC_{50} = 57 \pm 3$  nM), shortening the chain by two alkyl carbons to yield roscovitine (the parent scaffold from which CR8 was developed) preserved the CR8-like binding mode at the interface but gave rise to much poorer recruitment ( $EC_{50} = 703 \pm 144$  nM). Strikingly, even an octyl chain (DS06) protruding into DDB1 supported the formation of an analogous complex, albeit with a lower affinity ( $EC_{50} = 1162 \pm 115$  nM), likely through van der Waals contacts with Arg928 (**Figure 3.1c, d**). These findings showed that the ternary complex is surprisingly permissive to changes in CR8, leading us to ask whether simply filling the kinase pocket would be sufficient to induce the interaction. For this, we tested the di-substituted purine core alone, with only a methyl in R<sub>1</sub> (DS23), and found that it did not appreciably support the binding, demonstrating that more extensive engagement of DDB1 residues by the compound is required for molecular glue activity (**Figure 3.1c, d**). Hence, we set out to explore broader R<sub>1</sub> modifications to define the steric and spatial constraints of this cavity and dissect the contribution of specific contacts ( $\pi$ -cation interactions, hydrogen bonding) to complex formation.



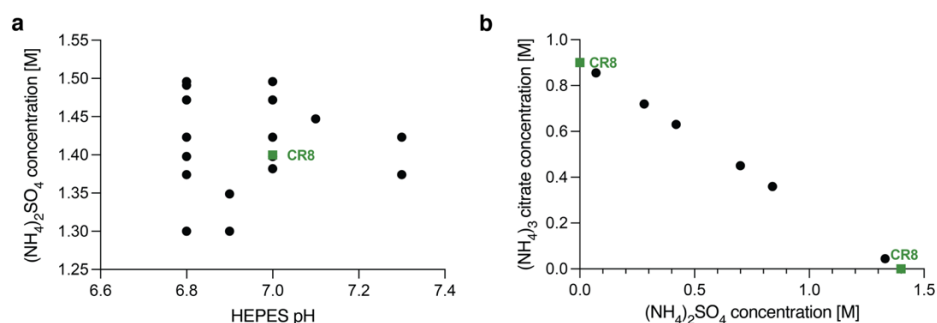
**Figure 3.1. Modifications of the CR8 scaffold preserve its molecular glue activity.** **a**, Crystal structure of the DDB1-CR8-CDK12-cyclin K complex (PDB 6TD3)<sup>294</sup>. The zoomed panel depicts the binding mode of CR8, with the phenylpyridine moiety (*gluing moiety*; shown in pale red) engaging DDB1 (shown in surface representation). **b**, Chemical structure of the 2,6,9-trisubstituted purine core. The R<sub>1</sub> group is referred to as the *gluing moiety* and is coloured red throughout. **c**, Chemical structures of CR8, DS16, DS11, roscovitine, DS06, DS23 and ternary complex crystal structures of the DDB1-CDK12 interfaces induced by those compounds. **d**, *In vitro* TR-FRET complex formation assay for compound shown in (c) and DMSO. **e**, *In vitro* TR-FRET complex formation assay for DS08, DS15, CR8 and DMSO. **f**, Chemical structures of DS08 and DS15 and ternary complex crystal structures of interfaces induced by those compounds. Conformational restriction of the *gluing moiety* of DS15 is not entropically favourable, likely contributing to the lower ternary affinity for DS15 than for DS08. **g**, Overlay of ternary complex structures from (d, f). (c, d) Data represent the mean ± s.d. (n ≥ 2). (d, f) Interactions are represented by dashed lines. Hydrogen bonds to the hinge region are shown in pink, other hydrogen bonds in yellow, aromatic H-bonds in grey, and π-cation interactions in green. Density maps and interaction distances can be found in Supplementary Figures S3.1 and S3.2.

**Table 3.1.** Data collection and refinement statistics for structures presented in **Figure 3.1**.

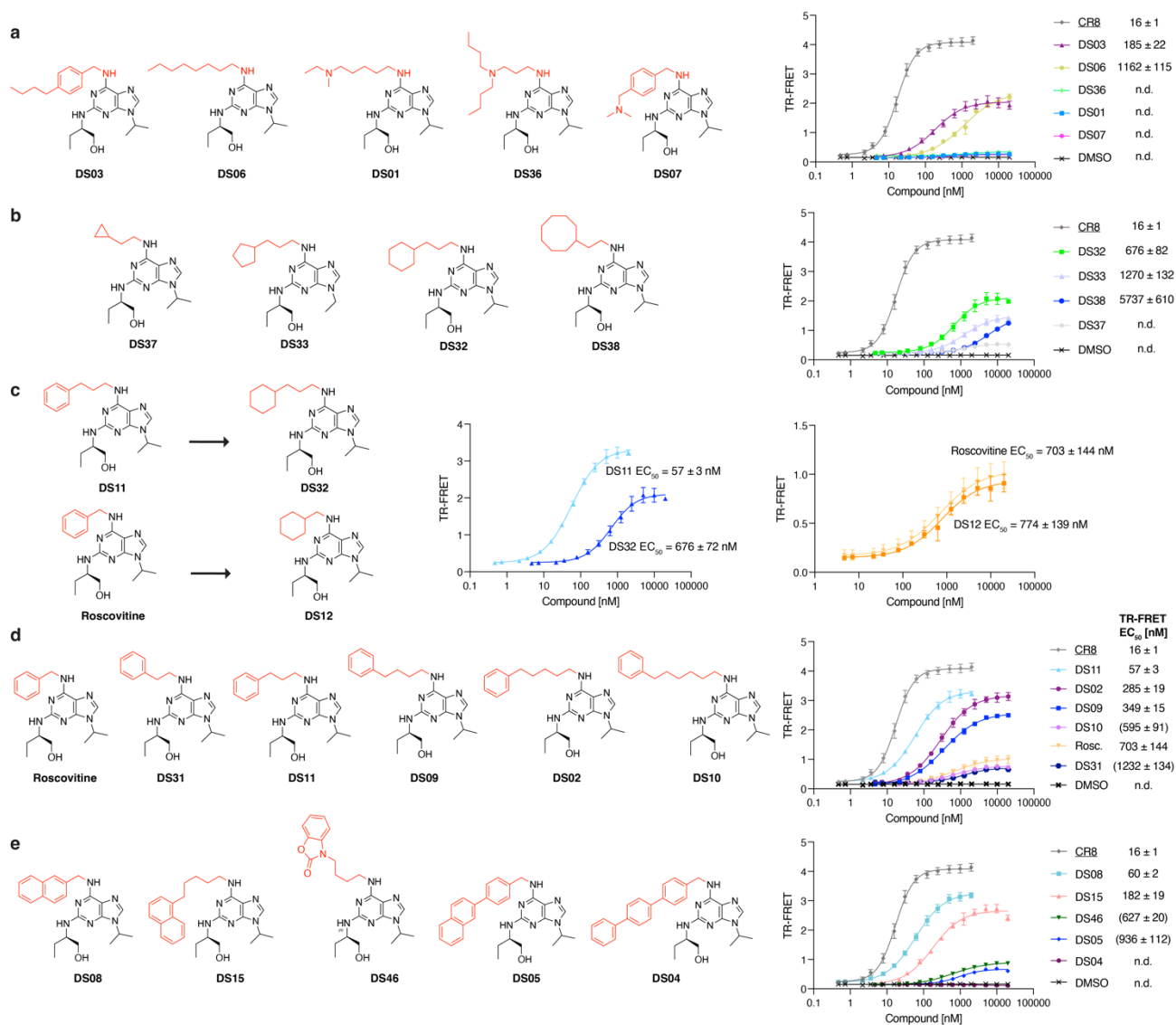
Ligand	Roscovitine	DS06	DS08	DS11	DS15	DS16
<b>Data collection</b>						
<b>Space group</b>	P 31 2 1	P 31 2 1	P 31 2 1	P 31 2 1	P 31 2 1	P 31 2 1
<b>Cell dimensions</b>						
<b>a, b, c (Å)</b>	249.461	249.379	248.901	249.356	249.635	249.675
	249.461	249.379	248.901	249.356	249.635	249.675
	216.033	218.443	218.059	217.993	218.865	218.54
<b>a, b, g (°)</b>	90 90 120	90 90 120	90 90 120	90 90 120	90 90 120	90 90 120
<b>Resolution (Å)</b>	216.037-3.505	215.946- 3.618	215.488-3.406	215.949-3.402	216.19-3.359	216.225-3.082
	(3.707-3.505)	(3.784- 3.618)	(3.473- 3.506)	(3.577-3.402)	(3.548-3.359)	(3.251-3.082)
<b>Rsym or Rmerge</b>	0.45 (6.23)	0.243 (4.866)	0.173 (3.237)	0.234 (3.985)	0.248 (4.694)	0.17 (4.272)
<b>Rpim</b>	0.1 (1.385)	0.078 (1.438)	0.055 (0.995)	0.052 (0.871)	0.055 (1.032)	0.038 (0.94)
<b>I / σI</b>	9.5 (1)	11.6 (0.8)	15.5 (1.1)	11.5 (0.9)	8.1 (0.8)	16.7 (0.9)
<b>CC1/2</b>	0.997 (0.3)	0.999 (0.41)	0.999 (0.514)	0.999 (0.407)	0.998 (0.37)	0.999 (0.374)
<b>Completeness (ellipsoidal, %)</b>	95.5 (59.1)	95.8 (59.8)	91.2 (52.8)	94.4 (68.8)	96.1 (64.5)	96.7 (70.5)
<b>Redundancy</b>	21 (21.2)	21.1 (22.6)	20.8 (21.4)	20.9 (21.6)	21 (21.6)	21 (21.6)
<b>Refinement</b>						
<b>Resolution (Å)</b>	68.32 - 3.51	71.98 - 3.62	68.85 - 3.41	53.99 - 3.402	60.19 - 3.362	60.17 - 3.082
	(3.636 - 3.51)	(3.75 - 3.62)	(3.615 - 3.41)	(3.523 - 3.402)	(3.482 - 3.362)	(3.192 - 3.082)
<b>No. reflections</b>	82808 (1944)	89329 (2578)	92840 (8099)	91146 (2233)	95319 (2080)	122705 (2868)
<b>Rwork / Rfree</b>	0.1923 / 0.2345	0.1760 / 0.2128	0.1867 / 0.2129	0.1856 / 0.2260	0.1822 / 0.2187	0.2119 / 0.2352
<b>No. atoms</b>	33889	33887	33939	33953	34033	34002
<b>Protein</b>	33736	33736	33736	33736	33736	33736
<b>Ligand/ion</b>	231	259	292	319	405	356
<b>Water</b>	0	0	0	0	0	0
<b>B-factors</b>	153.5	162.70	147.34	143.01	146.03	116.85
<b>Protein</b>	153.5	162.66	147.24	142.95	145.88	116.74
<b>Ligand/ion</b>	153.27	171.43	164.26	152.2	163.18	130.55
<b>Water</b>						
<b>R.m.s. deviations</b>						
<b>Bond lengths (Å)</b>	0.006	0.005	0.004	0.007	0.006	0.006
<b>Bond angles (°)</b>	0.93	0.87	0.78	0.95	0.9	0.92



**Figure 3.2. The mapping of interfacial pocket dimensions and key interactions with derivative series.** **a**, Fluorescent label positions chosen for the optimized TR-FRET assay. **b**, Schematic of the compound titration used for ternary complex formation assessment. **c**, Optimized *in vitro* TR-FRET complex formation assay for CR8 and roscovitine reveals a large difference in activity between the two compounds in accordance with the lack of cyclin K degradation activity of roscovitine in cells<sup>1</sup>. The large difference in  $EC_{50}$  values between CR8 and roscovitine indicates that the assay is an appropriately sensitive readout for the evaluation of closely related compounds. **d**, Illustration of the CDK12 loop (aa 731-743) that encloses the active site, with the Ile733 sidechain oriented towards the compound. This loop is omitted from most figure panels for clarity



**Figure 3.3.** Illustration of crystallization conditions that yielded best diffracting crystals resulting in the presented structures. Each data point corresponds to a different compound and the published conditions for CR8 are marked in green. **a**, Fine screen with the concentration of ammonium sulphate (precipitate) and pH of the HEPES buffer varied. **b**, Fine screen with two precipitants (ammonium citrate and ammonium sulphate) with their concentrations varied. Crystallisation screening is illustrated in detail in **Figure 6.2** in the Methods section.



**Figure 3.4. Aliphatic moieties do not support effective CDK12-DDB1 complex formation.** **a**, Chemical structures of a series of derivatives bearing aliphatic chains at the  $R_1$  position (left). *In vitro* TR-FRET complex formation assay for the compounds shown, DMSO, and CR8 (right). **b**, Chemical structures of a series of derivatives bearing saturated rings at the  $R_1$  position (left). *In vitro* TR-FRET complex formation assay for the compounds shown, DMSO, and CR8 (right). **c**, The impact of aromaticity on ternary complex formation. Chemical structures of two aromatic-aliphatic pairs and the associated *in vitro* TR-FRET compound titration results. Comparing the two matched pairs, the cyclohexane-bearing DS32 showed ten-fold weaker binding than DS11, while the DS12-roscovitine pair with equivalent rings but on a shorter chain showed comparably poor binding. This is likely due to the shorter chain in roscovitine only supporting complex formation through dispersion forces even when an aromatic ring is present<sup>2</sup>. **d**, Chemical structures of a series of derivatives bearing an alkyl phenyl  $R_1$  group with the alkyl linker varying in length (left). *In vitro* TR-FRET complex formation assay for the compounds shown, DMSO, and CR8 (right). **e**, Chemical structures of several derivatives containing fused or multiple rings (left). *In vitro* TR-FRET complex formation assay for the compounds shown, DMSO, and CR8 (right). Crystal structures with compounds DS08 and DS15 are displayed in **Figure 3.1f**. (a-e) Data represent the mean  $\pm$  s.d. ( $n \geq 2$ ).

### 3.3.2 Pi-cation interactions are important for robust complex formation

To explore the importance of interactions between the DDB1 Arg928 and the solvent-exposed arene moiety of the ligand, we systematically surveyed compounds that feature aliphatic chains (DS01, 36) or rings (DS12, 32, 33, 37, 38) as their gluing moieties and found these did not support robust complex formation (**Figure 3.4a-c**).

To investigate the optimal arene placement for  $\pi$ -cation contacts to Arg928, while also probing the dimensions and steric constraints of the DDB1-compound interface, we mapped the cavity with a derivative series bearing phenyl gluing moieties on alkyl chains of varying length (**Figure 3.4d**). DS11 was a robust glue (**Figure 3.1c, d**) and while derivatives with longer chains (DS09, DS02, DS10; four-six carbon chain) were somewhat tolerated (5-10-fold lower affinity than DS11), shorter chains in roscovitine or DS31 (one-two carbon chain) showed more deleterious effects (12-22-fold lower affinity) (**Figure 3.4d**). We then probed the impact of larger  $\pi$ -systems by introducing naphthalene-containing gluing moieties. Compound DS08 served as an efficient complex inducer ( $EC_{50} = 60 \pm 2$  nM), likely due to favourable interactions of the extended  $\pi$  system with Arg928<sup>299,300</sup>, while a naphthalene on a longer chain (DS15) was only accommodated with a suboptimal geometry ( $EC_{50} = 182 \pm 19$  nM) (**Figure 3.1e, f**). Rigidifying the linker by adding an additional phenyl ring (DS05) led to very poor DDB1 recruitment, while introducing three consecutive rings (DS04) entirely abolished binding (**Figure 3.4e**). These findings underscore the importance of the correct positioning of the arene for optimal Arg928 interactions and identify steric constraints for bulky compounds that lack conformational plasticity.

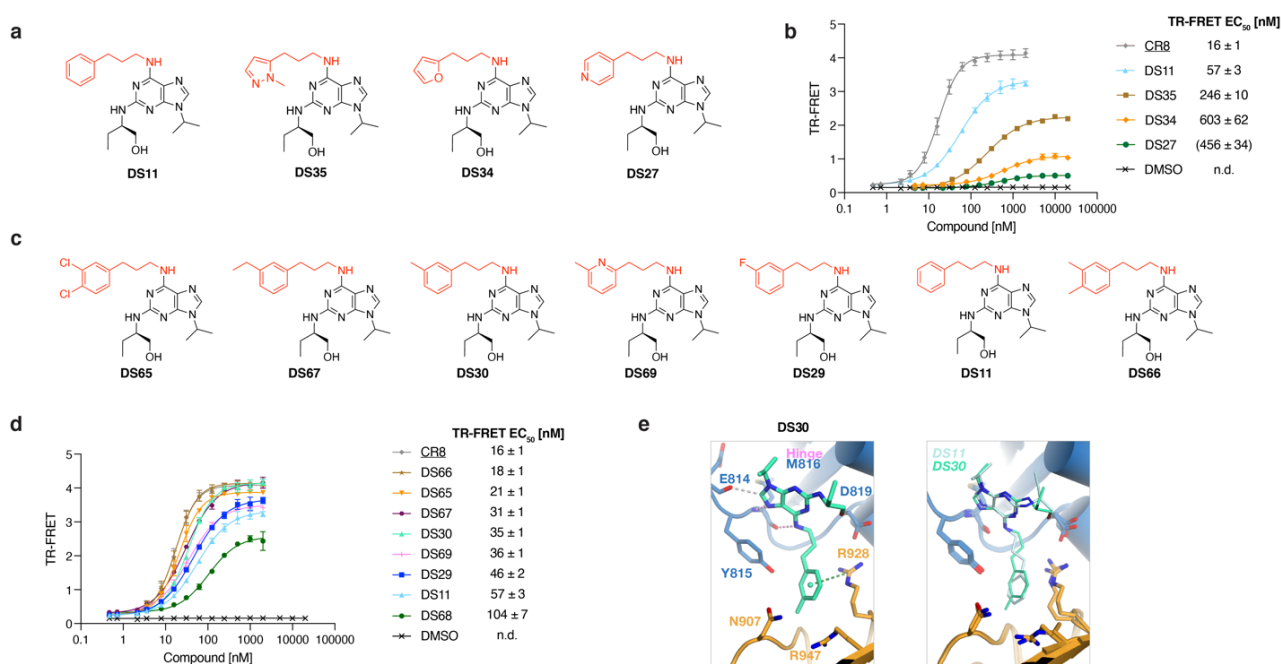
We further assessed the impact of the  $\pi$ -cation interaction between the compound and DDB1 Arg928 by comparing DS11 with compounds bearing pyridine (DS27), furan (DS34) or methylpyrazole (DS35) gluing moieties. The phenyl ring in DS11 was strongly preferred over electron-poor (pyridine) but also electron-rich (furan, methylpyrazole) heterocycles (**Figure 3.5a, b**). Moreover, we saw that complex formation could be subtly improved by simple derivatisation of DS11, with a methyl substitution of the phenyl ring (DS30) leading to slightly higher affinity ( $EC_{50} = 35 \pm 1$  nM) and the binding mode preserved (**Figure 3.5c-e**). Notably, dimethyl (DS66;  $EC_{50} = 18 \pm 1$  nM) and dichloro (DS65;  $EC_{50} = 21 \pm 1$  nM) derivatives both showed CR8-level *in vitro* recruitment of CDK12 to DDB1 (**Figure 3.5c, d**). We also explored exchanging the phenyl rings in the R<sub>1</sub> of CR8 for heterocycles or substituted arenes (*e.g.* DS43 or DS25) and observed only small changes in ternary complex affinity (**Figure 3.6a-c**). This identifies interactions between Arg928 and diverse aromatic groups in the ligand's gluing moiety as key mediators of DDB1-CDK12 molecular glue degrader activity.

### 3.3.3 Hydrogen bonding interactions of the gluing moiety more weakly modulate affinity

Next, we focussed on putative hydrogen bonding interactions of the gluing moiety. As the nitrogen in the CR8 pyridine ring could serve as a potential HBA, we systematically explored the importance of its presence and position in the gluing moiety. Changing the phenylpyridine to biphenyl (CR8 to DS16) (**Figure 3.1c, d**) or methylpyridine to tolyl (DS69 to DS30) (**Figure 3.5c, d**) had no effect on *in vitro* activity, suggesting that the HBA is dispensable, or that its absence can be compensated by the more electron-rich nature of benzene compared to pyridine and hence better  $\pi$ -cation interactions with Arg928. We also tested whether changing the nitrogen position in the ring impacts ternary complex formation and found that 4-pyridyl (DS44) and 3-pyridyl (DS45) were two- to three-fold poorer recruiters than CR8 (2-pyridyl) (**Figure 3.6d, e**).

Our structural evaluation highlighted the CDK12 residue Tyr815 in proximity of the binding pocket as a potential additional hydrogen-bonding contact. This residue is only present in CDK12/13 and hence its engagement by the ligand could confer specificity within the CDK family (**Figure 3.6f**). In an attempt to leverage this residue, we designed WX3 bearing a 2-pyridinone ring instead of the first phenyl ring of CR8 (**Figure 3.6a, g**). Crystallographic analysis showed the C=O functionality within hydrogen-bonding distance to Tyr815 and the N-H interacting with the Met816 carbonyl, yet the binding affinity did not improve ( $EC_{50} = 21 \pm 1$  nM) (**Figure 3.6b, g**). Other modifications, such as installing a fluorine at this position (DS24), led to decreased affinity (**Figure 3.6a, b, g**).

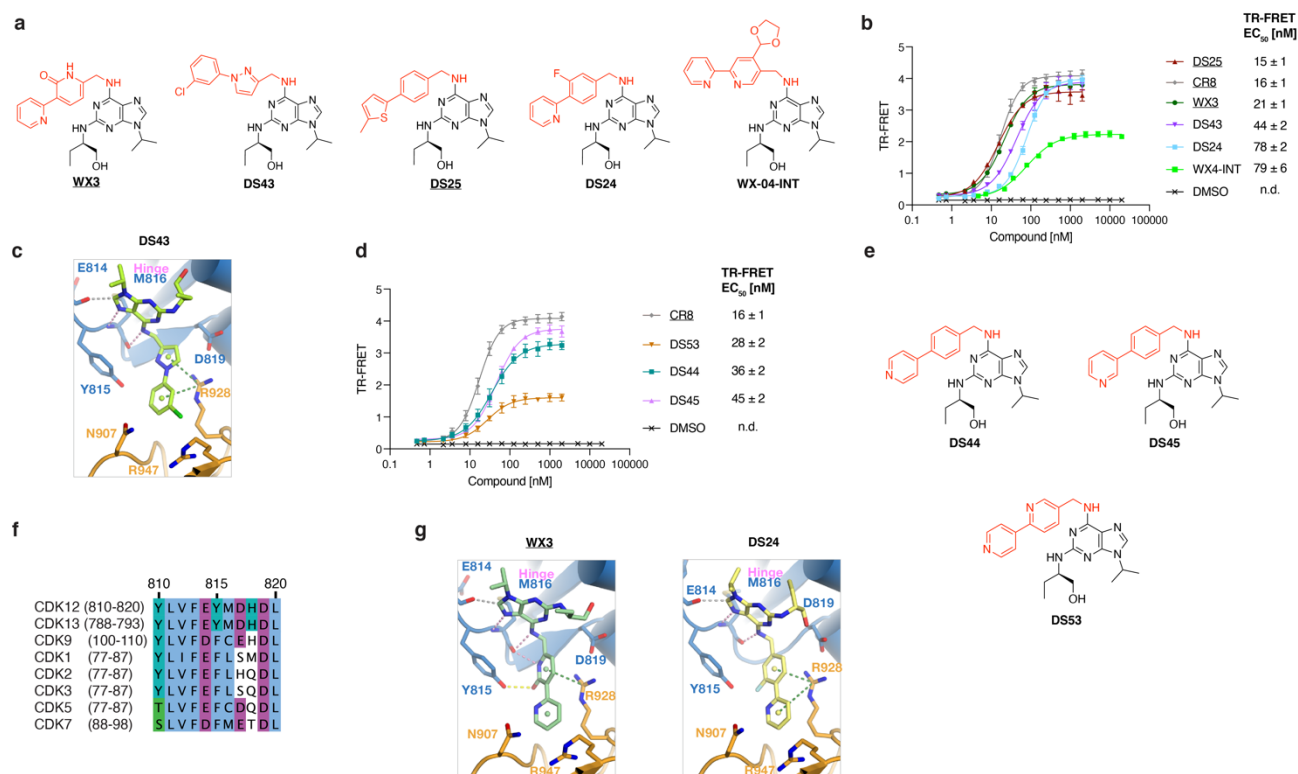
In summary, the exploration of CR8 SAR around the R<sub>1</sub> position demonstrated that a surprisingly wide range of gluing moieties (arenes and heteroarenes of varying size, but also aliphatic groups – albeit weakly) can engage the Arg928 side chain and therefore facilitate DDB1-CDK12 interactions. We show that while the HBA is dispensable, appropriate steric and electronic properties for effective  $\pi$ -cation interactions with DDB1 are required for high-affinity complex formation. The position of DDB1 Arg928 remains relatively static in all structures determined, likely due to its interaction with CDK12 Asp819 anchoring it in this position (**Figure 3.1g**). We conclude that diverse substituents can be accommodated in this  $\sim 370$  Å<sup>3</sup> CDK12-DDB1 interfacial cavity in disparate ways, with each proficient molecular glue compound engaging DDB1 Arg928.



**Figure 3.5. Miscellaneous modifications of CR8-like scaffolds.** **a**, Chemical structures of derivatives bearing heterocyclic gluing moieties. **b**, *In vitro* TR-FRET complex formation assay for compounds shown in **(a)**, DMSO, and CR8. **c**, Chemical structures of several compounds derived from simple modifications of the simplified CR8-like scaffold DS11. **d**, *In vitro* TR-FRET complex formation assay for compounds shown in **(c)**, DMSO, and CR8. **e**, Crystal structure of the ternary complex formed with DS30 (chemical structure in **(c)**) (left) and overlay of the binding mode of DS11 and DS30 (right). Interactions are represented by dashed lines. Hinge hydrogen bonds are shown in pink, aromatic H-bonds in grey, other hydrogen bonds in yellow, and  $\pi$ -cation interactions in green.



Multiple sequence alignment of different human CDKs highlighting a unique tyrosine residue in CDK12/13 (Tyr815 in CDK12). **I**, Structures of interfaces induced by WX3 and DS24. (**b, d, g, i**) Data represent the mean  $\pm$  s.d. ( $n \geq 2$ ). (**e, h, l**) Interactions are represented by dashed lines. Hinge hydrogen bonds are shown in pink, aromatic H-bonds in grey, other hydrogen bonds in yellow, and  $\pi$ -cation interactions in green.



**Figure 3.6. Other R<sub>1</sub> modifications of the CR8 scaffold.** **a**, Chemical structures of various CR8 derivatives. **b**, *In vitro* TR-FRET complex formation assay for compounds shown in (**a**), DMSO, and CR8. **c**, Ternary complex crystal structures with DS43. **d**, *In vitro* TR-FRET complex formation assay for compounds shown in (**e**), DMSO, and CR8. **e**, Chemical structures of CR8 derivatives where the location of the purine nitrogen is varied. **f**, Multiple sequence alignment of different human CDKs highlighting a unique tyrosine residue in CDK12/13 (Tyr815 in CDK12). **g**, Structures of interfaces induced by WX3 and DS24. (**b, d**) Data represent the mean  $\pm$  s.d. ( $n \geq 2$ ). (**c, g**) Interactions are represented by dashed lines. Hinge hydrogen bonds are shown in pink, aromatic H-bonds in grey, other hydrogen bonds in yellow, and  $\pi$ -cation interactions in green.

**Table 3.2.** Data collection and refinement statistics for structure presented in **Figure 3.5** and **Figure 3.6**.

Ligand	DS24	DS30	DS43	WX3
<b>Data collection</b>				
Space group	P 31 2 1	P 31 2 1	P 31 2 1	P 31 2 1
Cell dimensions				
a, b, c (Å)	249.8 249.8 218.078	248.058 248.058 221.169	250.162 250.162 218.849	249.143 249.143 217.39
a, b, g (°)	90 90 120	90 90 120	90 90 120	90 90 120

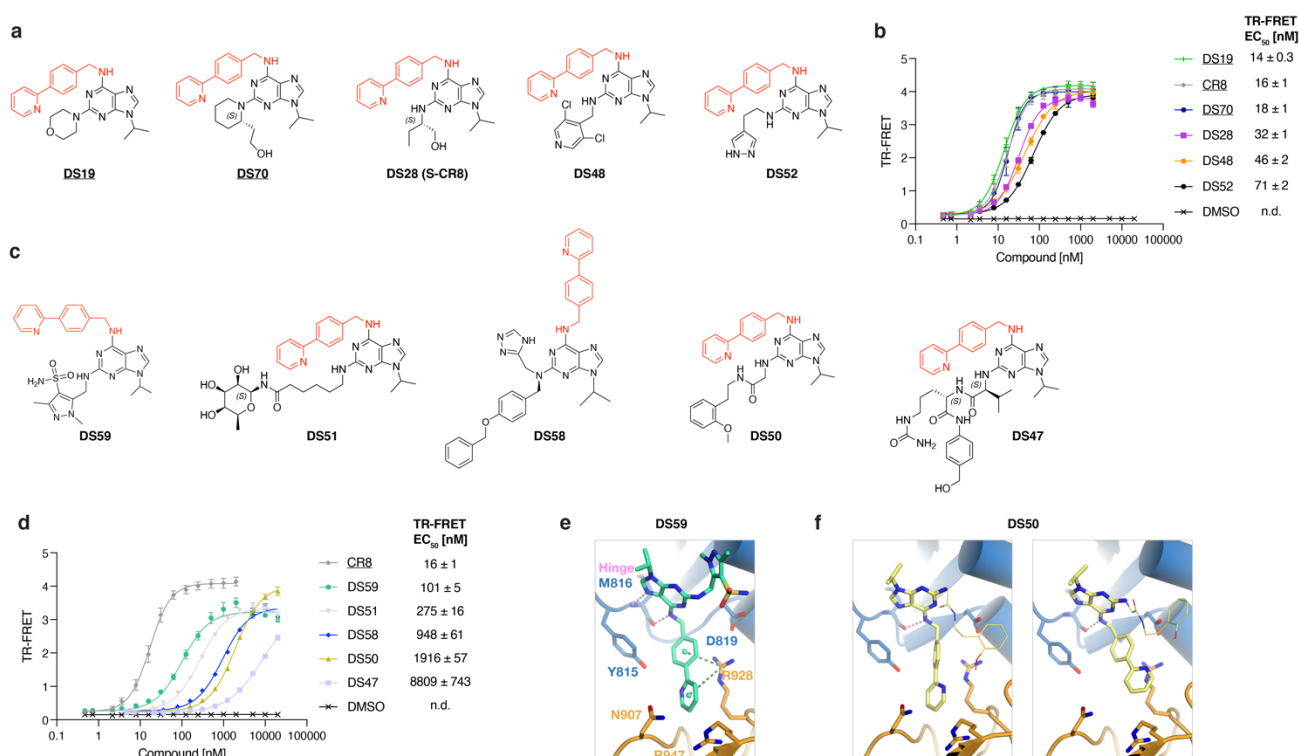
<b>Resolution (Å)</b>	216.333-3.58 (3.83-3.58)	214.825-3.408 (3.586-3.408)	216.647-3.199 (3.393-3.199)	215.765-3.789 (4.019-3.789)
<b>Rsym or Rmerge</b>	0.274 (4.361)	0.188 (4.171)	0.167 (3.462)	0.564 (5.81)
<b>Rpim</b>	0.061 (0.986)	0.042 (0.915)	0.037 (0.753)	0.125 (1.291)
<b>I / σI</b>	10.1 (0.9)	13.6 (0.9)	14.6 (1)	7 (0.7)
<b>CC1/2</b>	0.998 (0.329)	0.999 (0.372)	0.999 (0.463)	0.994 (0.294)
<b>Completeness (ellipsoidal, %)</b>	95 (56.7)	96 (63.4)	96.4 (69.6)	95.5 (64.4)
<b>Redundancy</b>	21.1 (20.5)	21 (21.7)	21 (22.1)	21.2 (21.1)
<b>Refinement</b>				
<b>Resolution (Å)</b>	54.08 - 3.58 (3.708 - 3.58)	54.58 - 3.411 (3.533 - 3.411)	54.3 - 3.2 (3.315 - 3.2)	37.21 - 3.79 (3.925 - 3.79)
<b>No. reflections</b>	72805 (733)	92797 (2388)	105010 (1931)	63860 (1063)
<b>Rwork / Rfree</b>	0.1826 / 0.2299	0.1885 / 0.2268	0.1904 / 0.2211	0.1912 / 0.2338
<b>No. atoms</b>	33965	33967	33910	33895
<b>Protein</b>	33736	33595	33736	33736
<b>Ligand/ion</b>	313	468	285	243
<b>Water</b>	0	0	0	0
<b>B-factors</b>	166.05	151.34	130.37	156.45
<b>Protein</b>	165.92	151.08	130.34	156.49
<b>Ligand/ion</b>	185.28	174.46	137.15	148.61
<b>Water</b>				
<b>R.m.s. deviations</b>				
<b>Bond lengths (Å)</b>	0.006	0.006	0.003	0.006
<b>Bond angles (°)</b>	0.98	0.92	0.64	0.88

### 3.3.4 Diverse R<sub>2</sub> modifications are tolerated for DDB1-CDK12 complex formation

Given the tolerance of the CR8 gluing moiety to chemical diversification, we next explored modifications of other positions on this scaffold. As the R<sub>3</sub> group faces into the kinase pocket and would only be expected to govern binary binding, we instead focussed on the aminobutanol moiety at the R<sub>2</sub> position, which in the CR8-induced complex is largely solvent-exposed. The inversion at this stereocentre (S-CR8; DS28; EC<sub>50</sub> = 32 ± 1 nM) or introducing a morpholino at the R<sub>2</sub> position (DS19; EC<sub>50</sub> = 14 ± 0.3 nM) gave rise to compounds with comparable activity to CR8 (R-CR8 is referred to as CR8 throughout) (**Figure 3.7a, b** and **Figure 3.8**), suggesting that hydrogen-bonding interactions observed between R<sub>2</sub> and the CDK12 backbone do not strongly contribute to binding. Installation of a hydroxyethyl piperidine functionality, present in the potent CDK inhibitor dinaciclib<sup>301</sup>, at this position (DS70) also yielded a potent cyclin K molecular glue (EC<sub>50</sub> = 18 ± 1 nM), while substituting with a dichloropyridine (DS48) or pyrazole (DS52) led to a few-fold lower affinity (**Figure 3.7a, b**). In pursuit of more structurally diverse compounds, we performed R-group docking at this position using Glide (Schrödinger) and selected several compounds stipulated to engage in additional interactions (**Figure 3.7c**). We crystallised a ternary complex with DS59, bearing an imidazole sulphonamide, but only observed

medium activity ( $EC_{50} = 101 \pm 5$  nM) (**Figure 3.7c-e**). Compound DS50, with the R<sub>2</sub> taking the role of a second gluing moiety contacting DDB1 in the docked pose, yielded a crystal structure with the R<sub>2</sub> substituent in multiple conformations (**Figure 3.7c, d, f**).

The examination of the CR8 SAR showed that this scaffold is amenable to extensive modifications. While changes at the R<sub>1</sub> position are the most consequential for gain-of-function glue activity, derivatisation at R<sub>2</sub> further affects complex formation, and both R<sub>1</sub> and R<sub>2</sub> can be used to tune properties of the compound. Taken together, substantial changes in the compound structure, size, and the geometry of the key interaction with DDB1 Arg928 are accommodated with surprisingly small penalties in glue-induced binding affinity.



**Figure 3.7. Modification of the R<sub>2</sub> group of the CR8 scaffold.** **a**, Chemical structures of derivatives with various R<sub>2</sub> substituents, e.g. inspired by SR-4835 (DS19), dinaciclib (DS70) or 21195 (DS48). **b**, *In vitro* TR-FRET complex formation assay for compounds shown in (a), DMSO, and CR8. **c**, Chemical structures of compounds obtained through R-group docking at the R<sub>2</sub> position. **d**, *In vitro* TR-FRET complex formation assay for compounds shown in (c), DMSO, and CR8. **e**, Crystal structure of the ternary complex formed with DS59 (chemical structure in (c)). Interactions are represented by dashed lines. Hinge hydrogen bonds are shown in pink, aromatic H-bonds in grey, and  $\pi$ -cation interactions in green. Possible  $\pi$ -cation interactions between the diazole and proximal CDK12 lysine residues (K756, K861) were omitted for clarity. **f**, Crystal structure of the ternary complex formed with DS50 (chemical structure in (c)). At the contour level of  $1\sigma$  the R<sub>2</sub> substituent could not be unambiguously fit into the density and wire representation marks the part of the molecule set to zero occupancy. Two probable conformations are shown. (**b, d**) Data represent the mean  $\pm$  s.d. ( $n \geq 2$ ).

**Table 3.3.** Data collection and refinement statistics for structures presented in **Figure 3.7**, **Figure 3.9**, and **Figure 3.10**.

Ligand	DRF053	DS50	DS55	DS59	DS61
Data collection					
Space group	P 3 1 2 1	P 3 1 2 1	P 3 1 2 1	P 3 1 2 1	P 3 1 2 1

<b>Cell dimensions</b>					
<b>a, b, c (Å)</b>	249.436	250.266	249.272	249.522	249.701
<b>a, b, g (°)</b>	218.342	217.643	220.107	218.546	218.495
<b>Resolution (Å)</b>	216.018-3.505 (3.697-3.505)	216.737-3.643 (3.812-3.643)	215.874-3.446 (3.678-3.446)	216.091-3.253 (3.43-3.253)	216.248-3.246 (3.416-3.246)
<b>Rsym or Rmerge</b>	0.154 (3.927)	0.523 (5.933)	0.214 (4.134)	0.239 (4.223)	0.191 (4.157)
<b>Rpim</b>	0.034 (0.876)	0.115 (1.262)	0.048 (0.933)	0.053 (0.936)	0.043 (0.928)
<b>I / σI</b>	14.7 (0.9)	7.9 (0.8)	12.9 (1)	12.4 (0.9)	15.3 (0.9)
<b>CC1/2</b>	0.999 (0.406)	0.993 (0.273)	0.999 (0.401)	0.998 (0.379)	0.999 (0.392)
<b>Completeness (ellipsoidal, %)</b>	96.1 (63.9)	96.1 (61.5)	95.1 (59.8)	94.2 (66.9)	94.8 (67.9)
<b>Redundancy</b>	21 (21)	21.4 (23)	21 (20.5)	20.8 (21.3)	20.8 (21)
<b>Refinement</b>					
<b>Resolution (Å)</b>	60.11 - 3.505 (3.63 - 3.505)	97.01 - 3.643 (3.774 - 3.643)	60.68 - 3.451 (3.574 - 3.451)	60.14 - 3.261 (3.377 - 3.261)	60.41 - 3.253 (3.369 - 3.253)
<b>No. reflections</b>	84844 (1634)	79775 (2354)	76905 (1074)	101831 (2338)	104899 (2675)
<b>Rwork / Rfree</b>	0.1887 / 0.2300	0.1967 / 0.2412	0.2218 / 0.2492	0.2129 / 0.2331	0.1865 / 0.2187
<b>No. atoms</b>	33938	34102	33862	33923	33931
<b>Protein</b>	33736	33736	33736	33736	33736
<b>Ligand/ion</b>	289	570	192	277	285
<b>Water</b>	0	0	0	0	0
<b>B-factors</b>	171.79	139	152.26	130.19	135.35
<b>Protein</b>	171.75	139.05	152.34	130.16	135.33
<b>Ligand/ion</b>	179.67	134.94	131.56	135.34	139.52
<b>Water</b>					
<b>R.m.s. deviations</b>					
<b>Bond lengths (Å)</b>	0.005	0.006	0.006	0.003	0.004
<b>Bond angles (°)</b>	0.91	0.95	0.9	0.71	0.7

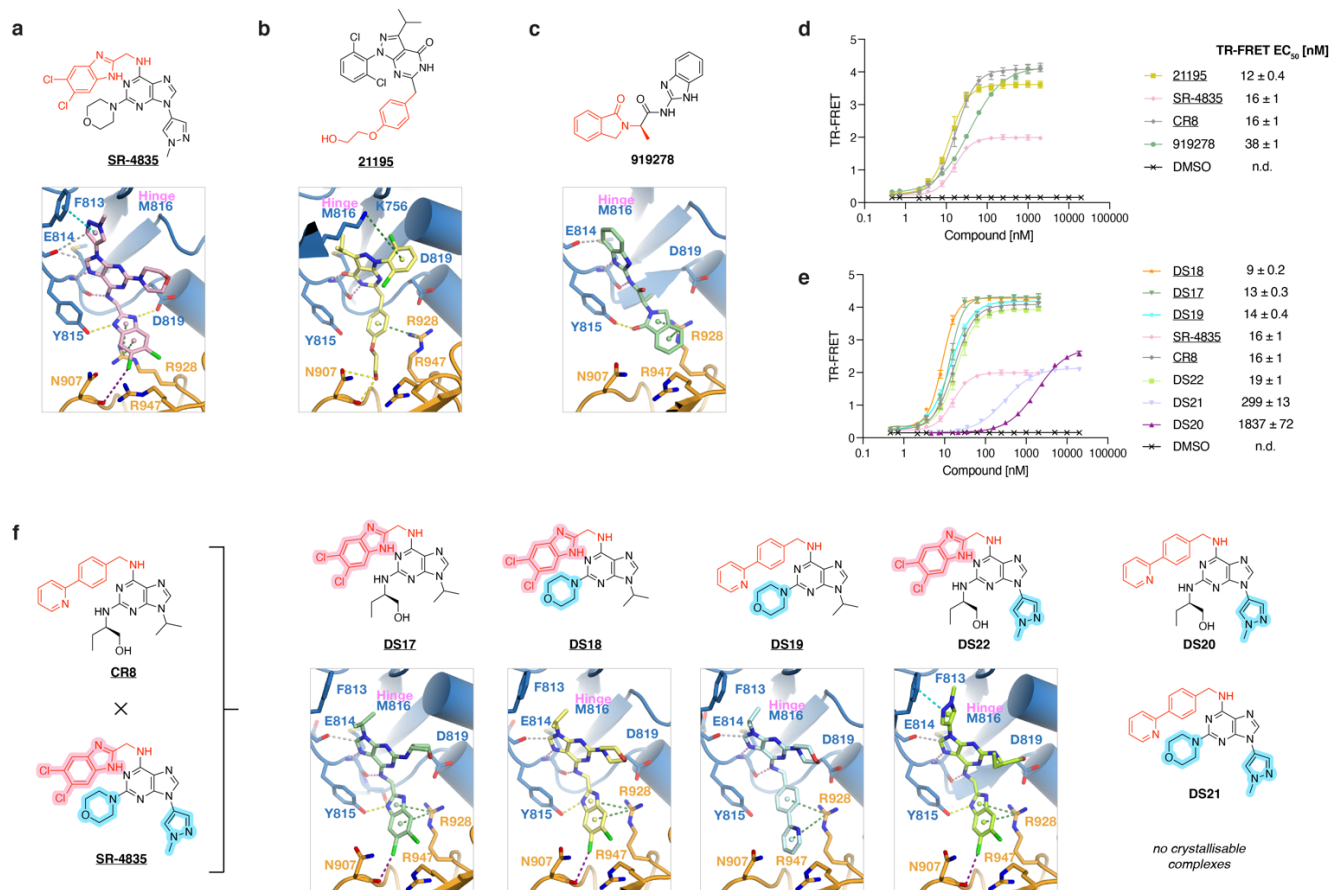
### 3.3.5 CDK12-selective inhibitors can be cyclin K degraders

Having established that gluing activity is common among CR8 derivatives, we next aimed to explore whether purine-based compounds with more diverse substituents might also show gain-of-function molecular glue activity. For this, we turned to published kinase inhibitors (**Figure 3.8a-d**, **Figure 3.9**). We first focussed on the CDK12-selective inhibitor SR-4835<sup>242,302</sup>, a compound with a CR8-like 2,6,9-trisubstituted purine structure but with different R<sub>1-3</sub> substituents. SR-4835 potently recruited DDB1 to CDK12-cyclin K in our assay (EC<sub>50</sub> = 16 ± 1 nM) (**Figure 3.8a, d**), which is consistent with a recent report that it degrades cyclin K<sup>290</sup>. Crystallisation of the ternary complex revealed that SR-4835 binds in a manner similar to CR8, yet it induces conformational changes in the N-lobe of the kinase, most notably in the aa 731-743 loop, with small variations propagating

across the complex (**Figure 3.9a**) ( $\text{RMSD}_{\text{CR8}} = 1.6 \text{ \AA}$ ). The imidazole ring of the gluing moiety interacts with DDB1 Arg928 and is additionally positioned by hydrogen-bonding interactions with Tyr815 and Asp819 of CDK12 (**Figure 3.8a**). As Tyr815 is a conserved phenylalanine in other CDKs, this interaction likely explains the high inhibitory selectivity of SR-4835 for CDK12/13. We also observed additional CDK12-ligand contacts, with  $\pi$ - $\pi$  interactions between methylpyrazole and the kinase gatekeeper residue Phe813, a putative halogen bond to DDB1 Asn907, and two  $\text{C}_{\text{Ar}}\text{-H}\cdots\text{O}$  interactions with the backbone carbonyl of Glu814.

To better understand the contributions of the SR-4835 structural features to its activity, we evaluated its derivatives (**Figure 3.9b-e**). At  $\text{R}_1$ , an unsubstituted benzimidazole gluing moiety (DS55) showed a similar binding mode and only slightly impaired recruitment ( $\text{EC}_{50} = 36 \pm 2 \text{ nM}$ ), while exchanging the benzimidazole for an indole (DS56) largely abolished activity ( $\text{EC}_{50} = 2114 \pm 252 \text{ nM}$ ), highlighting the importance of the two hydrogen-bonding interactions made possible by the imidazole moiety in SR-4835 (**Figure 3.9c-e**). We also synthesised six compounds that are hybrids of SR-4835 and CR8 (DS17-DS22), yielding a series bearing all possible combination of the  $\text{R}_{1-3}$  substituents (**Figure 3.8e, f**). While DS17-19, DS22, CR8 and SR-4835 promoted very robust *in vitro* complex formation and yield crystallisable complexes, compounds DS21 and DS20, both of which possess the phenylpyridine gluing moiety of CR8 and the SR-4835-derived methylpyrazole in  $\text{R}_3$ , performed considerably poorer in TR-FRET, with the resulting complexes refractory to structural characterisation. We conclude that these two bulky substituents cannot be accommodated at the same time, while all other combinations permit effective complex formation, with DS17-19 being among the most potent recruiters in our compound series (**Figure 3.8e, f** and **Figure 3.9b**).

Changing the position of one nitrogen atom in the purine ring system was previously reported to increase potency for CR8<sup>237,295,303,304</sup>. We therefore evaluated the effect of such scaffold hopping in the case of SR-4835, while keeping the  $\text{R}_{1-3}$  substituents unchanged (DS74). The resulting compound DS74, however, performed comparably to SR-4835 in TR-FRET ( $\text{EC}_{50} = 20 \pm 1 \text{ nM}$ ) (**Figure 3.9f, g**). Multiple structurally related compounds were disclosed in a recent patent (WO2021116178)<sup>305</sup>, and we chose to evaluate three of these reported degraders (DS71-73 corresponding to P133, P25, and P342 respectively) based on their reported potency and similarity to other small molecules in our series (**Figure 3.9f**). TR-FRET measurements revealed these molecules facilitate complex formation *in vitro* with affinities comparable to CR8 and SR-4835 ( $12 \text{ nM} < \text{EC}_{50} < 28 \text{ nM}$ ), with DS72 performing the best (**Figure 3.9g**). Compound DS71 and DS72 feature a difluorobenzimidazole gluing moiety previously reported to enhance CDK12 binding<sup>221</sup>. We investigated whether a simple substitution of the CR8 phenylpyridine for this moiety would result in higher potency (DS64) but found that it decreased the *in vitro* activity ( $\text{EC}_{50} = 41 \pm 2 \text{ nM}$ ) (**Figure 3.9f, g**). Thus, purine scaffold hopping does not independently enhance gluing activity but can yield highly potent glues (DS71-73) in combination with appropriate  $\text{R}_1$ - $\text{R}_3$  substituents.

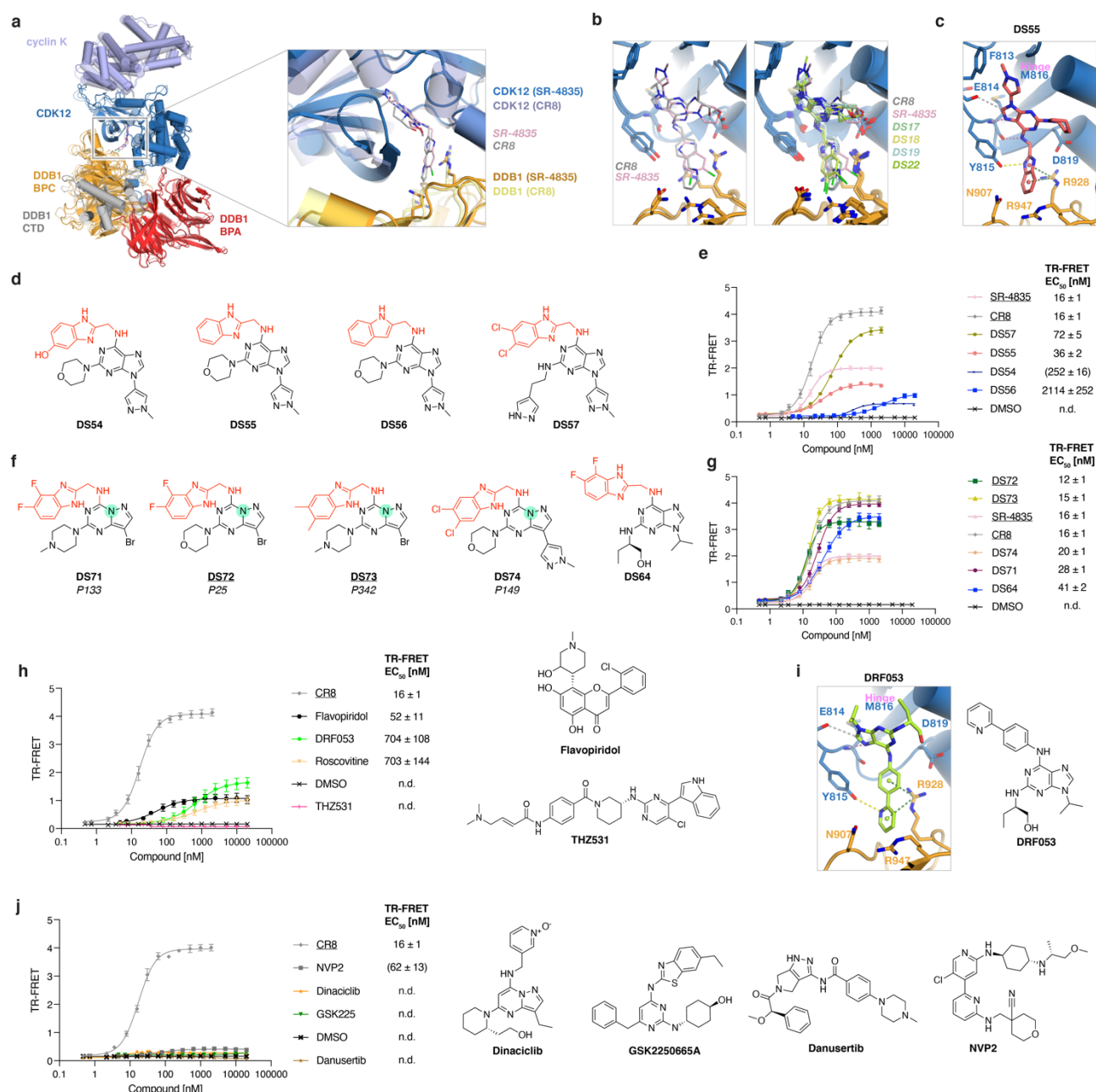


**Figure 3.8. Published CDK inhibitors have cryptic molecular glue degrader activity.** **a**, Chemical structure of SR-4835 and ternary complex crystal structure of the DDB1-CDK12 interface induced by the compound. **b**, Chemical structure of 21195 and ternary complex crystal structure of the DDB1-CDK12 interface induced by the compound. **c**, Chemical structure of 919278 and ternary complex crystal structure of the DDB1-CDK12 interface induced by the compound. **d**, *In vitro* TR-FRET complex formation assay for compound shown in (a-c), CR8, and DMSO. **e**, *In vitro* TR-FRET complex formation assay for DS17-22, which are hybrids of CR8 and SR-4835. **f**, Chemical structures of CR8/SR-4835 hybrid compounds and ternary complex crystal structures of the DDB1-CDK12 interfaces induced by DS17, DS18, DS19, and DS22. (**d**, **e**) Data represent the mean  $\pm$  s.d. ( $n \geq 2$ ). (**a-c**, **f**) Interactions are represented by dashed lines. Hinge hydrogen bonds are shown in pink, other hydrogen bonds in yellow, aromatic H-bonds in grey,  $\pi$ -cation interactions in green,  $\pi$ - $\pi$  interactions in cyan and halogen bonds in purple.

**Table 3.4.** Data collection and refinement statistics for structures presented in **Figure 3.8**.

Ligand	SR-4835	21195	919278	DS17	DS18	DS19	DS22
<b>Data collection</b>							
<b>Space group</b>	P 31 2 1	P 31 2 1	P 31 2 1	P 31 2 1	P 31 2 1	P 31 2 1	P 31 2 1
<b>Cell dimensions</b>							
	249.938	248.709	250.357	249.447	249.896	249.689	250.401
<b>a, b, c (Å)</b>	249.938	248.709	250.357	249.447	249.896	249.689	250.401
	219.302	218.615	217.567	218.45	218.57	217.939	215.706
<b>a, b, g (°)</b>	90 90 120	90 90 120	90 90 120	90 90 120	90 90 120	90 90 120	90 90 120
<b>Resolution (Å)</b>	216.453- 3.134	215.366- 3.245	216.791- 3.193	216.027- 2.984	216.416- 3.127	216.169- 3.42 (3.621- 3.42)	152.931- 3.079

	(3.333- 3.134)	(3.396- 3.245)	(3.361- 3.193)	(3.138- 2.984)	(3.317- 3.127)		(3.215- 3.079)
<b>Rsym or Rmerge</b>	0.172 (3.727)	0.157 (4.443)	0.191 (4.11)	0.165 (3.756)	0.201 (3.627)	0.238 (3.982)	0.173 (4.747)
<b>Rpim</b>	0.038 (0.824)	0.05 (1.35)	0.061 (1.245)	0.037 (0.822)	0.045 (0.795)	0.076 (1.218)	0.039 (1.032)
<b>I / sI</b>	15.1 (1)	16.4 (0.9)	14.9 (0.9)	16.4 (1)	12.3 (1)	10.8 (0.9)	17.2 (0.8)
<b>CC1/2</b>	0.999 (0.445)	0.999 (0.385)	0.999 (0.443)	0.999 (0.401)	0.999 (0.441)	0.998 (0.435)	0.999 (0.328)
<b>Completeness (ellipsoidal, %)</b>	96.4 (70.1)	96.1 (61.6)	96.5 (67.6)	96.6 (68.2)	96.1 (65.1)	96 (65)	96.1 (64)
<b>Redundancy</b>	21 (21.4)	21.1 (22.0)	21.1 (22.1)	21 (21.8)	20.9 (21.7)	21.0 (21.5)	21 (22.1)
<b>Refinement</b>							
	60.58 - 3.134 (3.246 - 3.134)	72.87 - 3.245 (3.361 - 3.245)	68.77 - 3.193 (3.307 - 3.193)	60.12 - 2.984 (3.091 - 2.984)	60.2 - 3.13 (3.242 - 3.13)	48.66 - 3.42 (3.542 - 3.42)	60.05 - 3.091 (3.201 - 3.091)
<b>Resolution (Å)</b>							
<b>No. reflections</b>	108137 (1775)	110924 (3254)	130192 (2849)	139125 (3433)	112666 (1796)	105712 (1556)	128680 (4328)
<b>Rwork / Rfree</b>	0.2104 / 0.2335	0.1972 / 0.2237	0.1996 / 0.2251	0.1945 / 0.2217	0.1970 / 0.2239	0.1882 / 0.2191	0.2209 / 0.2426
<b>No. atoms</b>	33952	33999	33925	33919	34034	33940	33908
<b>Protein</b>	33796	33736	33736	33736	33736	33736	33736
<b>Ligand/ion</b>	231	347	282	375	364	297	238
<b>Water</b>	0	0	0	0	0	0	0
<b>B-factors</b>	131.29	141.48	122.92	110	113.99	151.57	115.15
<b>Protein</b>	131.35	141.36	122.88	110.04	113.8	151.54	115.17
<b>Ligand/ion</b>	118.4	156.40	130.80	102.77	135.9	157.28	111.2
<b>Water</b>							
<b>R.m.s. deviations</b>							
<b>Bond lengths (Å)</b>	0.007	0.005	0.005	0.007	0.007	0.006	0.006
<b>Bond angles (°)</b>	1.01	0.84	0.90	0.94	0.99	0.93	0.93



**Figure 3.9. Evaluation of reported molecules and their derivatives for cryptic molecular glue activity.** **a**, SR-4835 causes conformational changes in the N-lobe of CDK12 as compared to CR8. The overlay depicts the SR-4835 structure in darker colours (dark blue for CDK12, orange for DDB1) and CR8 in lighter colours (light blue for CDK12 and yellow for DDB1). SR-4835 compound is shown as sticks in pink and CR8 in grey. **b**, Overlay of the CR8 and SR-4835 complex structure alone (left) or including DS17, DS18, D19 and DS22 (right). **c**, Crystal structure of the ternary complex formed with DS55 (chemical structure in **d**). **d**, Chemical structures of SR-4835 derivatives. **e**, *In vitro* TR-FRET complex formation assay for compounds shown in **(d)** as well as CR8 and DMSO. **f**, Chemical structures of compounds disclosed in patent WO2021116178<sup>305</sup> (the original name quoted below each structure) and our derivative DS64 bearing the same gluing moiety as DS71 and DS72. Scaffold hopping in the purine core is emphasized with a green highlight. **g**, *In vitro* TR-FRET complex formation assay for compounds shown in **(f)**, DMSO, and CR8. **h**, *In vitro* TR-FRET complex formation assay for roscovitine, DRF053, flavopiridol, THZ531, DMSO, and CR8. **i**, Crystal structure of the ternary complex formed with DRF053 and the chemical structure of DRF053. **j**, *In vitro* TR-FRET complex formation assay for various known inhibitors. **(e, g, h, j)** Data represent the mean ± s.d. (n ≥ 2). **(c, i)** Interactions are represented by dashed lines. Hinge hydrogen bonds are shown in pink, aromatic H-bonds in grey, and π-cation interactions in green.

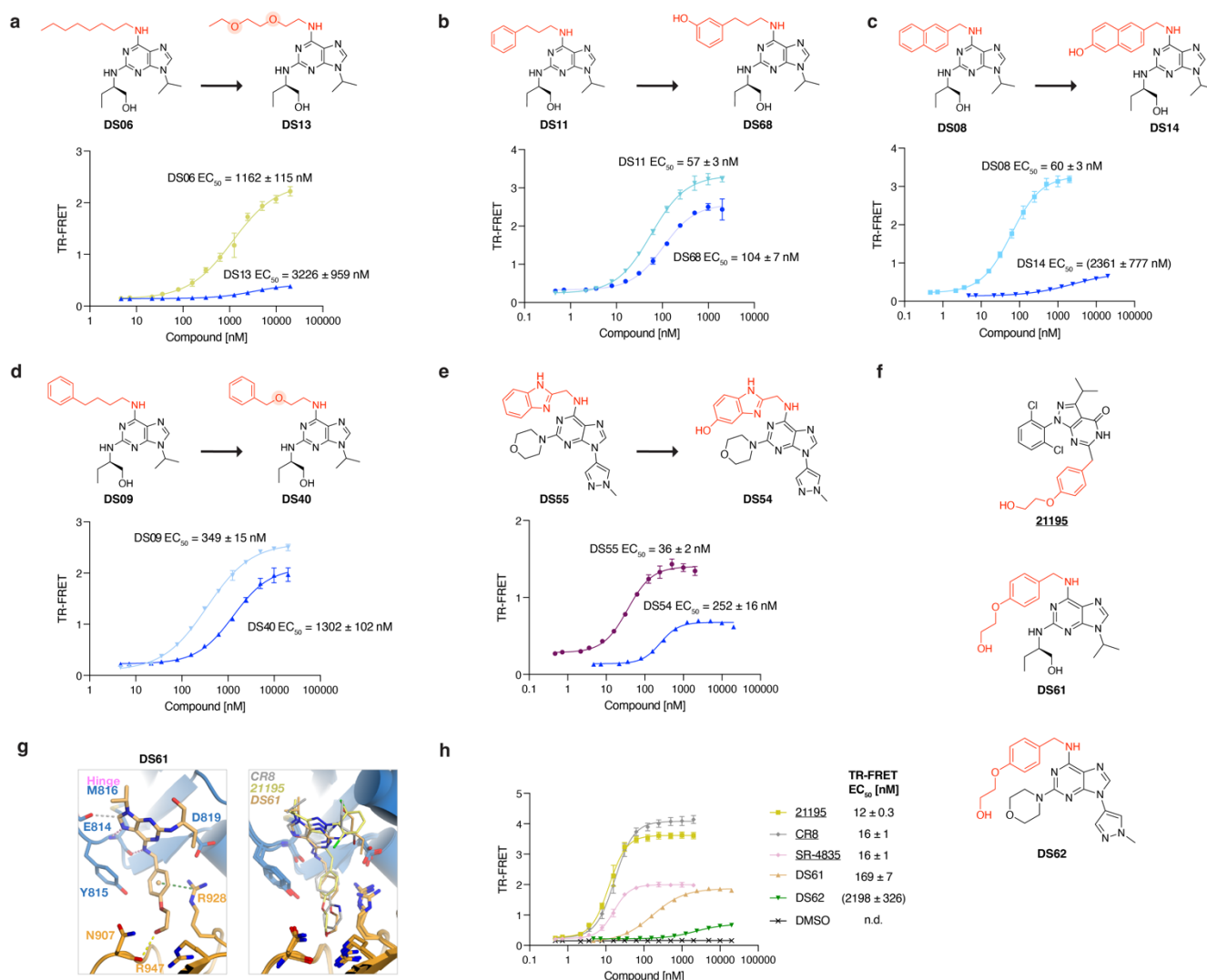


### 3.3.6 Hydrophilic moieties can engage DDB1 to degrade cyclin K

To identify more diverse DDB1-CDK12 molecular glues and therefore further interrogate the SAR, we performed virtual screening of the ZINC kinase inhibitor library against the DDB1-CDK12 interface<sup>306</sup>. We reasoned that kinase binders constitute a promising pool of putative degraders, as their distinct solvent-facing moieties could leverage diverse exit vectors for DDB1 engagement or show unique interactions. One prominent hit was the CRK inhibitor 21195 (RGB-286147)<sup>307</sup> (**Figure 3.8b**). Crystallographic analysis showed that, despite a seemingly different chemical structure, 21195 displays a similar binding mode to CR8 and potently promotes ternary complex formation ( $EC_{50} = 12 \pm 0.4$  nM) (**Figure 3.8b, d**). The pyrazolopyrimidone core is rotated by 90 degrees with respect to the CR8 purine to allow favourable hinge interactions while the three substituents occupy equivalent positions to R<sub>1-3</sub> of CR8 (**Figure 3.8b**). Remarkably, the gluing moiety of 21195, while still bearing a benzene ring capable of the key  $\pi$ -cation interaction with Arg928, has an otherwise hydrophilic character, with a hydroxy group pointing directly towards DDB1 and hydrogen-bonding with a DDB1 Asn907 (**Figure 3.8b**).

To test more broadly whether hydrophilic moieties can recruit DDB1 we derivatised several degraders with hydroxy substituents or PEG groups, which led to effects ranging from a few-fold decrease to complete loss of activity across multiple R<sub>1</sub> scaffolds (**Figure 3.10a-e**). We also synthesised additional hybrid compounds by transplanting the hydrophilic gluing moiety of 21195 onto the CR8 (DS61) or SR-4835 (DS62) scaffold (**Figure 3.10f-h**) and saw that while DS62 was incompatible with the DDB1-CDK12 interface, DS61 displayed molecular glue activity ( $EC_{50} = 169 \pm 7$  nM) (**Figure 3.10h**). The ternary crystal structure with DS61 revealed that the hydrophilic gluing moiety again engages in hydrogen bonding with Asn907 in addition to interactions with Arg928, albeit with a less optimal geometry (**Figure 3.10g**).

These results demonstrate that kinase-engaging scaffolds with distinct exit vectors can act as efficient molecular glues, and that hydrophilicity of the gluing moiety does not preclude robust DDB1-CDK12 interactions. The large DDB1-CDK12 protein-protein interface as well as compensatory interactions engaging hydrophilic moieties of the compound are sufficient to drive high-affinity association despite large expected desolvation penalties.



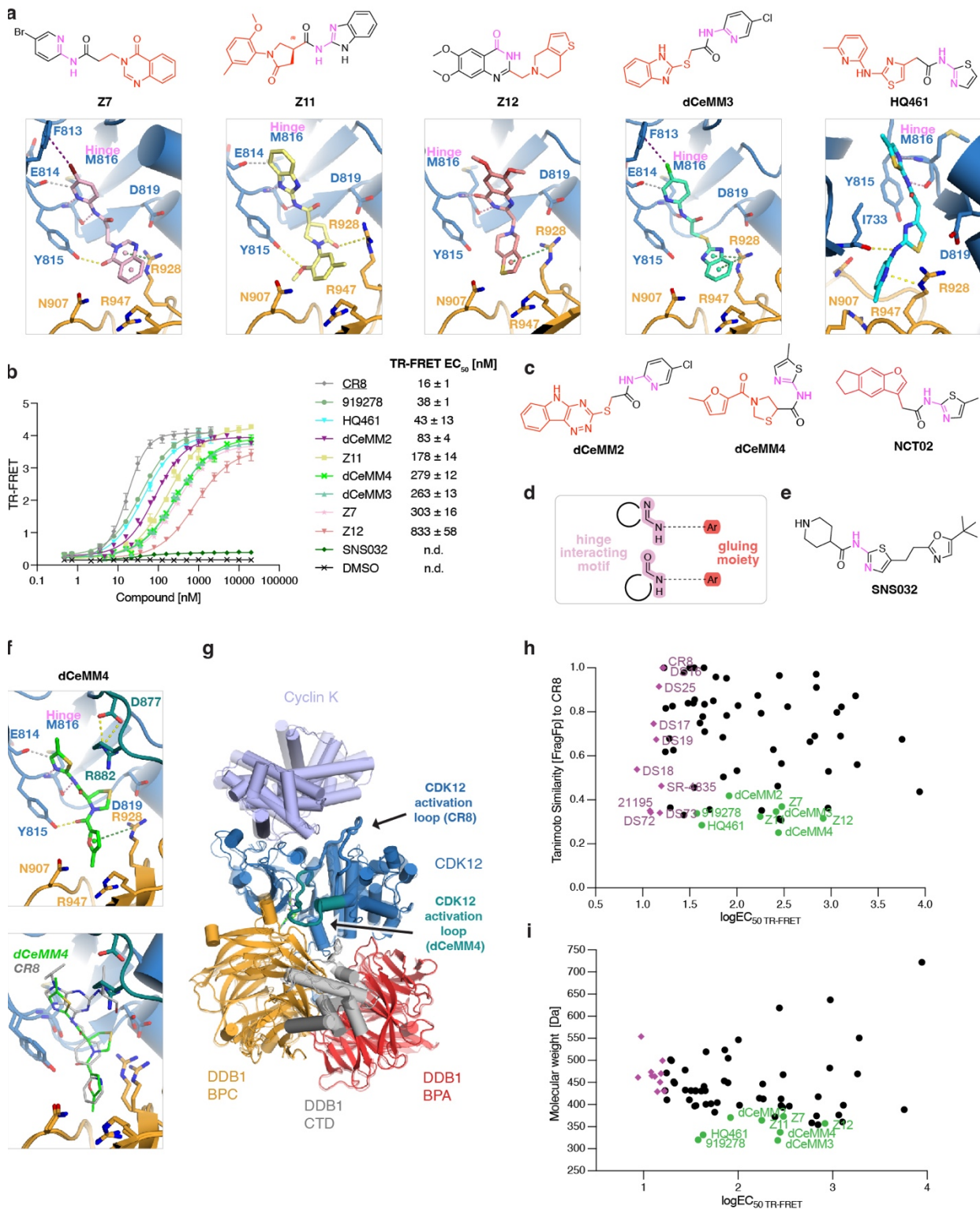
**Figure 3.10. Hydrophilicity of the gluing moiety does not preclude complex formation. (a-e)** Examples of the impact of increasing the hydrophilicity of the gluing moiety on multiple scaffolds including the relevant chemical structures and the associated *in vitro* TR-FRET compound titration results. **a**, DS06 and DS13. **b**, DS11 and DS68. **c**, DS08 and DS14. **d**, DS09 and DS40. **e**, DS55 and DS54. **f**, Chemical structures of 21195 and compounds resulting from hybridization of CR8 and SR-4835 with this inhibitor. **g**, Crystal structure of the ternary complex formed with DS61 (left) and the overlay of complexes induced by DS61, 21195, and CR8 (right). Interactions are represented by dashed lines. Hinge hydrogen bonds are shown in pink, other hydrogen bonds in yellow, aromatic H-bonds in grey, and  $\pi$ -cation interactions in green. **h**, *In vitro* TR-FRET complex formation assay for compounds shown in (f), DMSO, and CR8. (a-e, h) Data represent the mean  $\pm$  s.d. ( $n \geq 2$ ).

### 3.3.7 Specific fingerprint defines small, fragment-like cyclin K degraders

Having observed that CDK12 binders beyond CR8 derivatives display molecular glue activity through facilitation of an analogous CDK12-DDB1 complex, we investigated more structurally dissimilar CDK12 inhibitors for such gain-of-function activity. We focussed on compound 919278 (**Figure 3.8c**), for which a discrepancy between its weak CDK12 binding affinity (5.6  $\mu\text{M}$ ) and pronounced downstream transcriptional effects<sup>202</sup> prompted us to explore whether 919278 could function as a cyclin K degrader. Strikingly, despite its low molecular weight and lack of chemical similarity to CR8, SR-4835, or 21195, we found that compound 919278 recruits DDB1 to CDK12-cyclin K *in vitro* in an analogous manner (**Figure 3.8c, d**). Structural characterisation demonstrated that notwithstanding its smaller size, the compound satisfies the key hinge contacts, shows a  $\text{C}_{\text{Ar}}\text{-H}\cdots\text{O}$  interaction with Glu814 akin to that of CR8, and interacts with CDK12 Tyr815 through hydrogen bonding (**Figure 3.8c**). Crucially, its isoindolinone gluing moiety protrudes out into the interface engaging DDB1 Arg928 via  $\pi$ -cation interactions as seen for previous scaffolds (**Figure 3.8c**). Hence, the diversity observed among DDB1-CDK12 glues extends beyond typical purine-based kinase inhibitor scaffolds.

Interestingly, multiple similar, small cyclin K degraders have recently been published<sup>285,290,291</sup>. Structural analysis of several more (dCeMM3, Z7, Z11, Z12, HQ461; **Figure 3.11**) revealed that all these compounds promote the formation of a highly similar ternary complex, sustain the hinge hydrogen bonds and the  $\pi$ -cation interaction with DDB1 Arg928 (suboptimal for Z11 and dCeMM3; for HQ461 and Z11 we observe hydrogen bonding with the side chain instead), while also displaying interactions with additional residues (**Figure 3.11a**). Those include CDK12 Tyr815 (hydrogen bonds for Z7 and Z11), Glu814 ( $\text{C}_{\text{Ar}}\text{-H}\cdots\text{O}$  interaction for Z7, Z11, dCeMM3), Phe813 (*halogen bonds* for Z7 and dCeMM3), and Ile733 (hydrogen bond for HQ461) (**Figure 3.11a**).

These data suggest that DDB1-CDK12 complex formation and cyclin K degradation can be achieved with small, almost fragment-like compounds. Furthermore, the diverse set of small-molecule bound DDB1-CDK12 structures (**Figure 3.11a-c**) identifies a minimal fingerprint for a cyclin K degrader leveraging a hinge-interacting acceptor-donor motif common to kinase inhibitors, and a gluing moiety bearing an aromatic system extending from the hydrogen bond donor (**Figure 3.11d**). The structurally related inhibitor SNS032 (**Figure 3.11e**) serves as a negative example, for which a non-aromatic piperidine is expected to point towards DDB1 in the described CDK binding mode<sup>231</sup>, an observation consistent with SNS032 displaying no gain-of-function activity (**Figure 3.11b**).



**Figure 3.11. Low molecular weight cyclin K glues.** **a**, Chemical structure of Z7, Z11, Z12, dCeMM3, and HQ461 and ternary complex crystal structures of the DDB1-CDK12 interfaces induced by each compound. The HQ461 compound could in theory bind at the interface in two directions (see fingerprint in **(d)**) yet the density, while somewhat ambiguous, suggests the methylpyridine moiety pointing towards DDB1. **b**, *In vitro* TR-FRET complex formation assay for compounds shown in **(a-e)**. Data represent the mean  $\pm$  s.d. ( $n \geq 2$ ). **c**, Chemical structures of dCeMM2, dCeMM4 and NCT02, which are cyclin K degraders. **d**, The fingerprint of a cyclin K degrader. **e**, Chemical structure of SNS032, which is not a cyclin K degrader despite binding CDK12. **f**, Ternary complex crystal structure with dCeMM4 (top) and overlay of dCeMM4 and CR8 (bottom). **g**, DDB1-CDK12-cyclin K complex architecture, with a conformational change in a CDK12 activation loop

(thick ribbon) induced by dCeMM4. While a closed-loop kinase conformation is often associated with an inactive kinase state whereby the Asp-Phe-Gly (DFG) motif flips, here no DFG flip was observed. **h**, Diversity of cyclin K degraders illustrated through a plot of the compounds' Tanimoto similarity to CR8 and their ternary complex formation affinity. The ten compounds most active *in vitro* are shown in purple and the low molecular weight cyclin K degraders described above are coloured green. **i**, As in (**h**) but showing the compounds' molecular weight and their ternary complex formation affinity. (**a, f**) Interactions are represented by dashed lines. Hinge hydrogen bonds are shown in pink, other hydrogen bonds in yellow, aromatic H-bonds in grey,  $\pi$ -cation interactions in green, and halogen bonds in purple.

**Table 3.5.** Data collection and refinement statistics for the structures presented in **Figure 3.11**.

Ligand	HQ461	Z7	Z11	Z12	dCeMM4	dCeMM3
<b>Data collection</b>						
<b>Space group</b>	P 31 2 1	P 31 2 1	P 31 2 1	P 31 2 1	P 31 2 1	P 31 2 1
<b>Cell dimensions</b>						
	249.653	249.626	249.329	249.618	247.861	249.773
<b>a, b, c (Å)</b>	249.653	249.626	249.329	249.618	247.861	249.773
	216.85	218.222	218.681	220.667	220.715	218.027
<b>a, b, g (°)</b>	90 90 120	90 90 120	90 90 120	90 90 120	90 90 120	90 90 120
	216.206-	216.182-	215.925-	216.176-	214.654-	153.558-
	3.527	3.198	3.246 (3.45-	3.298	3.417	3.854
<b>Resolution (Å)</b>	(3.737-	(3.386-	3.246)	(3.484-	(3.611-	(3.995-
	3.527)	3.198)		3.298)	3.417)	3.854)
<b>Rsym or Rmerge</b>	0.216	0.147	0.177	0.223	0.158	0.766
	(4.155)	(3.571)	(4.066)	(4.301)	(4.128)	(7.594)
<b>Rpim</b>	0.048	0.033	0.04 (0.896)	0.05 (0.946)	0.035	0.17 (1.627)
	(0.927)	(0.789)			(0.906)	
<b>I / sI</b>	12.7 (0.9)	16.9 (1)	14.4 (1)	11.9 (0.9)	15.2 (0.9)	8 (0.8)
<b>CC1/2</b>	0.999	0.999	0.999	0.999	0.999 (0.39)	0.994
	(0.363)	(0.447)	(0.414)	(0.388)		(0.328)
<b>Completeness (ellipsoidal, %)</b>	95.4 (61.9)	96.3 (66.7)	95.9 (63.9)	96.1 (65.1)	96.2 (67.1)	95.5 (55.2)
<b>Redundancy</b>	21 (21)	21 (21.4)	21 (21.5)	21 (21.6)	21 (21.6)	21.1 (22.6)
<b>Refinement</b>						
	60.09 -		57.76 - 3.25		60.04 -	76.78 -
	3.531	60.35 - 3.2	(3.366 -	54.67 - 3.3	3.422	3.854 (3.992
<b>Resolution (Å)</b>	(3.657 -	(3.314 - 3.2)	3.25)	(3.418 - 3.3)	(3.545 -	- 3.854)
	3.531)				3.422)	
<b>No. reflections</b>	79144	106711	97986	100826	89556	69417
	(1316)	(2142)	(1754)	(2123)	(1943)	(3317)
<b>Rwork / Rfree</b>	0.1957 /	0.1962 /	0.1935 /	0.1806 /	0.2196 /	0.1838 /
	0.2350	0.2251	0.2191	0.2206	0.2417	0.2287
<b>No. atoms</b>	33896	34000	34032	34041	33893	33870
<b>Protein</b>	33736	33736	33736	33736	33736	33736
<b>Ligand/ion</b>	240	303	356	362	237	177
<b>Water</b>	0	0	0	0	0	0

<b>B-factors</b>	162.31	136.93	146.49	145.22	156.62	147.97
<b>Protein</b>	162.25	136.77	146.32	145	156.55	147.88
<b>Ligand/ion</b>	175.64	157.47	166.43	169.36	171.04	169.7
<b>Water</b>						
<b>R.m.s. deviations</b>						
<b>Bond lengths (Å)</b>	0.005	0.007	0.005	0.005	0.004	0.004
<b>Bond angles (°)</b>	0.84	0.97	0.86	0.89	0.6	0.68

### 3.3.8 Conformational change triggered by dCeMM4 shows PPI plasticity

While all other compounds yielded highly similar overall ternary complexes (max RMSD 1.7 Å), the dCeMM4-induced complex was distinctive (RMSD 3.1 Å). The compound is positioned at the interface in the expected manner, with  $\pi$ -cation interaction of the furan ring with DDB1 Arg928, hydrogen bonding of a carbonyl with CDK12 Tyr815 and  $C_{Ar}-H\cdots O$  interaction with Glu814 (**Figure 3.11f**). However, further inspection of the compound interface revealed that Arg882 of CDK12 is brought into the vicinity of the pocket and stabilised in position by interactions with Asp819. Such a conformational change is not feasible with other compounds due to steric hindrance, *e.g.* with the aminobutanol moiety of CR8 (**Figure 3.11f**). Remarkably, this local change is accompanied by a rearrangement of the entire activation loop of the kinase, resulting in a so-called closed-loop conformation. In two of the three molecules in the asymmetric unit (chains B, E but not chain H) this loop (aa 876-898), although flexible, is clearly flipped and at lower contour levels appears to be pointing towards DDB1 (**Figure 3.11g**). While not strictly required for compound binding, the observed conformational change illustrates inherent plasticity not only in protein-ligand interactions but also within the extensive DDB1-CDK12 interface itself.

### 3.3.9 *In vitro* versus cellular activity

A fundamental question in the molecular glue degrader field has been the relationship between ternary complex formation and cellular degradation. Based on several examples, a step function-like relationship was postulated, whereby a few-fold difference in *in vitro* affinity toggles target degradation in cells<sup>113,294</sup>. To investigate this relationship for our large set of cyclin K degraders we evaluated all compounds using a dual-colour cyclin K-eGFP stability reporter in HEK293T cells as previously described<sup>294</sup> (**Figure 3.12a, Supplementary Table S3.1**). Based on this assay, we were able to distinguish potent degraders (such as CR8 and 21195) from weaker (*e.g.* 919278) or non-degrading (*e.g.* dinaciclib) compounds (**Figure 3.12b**). The ten best-performing degraders were DS73, DS71, DS72, DS17, DS18, SR-4835, WX3, CR8, 21195, DS28, a list that overlaps remarkably well (8/10) with our *in vitro* data. We then examined the correlation of cyclin K reporter degradation ( $DC_{50}$ ) and *in vitro* ternary affinity (TR-FRET  $EC_{50}$ ) in more detail and across the entire compound set, only including compounds with unambiguous cellular degradation activity ( $n = 42$ ) (**Figure 3.12c**). We observed a linear correlation between  $\log DC_{50}$  and  $\log EC_{50} (TR-FRET)$  (R squared = 0.6) (**Figure 3.12c**), whereas the non-log data could be described with a Hill-type equation (R squared = 0.7) (**Figure 3.13a-c**).

The overall  $EC_{50}(\text{TR-FRET}) - DC_{50}$  correlation observed (**Figure 3.12c**, **Figure 3.13a, b**) suggests that the data from *in vitro* complex formation assays are largely predictive of the extent of in-cell degradation. The sigmoidal relationship further illustrates that small differences in *in vitro* affinity translate to larger discrepancies in cellular degradation, supporting the notion of an apparent affinity threshold<sup>33</sup> and fine-tuning the description of this phenomenon. We acknowledge that while TR-FRET measurements allow us to establish an approximate rank order of *in vitro* compound activities, affinities tighter than 50 nM lie at or below the dynamic range of our assay, likely leading to  $EC_{50}$  overestimation for the best compounds. Performing the assay at lower protein concentrations for selected compounds revealed sub-nanomolar dissociation constants for the top derivatives (*e.g.* DS17), but presented experimental challenges with a much lower TR-FRET assay window (**Figure 3.13c** and **Methods**). While the observed relationship between *in vitro* complex formation and cellular degradation data indicates a clear correlation, effective quantification of sub-nanomolar affinities is expected to further improve this description (**Figure 3.13b, c**).

### 3.3.10 Cyclin K degrader diversity

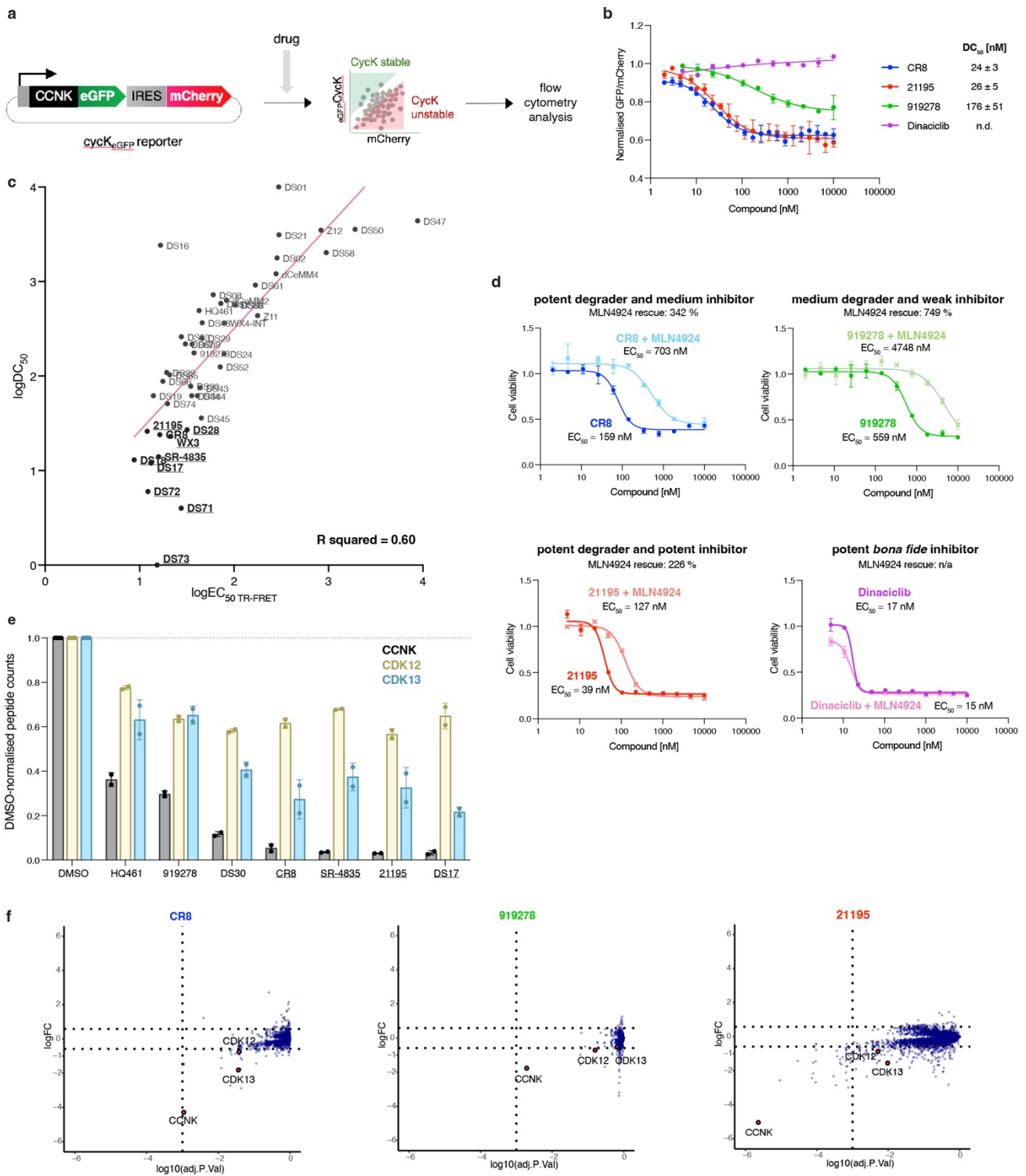
Given the surprisingly tolerant SAR we observed for the tested cyclin K degraders, we aimed to better gauge the extent of chemical diversity among our molecular glue compound set. For this, we explored the distribution of various physicochemical properties and computed the Tanimoto coefficients describing the similarity to CR8 for each compound (**Figure 3.13d** and **Figure 3.11h, i**). We found that cyclin K degraders that display *in vitro* activity span molecular weights from 317 to 722 Da and clogP values between 0.4 and 5.1 (**Figure 3.13d**). While the compounds are highly diverse, the vast majority of the evaluated small molecules is conventionally drug-like, as expected for kinase inhibitor-derived or small, almost fragment-like compounds. Notably, the compound's Tanimoto similarity to CR8 and its *in vitro* activity show no clear correlation, with the ten best DDB1-CDK12 molecular glues (DS18, 21195, DS72, DS17, DS19, DS25, DS73, SR-4835, CR8, DS16) ranging from CR8-like (1.0) to distinct (0.33) (**Figure 3.11h**), further supporting that diverse chemistries facilitate the formation of this complex.

We next set out to examine the differences in the mode of action among these diverse compounds. To probe the ubiquitin ligase-dependent cytotoxicity of the degraders, we measured the viability of HEK293T cells after 72 h of compound exposure with and without pre-treatment with 100 nM MLN4924, a neddylation inhibitor, for 2 h (**Figure 3.12d**, **Supplementary Table S3.1**, **Supplementary Figure 4**). The rescue of cytotoxicity by MLN4924 indicates that changes in cell viability are due to neddylation-dependent degradation of cyclin K, as would be expected from the DDB1-mediated gluing effect, whereas cytotoxicity due to *e.g.* CDK inhibition would not be altered by MLN4924. While for some compounds cyclin K degradation appears to be the dominant contribution to cytotoxicity (*e.g.* 919278), others (such as 21195) work partly by different mechanisms likely involving conventional kinase inhibition (**Figure 3.12d**) (see below).

We further utilised the results of multiple assays –  $\log DC_{50}$ ,  $\log EC_{50}$  TR-FRET,  $\log EC_{50}$  cell viability, and a descriptor of the rescue of degrader toxicity upon MLN4924 co-treatment (% MLN rescue) – as inputs for principal component analysis (PCA). Distinct clusters of cyclin K molecular glue degraders emerged, with the best degraders placed into multiple subgroups (**Figure 3.13e**). The cyclin K degrader set therefore not only features compounds with considerable diversity in size and physicochemical properties (**Figure 3.11h**, **Figure 3.13d**) but also a diverse range of cellular properties.

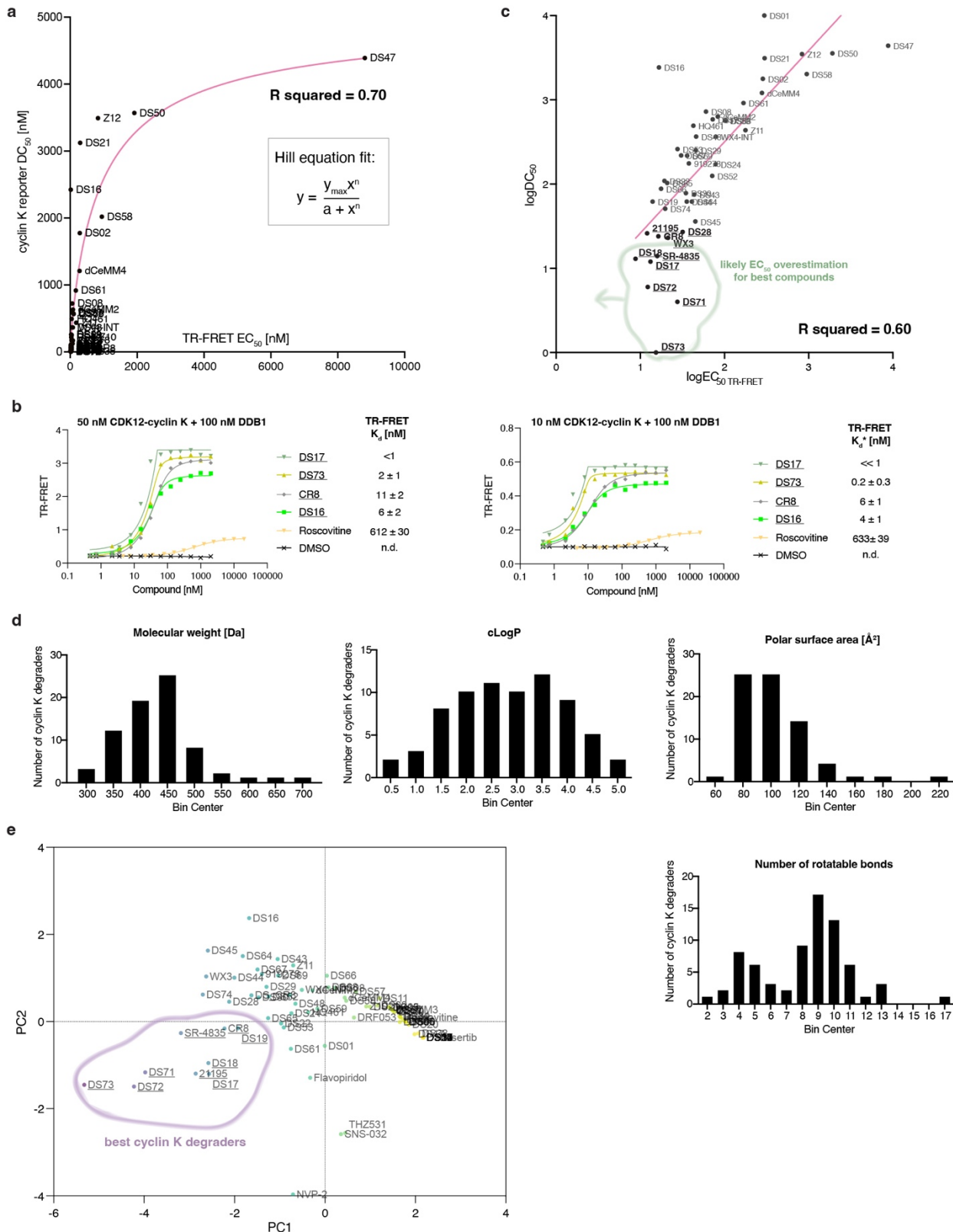
To better define how different compounds give rise to diverse cellular responses, we next assessed the extent of cellular CDK12, CDK13 and cyclin K degradation these small molecules cause by mass spectrometry. We performed quantitative label-free proteomics on MDA-MB-231 cells treated with several representative compounds to assess the depth and specificity of cyclin K degradation (**Figure 3.12e, f** and **Figure 3.14**). These experiments identified three of the six compounds tested (DS17, 21195 and SR-4835) to be more effective at depleting cyclin K than CR8 at the tested concentration (**Figure 3.12e, f**). We further observed that all compounds selectively degrade cyclin K and that the extent of depletion correlates well with the cyclin K stability reporter and TR-FRET complex formation data (**Figure 3.14b**; R squared = 0.84). Interestingly, while CDK12 showed only a mild decrease of abundance with all compounds, CDK13 depletion was more pronounced and appeared to scale with that of cyclin K (**Figure 3.12e**).



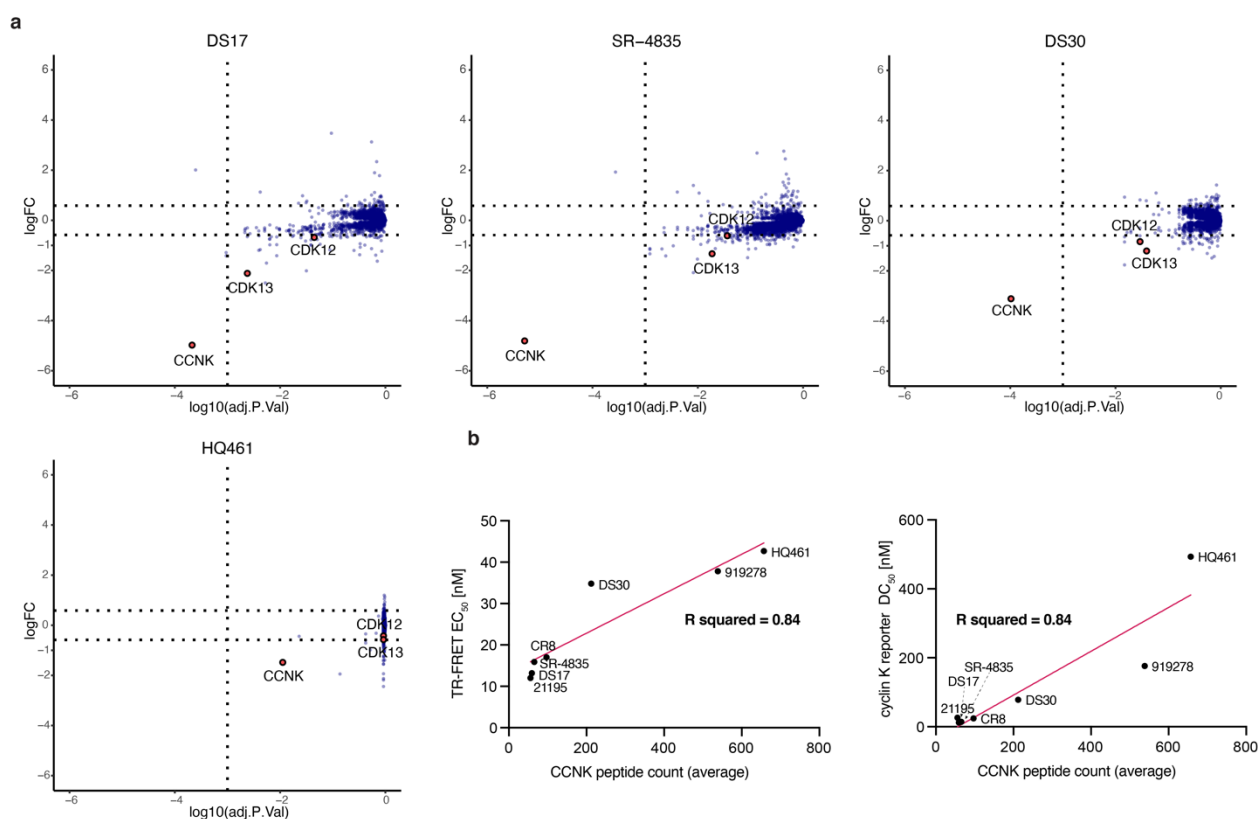


**Figure 3.12. Diverse cyclin K molecular glue degraders give rise to unique cellular responses.** **a**, Schematic of the cyclin K dual-colour reporter assay. **b**, HEK293T cycK<sub>eGFP</sub> reporter assay results for four example compounds. Data represent the mean ± s.d. (n ≥ 2). **c**, Correlation of *in vitro* complex formation affinity (logEC<sub>50</sub> TR-FRET) and cycK<sub>eGFP</sub> reporter results (logDC<sub>50</sub>) modelled with a linear regression (R squared = 0.60). The *in vitro* TR-FRET EC<sub>50</sub> values for the best compounds (bold) are overestimated, which negatively impacts the correlation. The equivalent plot in linear scale is shown in EDF6a. **d**, Viability assay in HEK293T cells for four example compounds, with curves corresponding to treatment with the individual drug or additional pre-treatment with 100 nM of the neddylation inhibitor MLN4924. Data represent the mean ± s.d. (n ≥ 2). **e**, Representation of the average peptide counts of cyclin K (dark grey), CDK12 (yellow), and CDK13 (blue) across all runs. Data represent the mean ± s.d. (n=4 for DMSO and n=2 for each compound treatment). The

corresponding volcano plots can be found in **Figure 3.14a. f**, MDA-MB-231 cells were exposed to 1  $\mu\text{M}$  CR8, 919278, 21195, or DMSO for 5 hours followed by whole proteome quantification using label-free mass spectrometry (mean log<sub>2</sub> fold change, p value calculated by a moderated t-test, n=4 (DMSO), n=2 (CR8, 919278, 21195)).



**Figure 3.13. Further characterisation of molecular glue degrader compounds.** **a**, Correlation of *in vitro* TR-FRET assay data (TR-FRET EC<sub>50</sub> [nM]) with cellular cyclin K<sub>eGFP</sub> reporter assay results (DC<sub>50</sub> [nM]). Only compounds with unambiguous activity <10  $\mu$ M in the reporter assay were included (n = 42). The relationship can be described with a Hill equation, yielding an R squared of 0.70. The Hill equation is used solely to describe the mathematical relationship between the two experimental datasets and not for its biological meaning. **b**, Correlation of *in vitro* complex formation affinity (logEC<sub>50</sub> TR-FRET) and cycK<sub>eGFP</sub> reporter results (logDC<sub>50</sub>) modelled with a linear regression (R squared = 0.60). The *in vitro* TR-FRET values for the best compounds (bold) are overestimated and the green cloud indicates how the data points would be predicted to shift upon accurate affinity quantification, further improving the observed correlation (*c.f.* panel **c**). **c**, *In vitro* TR-FRET complex formation assay performed as a compound titration with 50 nM CDK12-cyclin K (left) or 10 nM CDK12-cyclin K (right) performed for several chosen compounds. Data represent the mean  $\pm$  s.d. (n  $\geq$  2). Data were fitted with a quadratic equation appropriate for a case where the expected K<sub>d</sub> value is lower than the protein concentration used<sup>308</sup>. The spurious fit for DS17 reflects its tight affinity being outside of the dynamic range of the assay. K<sub>d</sub>\* correspond to the value determined with 10 nM CDK12-cyclin K. Lowering the protein concentration resulted in a much smaller assay window but yielded K<sub>d</sub> values in the sub-nanomolar range for the top compounds (DS17, DS73), while showing no difference for the weak recruiter roscovitine, indicating that the tightest glues lie below the limit of detection of the TR-FRET assay (see **Methods**). **d**, Histograms displaying the molecular weight, clogP, polar surface area, and the number or rotatable bonds distributions across the cyclin K degrader compound set. **e**, PCA of the combined multi-assay data for cyclin K degraders.



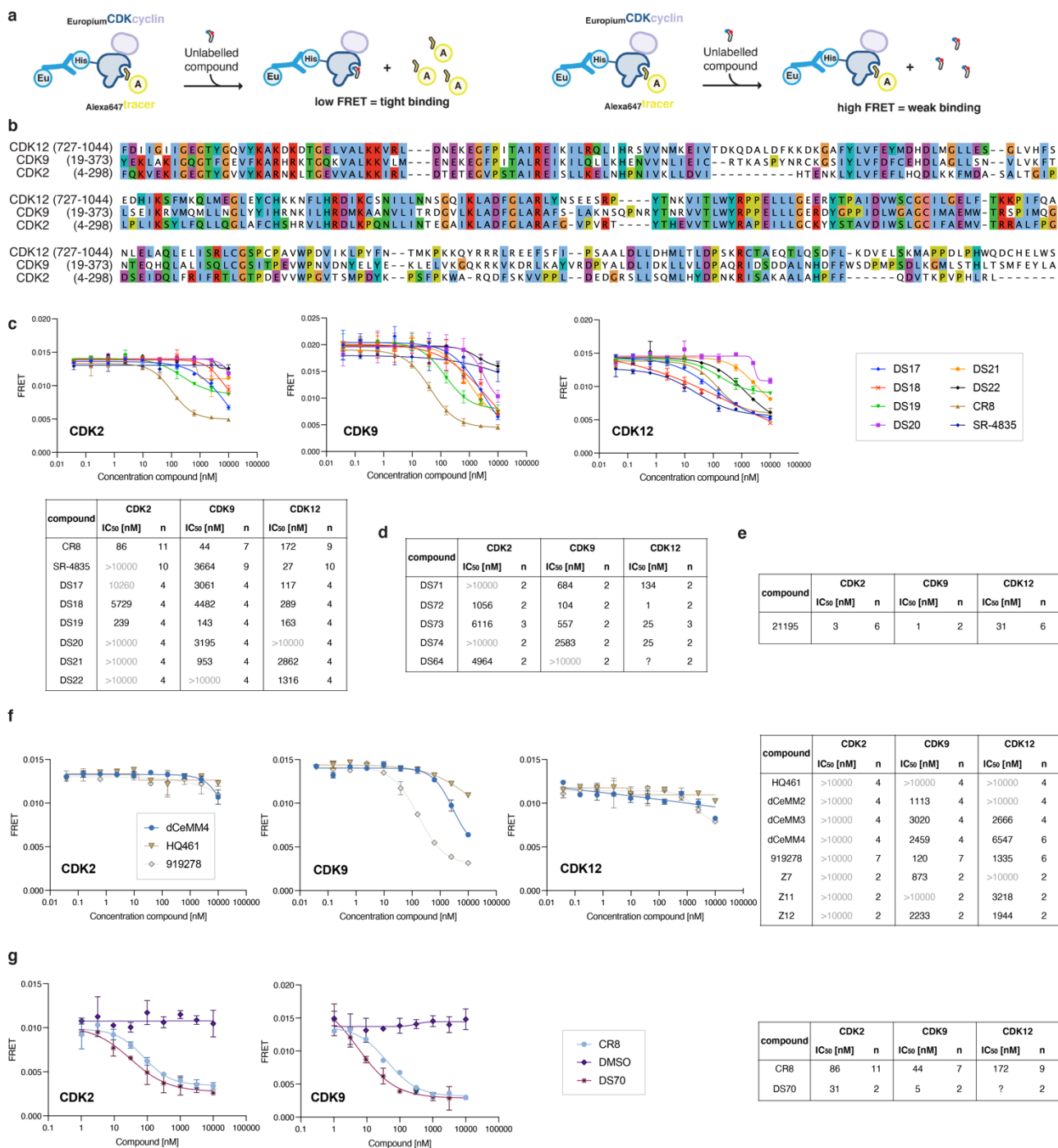
**Figure 3.14. Proteomics reveals large differences in the extent of cyclin K depletion among the degrader set.** **a**, Label-free proteomics for DS17, 21195, SR-4835, CR8, DS30, 919278, and HQ461 (n=2) versus DMSO (n=4). MDA-MB-231 cells were treated with 1  $\mu$ M drug for 5 h. **b**, Correlation of the average cyclin K peptide count with the TR-FRET EC<sub>50</sub> (left) or with the cyclin K<sub>eGFP</sub> reporter DC<sub>50</sub> (right).

### 3.3.11 Kinase inhibitory selectivity can be decoupled from glue activity

Next, we examined the kinase inhibitory selectivity of these degraders. While CR8 has a dual CDK inhibitor/cyclin K degrader mode of action, the diverse scaffolds presented herein, including CDK12-selective drugs (SR-4835 and derivatives) or fragment-like compounds (919278, Z7, Z11, Z12, dCeMM2-4, HQ461) principally offer a path to tune inhibitory selectivity or even decouple kinase inhibition from molecular glue action (**Figure 3.15**). We therefore measured CDK inhibitory activity of the compounds using Lanthascreen, an assay which monitors the displacement of a fluorescent tracer from the active site of an epitope-tagged kinase (**Figure 3.15a**). We performed the assay with CDK12, as well as with CDK9 and CDK2, which are representative transcriptional and cell-cycle CDKs that are similar in sequence to the kinase domain of CDK12 (**Figure 3.15b**, **Supplementary Figure S3.3**).

As SR-4835 is CDK12/13-selective while CR8 inhibits CDK1/2/5/7/9/12/13, we used the Lanthascreen assay to evaluate the CR8/SR-4835 hybrid compounds (DS17-22) and therefore investigate how the CDK family selectivity is tuned by compound modifications (**Figure 3.8**, **Figure 3.15c**). While SR-4835, DS17, and DS18 were CDK12-selective, CR8 and DS19 were not, which links potent pleiotropic CDK inhibition to the co-existence of the phenylpyridine and the isopropyl substituents on the purine scaffold (**Figure 3.15c**). This suggests that the modification of the gluing moiety from phenylpyridine to dichlorobenzimidazole (CR8 vs DS17 or DS19 vs DS18) is sufficient both to increase the gluing potency and to reduce off-target CDK inhibitory effects of CR8 (**Figure 3.15c**).

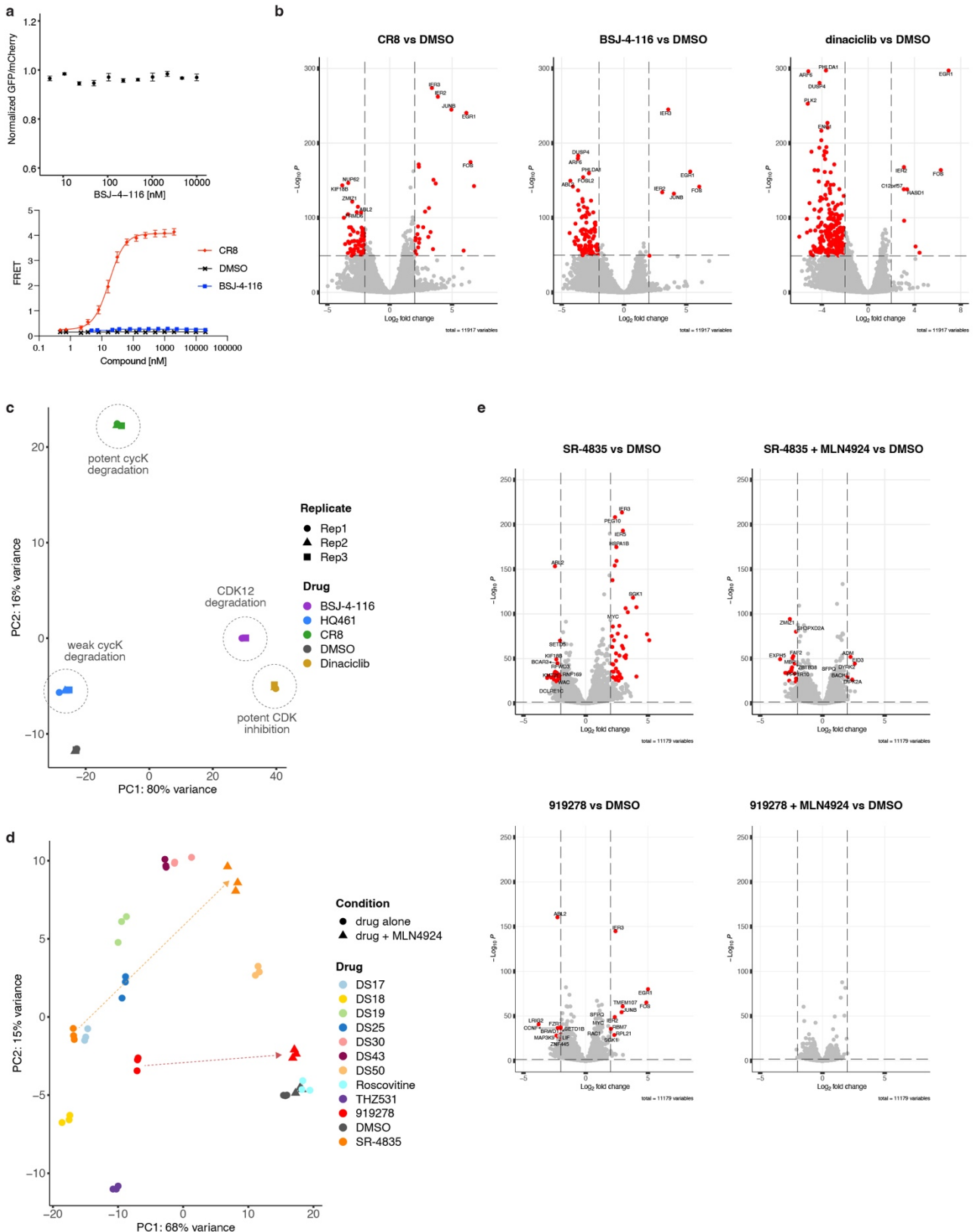
Probing of how small cyclin K degraders promote the formation of the DDB1-CDK12-cyclin K complex revealed a less optimal engagement of the ATP pocket by these compounds than by purine-based drugs. We therefore asked whether these compounds show CDK inhibitory activity and observed that while CR8 displayed  $IC_{50}$  values of 40-200 nM against these CDKs, the *small* cyclin K degraders showed much weaker binding (**Figure 3.15f**). None of the compounds appreciably bound CDK2 and while compound 919278 showed some inhibition of CDK9/12 ( $IC_{50\ CDK9} = 120$  nM,  $IC_{50\ CDK12} = 1335$  nM), dCeMM2-4 and Z11-12 displayed only micromolar  $IC_{50}$ s and HQ461 demonstrated no inhibition up to 10  $\mu$ M (**Figure 3.15f**, **Supplementary Figure S3.3**). These results show that cyclin K degraders do not necessarily require traditional kinase inhibitory properties for molecular glue activity. Kinase inhibitory selectivity can thus be decoupled and tuned separately from ternary complex affinity optimisation. Notably, the diversity of cyclin K degraders stems not only from the many distinct ways by which the CDK12-DDB1 interface can be effectively bridged by gluing moieties, but also from the plethora of ways the feature-rich ATP pocket of CDK12 can be engaged by small molecules. This in turn gives rise to compounds with a varying relative extent of inhibitory and gain-of-function degradative properties.



**Figure 3.15. Cyclin K degraders do not require traditional kinase inhibitory properties for activity.** **a**, Schematic of the commercial *in vitro* Lanthascreen assay. **b**, Multiple sequence alignment of CDK12, CDK9, and CDK2 omitting the C-terminal extension of CDK12. **c**, Selectivity profiling of CR8/SR-4835 hybrid compounds across CDK2/9/12 using Lanthascreen. Representative TR-FRET curves (top) and a table listing the average IC<sub>50</sub> values and total number (n) of measurements (bottom). **d**, Selectivity profiling results for compounds derived from a recent patent and related derivatives. **e**, Selectivity profiling of compound 21195. **f**, Representative TR-FRET curves (left) and a table listing the average IC<sub>50</sub> values and total number (n) of measurements (right) for low molecular weight cyclin K degraders. **g**, Representative TR-FRET curves (left) and a table listing the average IC<sub>50</sub> values and total number (n) of measurements (right) for CR8 and DS70, a hybrid of CR8 and dinaciclib bearing the hydroxyethyl piperidine functionality present in this potent CDK inhibitor. DS70 shows more pronounced CDK engagement than CR8. (c-g) Data represent the mean ± s.d. (n ≥ 2). Additional replicates and further data can be found in **Supplementary Figure S3.3**.

### 3.3.12 Cyclin K degraders and CDK12 inhibitors show distinct transcriptional signatures

Finally, we asked whether a cyclin K degrader effects a different transcriptional response than a CDK12 degrader or a CDK12 inhibitor. For this, we performed RNAseq in MDA-MB-231 cells, with triple negative breast cancer being a highly relevant therapeutic context for CDK12/13-cyclin K inactivation<sup>242,249</sup> (**Figure 3.16**). We observed that while a CDK12 inhibitor (dinaciclib) and CDK12 degrader (PROTAC BSJ-4-116 that does not degrade cyclin K<sup>243</sup>; **Figure 3.16a**) clustered together following PCA, both the potent (CR8) and the weak degrader (HQ461) clearly clustered separately (**Figure 3.16b, c**). The degradation of cyclin K therefore results in different transcriptional signatures than degradation or inhibition of CDK12. Supporting this notion, mRNA sequencing of cells treated with a larger set of compounds followed by PCA revealed specific clusters relating to their mechanism of action. The relative extent of cyclin K degradation (and the associated CDK12/13 depletion) versus sole CDK inhibition further differentiates the observed PCA clusters (**Figure 3.16d**). Moreover, the analysis of related compounds (*e.g.* CR8 versus DS17-19 and SR-4835) illustrates how subtle modifications impacting both inhibitory selectivity and degrader potency can further tune these signatures (**Figure 3.16d**). We further observed that co-treatment of cells with MLN4924 rescued the cyclin K degradation-related perturbation, resulting in a shift to a DMSO-like transcriptional state for compounds with limited CDK12 inhibition (919278) or elsewhere for potent CDK12/13 inhibitors (SR-4835) (**Figure 3.16d, e**). Therefore, cyclin K degrader compounds range from multi-CDK inhibitors/cyclin K degraders (*e.g.* CR8), to CDK12/13-selective inhibitors and cyclin K degraders (*e.g.* SR-4835, DS17), and degraders with little-to-no CDK inhibition (*e.g.* HQ461), all of which differ in their cellular activity, thus offering distinct therapeutic opportunities.



**Figure 3.16. RNA sequencing reveals distinct transcriptional signatures for CDK12 inhibition, CDK12 degradation, and cyclin K degradation.** **a**, Cyclin K reporter degradation assay (top) and *in vitro* TR-FRET complex formation assay (bottom) showing no cyclin K molecular glue degrader activity for the CDK12 PROTAC BSJ-4-116. Data represent the mean  $\pm$  s.d. ( $n \geq 2$ ). **b**, Volcano plots depicting the RNAseq results for CR8 (cyclin K degrader), BSJ-4-116 (CDK12 degrader), and dinacliclib (CDK inhibitor). **c**, PCA of the RNAseq data for CR8, BSJ-4-116, dinacliclib, HQ461, and DMSO

(n=3). **d**, PCA analysis of the RNAseq data for a larger selection of compounds (n=3). For 919278 and SR-4835, two conditions were assessed: compound alone (circle) or compound + MLN4924 (triangle). Co-treatment with the neddylation inhibitor resulted in large shifts of the resulting points (dashed arrows). **e**, Volcano plots depicting the RNAseq results for 919278 and SR-4835, either alone or in combination with MLN4924.

### 3.4 Discussion

In this study, we extensively surveyed the structure-activity relationship of cyclin K molecular glue degraders and, in the process, identified multiple novel scaffolds that function as cyclin K degraders, including published kinase inhibitors. A total of 94 compounds were evaluated, 64 of these were found to trigger CDK12-DDB1 complex formation with an affinity of <1  $\mu$ M, among which 40 had <100 nM activity. Crystallographic dissection of 29 compound-induced ternary complexes demonstrated that despite considerable chemical diversity of the compounds (**Figure 3.11h, i** and **Figure 3.13d**), the overall DDB1-CDK12/13 complex architecture is preserved.

All the identified glues contact DDB1 Arg928, a residue we previously showed to be essential for complex formation<sup>294</sup>. Arginine residues are principally able to engage in polyvalent, low-selectivity ligand interactions and arene-arginine contacts have been shown to involve a mix of electrostatic and dispersion attractions<sup>309–312</sup>. Accordingly, we find that gluing moieties contact DDB1 preferentially through  $\pi$ -cation interactions (*e.g.* DS08, **Figure 3.1f**), which can support drug interactions also over large distances ( $\sim$ 6 Å) and variable geometries<sup>312</sup>. The compounds can additionally contact Arg928 via hydrogen-bonding (*e.g.* Z11 and HQ461, **Figure 3.11a**) and hydrophobic interactions (*e.g.* DS06, **Figure 3.1d**). The low-affinity DDB1-CDK12 interaction primes the complex, while the ligand, through filling the kinase pocket and bridging over to DDB1, drives robust association. This, together with the fact that many distinct scaffolds can bind the kinase pocket and multiple different features of gluing moieties effectively foster DDB1 interactions, offers a rationale for the unusually broad chemical diversity observed.

We put forward several conceptual derivatisation strategies as well as a large set of degrader compounds with diverse chemical properties, many of which can serve as advantageous starting scaffolds for medicinal chemistry optimisation of clinically relevant cyclin K degraders. We also report small molecules that are more potent cyclin K degraders and show better kinase inhibitory selectivity profiles than CR8 (*e.g.* DS17). As cyclin K degraders phenotypically differ from CDK12 degraders or inhibitors (**Figure 3.16**), the compounds discussed herein offer unique therapeutic opportunities, further enabling the therapeutic pursuit of CDK12/13-cyclin K as emerging targets in oncology<sup>190,242</sup>. Our results demonstrate that the inhibitory versus degradative properties of the molecular glue degrader can be directly tuned, and that degradation can be gradually decoupled from inhibition. This is most evident for smaller compounds such as HQ461, which can provide an important starting point for the development of a *de facto* interface stabiliser that predominantly triggers cyclin K degradation not kinase inhibition (**Figure 3.11**). Thus, the data presented herein reveal design principles for small molecules that robustly inactivate CDK12/13-cyclin K and have properties that can differ from traditional kinase inhibitor scaffolds. More broadly, our results show that low-affinity interactions can be effectively strengthened by diverse chemical matter. The more extensive and complementary the interface and the larger the interfacial



cavity, the more solutions will exist to achieve gain-of-function glue activity with small molecules. Hence, low-affinity interfaces that encompass a defined ligand-binding pocket offer ideal starting points for molecular glue development.

We predict that many more compounds and scaffolds can function as cyclin K degraders if they (i) bind the CDK12/13 active site, even if very weakly, (ii) are shape- and size-compatible with the CDK12/13-DDB1 interface and its *circa* 370 Å<sup>3</sup> cavity<sup>313</sup>, and (iii) engage DDB1 residue Arg928 with various interaction geometries. We did not observe the targeting of other kinases in our studies beyond CDK12/13-cyclin K (**Figure 3.12f**, **Figure 3.14a**) and, notably, CDK13 degradation was more pronounced than that of CDK12 and appeared to scale with the loss of cyclin K (**Figure 3.12e**). The specificity for the CDK12/13-DDB1 pairing therefore appears to be primed predominantly by the protein-protein interface (~2100 Å<sup>2</sup>), featuring a CDK12/13-unique C-terminal tail<sup>294</sup>, and this complementarity is underscored by their ~50 μM binding affinity in the absence of compound<sup>294</sup>. Nonetheless, additional compound-mediated interactions are required - not for specificity, but for sufficient stabilisation of the ternary complex for cyclin K ubiquitination and degradation.

Our data further illustrate that small differences in ternary complex affinity translate to larger disparities in cellular degradation, which builds on an earlier concept of an affinity threshold (**Figure 3.12c**, **Figure 3.13a-c**)<sup>113,294</sup>. The correlation between cellular degradation and *in vitro* complex formation observed among a large compound set suggests that biochemical ternary affinity is predictive of in-cell degradation. Our data also demonstrate that cellular degrader screening for cell viability +/- MLN4924 (**Figure 3.12d**) is feasible and informative for degraders and is akin to compound screening in wild-type versus hyponeddylated cellular models<sup>285,294</sup>.

Taken together, these results offer generalisable learnings for other classes of molecular glues. We propose that the absolute size of the protein-protein interface, as well as the relative contribution of the small molecule to the interface determines both the SAR and the substrate specificity behaviour. For IMiDs and CRBN, the compound contributes *circa* 40 % of the relatively small interface area and therefore drives both specificity and affinity with tightly defined SAR<sup>293,297,298</sup>. In contrast, for DDB1-CDK12-cyclin K, where the extensive protein-protein interface with a *ca.* 20 % compound contribution allows for much greater variability in protein-ligand interactions, yielding a tolerant SAR but limiting the compound's contribution to neosubstrate specificity<sup>293</sup>.

## Chapter 4 : Fluorescence-based binding studies

This chapter contains both unpublished (section 4.1) and predominantly published (section 4.2) data.

**Section 4.1** offers a general introduction to TR-FRET and details TR-FRET assay optimisation for the DDB1-CDK12-cyclin K system (see **Chapter 2** for the initial version of the assay and **Chapter 3** for the final assay setup). I planned and performed experiments presented in this section, some of which were undertaken together with Vivian Focht, a Master student that I supervised at the time. I would also like to acknowledge the help of Sandra Romero-Lopez and Ulrich Hassiepen (Novartis) who provided helpful advice during the assay optimisation process.

**Section 4.2** focuses on fluorescence polarisation experiments to study protein-DNA and protein-nucleosome interaction and is based on two published manuscripts. **Section 4.2.1** introduces the technique and in **section 4.2.2**, I present a short summary of a published manuscript describing the interplay between OCT4-SOX2 and the nucleosome, work which was spearheaded by Alicia Michael (Thomä lab, FMI), Ralph Grand and Luke Isbel (Schübeler lab, FMI). My contributions to this study include the optimisation and performance of fluorescence polarisation assays to monitor TF-DNA and TF-nucleosome interactions, as well as protein and DNA purification and nucleosome reconstitutions. The fluorescence polarisation assays I performed are described in more detail in **section 4.2.2.4** (comprising both published and unpublished data).

### Mechanisms of OCT4-SOX2 motif readout on nucleosomes

A. Michael\*, R. Grant\*, L. Isbel\*, S. Cavadini, **Z. Kozička**, R. Bunker, A. Schenk, A. Graff-Meyer, G. Pathare, J. Weiss, S. Matsumoto, L. Burger, D. Schübeler<sup>†</sup>, N. Thomä<sup>†</sup>  
*Science* (2020); <https://doi.org/10.1126/science.abb0074>

**Section 4.2.3** features a summary of another published article focusing on cGAS-nucleosome interactions, an effort led Ganesh Pathare (Thomä lab, FMI) and Alexeine Decout (Ablasser lab, EPFL). I contributed to this work by establishing a fluorescence polarisation assay for cGAS-DNA and cGAS-nucleosome interactions as well as performing the experiments together with Joscha Weiss, a Master student at the time (Thomä lab, FMI). These efforts (comprising both published and unpublished data) are described in more detail in **sections 4.2.3.3 and 4.2.3.4**.

### Structural mechanism of cGAS inhibition by the nucleosome

G. Pathare, A. Decout, S. Glück, S. Cavadini, K. Makasheva, R. Hovius, G. Kempf, J. Weiss, **Z. Kozička**, B. Guey, P. Melenec, B. Fierz, N. Thomä<sup>†</sup>, A. Ablasser<sup>†</sup>  
*Nature* (2020), <https://doi.org/10.1038/s41586-020-2750-6>

The published manuscripts in PDF format can be found in the **Appendix** of this thesis.

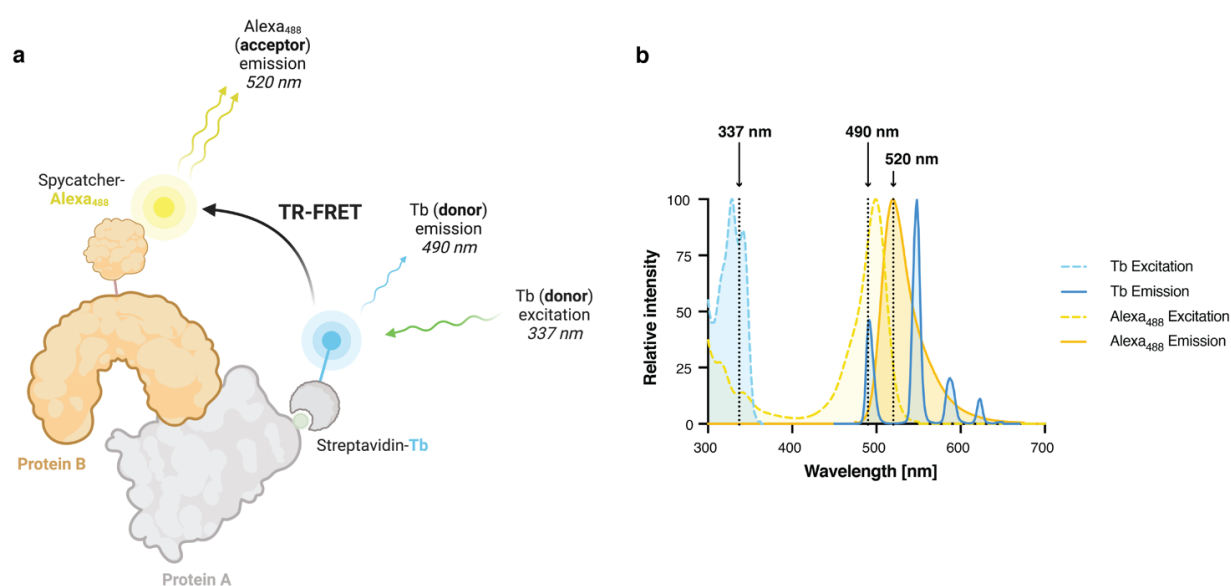
## 4.1 *In vitro* TR-FRET optimisation for molecular glue degrader screening and evaluation

### 4.1.1 The principles of TR-FRET

Fluorescence is the process of emission of light by a molecule that has absorbed initial excitatory radiation, and this phenomenon can be leveraged in quantitative assays that enable the study of protein-protein and protein-ligand interactions. Fluorescence-based *in vitro* experiments are of paramount importance in biology and due to their sensitivity, reproducibility, and scalability, they are particularly widely used in drug discovery. Two commonly used techniques are time-resolved Förster resonance energy transfer (TR-FRET) and fluorescence polarisation (FP). This subsection provides a general overview of TR-FRET, while the following ones detail the assay optimisation process for the evaluation of cyclin K molecular glue degraders.

FRET describes a nonradiative energy transfer mechanism that can occur between two matched fluorophores, termed donor and acceptor (**Figure 4.1a**)<sup>315</sup>. When excited by light at its fluorescence excitation wavelength, a fluorescent donor can transfer this energy by a long-range dipole-dipole coupling mechanism to an acceptor fluorophore, and thereby return to its ground electronic state. The efficiency of such an energy transfer is inversely proportional to the sixth power of the distance (**Equation 4.1**), making FRET a highly sensitive metric of whether two fluorophores are in proximity<sup>315</sup>. The fluorophore pair must be selected such that there is sufficient overlap between the donor emission and acceptor excitation spectra, and spectra for an example pair (terbium (Tb) donor and Alexa<sub>488</sub> acceptor) are displayed in **Figure 4.1b**.

In biological assays, this can be leveraged via fluorescent labelling of two molecules (be it proteins, nucleic acids, or ligands) with a suitable donor and acceptor. The interactions of the molecules can then be monitored through the simultaneous measurement of both donor and acceptor fluorescence intensity following donor excitation. A low ratio of acceptor/donor fluorescence signifies little-to-no energy transfer, while a high acceptor/donor fluorescence ratio is indicative of robust FRET and therefore complex formation.



**Figure 4.1.** **a**, Proximity-based energy transfer in TR-FRET. Image created with BioRender.com **b**, Emission and excitation spectra of terbium chelate donor and its matched acceptor fluorophore Alexa<sub>488</sub>. The arrows indicate the

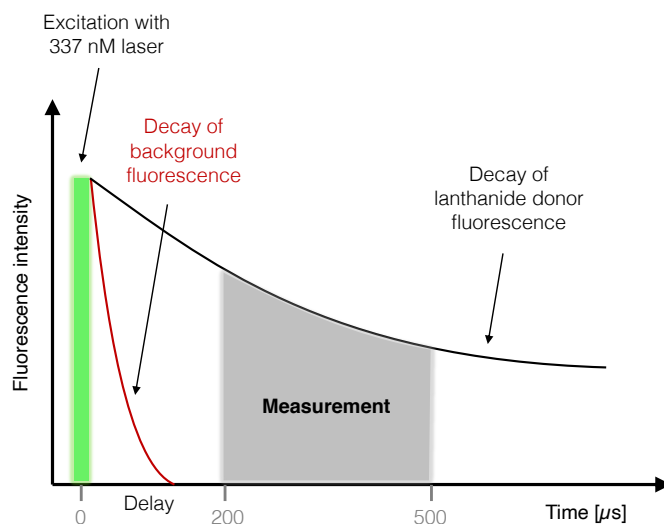
wavelength of laser excitation (337 nm) and of emission measurements (490 and 520 nm). Spectral data were obtained from Thermo Fischer.

$$k_{ET} = \frac{1}{\tau_D^0} \times \left(\frac{R_0}{R}\right)^6$$

**Equation 4.1.** The energy transfer rate ( $k_{ET}$ ), as described by Theodor Förster<sup>315</sup>, is inversely proportional to the donor lifetime ( $\tau_D^0$ ) and the sixth power of donor-acceptor distance ( $R$ ). It is also proportional to the sixth power of the Förster radius ( $R_0$ ), which is the distance between two fluorophores at which the FRET efficiency is 50% (usually in the range of 20-90 Å). The Förster radius is a specific property of each FRET pair and contains information on the dipole orientation and the spectral overlap between the donor and acceptor.

Time-resolved FRET (TR-FRET) combines time-resolved fluorimetry with FRET to afford an optimised proximity assay. It utilises long-lived lanthanide fluorophores as donors and employs temporal gating of fluorescence intensity measurements to allow for more sensitive detection of association and dissociation events (**Figure 4.2**). Chelates or cryptates of lanthanides such as terbium (Tb) or europium (Eu) are used as they have millisecond-long excited state lifetimes<sup>316</sup>. Hence, the fluorescent intensity measurement can be delayed by *ca.* 70-200  $\mu$ s to allow for the background fluorescence to decay and interference from scattered light to be minimised, leading to improved sensitivity. Additionally, lanthanides have narrow emission peaks and large Stokes shifts, which further limits the crosstalk between excitation and emission wavelengths and thereby improves the signal-to-noise ratio<sup>316</sup>.

In practical terms, the assay is performed by titrating the acceptor-labelled component into the donor-conjugated biomolecule in a microplate, followed by donor excitation and simultaneous donor and acceptor emission measurements using a plate reader equipped with suitable filters. TR-FRET compares favourably to other methods of protein-protein or protein-ligand interaction measurements. Firstly, the measurement is ratiometric in nature as both donor and acceptor signal are recorded throughout. This reduces assay variability and stands in contrast to methods such as Alphascreen, whereby only acceptor signal is recorded<sup>317</sup>. Additionally, unlike biolayer interferometry (BLI) or surface plasmon resonance (SPR), the method does not require the immobilisation of any components on chips or beads, leading to more entropic freedom. However, the limitations of TR-FRET include assay interference caused by compounds with fluorescent or quenching properties<sup>318</sup> as well as that caused by the fluorescent labels themselves, often conjugated via a relatively bulky protein tag (e.g. Streptavidin or SpyCatcher) that could affect the studied interaction. The latter can be curtailed through incorporation of flexible linkers or repositioning of the tags in the hope of minimising any steric hindrance. Moreover, such interference can be circumvented by employing label-free methods such as ITC, yet such methods utilise several orders of magnitude more protein sample than TR-FRET per titration.



**Figure 4.2.** Time-gated measurement of fluorescence intensity in TR-FRET. The image is partly based on <sup>319</sup>.

### 4.1.2 Original assay setup

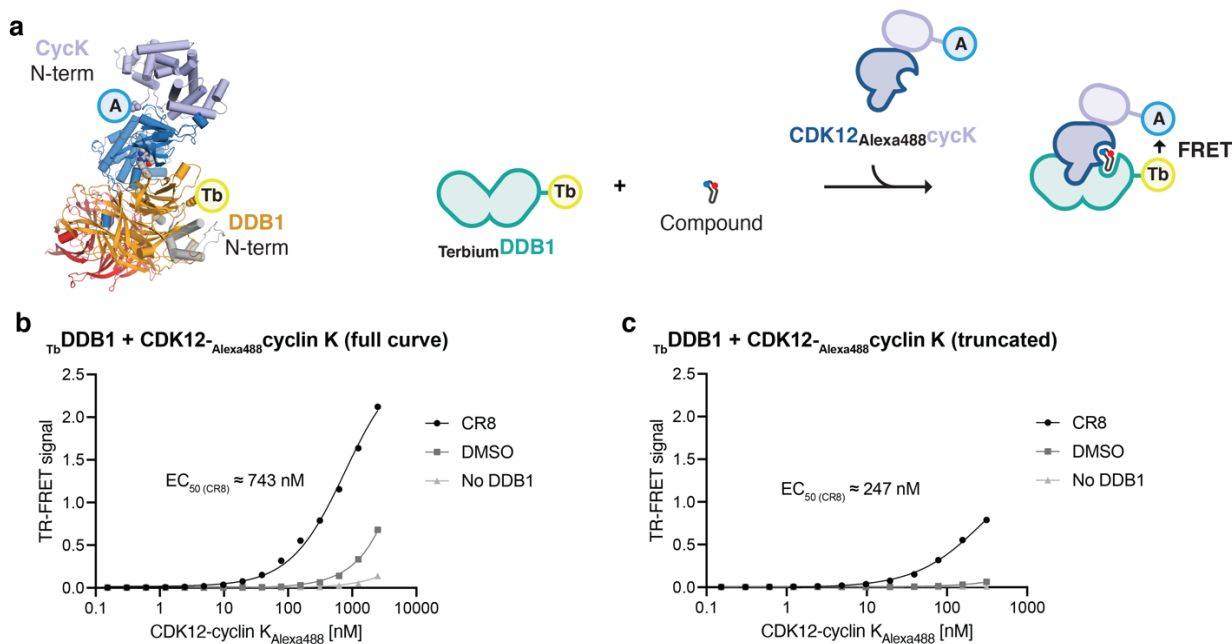
Due to its many advantages, TR-FRET has become a mainstay tool for *in vitro* evaluation of ternary complex formation in targeted protein degradation<sup>320–322</sup>. In the published studies on cyclin K degraders (**Chapter 2**), we utilised a TR-FRET assay to monitor the interaction between CDK12-cyclin K and DDB1 in the presence of compounds. Our later work on CR8 derivatives (**Chapter 3**), however, necessitated a highly optimised assay to be developed for the robust evaluation of closely related molecular glue degraders. The following subchapter describes the assay optimization process. While the details pertain specifically to the DDB1-CR8-CDK12-cyclin K complex, many learnings are expected to be more broadly applicable to TR-FRET assay optimization for diverse interactors.

In this project, the TR-FRET assay is employed to quantify the binding constant (ternary complex formation affinity) for the compound-mediated interaction between DDB1 and CDK12-cyclin K. However, one important caveat in this system is the measurable affinity between DDB1 and CDK12-cyclin K in the absence of a molecular glue compound (50  $\mu\text{M}$  as measured by ITC)<sup>294</sup>, a property that has been recently proposed to be a defining feature of all molecular glue degraders and thereby more widely applicable<sup>323</sup>. A second caveat, also common to most molecular glue degraders, is the fact that the glue compound displays affinity for one protein partner (in this case low affinity for CDK12; 5  $\mu\text{M}$  as measured by ITC)<sup>294</sup> and not the other (DDB1), and hence the formation of a DDB1-CR8-CDK12-cyclin K complex follows a defined order of binding. Both properties of the studied system have implications for the observed background signal, chosen titration mode, component ratios, and their order of addition.

Our original assay setup<sup>285,294</sup> produced data sufficient for establishing a compound-mediated interaction, evaluating the impact of protein mutations on affinity, as well as ranking different complex-inducing compounds. Yet, the binding constant reported in our original study ( $\text{EC}_{50} \approx 600 \text{ nM}$ ) (**Figure 2.8**) was inconsistent with the dissociation constant determined through ITC ( $K_{\text{app}} \approx 50 \text{ nM}$ ), a label-free affinity

measurement whereby DDB1 (100  $\mu\text{M}$ ) was titrated into a pre-formed complex of CR8 (30  $\mu\text{M}$ ) and CDK12-cyclin K (10  $\mu\text{M}$ ) (**Figure 2.10**)<sup>294</sup>. The original assay employed a common TR-FRET pair, the donor Tb and acceptor Alexa<sub>488</sub>, conjugated to the proteins of interest via biotin (Streptavidin-Tb) or using the SpyTag/SpyCatcher system (Alexa<sub>488</sub>) (fluorophore attachment is schematically depicted in **Figure 4.1a**). We initially sought to test multiple tag orientations, positioning the donor on the N-terminus of DDB1 and the acceptor on the N-terminus of cyclin K, and *vice versa*. However, we found that due to limited solubility, the CDK12-cyclin K complex was not amenable to enzymatic biotinylation required for the coupling of Streptavidin-Tb to the protein. Soluble CDK12-cyclin K complex could however be purified with a Spy tag and successfully coupled to an Alexa<sub>488</sub>-labelled SpyCatcher protein. DDB1 was therefore biotinylated on its N-terminus for conjugation of Streptavidin-Tb. Hence, the initial choice of tag positions (<sub>Tb</sub>DDB1 and CDK12-Alexa<sub>488</sub>cyclinK, illustrated in **Figure 4.3a**) was dictated predominantly by protein stability.

In a TR-FRET assay, the acceptor-bearing component is titrated into the donor-labelled molecule. Given the available constructs (<sub>donor</sub>DDB1 and CDK12-<sub>acceptor</sub>cyclinK), CDK12-Alexa<sub>488</sub>cyclin K (0-2.5  $\mu\text{M}$ ) had to be titrated into <sub>biotin</sub>DDB1 (50 nM) pre-mixed with a compound (10  $\mu\text{M}$ ; we utilised CR8 throughout the optimisation process), which produced TR-FRET signal, albeit one riddled with several issues (**Figure 4.3**). Most notably, as CR8 does not bind to DDB1 ( $K_{\text{app}} > 200 \mu\text{M}$  by ITC)<sup>294</sup> and binds to CDK12 ( $K_{\text{app}} \approx 5 \mu\text{M}$  by ITC)<sup>294</sup>, the relatively low affinity binary ligand-CDK12 binding event obstructs the observation of high-affinity ternary complex formation, leading to an overestimation of the ternary complex affinity ( $\text{EC}_{50}\text{s}$  ranging from 500 to 750 nM). At the highest CDK12 concentration of 2.5  $\mu\text{M}$ , the excess of CR8 is only four-fold, meaning that saturation conditions (all CDK12 is CR8-bound) are not met. Significantly higher compound concentrations could not be achieved due to its limited solubility in the assay buffer and poor assay tolerance of high DMSO concentrations. Additionally, when CDK12-Alexa<sub>488</sub>cyclin K titrations into <sub>biotin</sub>DDB1 were performed with the DMSO vehicle control, compound-independent complex formation was observed at high CDK12-cyclin K concentrations. The measured signal was significantly higher than spurious signal arising from fluorophore effects (the “no DDB1” curve in **Figure 4.3b** corresponds to a titration of CDK12-Alexa<sub>488</sub>cyclin K titration into Streptavidin-Tb alone) and could be markedly lowered by blocking the kinase active site with a covalent CDK12 inhibitor (**Figure 2.15a**). While weak compound-independent CDK12-DDB1 interaction was also apparent in ITC experiments ( $K_{\text{app}} \approx 50 \mu\text{M}$ )<sup>294</sup>, this basal binding led to difficulties in accurate fitting of the TR-FRET data due to biphasic behaviour. Restricting the accessible concentration range in order to dial out compound-independent effects yielded binding curves that do not reach saturation (**Figure 4.3c**).

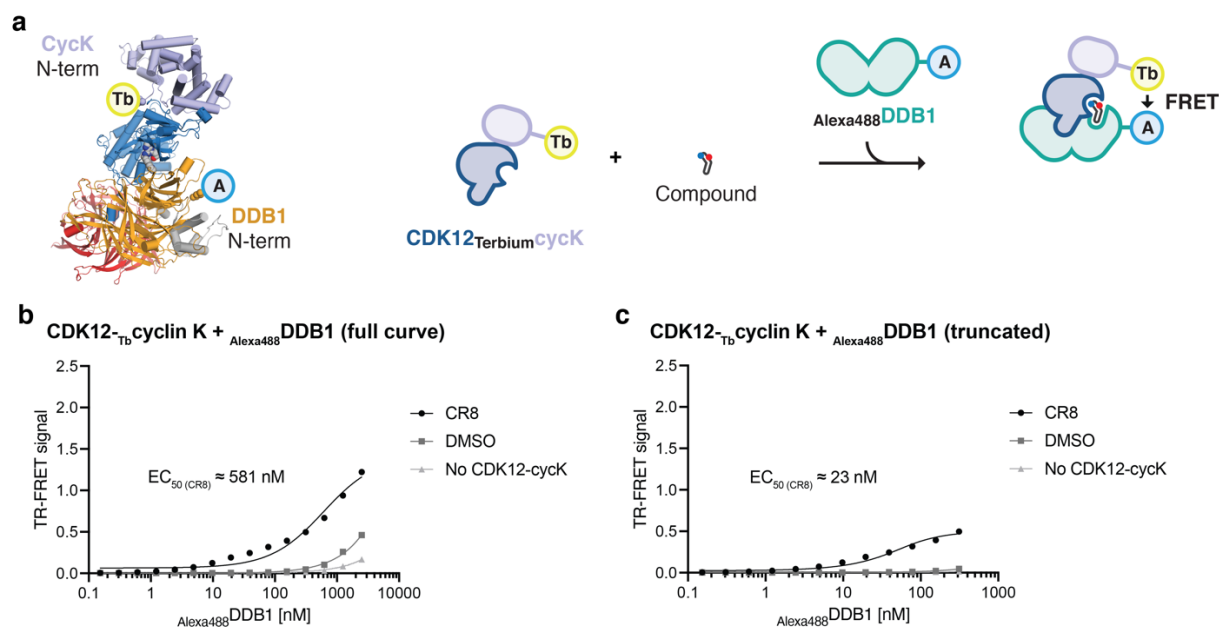


**Figure 4.3. Initial TR-FRET labelling setup.** **a**, Tag positions annotated on the ternary complex crystal structure (left). Schematic of a CDK12-cyclin K titration into DDB1 premixed with CR8 (termed the *forward titration* in subsequent sections) (right). **b**, Titration curve showing clear CR8-mediated binding but also compound-independent binding at high concentrations. **c**, Same titration as in (b) but with the concentration range restricted to curb compound-independent effects. However, the binding curve clearly does not reach saturation.

### 4.1.3 Tag position

To circumvent these limitations, we again explored additional tag positions. For this, the Avi tag for enzymatic biotinylation (to attach the Streptavidin-coupled Tb donor) was placed on either terminus of cyclin K, while the Spy tag (for Spycatcher-Alexa<sub>488</sub> coupling) was placed on the N-terminus of DDB1. While earlier attempts following this strategy were unsuccessful due to CDK12 precipitation upon overnight biotinylation, successful preparation of soluble biotinylated CDK12-cyclin K was achieved through optimization of the purification and labelling protocol (see **Methods**). Beyond resolving any potential steric hindrance and the hope to increase the dynamic range of the assay, this setup was thought to allow the resolution of the key limitation, which was the obscuration of the ternary binding event by the low-affinity binary binding of the compound to CDK12-cyclin K.

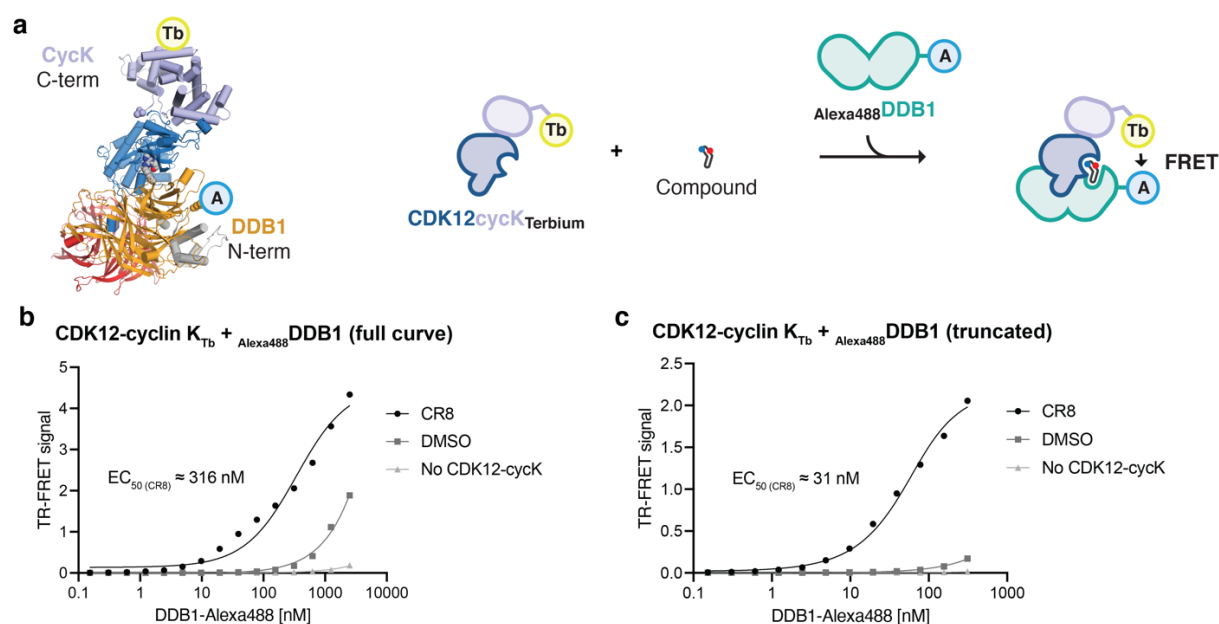
First, we successfully purified and labelled the CDK12-cyclin K complex with an Avi tag on the N-terminus of the cyclin (CDK12-T<sub>b</sub>cyclin K). Thus, we could now titrate DDB1<sub>Alexa488</sub> (0 - 2.5 μM highest concentration) into a pre-formed complex of CDK12-T<sub>b</sub>cyclin K (50 nM) with a saturating excess of CR8 (10 μM) (**Figure 4.4**). This strategy yielded a ternary complex binding constant ( $EC_{50} \approx 23$  nM) consistent with our label-free ITC measurements for this complex (**Figure 4.4c**). However, the biphasic behaviour persisted (**Figure 4.4b**) and the small assay window (low maximal TR-FRET signal intensity) suggested a suboptimal relative orientation of the two fluorophores.



**Figure 4.4. TR-FRET labelling setup with N-terminal biotin on cyclin K.** **a**, Tag positions annotated on the ternary complex crystal structure (left). Schematic of a DDB1 titration (0-2.5  $\mu$ M) into CDK12-cyclin K (50 nM) premixed with a saturating excess of CR8 (10  $\mu$ M) (termed the *forward titration* in subsequent sections) (right). **b**, Titration curve showing clear CR8-mediated binding but also compound-independent binding at high concentrations. **c**, Same titration as in **(b)** but with the concentration range restricted to curb compound-independent effects. However, biphasic behaviour persists, and the assay window is very low.

Hence, the Avi tag for Streptavidin-Tb conjugation was moved from the N-terminus to the C-terminus of cyclin K and a longer, flexible linker was introduced to prevent any conformational restriction (**Figure 4.5a**). This yielded a more robust assay window and a comparable  $EC_{50}$  value ( $EC_{50} \approx 31$  nM). While strong, compound-independent effects manifesting in biphasic curves were still observed for upon DDB1 titration (**Figure 4.5b**), restricting the concentration range yielded interpretable binding curves that appear to reach saturation (**Figure 4.5c**). Therefore, this tag combination (CDK12-cyclin K<sub>Terbium</sub> + Alexa488 DDB1) was ultimately chosen for future experiments. The comparison between the three labelling schemes tested is summarised in **Table 4.1**.





**Figure 4.5. Optimised TR-FRET labelling setup with C-terminal biotin on cyclin K.** **a**, Tag positions annotated on the ternary complex crystal structure (left). Schematic of a DDB1 titration (0-2.5  $\mu$ M) into CDK12-cyclin K (50 nM) premixed with a saturating excess of CR8 (10  $\mu$ M) (termed the *forward titration* in subsequent sections) (right). **b**, Titration curve showing clear CR8-mediated binding but also compound-independent binding at high concentrations. The assay window is strongly improved. **c**, Same titration as in (b) but with the concentration range restricted to curb compound-independent effects. The curve reaches saturation, and the dynamic range remains high.

**Table 4.1.** Side-by-side comparison of TR-FRET assay labelling schemes. Each panel consists of label position annotation on the structure (6TD3) and a schematic depiction of the performed titration. **a**, CDK12-Alexa488cyclin K + TerbiumDDB1. **b**, CDK12-TerbiumCyclin K + Alexa488DDB1. **c**, CDK12-cyclin K<sub>Terbium</sub> + Alexa488DDB1.

	CDK12-Alexa488cyclin K + TerbiumDDB1	CDK12-Terbiumcyclin K + Alexa488DDB1	CDK12-cyclin K <sub>Terbium</sub> + Alexa488DDB1
Label positions			
CR8 EC <sub>50</sub>	247 nM*	23 nM	31 nM
Advantages	Robust purification and labelling protocol No CDK12 precipitation upon labelling	Binding constant consistent with ITC value Greater excess of compound over kinase Saturation of binding curves reached	Binding constant consistent with ITC value Greater excess of compound over kinase Saturation of binding curves reached High TR-FRET signal intensity

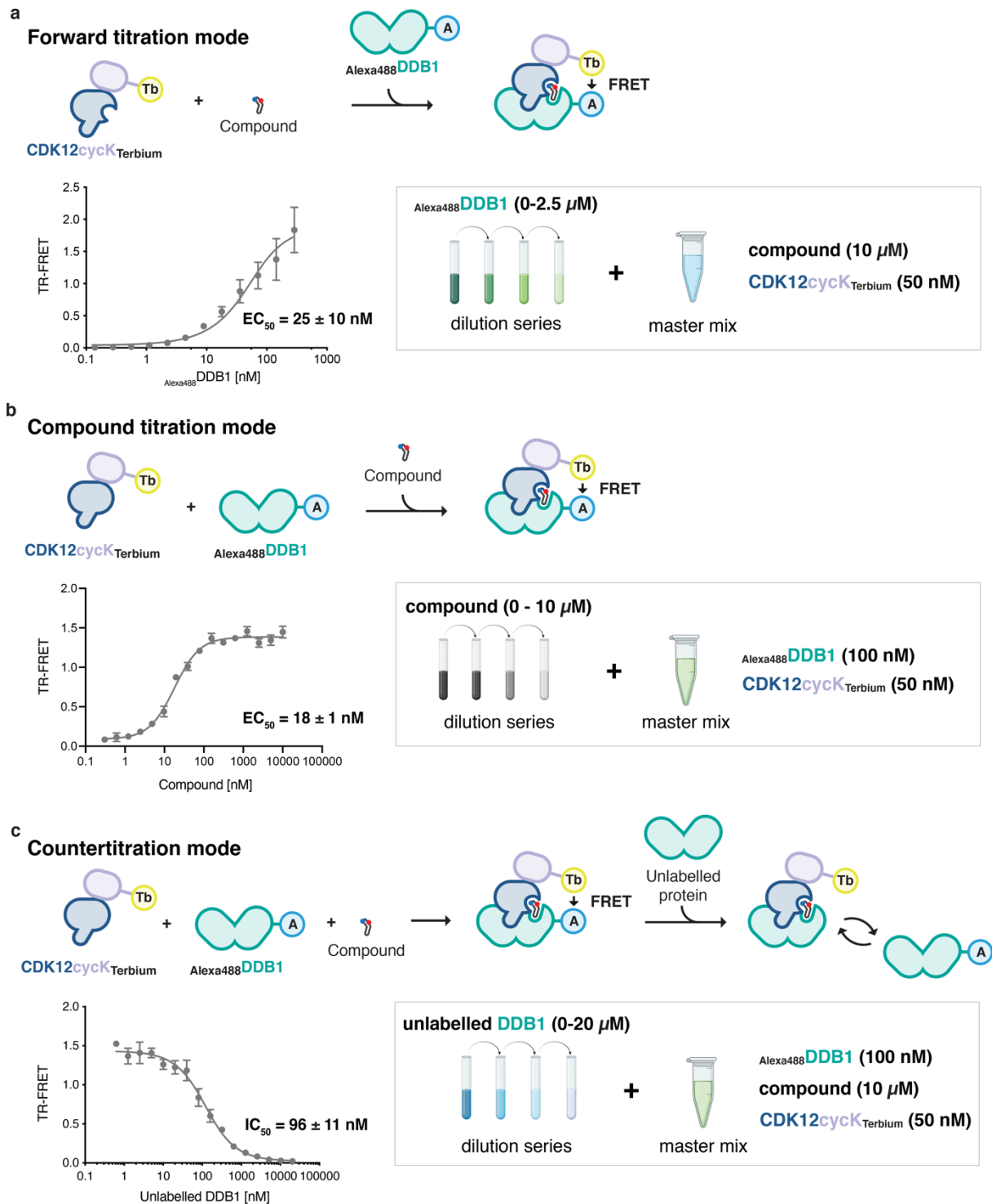
<b>Disadvantages</b>	Binding constant inconsistent with ITC value	Risk of CDK12 precipitation upon biotinylation	Risk of CDK12 precipitation upon biotinylation
	Compound-kinase interaction not saturated	Low TR-FRET signal intensity	Biphasic behaviour
	Biphasic binding behaviour	Biphasic behaviour	
	No curve saturation		
	Binding curves hard to interpret		

\*no saturation reached

#### 4.1.4 Titration mode

Subsequently, we explored additional titration modes, both as internal controls of assay quality and to further optimise the setup for compound evaluation. The three titration modes evaluated include *forward titration* (titration of DDB1 into CDK12-cyclin K premixed with an excess of CR8), *compound titration* (titration of CR8 into a pre-mixed CDK12-cyclin K and DDB1; of note this titration mode yielded only spurious signal when employed using the initial labelling strategy), and *countertitration* (titration of an unlabelled DDB1 into a pre-formed ternary complex), and are illustrated in **Figure 4.6**. The resulting binding affinities ( $EC_{50}(\text{forward}) = 25 \pm 10 \text{ nM}$ ,  $EC_{50}(\text{compound}) = 18 \pm 1 \text{ nM}$ ,  $IC_{50}(\text{counter}) = 96 \pm 10 \text{ nM}$ ) were of the same order of magnitude, confirming the reliability of the assay.

*Forward titration* and *compound titration* modes were considered for the evaluation of relative compound performance. Comparing the two assay modes, the forward titration appeared to be more prone to large errors towards higher titrant concentration and compound-independent biphasic behaviour, yielding less reliable fits. The compound titration mode showed more robust, saturating binding curves without any need for data exclusion and was therefore chosen for further compound characterisation. Additionally, the compound titration mode was favoured as it required much lower protein amounts.

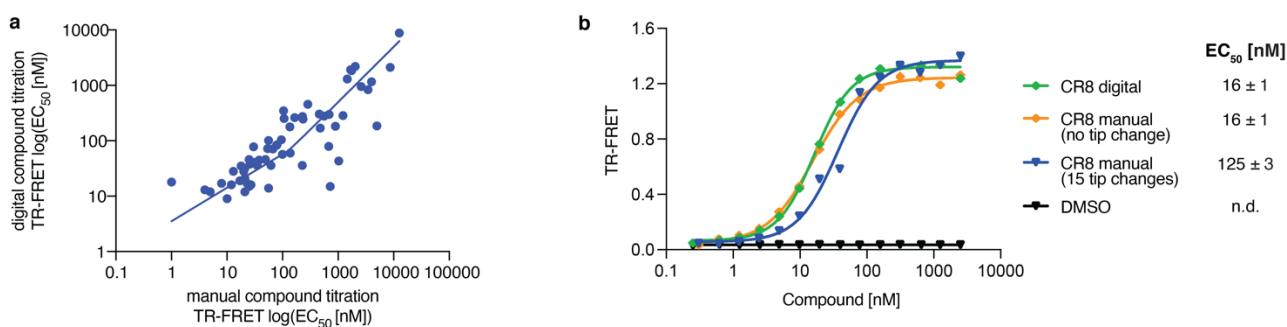


**Figure 4.6. Illustration of three different titration modes for molecular glue-induced ternary complex formation evaluation by TR-FRET.** **a**, Forward titration, whereby DDB1<sub>Alexa488</sub> (0-2.5 μM) is titrated into a pre-mixed solution of CDK12-cyclin K<sub>Terbium</sub> (50 nM) and CR8 (10 μM). **b**, Compound titration, where a dilution series of CR8 (0-10 μM) is added into a pre-mixed solution of DDB1<sub>Alexa488</sub> (100 nM) and CDK12-cyclin K<sub>Terbium</sub> (50 nM). **c**, Counter titration, where unlabelled DDB1 (0-20 μM) is titrated into a pre-mixed solution of DDB1<sub>Alexa488</sub> (100 nM), CDK12-cyclin K<sub>Terbium</sub> (50 nM) and CR8 (10 μM). (a-c) Data represent the mean ± s.d. (n ≥ 3). Titration schematics were created with BioRender.com.

### 4.1.5 Compound dispensing methods

Another important aspect of the assay optimization was the choice of a method for compound dilution and dispensing. All initial experiments were performed manually, with serial dilutions of each compound in assay buffer prepared from DMSO stocks using a Gilson or Eppendorf pipette in strips of polypropylene tubes (UltraFlux SNaStrip PCR tubes, SSI cat. no. 3240-00) and then transferred onto the assay plate using an E1-ClipTip electronic multichannel pipette (Thermo Fischer).

Alternatives to manual compound handling include acoustic and digital dispensing. At the late stage of the project, we acquired a D300e digital dispenser (Tecan) and could therefore compare manual and digital compound dispensing across the entire cyclin K degrader dataset (**Figure 4.7a**) While no obvious difference was noted for CR8 (**Figure 4.7b**) and the majority of investigated derivatives (overall linear correlation with R squared = 0.74, depicted in log scale in **Figure 4.7a**), several small molecules appeared to stochastically yield false binding curves when pipetted manually, most often in the form of horizontal lines corresponding to saturation at what were believed to be picomolar compound concentration (**Figure 4.8a**), precluding the confident assessment of their binding constants. Following extensive troubleshooting, the issue could be attributed to certain compounds adhering to plastic pipette tips when handled as aqueous solutions (but not as DMSO-only stocks) (**Figure 4.8**). Finding the root cause of this issue was impeded by the frequent changes of pipette tip manufacturers due to the recent worldwide shortage of such consumables, as it is possible that the interference is especially pronounced for tips of a particular brand. Notably, as tips sometimes become loose or spontaneously dislodge from the pipette, this would lead to the tip being changed, which would in turn give rise to interpretable data. Such a “deviation” from the protocol is not something that was noted under normal circumstances, adding to the stochastic nature of the interference and hindering the troubleshooting process.

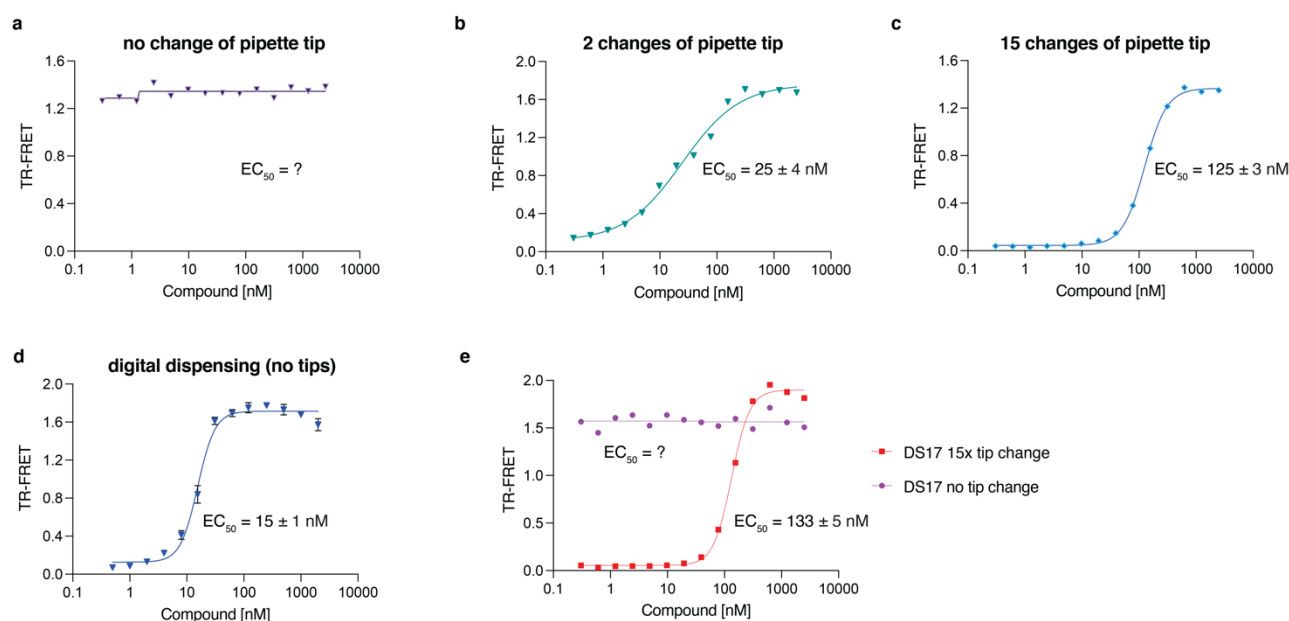


**Figure 4.7. a**, Correlation between the TR-FRET EC<sub>50</sub> values obtained through digital compound titrations (y axis) or manual titrations (x axis). The data were fit with a linear regression in linear scale (R squared = 0.74) but as the values span several orders of magnitude, both axes were subsequently log<sub>10</sub>-scaled for more informative illustration of the fitting. Hence, the fitting curve on the above graph appears non-linear. **b**, Digital or manual handling, and more specifically the number of tip changes upon preparation of the 16-point dilution series for the compound titration, do not affect the EC<sub>50</sub> value for CR8.

Problems that could result from tip-based compound handling are acknowledged in the high-throughput screening community but rarely formally reported on, with one important exception<sup>324</sup>. However, compound “stickiness” can greatly influence all their downstream evaluation. Such confounding behaviour is not

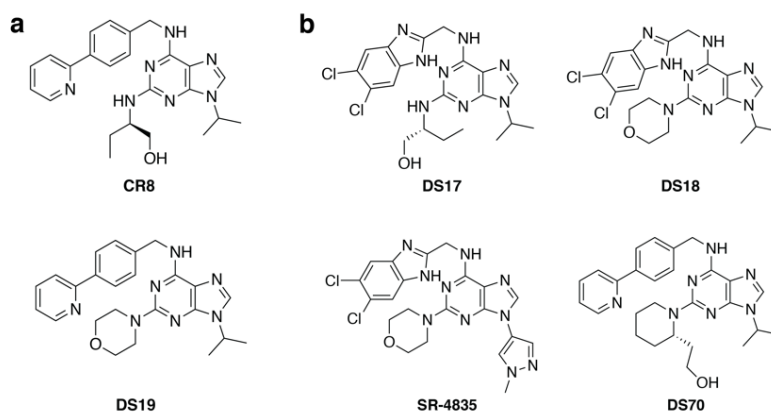
straightforward to predict or identify as it can give rise to both false negatives (*e.g.* if compounds firmly adhere to tubes and hence are depleted from the assay solution; such a case is detailed in <sup>324</sup>) and positives (if compounds (or their aggregates) weakly stick to plastic tips and hence are transferred at a much higher concentrations than expected in serial dilutions, see below). The referenced study reviews several direct comparisons for tip-based versus acoustic dispensing, and our data provide another dataset (**Figure 4.7a**).

**Figure 4.8** depicts titrations of compound DS17 (**Figure 4.9a**) into Alexa488DDB1 and CDK12-cyclinK<sub>Terbium</sub> obtained through different pipetting and dispensing schemes. A 16-point dilution series of the compound in buffer was either manually pipetted using an Eppendorf pipette with the same, single tip (**Figure 4.8a**), changing tips after the 2<sup>nd</sup> and 8<sup>th</sup> tube only (**Figure 4.8b**), changing tips after every tube (*i.e.* 15 times; **Figure 4.8c**), or pipetted with no tips, using the digital dispenser robot (**Figure 4.8d**), with all experiments performed using the same compound stock aliquot. The effect of such tip changes was corroborated in an independent experiment using a different compound and protein stock solutions, pipette, and tips (**Figure 4.8e**).



**Figure 4.8. Impact on the dispensing method on compound titrations.** **a**, TR-FRET curve resulting from a manual 16-point compound dilution in the assay buffer performed with an Eppendorf pipette in two strips of PCR tubes without tip changes throughout. **b**, As in (**a**) but with a tip change after the first and eighth tube, respectively. **c**, As in (**b**) but with tip changes after every tube. **d**, Digital dispensing of the compound, which utilises no tips. The stock solution in DMSO is directly pipetted onto the chip of the D300e robot. Data represent  $\pm$ s.d from two technical replicates. **e**, An additional replicate of a DS17 titration with the compound serial dilution performed with no tip change (purple) versus with tip changes after each tube (red) by another experimenter (Master student Vivian Focht).

Interestingly, the above-described discrepancy only occurred for some derivatives and not for other, closely related compounds (**Figure 4.9**). While no common structural determinant of the observed behaviour could be identified, we note for instance that the majority of compounds with the dichlorobenzimidazole moiety (present in DS17, 18, and SR-4835) display this interference pattern.



**Figure 4.9.** **a**, Chemical structures of compounds with no noted issues upon manual dilution. **b**, Examples of compounds closely related to those in **(a)** yet giving rise to spurious titration results upon manual handling.

In all the following experiments, the compounds were dispensed digitally using the D300e robot (Tecan).

#### 4.1.6 FRET pair

To explore whether the dynamic range of the assay could be further improved by employing a different TR-FRET pair, we tested multiple different donors (Tb and Eu) and acceptors (Alexa<sub>488</sub>, Alexa<sub>647</sub>, and Cy5) (**Table 4.2**). For this, individual batches of purified CDK12-cyclin K were biotinylated and incubated with Streptavidin-Tb or Streptavidin-Eu. The S50C mutant of the SpyCatcher protein was purified and labelled with Alexa<sub>488</sub>, Alexa<sub>647</sub> or Cy5 and conjugated to separate batches of <sub>spy</sub>DDB1. The exploration of multiple TR-FRET donors was motivated by their distinct properties. The proprietary lanthanide chelates and cryptates (Thermo Fischer) differ in their tolerance to chemical and buffer conditions and have distinctive spectral properties. Additionally, terbium has been reported to have a higher quantum yield and molar extinction coefficient than europium, leading to potentially improved screening performance<sup>325</sup>. Different TR-FRET pairs are also characterised by a varying separation between donor and acceptor excitation and emission peaks and could therefore offer a distinct signal-to-noise ratio and dynamic range.

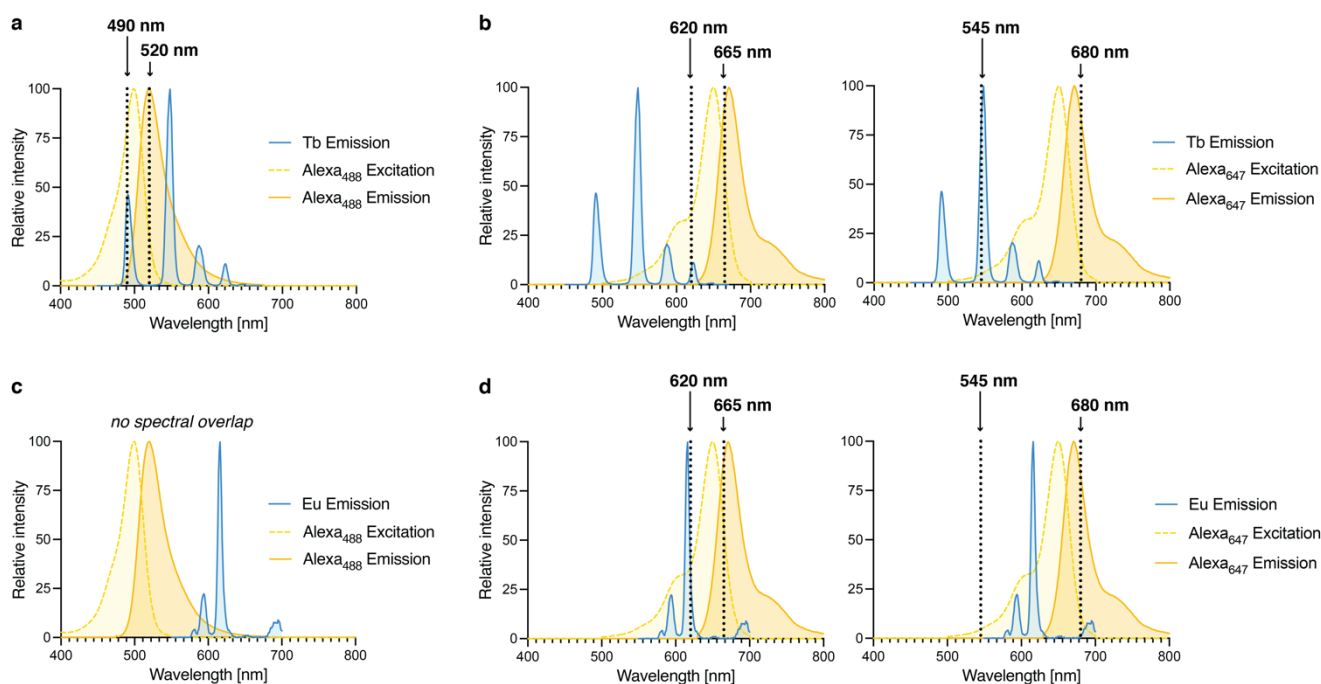
While complex formation with all possible combinations of the fluorescently-labelled CDK12-cyclin K and DDB1 was evaluated for completeness (**Figure 4.11**), only some pairs exhibited favourable spectral overlap and hence yielded strong signal upon CR8 titration (**Table 4.2**, **Figure 4.11**, **Figure 4.10**). The assay window varied strongly but, importantly, the CR8 EC<sub>50</sub> value was consistent across all the tested sets (11 nM > EC<sub>50</sub> > 19 nM).

Another variable to consider is the choice of wavelengths at which donor and acceptor emissions are read out. For the pairing of Tb with Alexa<sub>647</sub> or Cy5, two different filter sets were employed. Conventionally, such TR-FRET measurements are performed using a filter set with excitation at 337 nm and emissions at 665 nm (Alexa<sub>647</sub>) and 620 nm (Tb) (**Figure 4.10b**). However, a different, more pronounced and more spectrally

distant peak of Tb emission (at 545 nM) can be sampled instead using a 337-545-680 filter cube (kindly provided by Uli Hassiepen, Novartis).

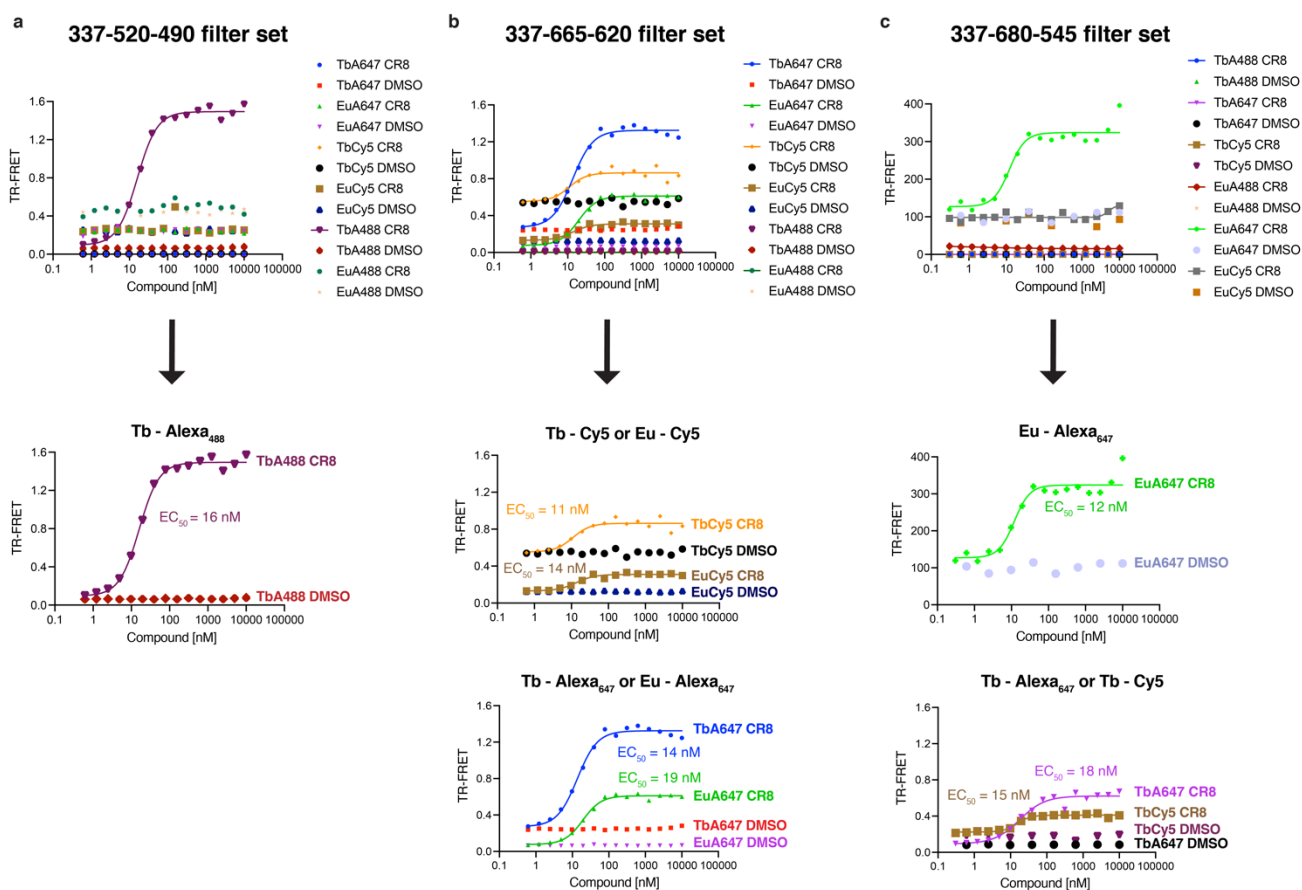
**Table 4.2.** Effective combinations of TR-FRET donors, acceptors, and filter sets for FRET signal measurement. The excitation and emission peaks expected to be read out are underlined. Asterisk marks a combination where the filter set should in principle not yield any signal, yet meaningful titrations are observed.

Donor	Donor excitation → emission wavelength [nm]	Acceptor	Acceptor excitation → emission wavelength [nm]	Filter wavelengths [nm] for TR- FRET measurement	CR8 EC <sub>50</sub> [nM]	CR8 FRET <sub>max</sub>	DMSO FRET <sub>max</sub>
Tb	<u>337</u> → <u>490</u> , 550, 590, 625	Alexa488	501 → <u>522</u>	337-520-490	16	1.57	0.08
Tb	<u>337</u> → 490, 550, 590, <u>625</u>	Alexa647	614, 653 → <u>673</u>	337-665-620	14	1.38	0.28
	<u>337</u> → 490, <u>550</u> , 590, 625			337-680-545	18	0.67	0.08
Tb	<u>337</u> → 490, 550, 590, <u>625</u>	Cy5	609, 652 → <u>672</u>	337-665-620	11	0.94	0.59
	<u>337</u> → 490, <u>550</u> , 590, 625			337-680-545	15	0.43	0.18
Eu	<u>337</u> → 595, <u>615</u>	Alexa647	614, 653 → <u>673</u>	337-665-620	19	0.60	0.06
	<u>337</u> → 595, 615			337-680-545*	12*	331*	114*
Eu	<u>337</u> → 595, <u>615</u>	Cy5	609, 652 → <u>672</u>	337-665-620	14	0.33	0.10



**Figure 4.10.** Emission spectra of TR-FRET donors (Tb, Eu) and excitation spectra of TR-FRET acceptors (Alexa<sub>488</sub>, Alexa<sub>647</sub>, Cy5). Vertical dotted lines correspond to the wavelengths at which fluorescence is measured during the experiment. **a**, Experiments with the Tb-Alexa<sub>488</sub> FRET pair can be performed with the 337-520-490 filter set to follow Alexa<sub>488</sub> emission (520 nm) and Tb emission (490 nm). **b**, Assays with the Tb-Alexa<sub>647</sub> FRET pair can be performed with a 337-665-620 filter cube to follow Alexa<sub>647</sub> emission (665 nm) and Tb emission (620 nm) or using a 337-680-545 filter set to follow Alexa<sub>647</sub> emission (680 nm) and a better separated peak of Tb emission (545 nm). **c**, Eu and Alexa<sub>488</sub> do not constitute a usable FRET pair due to no overlap between the Eu emission spectrum and the Alexa<sub>488</sub> excitation spectrum. **d**, As in **(b)** but for Eu-Alexa<sub>647</sub>. The rightmost panel illustrates how no signal was expected for the measurement using the 337-680-545 filter cube. **(c, d)** As Alexa<sub>647</sub> and Cy5 have closely similar spectra, only Alexa<sub>647</sub> is depicted for brevity. Spectral data for individual fluorophores were obtained from Thermo Fischer.

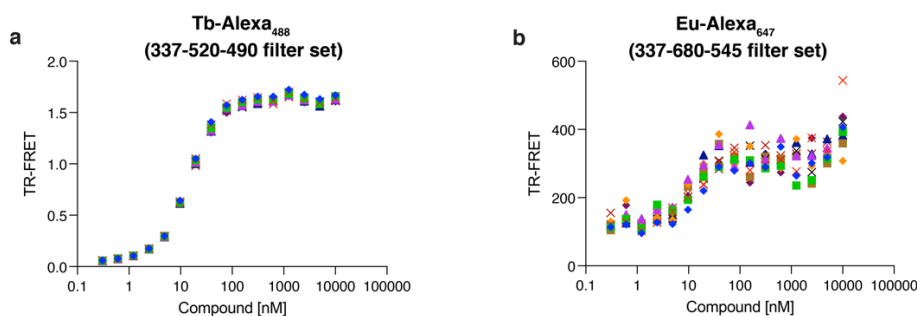




**Figure 4.11.** **a**, Titration of CR8 or DMSO vehicle control into DDB1 and CDK12-cyclin K labelled with all combinations of TR-FRET donor and acceptor, with the signal measured using the 337-520-490 filter set (top panel). Tb-Alexa488 constitute the best FRET pair among the tested combinations (bottom panel). **b**, Titration of CR8 or DMSO vehicle control into DDB1 and CDK12-cyclin K labelled with all combinations of TR-FRET donor and acceptor, with the signal measured using the 337-665-620 filter set (top panel). Several FRET pairs yielded robust signal (bottom panels). **c**, Titration of CR8 or DMSO vehicle control into DDB1 and CDK12-cyclin K labelled with all combinations of TR-FRET donor and acceptor, with the signal measured using the 337-680-545 filter set (top panel). Several FRET pairs yielded robust signal (bottom panels). (**a-c**) The shown data represent the mean signal across ten consecutive measurements of the same plates.

Notably, a setup that was not expected to produce any meaningful readings (the donor Eu and acceptor Alexa<sub>647</sub> with the 337-680-545 filter) yielded the highest FRET signal, with the maximum readout (CR8 FRET<sub>max</sub>) of 331 (*c.f.* 1.57 for the second highest-scoring setup; **Table 4.2**). As the TR-FRET signal is a ratio of signal at 680 nM to that at 545 nM, this can be likely attributed to the measurement at 545 nM yielding extremely low values, as neither Eu nor Alexa<sub>647</sub> emit in this region of the spectrum, effectively invalidating the ratiometric nature of the measurement and likely leading to added noise.

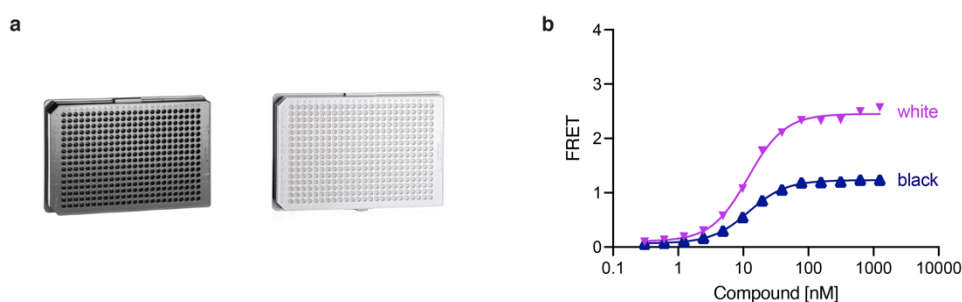
We assessed the time stability of the signal resulting from the above-described combinations, focusing the two FRET pairs giving rise the highest assay window, and we observed strong differences in signal stability (**Figure 4.12**), with the Eu - Alexa<sub>647</sub> pair measured using the 337-680-545 filter yielding large deviations across repeated measurements. Hence, we chose to proceed with the initial combination of the Tb-Alexa<sub>488</sub> FRET pair and the original 337-520-490 filter for all the following experiments.



**Figure 4.12.** Signal stability evaluated through ten consecutive measurements of the same plate every two minutes. Each colour corresponds to an individual timepoint. **a**, The Tb-Alexa<sub>488</sub> FRET pair used throughout shows reliable signal with no changes over time. **b**, The Eu-Alexa<sub>647</sub> pair data show very large variability and hence this setup was not considered for further experiments.

### 4.1.7 Plate type

Both black and white microplates can be used for FRET experiments. Our previous assays were performed solely in black 384-well microtiter plates as they are known to absorb light and reduce well-to-well crosstalk and background signal, and black plates are also utilised for other fluorescence-based techniques such as FP. However, as TR-FRET uses donors with long half-lives and therefore incorporates a delay time between donor excitation and signal measurement, any background fluorescence is expected to fade before data collection, likely making this precaution superfluous, with black plastic unnecessarily quenching the observed signal. As white plates maximally reflect the light, we tested whether using such microplates would result in higher raw signals. Indeed, pipetting the same CR8 titration into a black and a white Greiner 384-well microplate (*cat. no.* 784076 and 784075, respectively; **Figure 4.13a**) at a final volume of 8  $\mu$ L demonstrated that white plates give rise to a larger assay window (**Figure 4.13b**).

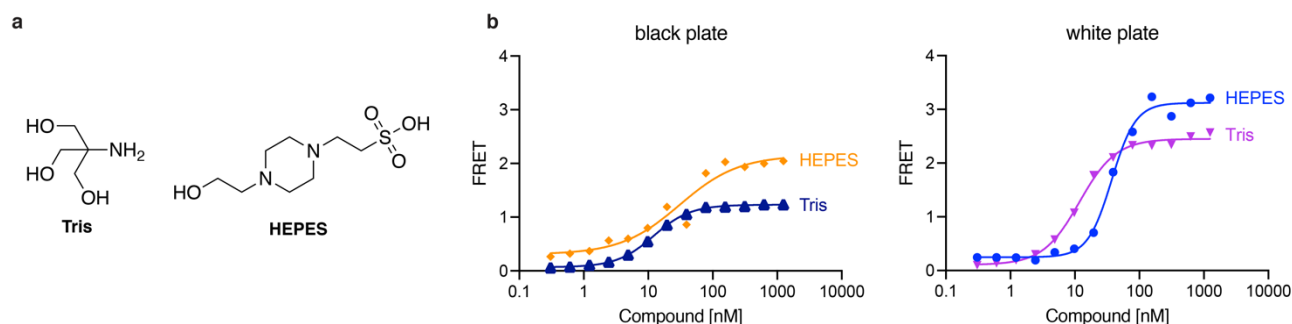


**Figure 4.13. Plate optimization.** **a**, Greiner 384-well microplates with catalogue numbers 784076 (black) and 784075 (white). **b**, TR-FRET curves showing the same CR8 titration pipetted into two different plates. Same EC<sub>50</sub> (12 vs 15 nM for black and white plate, respectively) but a larger assay window for the white plate.

### 4.1.8 Buffer components

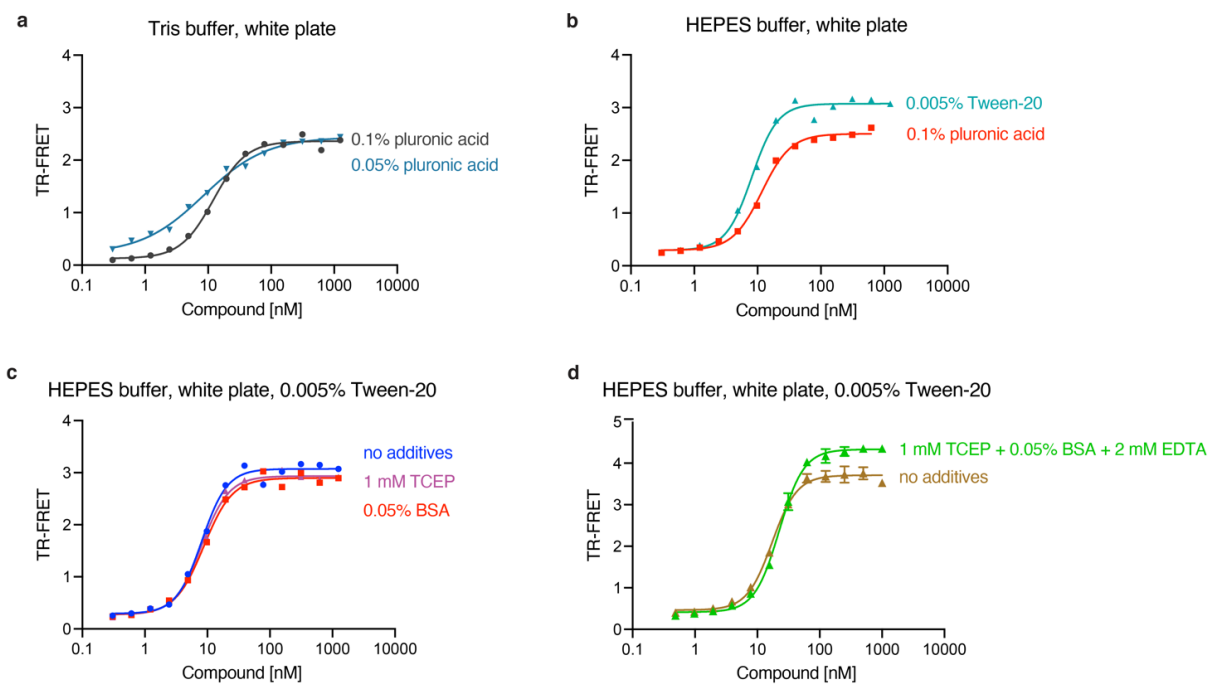
All the above-described experiments were performed in a buffer containing Tris 7.5 (50 mM), NaCl (150 mM), pluronic acid (F-68, 0.1 %), and DMSO (0.5%). During the assay optimization process, we evaluated the impact of various buffer components on the dynamic range and signal stability of the TR-FRET CR8 compound

titration. First, we compared the buffering agents HEPES (pH 7.4, 50 mM) and Tris (pH 7.5, 50 mM). As depicted in **Figure 4.14a**, Tris has a free amine moiety while HEPES is an inherently more inert, zwitterionic buffer. We observed that HEPES-based buffers give rise to a considerably larger assay window when measured both in black and in white plates (**Figure 4.14b**). We therefore decided to conduct all further experiments in a HEPES-based buffer and in white microtiter plates.



**Figure 4.14. Buffering components.** **a**, Chemical structures of Tris (free base) and HEPES. **b**, CR8 titration into Alexa488DDB1 and CDK12-cyclin K<sub>Terbium</sub> in a buffer consisting of 150 mM NaCl, 0.5% DMSO, 0.1% pluronic acid, and either 50 mM HEPES pH 7.4 or 50 mM Tris pH 7.5. The same reactions were pipetted into black (left panel) or white (right panel) 384-well microplates.

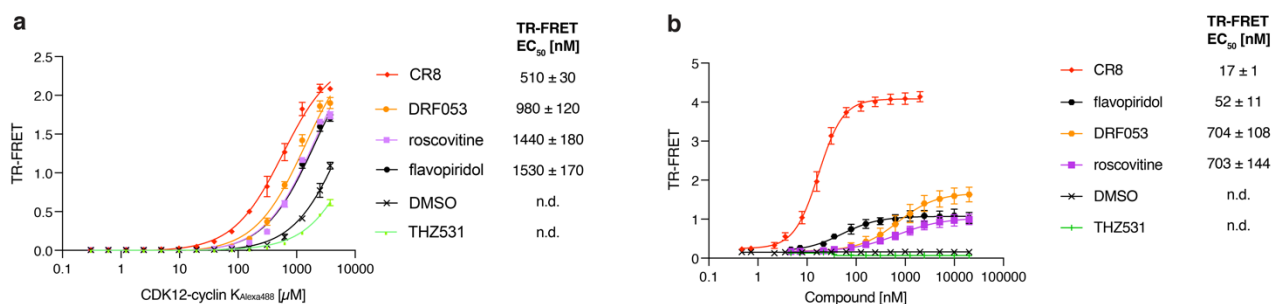
In TR-FRET, a detergent is added to improve the manual handling of low-microliter volumes, as well as to solubilise the assay components and reduce their attachment to the plastic. Our initial setup used pluronic acid (F-68, *Gibco*), a non-ionic surfactant, at 0.1% (v/v). Lowering the concentration to 0.05 % considerably affected the CR8 titration curve (**Figure 4.15a**) and its omission yielded exceedingly noisy data (*not shown*). Another non-ionic surfactant often used in HTS is Tween-20. We tested whether Tween-20 at 0.005 % (v/v; concentration chosen to be ten-fold lower than the critical micelle concentration) would perform better than pluronic acid at 0.1% (v/v) with the HEPES-based buffer and found that this substitution was favourable, giving rise to a larger dynamic range of the assay (**Figure 4.15b**).



**Figure 4.15.** **a**, CR8 titration into  $Alexa488DDB1$  and  $CDK12$ -cyclin  $K_{Terbium}$  in a buffer consisting of 50 mM Tris pH 7.5, 150 mM NaCl, 0.5 % DMSO and either 0.1 % or 0.05 % pluronic acid. **b**, CR8 titration into  $Alexa488DDB1$  and  $CDK12$ -cyclin  $K_{Terbium}$  in a buffer consisting of 50 mM HEPES pH 7.4, 150 mM NaCl, 0.5% DMSO and either 0.1 % pluronic acid or 0.005 % Tween-20. **c**, CR8 titration into  $Alexa488DDB1$  and  $CDK12$ -cyclin  $K_{Terbium}$  in a buffer consisting of 50 mM HEPES pH 7.4, 150 mM NaCl, 0.5 % DMSO and 0.005 % Tween-20, and either no additional additives or 1 mM TCEP or 0.05 % BSA. **d**, CR8 titration into  $Alexa488DDB1$  and  $CDK12$ -cyclin  $K_{Terbium}$  in a buffer consisting of 50 mM HEPES pH 7.4, 150 mM NaCl, 0.5 % DMSO and 0.005 % Tween-20, and either no additional additives or 2 mM EDTA, 1 mM TCEP, and 0.05 % BSA. (**a-d**) All assays were performed in white 384-well microplates.

In a final buffer optimization step, multiple additives were evaluated. The addition of reducing agent (TCEP at 1 mM) or BSA (at 0.05% (w/v), protease and essential globulin-free, Sigma Aldrich *cat. no. A-7030*) did not affect the assay window or the CR8 binding curve (**Figure 4.15c**). However, both additives were incorporated as they could impact the long-term stability of the assay components in case prolonged incubation times are required during experiments. A chelator (EDTA at 2 mM) was also added to scavenge any free metal ions that could persist following compound synthesis or protein purification. **Figure 4.15d** shows that the simultaneous incorporation of these three additives had a positive effect on the assay window.

The overall difference in the assay dynamic range and data quality resulting from the initial and optimised TR-FRET experimental design is illustrated with a ternary complex DDB1-CDK12-cyclin K assay with several inhibitors (**Figure 4.16**). The initial and final experimental conditions are additionally summarised in **Table 4.3**.



**Figure 4.16.** Improvement in the TR-FRET assay results. **a**, Original forward titrations showing ternary complex formation with several published CDK inhibitors as featured in <sup>294</sup>. **b**, Optimised compound titrations with the same compound set. The assay can better differentiate between the compounds' ability to promote complex formation. (**a**, **b**) Data represent mean±s.d. from three technical replicates.

**Table 4.3.** Comparison of the initial and final, optimised TR-FRET setup.

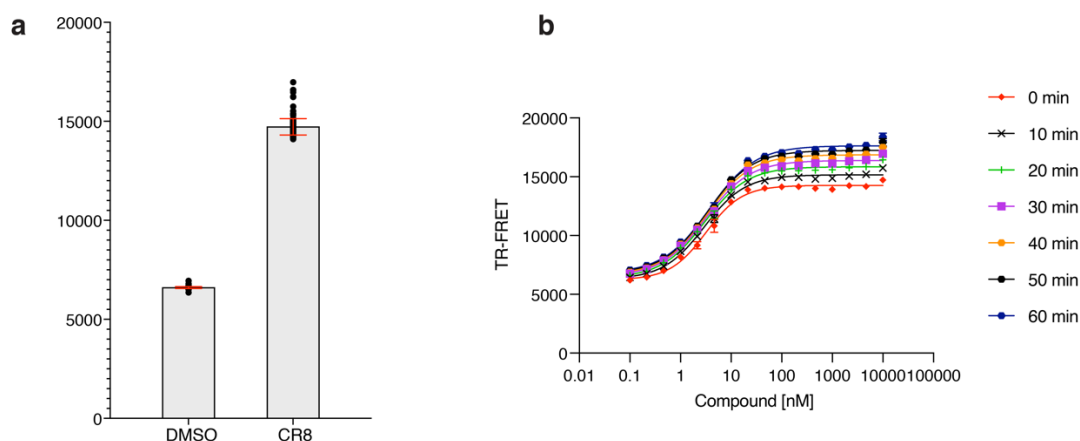
Setup	Label positions	Titration mode	Buffer composition	Plate type	Compound dispensing
Initial (Chapter 2)	CDK12-Alexa488cyclin K and TerbiumDDB1	Forward titration	50 mM Tris pH 7.5 150 mM NaCl 0.1 % Pluronic F-68 0.5 % DMSO	Black (Greiner, cat. no. 784076)	Manual
Final (Chapter 3)	CDK12-cyclin K_Terbium and Alexa488DDB1	Compound titration	50 mM HEPES pH 7.4 150 mM NaCl 0.005 % Tween 20 0.5 % DMSO 0.05 % BSA 1 mM TCEP 2 mM EDTA	White (Greiner, cat. no. 784075)	Digital

#### 4.1.9 Assay quality control and miniaturisation

Finally, we performed additional steps to verify whether the optimised TR-FRET setup could be used for HTS of larger libraries.

For this, we tested whether the assay could be effectively performed in a 1536-well format and in a 4  $\mu$ L volume (as opposed to the 384-well plates and volume of 8  $\mu$ L per well employed so far). First, a high-low experiment was performed, whereby many replicates of a positive (here: CR8; n=198) and negative control (here: DMSO; n=198) are measured to determine the assay quality and dynamic range. For the handling of 1536-well plates, liquid-dispensing systems at Novartis were used under the supervision of Sandra Romero-

Lopez. Firstly, CR8 was dispensed using a D300e (Tecan) digital dispenser as described above, with the exception that now the order of addition was reversed due to the constraints of the HTS pipeline. The assay solution (containing Alexa488DDB1, CDK12-cyclin K<sub>biotin</sub>, and Streptavidin-Tb in assay buffer) was then plated using a CERTUS FLEX liquid dispenser. TR-FRET measurements performed using the Pherastar (BMG) demonstrated a robust assay window and limited variability within each condition (**Figure 4.17a**).



**Figure 4.17. Assay miniaturization.** CR8 (1  $\mu$ M) or DMSO were added to the assay solution containing 100 nM Alexa488DDB1, 50 nM CDK12-cyclin K<sub>biotin</sub>, and 4 nM Streptavidin-Tb in the final assay buffer (**Table 4.3**). b, Titration of CR8 into the assay solution, with consecutive readouts every 10 minutes. (a, b) Both experiments were performed in a 1536-well white plate with 4  $\mu$ L per well. The y-axis range is indicative solely of the Pherastar plate reader signal multiplication settings and does not reflect a change of the assay window in this assay setup.

This dataset allowed us to evaluate the  $Z'$  factor, a common metric of HTS assay quality<sup>326</sup> (**Equation 4.2**). In the experiment shown in **Figure 4.17a**, the  $Z'$  factor was determined to be 0.83, a value indicative of a highly robust assay, indicating that assay hits can be identified with a high degree of confidence<sup>326</sup>.

$$Z' = 1 - \frac{(3\sigma_{c+} + 3\sigma_{c-})}{|\mu_{c+} - \mu_{c-}|}$$

**Equation 4.2.**  $Z'$  factor for assessment of assay quality.  $\sigma_{c+}$  and  $\sigma_{c-}$  refer to the standard deviation of the positive (+) and negative (-) control.  $\mu_{c+}$  and  $\mu_{c-}$  represent the mean of the positive (+) and negative (-) control, respectively<sup>326</sup>. The metric is reflective of both the dynamic range of the assay, and of the data variation. Zhang *et al.* describe assays with  $Z' > 0.5$  as excellent, with  $Z'=1$  corresponding to a theoretical ideal assay<sup>326</sup>.

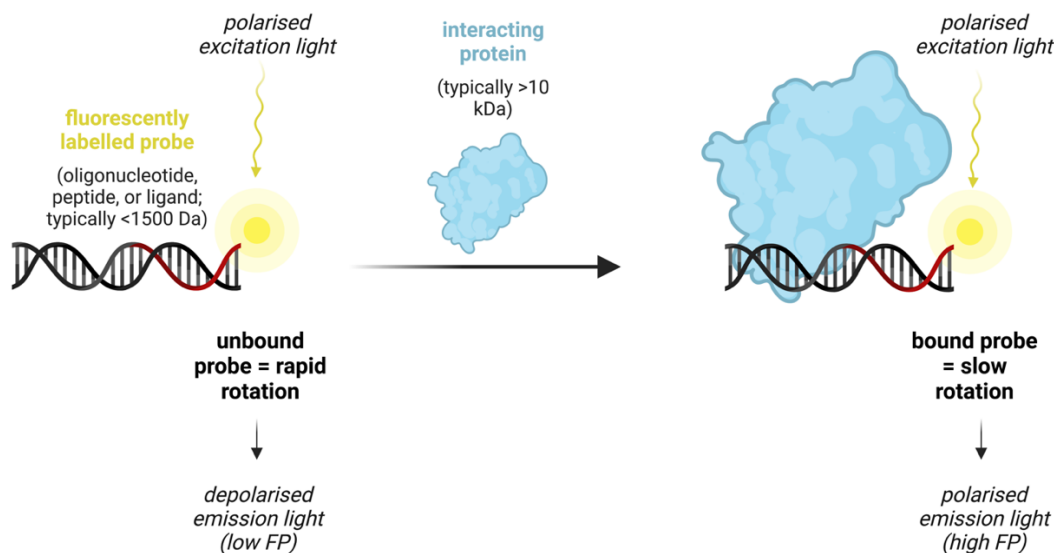
We further measured CR8 dose-response curves with the downscaled TR-FRET setup and the CR8 titration yielded EC<sub>50</sub> values in the expected range (7-10 nM) (**Figure 4.17**). The signal was read consecutively every 10 minutes to establish whether plate incubation is required for, or perhaps detrimental to, robust signal. We observed only minimal changes across the 60-minute measurement, meaning that a straightforward pipeline could likely be set up in the context of HTS assay.

Hence, the TR-FRET assay was optimised and successfully downscaled to a 1536-well format. The high  $Z'$  parameter and robust dynamic range strongly suggest that the assay can be used for HTS campaigns to identify compounds that promote the DDB1-CDK12-cyclin K interaction among larger libraries.

## 4.2 Fluorescence-based methods to study protein-DNA interactions

### 4.2.1 Fluorescence polarization

Fluorescence polarisation (FP) is another fluorescent-based technique that is frequently used to measure protein-ligand, protein-DNA, and protein-protein interactions. Originally described by Perrin<sup>327</sup>, the technique is based on the differential Brownian motion-driven rotation of molecules in solution depending on their mass, with larger masses moving slower<sup>328</sup>. Upon excitation by polarised light, a fluorescent molecule will emit light with a degree of polarisation inversely proportional to the rate of its molecular rotation. Thus, when a small fluorescently labelled species (*e.g.* peptide, oligonucleotide, or ligand; typically smaller than 1500 Da) is excited with polarised light, its rapid rotation will result in prompt depolarisation of the emitted light. However, when this small species binds a larger molecule (typically a protein larger than 10-15 kDa), its rotational movement becomes slower, with the emitted light therefore retaining more polarisation (**Figure 4.18**). The binding of a small, fluorescently labelled species to a larger unlabelled biomolecule can thus be monitored through changes in emitted light polarisation from low to high.



**Figure 4.18. The principle of FP.** An unbound fluorescent probe rotates rapidly, and thereby emits light that does not preserve the polarization of the excitatory light, leading to low FP signal. Upon binding to a much larger species, the fluorophore rotates more slowly, preserving some polarization of the excitation light (high FP signal).

The FP signal is proportional to the fraction of bound species and can be used to determine apparent  $K_d$  or  $IC_{50}$  values of the studied interaction. Quantitatively, FP is defined as the difference in the intensity of emitted light parallel ( $I_{||}$ ) and perpendicular ( $I_{\perp}$ ) to the plane of excitation light, normalised by the total fluorescence emission intensity (**Equation 4.3**)<sup>328</sup> and is commonly expressed in millipolarisation units (mP). The FP value

is therefore independent of the absolute intensities measured at either orientation ( $I_{\parallel}$ ,  $I_{\perp}$ ) and hence of the concentration of the fluorophore. While this relationship holds in principle, deviations can be observed upon aggregation of the fluorescent probe at high concentrations or in case the fluorophore's quantum yield (intrinsic fluorescence intensity) changes upon protein binding, and such aberrations complicate the interpretation of the results<sup>329</sup>. Hence, the nature of the fluorophore can affect the studied interaction in ways that are challenging to predict.

$$FP = \frac{I_{\parallel} - I_{\perp}}{I_{\parallel} + I_{\perp}}$$

**Equation 4.3.** Fluorescence polarization (FP) is defined as the difference in the intensity of emitted light parallel ( $I_{\parallel}$ ) and perpendicular ( $I_{\perp}$ ) to the plane of excitation light, normalised by the total fluorescence emission intensity.

The dynamic range of the FP assay is determined by the change in molecular weight upon binding, but also by the fluorophore lifetime, and hence its choice should be carefully considered<sup>330</sup>. Other fluorophore properties such as quantum yield, extinction coefficient, and chemical stability should also be taken into account<sup>331</sup>.

From a practical standpoint, FP measurements are conducted black microtiter plates (usually 96- or 384-well) using a plate reader equipped with suitable polarised excitation and emission filters. The experiments require a fluorescent probe, and various fluorophores and conjugation methods can be employed depending on the chemical nature of the tracer and on the studied interaction. The main advantage of FP assays is their homogenous, mix-and-read nature. However, an important drawback of the method is a relatively restricted range of binding strengths that can be measured. This is because relative saturation of the tracer-protein interaction (large fraction bound) is required for a sufficient assay window, which for low-affinity interactions can lead to unattainably high protein concentrations being required for conclusive titrations<sup>332</sup>.

## 4.2.2 Mechanisms of OCT4-SOX2 motif readout on nucleosomes

This section features a summary of a published manuscript under the same title, including a more detailed description of some of the FP experiments that I contributed to this work (subsection 4.2.2.4, includes unpublished data). Specifically, my contribution spanned the design, optimisation, and execution of FP experiments for the investigation of the binding affinities of OCT and/or SOX2 for double-stranded DNA (dsDNA) and nucleosomes. I also performed histone purification and octamer reconstitution, DNA purification, nucleosome reconstitutions, and OCT4-SOX2 purifications. The methods pertaining to my contributions to this work are described in section 6.4, whereas the complete methodology for the study can be found in the published manuscript (**Appendix**).

### 4.2.2.1 Introduction

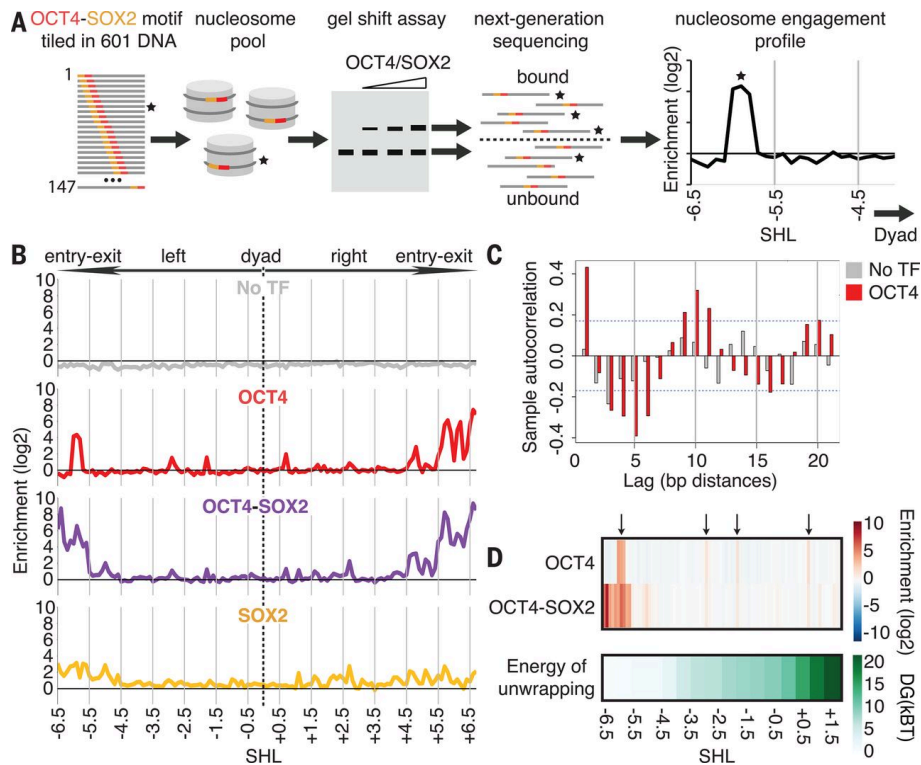
Transcription factors (TFs) regulate gene expression by interacting with specific motifs in chromatin, where its fundamental building blocks, nucleosomes, inherently restrict DNA accessibility<sup>333,334</sup>. Nucleosomes, composed



of histones (H2A, H2B, H3 and H4) and two DNA gyres, are predicted to restrict TF access to over 90% of nucleosomal DNA<sup>335,336</sup>. A subset of so-called pioneer TFs was defined based to their ability to recognize their motifs in chromatinised DNA - yet, it remains poorly understood how these TFs can bind nucleosome-occupied genomic loci. Multiple scenarios for pioneer factor engagement of chromatin have been proposed, including TF binding without alterations to the nucleosomal architecture, engagement with changes such as histone distortions, and binding facilitated by nucleosome unwrapping dynamics at the entry-exit sites<sup>337-340</sup>. In this study, we employed a novel quantitative assay, SeEN-seq, as well as structural and biophysical methods, to investigate how the key pluripotency pioneer factors OCT4-SOX2 engage nucleosomal DNA.

#### 4.2.2.2 SeEN-seq reveals the end-binding preference of OCT4-SOX2

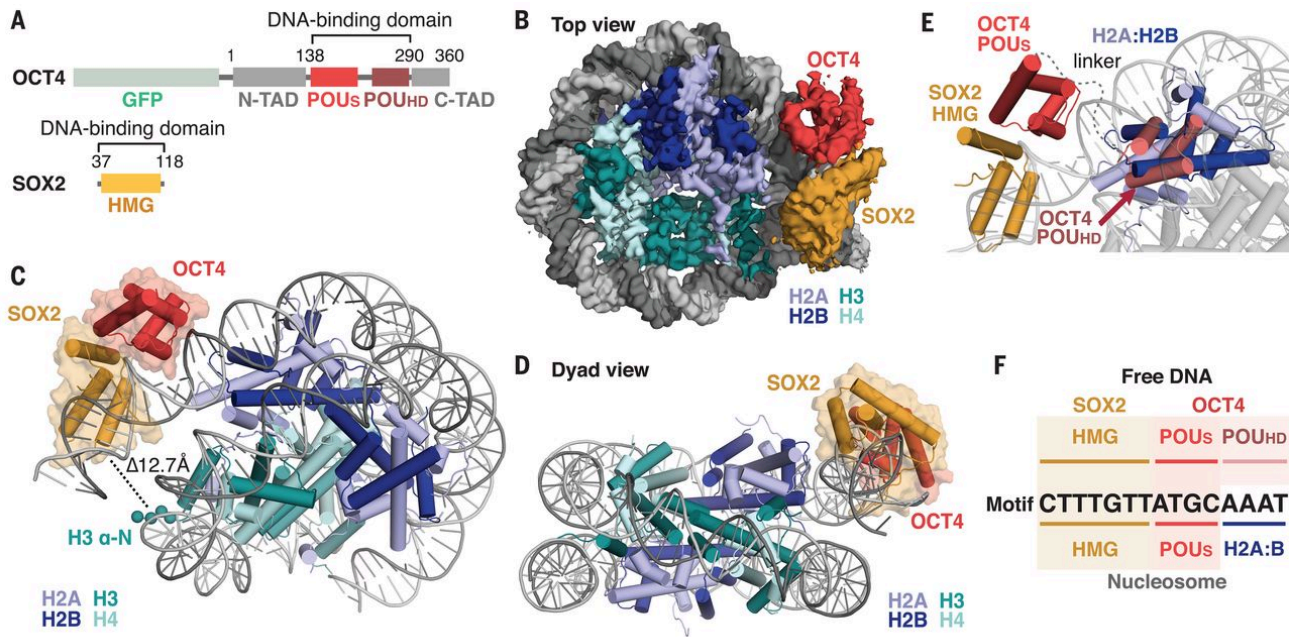
Around a nucleosome, each rotational and translational motif register places the TF in a distinct environment with respect to the DNA gyres and the histone octamer. To evaluate the relative affinity of OCT4/SOX2 at all possible nucleosomal positions in parallel, we developed a quantitative, sequencing-based Selected Engagement on Nucleosome sequencing (SeEN-seq) assay, which combines electron mobility shift assays (EMSAs) with a library of nucleosomes bearing different motif positions and next-generation sequencing (NGS) (**Figure 4.19a**). This technique allowed for the exhaustive evaluation of OCT4, OCT4-SOX2, and SOX2 binding preferences within a 601-nucleosome positioning sequence<sup>341</sup> containing a canonical OCT4 motif (**Figure 4.19b**). The assay revealed a clear OCT4 preference for nucleosomal DNA entry-exit sites, and a minor contribution of periodic binding (**Figure 4.19b, c**). While OCT4-SOX2 binding appears roughly symmetrical across the dyad, for OCT4 alone, enrichment in the right half of the nucleosome (SHL+4 to +6.5) is more pronounced (**Figure 4.19b**), likely due to the opposite motif orientation on either side of the dyad. SOX2 shows less differential enrichment in SeEN-seq, but it cooperatively strengthens OCT4 binding by up to 650-fold at discrete sites. While stronger binding near the DNA ends was observed throughout, local motif orientation and nucleosome context also modulated TF engagement (**Figure 4.19b, d**). Thus, spatial orientation of the motif, cooperativity and nucleosomal breathing dynamics all govern OCT4-SOX2 binding.



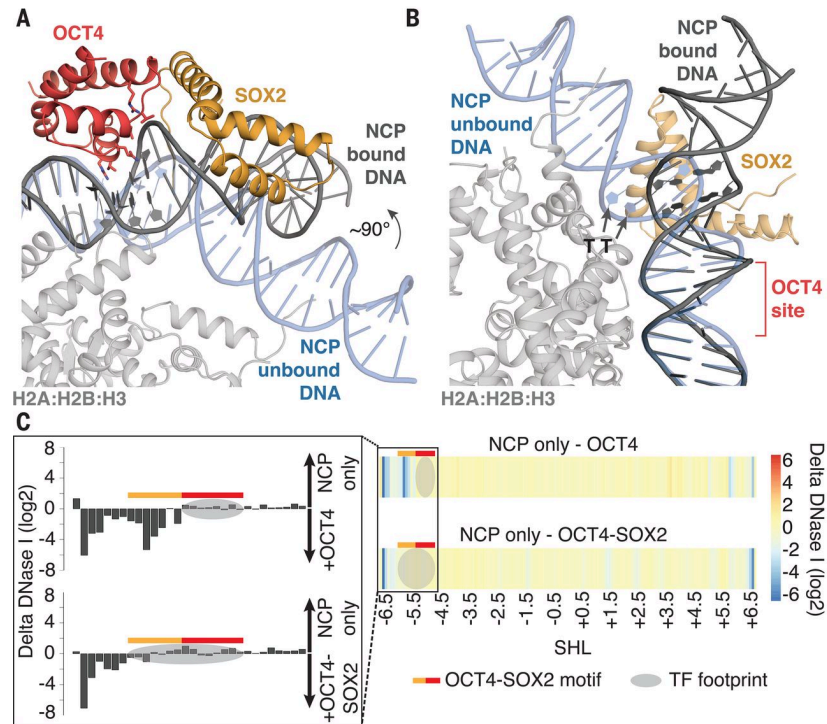
**Figure 4.19. SeEN-seq identifies preferred binding sites of OCT4 and OCT4-SOX2 across the nucleosome.** (A) Principle of SeEN-seq. A library of TF motif-containing nucleosome positioning sequences is assembled into nucleosomes and incubated with TF(s). TF-bound and unbound nucleosome complexes are separated by EMSA and sequenced, revealing position-specific enrichments. Stars indicate a specific example sequence at each step of the assay. (B) SeEN-seq enrichment ( $n = 3$  replicates) for nucleosome pool with or without TFs. Motif position is indicated by superhelix location (SHL) that describes where the minor groove faces away (SHLs  $\pm 1$ ,  $\pm 2$ , etc.) or toward (SHLs  $\pm 1.5$ ,  $\pm 2.5$ , etc.) the histone octamer. (C) Autocorrelation analysis of OCT4 enrichment; dashed lines indicate 95% confidence interval. (D) DNA nucleosome unwrapping energy<sup>338</sup> versus OCT4 and OCT4-SOX2 SeEN-seq enrichment profile. Arrows indicate enriched positions in OCT4. DG, delta Gibbs free energy; kBT, 0.6 kcal/mol, where  $T = 300$  K. Figure reproduced from<sup>342</sup>. *I have not contributed to these results (beyond protein purification).*

#### 4.2.2.3 Structural studies of OCT4-SOX2 on the nucleosome

We then pursued cryo-EM studies to visualise the OCT4-SOX2-nucleosome complex at two favoured positions (SHL-6 and SHL+6). At the SHL-6 motif location, the bound OCT4-SOX2 heterodimer causes partial DNA release away from the histone core, but without destabilising the octamer itself (Figure 4.20). OCT4 has a bipartite DNA binding domain (DBD) consisting of a POU-specific (POU<sub>S</sub>) and POU-homeodomain (POU<sub>HD</sub>), but only the POU<sub>S</sub> domain was visible and engaged to a partial OCT4 motif on the nucleosome<sup>343</sup> (Figure 4.20a, e, f). The DNA detachment from histones was attributed mainly to SOX2 inducing a *circa* 90° kink in a manner reminiscent of its engagement of free DNA, with the two factors synergistically releasing the DNA from the octamer (Figure 4.21a, b). As an orthogonal approach, we employed DNase I footprinting assays and confirmed that OCT4-SOX2 binding at the SHL-6 position results in DNA detachment from the histone octamer (Figure 4.21c).

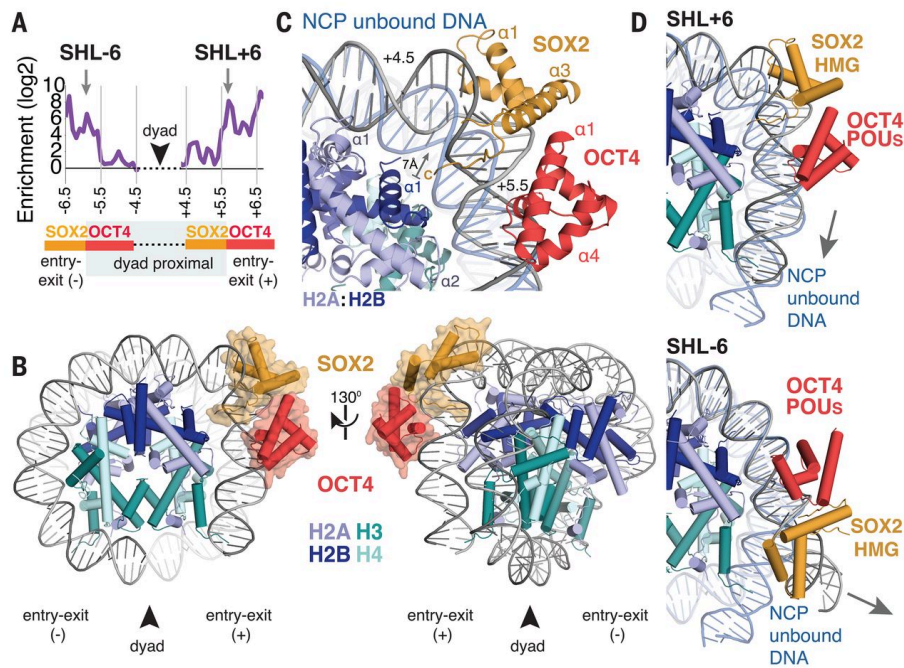


**Figure 4.20. Cryo-EM structure of OCT4-SOX2-NCP<sup>SHL-6</sup> complex.** (A) Domain schematic of OCT4 and SOX2 constructs. GFP, green fluorescent protein; TAD, transactivation domain. (B) Cryo-EM map of OCT4-SOX2-NCP<sup>SHL-6</sup> at 3.1-Å resolution. (C) Model of the OCT4-SOX2-NCP<sup>SHL-6</sup> complex. The H3 N-terminal helix and tail (shown as spheres) stabilise DNA at the entry-exit in a canonical nucleosome<sup>344</sup>. The presence of OCT4-SOX2 increases the distance between H3 (Arg41) and the nearest DNA base (T143) by 12.7 Å as compared with the unbound nucleosome. (D) Dyad view of the OCT4-SOX2-NCP<sup>SHL-6</sup> structure. (E) In the depicted OCT4-SOX2 arrangement, a model of POU<sub>HD</sub> engagement with its motif shows significant clash with the H2A:H2B dimer. Overlay of free DNA-bound structure (PDB: 1O4X) with the nucleosome-bound structure, aligned on the DNA. (F) Schematic of free DNA versus the observed nucleosome-binding mode of OCT4-SOX2. NCP: nucleosome core particle. Reproduced from<sup>342</sup>. *I have not contributed to these results (beyond NCP reconstitution and protein purification).*



**Figure 4.21. OCT4-SOX2 lifts the entry-exit DNA away from the histone core.** (A) Comparison of the unbound nucleosome DNA (blue) with the OCT4-SOX2 bound nucleosome structure (grey). The DNA is kinked  $\sim 90^\circ$  away from the histones. Residues at the OCT4-DNA interface are shown as sticks. POU<sub>S</sub> motif nucleotides are shown as ribose and base rings. (B) SOX2 kinks the nucleosomal DNA away from the histones. SOX2 uses conserved (Phe48 and Met49) residues to intercalate a TT base step, indicated by arrows. OCT4 is removed for clarity. (C) Difference in DNase I digestion across the nucleosome in the presence of OCT4 or OCT4-SOX2 (right). Zoom-in view around the OCT4-SOX2 motif (left). Reproduced from <sup>342</sup>. *I have not contributed to these results (beyond NCP reconstitution and protein purification).*

In a second structure, we visualised OCT4-SOX2 bound in a similar nucleosome environment (SHL+6), but with an inverted OCT4-SOX2 motif (**Figure 4.22**). At this register, OCT4 again was found to only engage a partial motif with its POU<sub>S</sub> domain, but instead of DNA release, OCT4-SOX2 binding was accommodated with only local distortions.



**Figure 4.22. OCT4-SOX2 bound at SHL +6 induces localized DNA distortion.** (A) Depiction of OCT4-SOX2 SeEN-seq profile illustrating symmetric positions used for structure determination (arrows). As the OCT4-SOX2 motif is tiled across the nucleosome, the orientation of the proteins in relation to the dyad is inverted. (B) OCT4-SOX2-NCP<sup>SHL+6</sup> model. (C) Details of SOX2-induced DNA kink. SOX2 binding locally distorts DNA near SHL +5. The C terminus of SOX2 approaches the histone core when bound in this orientation, distorting DNA away from the stabilizing H2B residues [N terminus (Arg34) and  $\alpha 1$ ]. The 7-Å movement indicated is calculated between the phosphate backbone at Cys123 in the bound and unbound DNA. SHLs are shown (SHLs +4.5 and +5.5). (D) Comparison of the DNA trajectory in the OCT4-SOX2-NCP<sup>SHL+6</sup> (SHL +6) (top) versus OCT4-SOX2-NCP<sup>SHL-6</sup> (SHL -6) (bottom) structures. The DNA trajectory at the entry-exit site of the TF-bound structures is indicated with an arrow. Reproduced from <sup>342</sup>. *I have not contributed to these results (beyond NCP reconstitution and protein purification).*

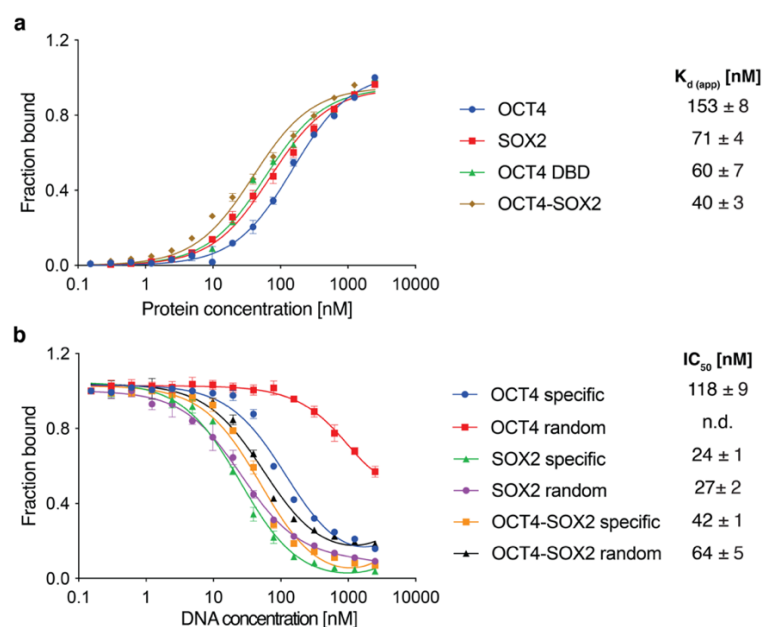
Hence, in both structures, OCT4 binds a partial DNA motif through its POU<sub>S</sub> domain, and together with SOX2 impacts the entire nucleosomal DNA structure to varying extents. Strikingly, the nucleosome responds to the same motif in a non-equivalent manner when it is presented in a different orientation, highlighting further structural implications of motif placement.

#### 4.2.2.4 Fluorescence polarisation measurements to study TF-DNA and TF-nucleosome interactions

The SeEN-seq assay revealed how the motif register impacts the binding affinity of OCT4 and/or SOX2 for nucleosomes (section 4.2.2.2). We sought to establish an orthogonal FP assay in order to corroborate the relative binding affinities for several chosen motif positions as well as to study the effect of protein mutations on these interactions.

For this, we used a 21-mer dsDNA probe containing the OCT4-SOX2 motif and labelled with fluorescein (Flc) (5'-Flc-GACCTTTGTTATGCAAATTA). We first tested the TF-probe interaction through forward titrations of each protein (OCT4, SOX2, OCT4-SOX2 complex or OCT4 DBD) into the tracer dsDNA (Figure 4.23a). We observed nanomolar interactions, with OCT4 alone binding more weakly than SOX2 or the

OCT4-SOX heterodimer. Interestingly, the OCT4 DBD was found to binds with a higher affinity than the full-length OCT4 protein. This observation is somewhat counterintuitive, yet we have observed such a pattern for other TFs (data not shown).

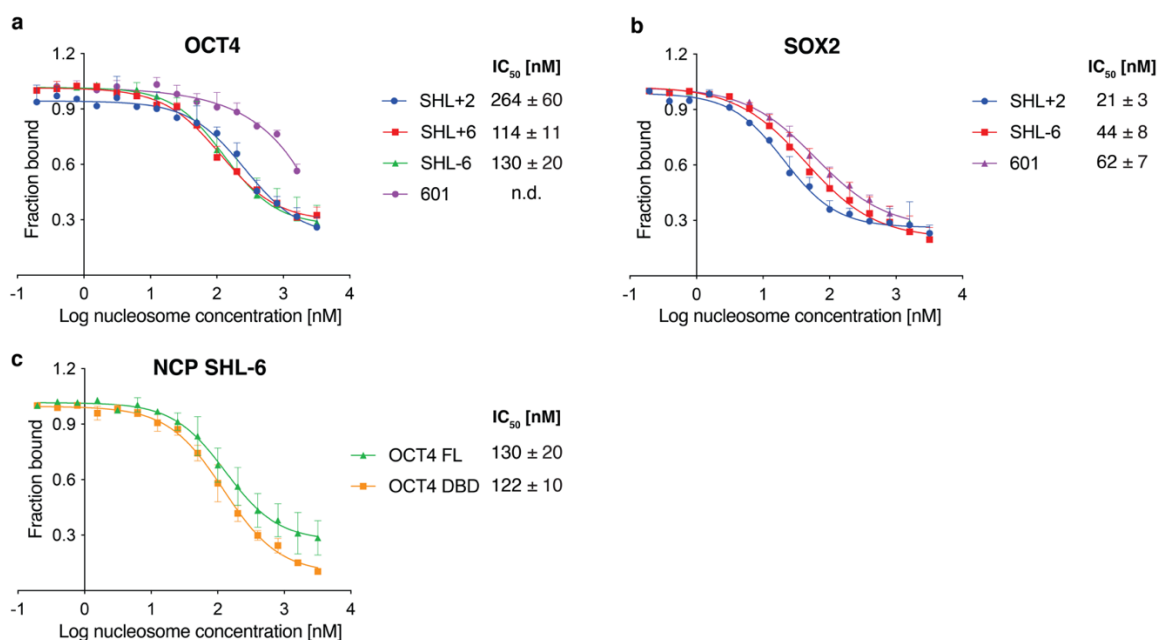


**Figure 4.23. a**, Fluorescence polarization forward titration experiments using 10 nM Flc-OCT4-SOX2 oligo (5'-Flc-GACCTTTGTTATGCAAATTAA) in the presence of increasing amounts of either full-length OCT4, OCT4 DBD, SOX2 (residues 37-118) or OCT4-SOX2 mixed at equimolar concentrations and titrated as indicated. For OCT4-SOX2, the concentration indicates the concentration of the heterodimer.  $K_d$  (app): apparent dissociation constant. **b**, 10 nM of a Flc-21-bp-DNA containing the OCT4-SOX2 motif was mixed with OCT4, SOX2 (residues 37-118) or OCT4-SOX2 and counter titrated with the indicated unlabelled oligonucleotides. Relative affinities are indicated as  $IC_{50}$  values. We note that these measurements are unable to distinguish between non-specific and specific binding events and were fitted using a total binding curve. All data include three technical replicates ( $n=3$ ) and are shown as mean  $\pm$  s.d.

Next, we wanted to investigate the sequence specificity of this interaction in a competition FP assay. In this assay format, the FP signal should decrease upon titration of an unlabelled competitor dsDNA. We performed counter titrations with a specific (identical to the tracer but unlabelled) or random (randomised sequence, no Flc) 21-bp dsDNA (**Figure 4.23b**). We observed that the binding of OCT4 shows clear dependence on the presence of its cognate motif, while SOX2 does not differentiate between random and motif-containing oligonucleotides and OCT4-SOX2 together display only a mild sequence preference.

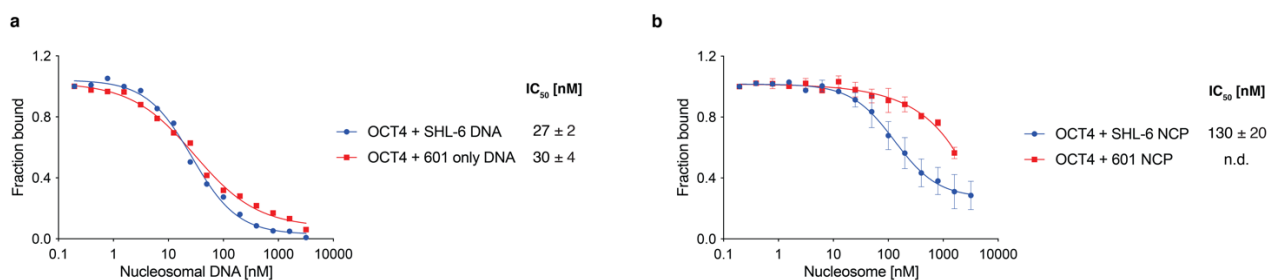
To validate the relative affinities of OCT4 and/or SOX2 binding at selected positions on the nucleosome, we next performed counter titrations with several nucleosomal substrates (**Figure 4.24**). We observed that OCT4 preferentially engages nucleosomes with the motif at the SHL $\pm$ 6 position than at an internal site (SHL+2), and only weakly interacts with a Widom 601 sequence nucleosome<sup>341</sup> with no motif (**Figure 4.24a**). Akin to its behaviour on free DNA, SOX2 showed only a mild preference for motif-containing over 601 nucleosomes and, interestingly, showed a two-fold higher affinity for the dyad-proximal SHL+2 motif position than for the SHL-6 location, in agreement with earlier reports<sup>339</sup> (**Figure 4.24b**). We also compared the binding

of OCT4 DBD and the full-length protein to the SHL-6 nucleosome and saw that unlike on free DNA, here they display roughly the same binding affinity (**Figure 4.24c**). Interestingly, the density for the OCT4 transactivation domains (TADs) was not observed in the cryo-EM structures, consistent with its lack of positive contribution to nucleosome binding in the assay.



**Figure 4.24.** **a**, 10 nM of a fluorescein (Flc) labelled 21-bp DNA containing the OCT4-SOX2 motif (5'-Flc-GACCTTTGTTATGCAAATTAA) were mixed with 300 nM OCT4 full-length protein and counter-titrated with nucleosomes. Relative affinities are indicated as IC<sub>50</sub> values. All data include two to three technical replicates and are shown as mean ± s.d. **b**, 10 nM of a Flc-21-bp-DNA containing the OCT4-SOX2 motif were mixed with 150 nM SOX2 (aa 37-118) and counter titrated with nucleosomes. Relative affinities are indicated as IC<sub>50</sub> values. All data include two to three technical replicates and are shown as mean ± s.d. **c**, 10 nM of a Flc-21-bp-DNA containing the OCT4-SOX2 motif were mixed with 300 nM full-length OCT4 or OCT4 DNA binding domain only (residues 134-290) and counter titrated with nucleosomes. All data include three technical replicates (n = 3) and are shown as mean ± s.d.

Pioneer factors have initially been postulated to have similar binding affinity on nucleosomes as on naked DNA<sup>345</sup>. More recent work from multiple laboratories, however, has shown that these factors generally exhibit higher affinity for free DNA or linker regions than for nucleosomes<sup>339</sup>. As described above, we found that OCT4 shows specific binding to both free DNA (21-mer oligonucleotide) and nucleosomes, while SOX2 displays a high-degree of nonspecific binding affinity to both free DNA (21-mer oligonucleotide) and nucleosomes. To directly compare the binding affinity to the naked vs nucleosomal DNA for OCT4-SOX2, we performed FP counter titration with long, 153-bp dsDNA that is not assembled into nucleosomes. Strikingly, OCT4 binds 153bp free DNA (601 sequence + *EcoRV* overhang), and the same DNA carrying an OCT4-SOX2 motif, with roughly equal affinity (**Figure 4.25a**); while for OCT4, the same DNA embedded in a nucleosome showed over ten-fold tighter binding of the motif-containing nucleosome compared to 601 alone (**Figure 4.25b**). Hence, nucleosomes appear to restrict non-specific TF binding, which is likely part of their function *in vivo*.

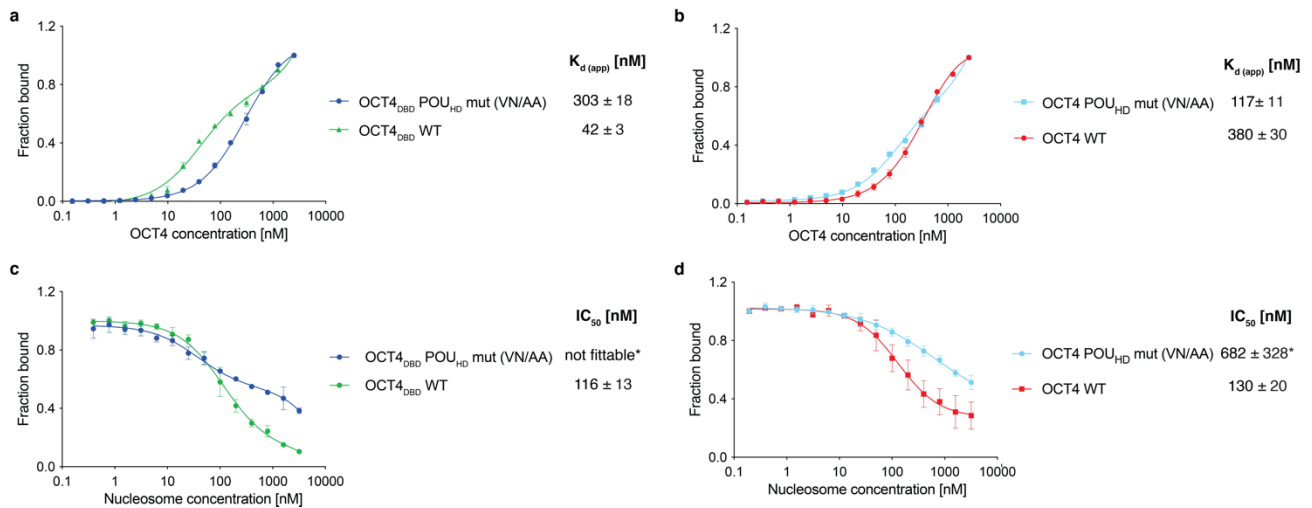


**Figure 4.25.** **a**, Fluorescence polarization measurements with 153bp 601 DNA with (SHL-6) and without (601 only) an OCT4-SOX2 motif. **b**, as in **(a)** but with SHL-6 and 601 DNA assembled on a nucleosome.

Finally, to further validate the POU<sub>S</sub> domain-only engagement of OCT4 observed in the structure, we attempted FP assays with relevant truncations and mutants. As the POU<sub>S</sub> domain-only construct of OCT4 only gave rise to insoluble protein, we employed constructs bearing POU<sub>HD</sub> mutations either in the isolated OCT4 DNA binding domain (DBD, residues 134-290) or the full-length protein as proxies for a POU<sub>S</sub>-only construct. Guided by the published structure of the OCT4 POU<sub>S</sub> and POU<sub>HD</sub> domains complexed to an isolated DNA duplex<sup>346</sup> and the high-resolution crystal structure of OCT1-SOX2 on naked DNA<sup>347</sup>, mutations in the POU<sub>HD</sub> DNA binding fold that make direct contacts with DNA (V277A, N280A) were introduced to reduce DNA binding affinity (VN/AA mutant).

On free DNA, this mutant exhibited a *circa* eight-fold decrease in binding affinity in the DBD context (**Figure 4.26a**), and three-fold lower binding in the context of the full-length protein (**Figure 4.26b**). Upon nucleosome counter titration, we observed biphasic binding behaviour in the case of the DBD mutant (**Figure 4.26c**), yet both in the DBD and in the full-length context we could conclude that the mutant shows only a mild binding impairment (**Figure 4.26c, d**). As TF binding on a nucleosome could not be saturated, the cited affinity differences should be viewed as a rough approximation. We therefore conclude that the OCT4 POU<sub>HD</sub> domain plays an important role in isolated duplex free-DNA binding, but perhaps less so at the single nucleosomal position we tested (SHL-6 motif location). This further supports the partial motif model describe in the study.

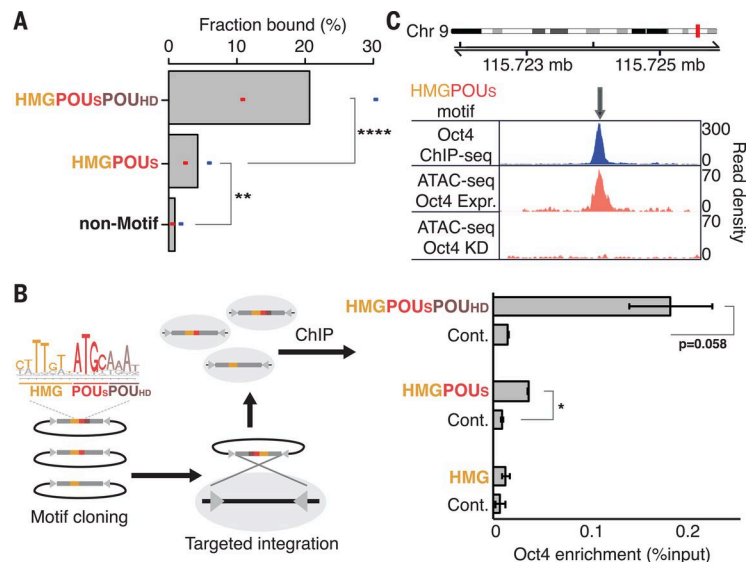




**Figure 4.26.** **a**, OCT4 POU<sub>HD</sub> domain mutant VN-AA shows a lower affinity to free DNA. Fluorescence polarization assay using a Flc-labelled 21-mer oligonucleotide containing the OCT4-SOX2 canonical motif was titrated with increasing amounts of the OCT4 DNA binding domain (aa 134-290) with the wild-type protein or a POU<sub>HD</sub> mutant (V277A, N280A).  $K_{d(app)}$  = apparent dissociation constant. **b**, As in **(a)** but using a full-length OCT4 and OCT4 mutant construct. **c**, Countertitration with NCP. **d**, as in **(c)** but using full-length OCT4 and OCT4 mutant construct.

#### 4.2.2.5 The OCT4-SOX2 partial motif is sufficient for chromatin opening in cells

Finally, by re-analysing published ChIP-seq datasets<sup>348,349</sup>, we show that the partial OCT4 motif, together with the SOX2 site, is sufficient to drive genomic binding (**Figure 4.27a, b**). Through ATAC-seq experiments, we also provide evidence that a partial motif is sufficient to allow chromatin opening in cells, indicating the *in vivo* relevance of OCT4 POU<sub>S</sub>-driven binding together with SOX2 (**Figure 4.27c**).



**Figure 4.27.** **The HMG-POUS partial motif is bound *in vivo* and opens chromatin.** **(A)** Fraction of the top thousand motifs, ranked by match to the position weight motif, that are bound by Oct4 (2-fold ChIP-seq enrichment). Mean and individual values from two separate datasets are shown<sup>348,349</sup>. Significant enrichment is detected for full and partial motifs compared with loci without a motif (lowest 1000 genomic matches to the Oct4-Sox2 motif). \*\* $P < 0.01$ ; \*\*\*\* $P < 0.0001$ . (Pearson's chi-square test, indicated for dataset with least significance). **(B)** Schematic representation of targeted motif

insertion and Oct4 ChIP-qPCR enrichment at the ectopic insertion locus and endogenous control (Cont.) locus. \* $P < 0.05$ ; error bars indicate SEM of at least two biological replicates. (C) Representative genomic region (chr9: 115723244 to 115725243) that shows binding of Oct4 to a partial HMG-POUS motif (top) and Oct4-dependent chromatin accessibility (bottom). Expr., Oct4-expressing cells; KD, Oct4-knockdown cells. Reproduced from <sup>342</sup>. *I have not contributed to these results.*

Overall, our study provides structural and mechanistic insights into how the pluripotency factors OCT4 and SOX2 engage their cognate motifs in the context of a nucleosome. Two cryo-EM TF-NCP structures reveal that the factors can be accommodated at different motif registers with a varying extent of DNA distortions, and that OCT4 only engages a partial motif at the studied sites. We further systematically analyse the motif position-dependence of TF-nucleosome interactions and validate our findings in cells, establishing how transcription factors can read out chromatinised motifs.

### 4.2.3 Structural mechanism of cGAS inhibition by the nucleosome

This section features a short summary of a published manuscript under the same title, including a more detailed description of the experiments that were my contribution to this work (subsections 4.2.3.3 and 4.2.3.4, including unpublished data). Specifically, my contribution spanned the design, optimisation, and execution of FP experiments for the investigation of the cGAS-nucleosome and cGAS-dsDNA binding affinities. The data presented in **Figure 4.33** was obtained by Alexiane Decout (Ablasser laboratory, EPFL) and some experiments in **Figure 4.34** were performed by a Master student at the time, Joscha Weiss. The methods pertaining to my contributions to this work are described in section 6.4.3, whereas the complete methodology for the study can be found in the published manuscript (**Appendix**).

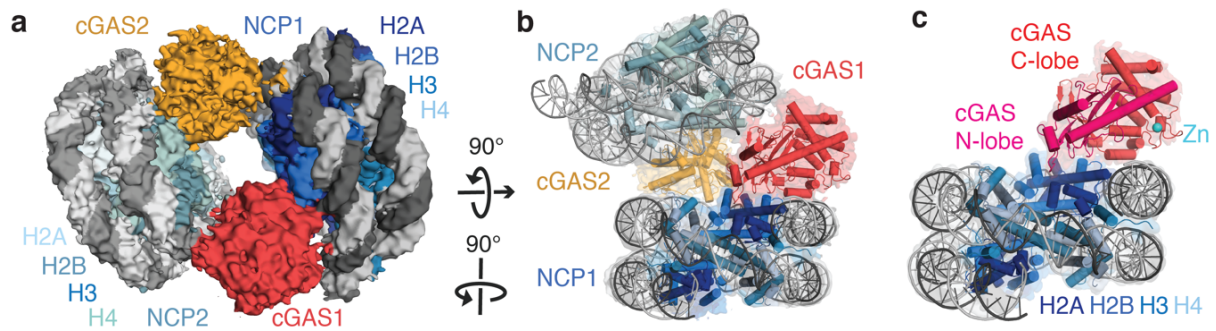
#### 4.2.3.1 Introduction

cGAS is a DNA sensor that triggers the innate immune response upon detection of foreign DNA, *e.g.* resulting from microbial infection<sup>350</sup>. Upon activation by dsDNA, cytosolic cGAS synthesises the second messenger 2'3'-cyclic GMP-AMP (cGAMP), which induces a pro-inflammatory response via the STING protein, leading to inflammatory cytokine and type I interferon (IFN) induction<sup>351–356</sup>. cGAS is also present in the nucleus where it tightly binds to chromatin<sup>357–359</sup>. Yet, nuclear cGAS is inactive and was postulated to be rendered such by chromatinised DNA through an incompletely understood mechanism<sup>360,361</sup>. In this work, we investigated the mechanism of cGAS inhibition by chromatin binding through structural studies.

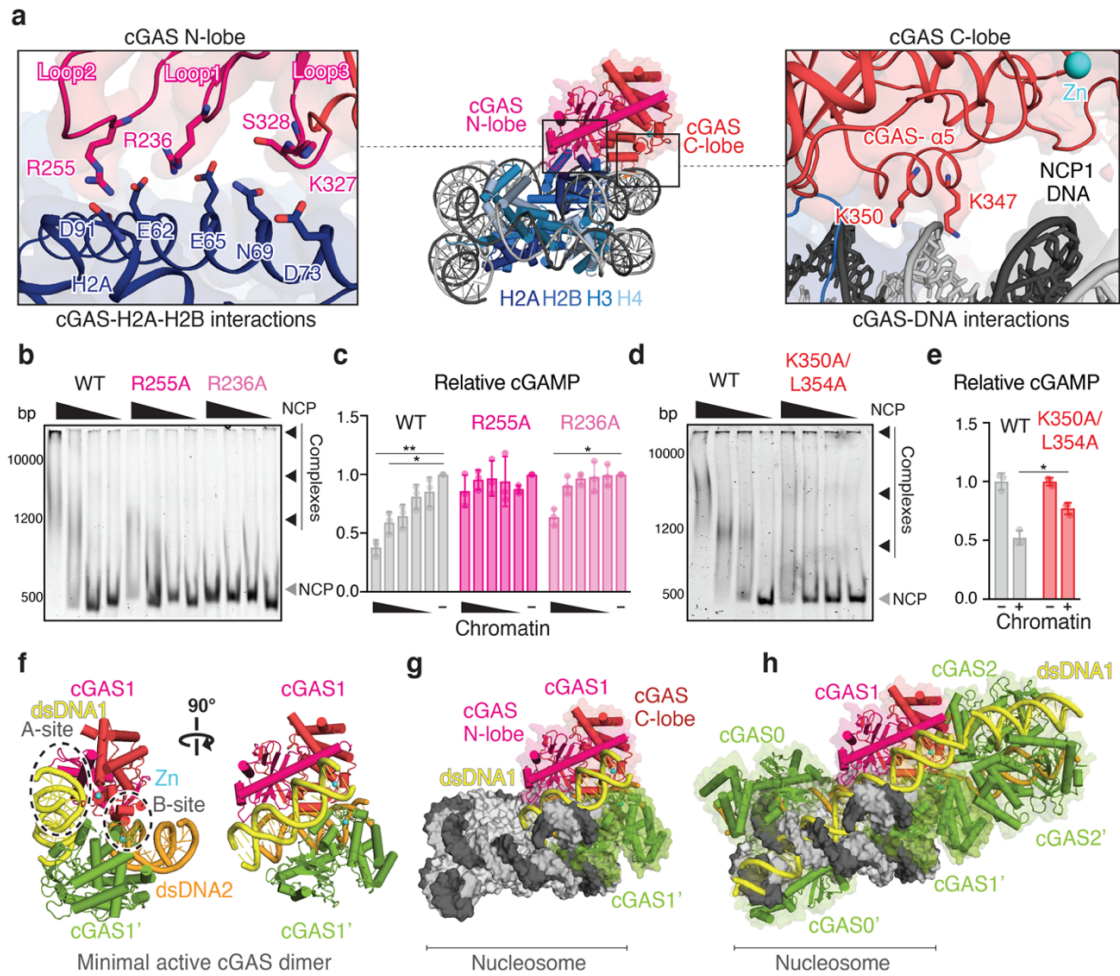
#### 4.2.3.2 Cryo-EM studies of cGAS-NCP complexes

For cryo-EM studies of cGAS-nucleosome complexes, WT or K394E mutant cGAS was mixed with a 147 bp 601 DNA<sup>341</sup> nucleosome core particle (NCP) and subjected to gradient centrifugation with chemical crosslinking (GraFix)<sup>362</sup>. For the cGAS K394E mutant, we obtained a 4.1 Å reconstruction revealing two NCPs organised in a NCP<sub>1</sub>-cGAS<sub>1</sub>-cGAS<sub>2</sub>-NCP<sub>2</sub> sandwich arrangement, with one cGAS-NCP pair (designated as

NCP<sub>1</sub>-cGAS<sub>1</sub>) considerably better resolved. In this 2:2 complex, each cGAS molecule interacts with the histone octamer of one NCP through H2A and H2B and with the nucleosomal DNA, while contacting the second nucleosome (*e.g.* cGAS<sub>1</sub> and NCP<sub>2</sub>) primarily through interactions with its nucleosomal DNA (**Figure 4.28a, b; Figure 4.29a**). A WT cGAS structure was solved at 5.1 Å, yielding the same overall arrangement. Given the structural similarity, the higher resolution cGAS K394E mutant was used for subsequent analysis.



**Figure 4.28.** **a**, 3D reconstruction of the complex containing two cGAS molecules, cGAS<sub>1</sub> (red) and cGAS<sub>2</sub> (orange), and two nucleosomal core particles, NCP<sub>1</sub> and NCP<sub>2</sub>, respectively. **b**, **c**, Ribbon diagrams of the NCP<sub>1</sub>-cGAS<sub>1</sub>-cGAS<sub>2</sub>-NCP<sub>2</sub> complex (**b**) and the cGAS<sub>1</sub>-NCP<sub>1</sub> complex (**c**) fit into corresponding electron-density maps. The two lobes of cGAS, N-lobe and C-lobe, are shown in pink and red, respectively. Adapted from <sup>363</sup>. *I have not contributed to these results.*

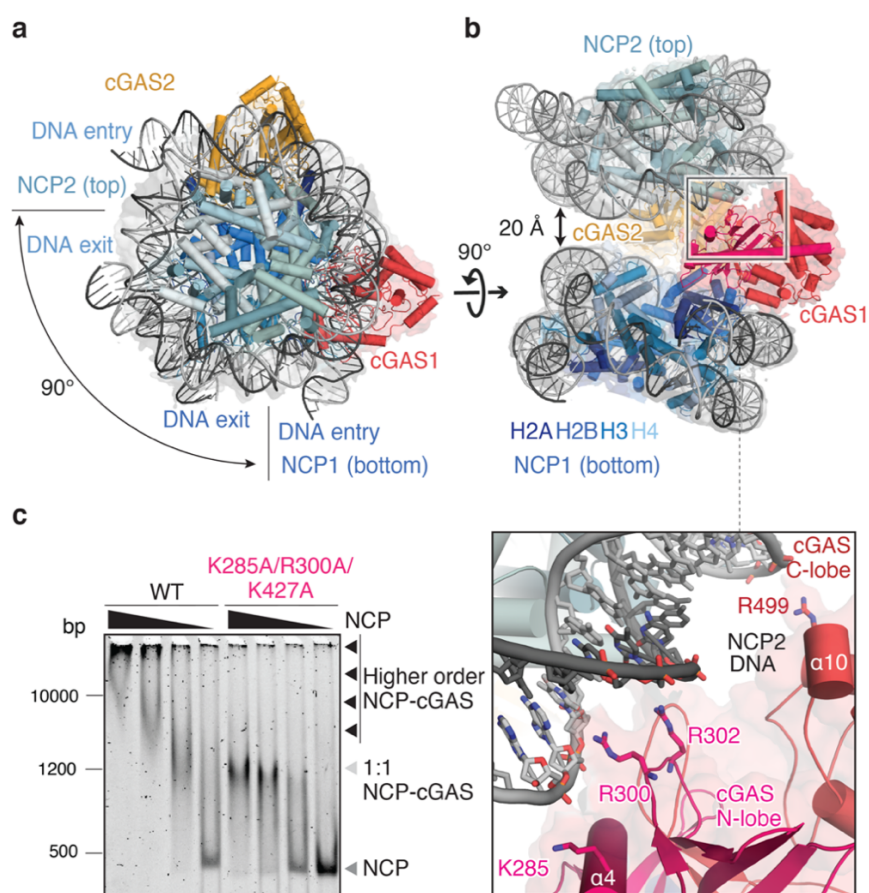


**Figure 4.29.** **a**, Magnified view of the cGAS1(K394E)–NCP1 complex bipartite interactions, cGAS–histone interactions (left) and cGAS–nucleosomal DNA interactions (right). **b**, EMSA gel showing the interaction of nucleosomes (40 ng/ $\mu$ l) with a concentration gradient of WT, R255A cGAS and R236A cGAS (from 50 to 6 ng/ $\mu$ l; 1:2 step dilutions); the black arrowheads indicate higher-order cGAS–NCP complexes. **c**, *In vitro* cGAMP synthesis of WT cGAS, R255A cGAS and R236A cGAS with or without a concentration gradient of chromatin (from 5 to 0.3125 nM; 1:2 step dilutions) normalised by cGAMP levels in the absence of chromatin for each mutant. **d**, EMSA gel showing the interaction of nucleosomes (40 ng/ $\mu$ l) with increasing concentrations of WT or K350A/L354A cGAS (from 100 to 12 ng/ $\mu$ l; 1:2 step dilution). **e**, *In vitro* cGAMP synthesis of WT and K350A/L354A hcGAS with or without chromatin (5 nM) normalised by cGAMP levels in the absence of chromatin for each individual mutant. Data are representative for three independent experiments showing similar results (**b**, **d**) or mean  $\pm$  s.d. of  $n = 3$  independent experiments (**c**, **e**). One-way analysis of variance (ANOVA) with post-hoc Dunnett multiple comparison test;  $**P = 0.0092$ ,  $*P = 0.092$  (WT) and  $*P = 0.0311$  (R236A) (**c**) and two-tailed Student's *t*-test;  $*P = 0.0192$  (**e**). The data points are from independent experiments. **f**, Overview of active hcGAS–DNA 2:2 complex with two distinct dsDNA-binding surfaces (A-site and B-site)<sup>364</sup> (Protein Data Bank (PDB) ID: 4LEY). **g**, Superposition of the hcGAS–dsDNA (**f**) and cGAS1–NCP1 complexes illustrating the incompatibility of DNA ligand binding (dsDNA1 in yellow) to cGAS in the nucleosome-bound configuration. **h**, Model based on superpositioning of the cGAS1–NCP1 complex onto DNA-bound cGAS oligomers as previously defined<sup>365</sup> (PDB: 5N6I). Reproduced from <sup>363</sup>. *I have not contributed to these results.*

Through focused refinement, the structure of one molecule of cGAS<sub>1</sub> bound to NCP<sub>1</sub> at 3.1 Å resolution was also determined, revealing more details of the interface. We saw extensive cGAS–NCP interactions mediated predominantly by the binding of three loops located in the N-lobe of cGAS to the acidic patch of H2A–H2B. Mutation of two key residues (R236A and R255A) that interact with the acidic patch fully abrogated

cGAS-nucleosome binding (**Figure 4.29b**). We also showed that these cGAS mutants (R255A and R236A) were no longer inhibited by chromatin in an *in vitro* enzymatic assay (**Figure 4.29c**).

The cGAS<sub>1</sub> molecule also binds to a second nucleosome (NCP<sub>2</sub>), predominantly through backbone contacts of conserved cGAS<sub>1</sub> residues E287, K299-R302, and K427 with the nucleosomal DNA (**Figure 4.30a, b**). The two NCPs are held ~20 Å apart with the DNA entry/exit sites of the two nucleosomes pointing roughly ~90° away (**Figure 4.30a**). EMSA revealed that while mutation of the NCP<sub>2</sub> interacting motifs on cGAS (K285, R300 and K427) still allowed for cGAS interaction with nucleosomes, it prevented the formation of any larger cGAS/NCP assemblies (**Figure 4.30c**), indicating that the secondary cGAS<sub>1</sub>/NCP<sub>2</sub> interface critically contributes to the formation of higher-order complexes (as modelled in **Figure 4.29f-h**).



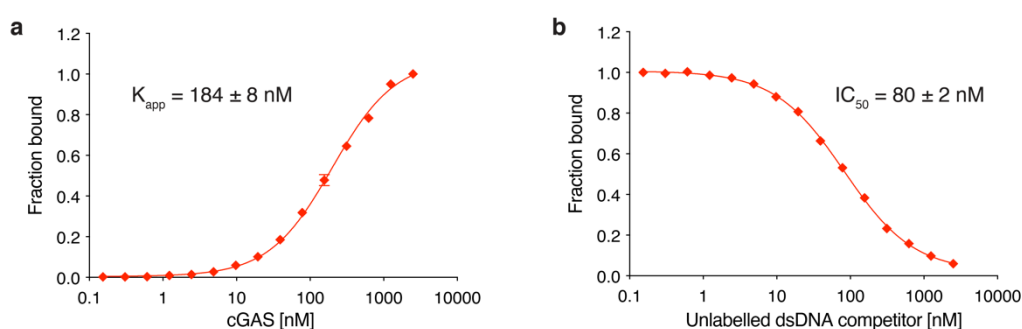
**Figure 4.30.** **a, b**, The NCP1–cGAS1–cGAS2–NCP2 di-nucleosomal arrangement is shown. A magnified view detailing the interactions between the N-lobe of cGAS1 (pink), the C-lobe of cGAS1 (red) and the nucleosomal DNA of NCP2 (grey) (**b**, bottom) is also displayed. **c**, EMSA gel showing the interaction of nucleosomes with increasing concentrations of WT or K285A/R300A/K427A cGAS (100 to 12 ng  $\mu\text{L}^{-1}$ ; 1:2 step dilution). The arrowheads highlight free nucleosomes (dark grey), complexed nucleosomes (black) and a putative 1:1 cGAS:NCP assembly (light grey). The experiment shown in **c** was independently repeated three times with similar results. Reproduced from <sup>363</sup>. *I have not contributed to these results.*

The minimal enzymatically active unit of cGAS is a 2:2 cGAS:DNA complex<sup>365–367</sup> (**Figure 4.29f**). The cGAS inactivation effected through nucleosome binding is therefore threefold. Firstly, due to steric hindrance from both the nucleosomal DNA and H2A-H2B, cGAS cannot engage dsDNA. Secondly, the key cGAS

residues essential for DNA binding and dimerization are unavailable due to nucleosome interactions. Finally, cGAS is effectively locked in a monomeric state as both histones and nucleosomal DNA sterically block its dimerization, an essential requirement for cGAS enzymatic activity<sup>364,365,367</sup>.

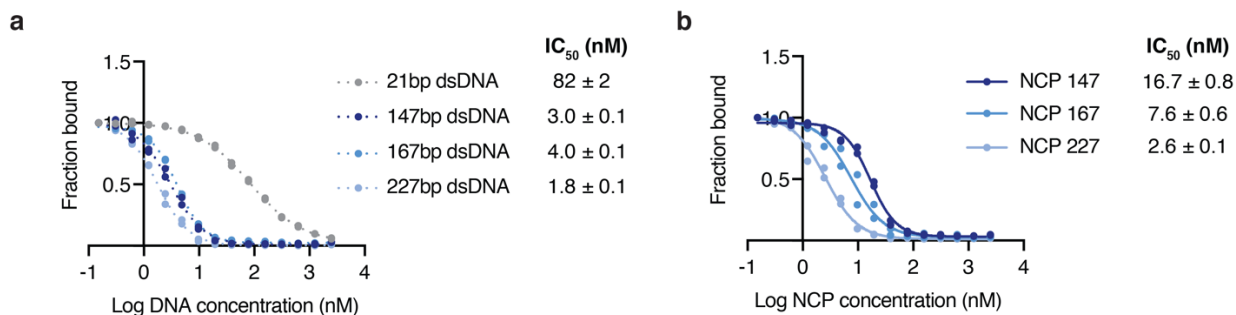
#### 4.2.3.3 Fluorescence polarisation studies to dissect the NCP linker contribution to cGAS binding

To study the underlying affinities of cGAS-DNA and cGAS-NCP interactions we aimed to establish an FP assay. We utilised a Flc-labelled dsDNA tracer (5'-Flc-GACCTTTGTTATGCAACCTAA-3') and first measured the tracer-cGAS interaction in a forward titration and observed a nanomolar binding affinity (**Figure 4.31a**). In a counter titration with an unlabelled oligonucleotide (**Figure 4.31b**) we saw clear displacement of the tracer and complete signal reduction, indicating a robust assay window for further experiments.



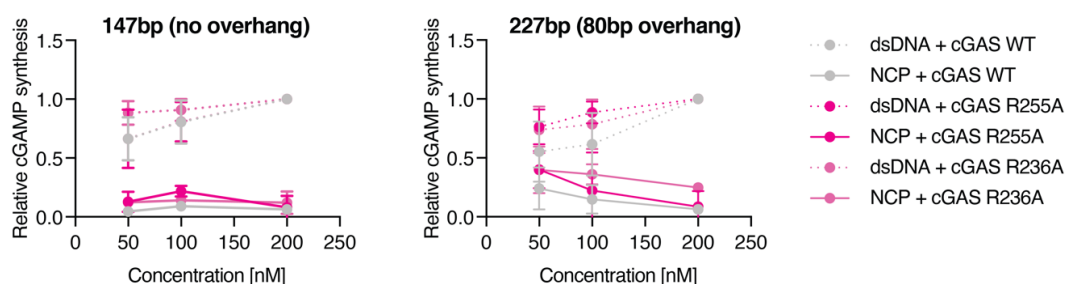
**Figure 4.31. Establishment of the FP assay for cGAS.** **a**, Forward titration of cGAS into a Flc-labelled tracer. **b**, Countertitration with an unlabelled version of the tracer reduces the FP signal to baseline. (**a,b**) Data represent mean  $\pm$  s.d. of  $n = 3$  technical replicates.

We then employed FP counter titrations to dissect the contributions of nucleosomal DNA versus linker DNA to cGAS binding and activation. These assays revealed that cGAS binds tighter to nucleosomes with long overhangs than to those without or with only short overhangs (**Figure 4.32a, b**). Through competitive titrations of a preassembled cGAS-tracer complex with various dsDNA competitors (21 bp, 147 bp, 167 bp, 227 bp), we showed that cGAS has a higher apparent affinity to longer dsDNA, likely due to the presence of more cGAS binding sites on longer substrates (**Figure 4.32a**). Counter titrations with nucleosomes with various linker lengths (NCP-147 with no linker, NCP-167 with a 20 bp linker, NCP-227 with an 80 bp linker) demonstrated that although NCP-147 is already tightly bound with an  $IC_{50}$  of 17 nM, the presence of longer linker DNA further increases the overall binding affinity of cGAS to an  $IC_{50}$  of 2.5 nM for NCP-227 (**Figure 4.32b**). We hypothesise that nucleosomes with overhangs, under condition where there is a large excess of nucleosome over short dsDNA fluorescent probe (present at 10 nM), have additional cGAS binding sites on the linker DNA and are thus better competitors. When used in stoichiometric ratios, however, the nucleosome overhangs do not trigger activation of WT cGAS and nucleosome-mediated inhibition of activity prevails. Accordingly, we expect that in a cellular context, the large excess of nucleosomes over cGAS results in sustained cGAS inhibition.



**Figure 4.32.** **a**, Fluorescein (Flc)-labelled 21 bp dsDNA tracer (10 nM) was mixed with cGAS protein (300 nM) and counter-titrated with unlabelled DNA. **b**, Fluorescein (Flc)-labelled 21 bp dsDNA tracer (10 nM) was mixed with cGAS protein (300 nM) and counter-titrated with unlabelled nucleosomes. **(a,b)** All replicates are explicitly shown. Adapted from <sup>363</sup>.

To corroborate the FP results, we assessed the catalytic activity of cGAS (WT) and cGAS acidic patch mutants, R236A and R255A, on nucleosomes with and without an 80bp dsDNA overhang. Whereas cGAS (WT) and cGAS mutants robustly synthesised cGAMP on naked dsDNA, they remained inactive in the presence of nucleosomes lacking a DNA overhang (**Figure 4.33**). Nucleosomal DNA is thus not a good substrate for cGAS activation. Nucleosomes carrying 80bp long linker DNA still failed to activate WT cGAS, but elicited activation of both cGAS R236A and R255A (**Figure 4.33**). WT cGAS is therefore typically not activated by nucleosomal or linker DNA but instead preferentially inhibited by binding to NCPs.

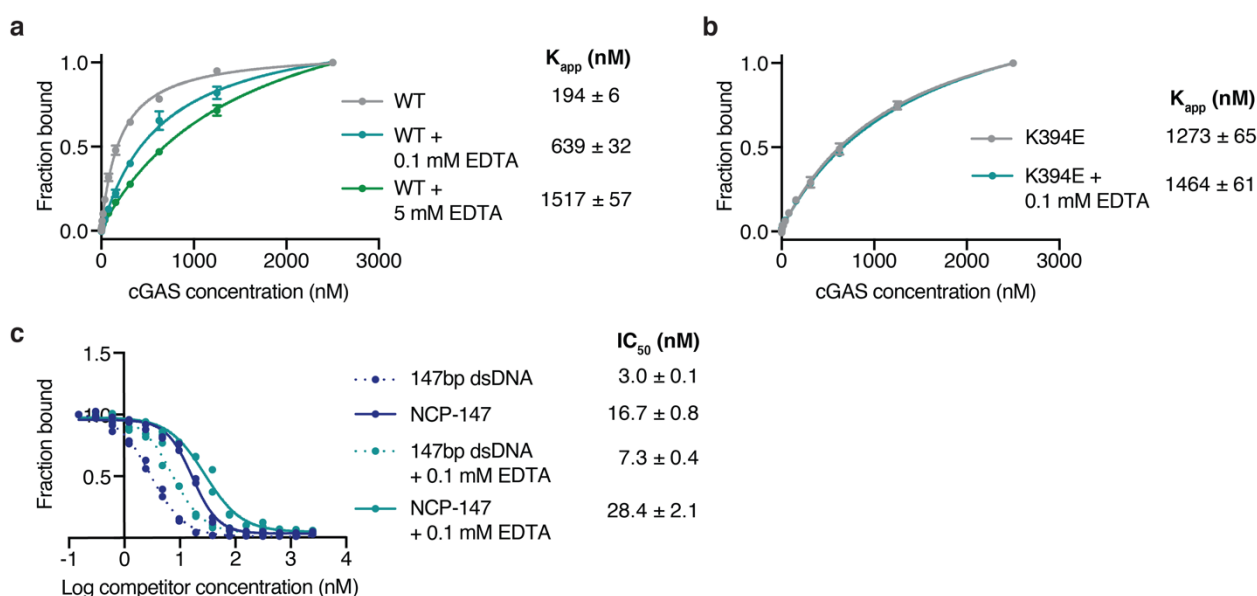


**Figure 4.33.** *In vitro* cGAMP synthesis of WT, R255A and R236A hcGAS (all 200 nM) with a concentration gradient of 147bp dsDNA or nucleosome (no DNA overhang) (left) or 227bp dsDNA or nucleosomes (80bp dsDNA overhang) (right) (from 200 nM to 50 nM; 1:2 step dilutions) normalised to cGAMP levels for 200 nM dsDNA for each individual mutant. Mean ± s.d. of  $n = 3$  independent experiments is shown. Adapted from <sup>363</sup>. *I have not contributed to these results.*

#### 4.2.3.4 Fluorescence polarisation studies to elucidate the role of a zinc finger motif

We further investigated whether, akin to the cGAS dimer interaction with foreign DNA, its interaction with the nucleosome requires the zinc finger motif. The structure of WT cGAS with the nucleosome revealed clear density around in the K394 residue, with the lysine sidechain approaching nucleosomal DNA from NCP<sub>1</sub>, and likely contributing to phosphate backbone interactions. Thus, the zinc finger motif, along with the K394-containing loop, could offer an additional nucleic acid binding site also in the context of chromatin. However, in our biochemical reconstitutions, we clearly observe cGAS-NCP complex formation both with the WT protein and with the K394E mutant, as evidenced by the structures obtained with both variants. Based on these more qualitative experiments, the K394-DNA interaction seems dispensable for cGAS binding to nucleosomes.

To dissect the contribution of K394 to cGAS-NCP complex formation more directly and to examine the role of zinc coordination, we performed additional fluorescence polarisation measurements (**Figure 4.34a-c**). We found that cGAS (K394E) binds dsDNA with an apparent affinity similar to that of WT cGAS pre-treated with EDTA to remove the zinc ion from the zinc-coordinating loop (**Figure 4.34a, b**). When WT cGAS, or WT EDTA-treated cGAS, were counter titrated with 147bp dsDNA, we found that zinc ion removal led to an over two-fold drop in  $IC_{50}$  relative to zinc-containing cGAS. A similar behaviour was observed for a nucleosomal substrate, which was bound by WT cGAS *circa* two-fold stronger than EDTA-treated cGAS (**Figure 4.34c**). We thus conclude that the zinc-coordination contributes to nucleosomal DNA binding in WT cGAS, although it is not essential. Zinc coordination plays a minor role in nucleic acid binding in the nucleosome-bound inhibited form of cGAS.



**Figure 4.34 a**, Forward titration experiments using 10 nM Flc-labelled probe in the presence of increasing amounts of either WT cGAS or cGAS pre-treated with 0.1-5 mM EDTA as indicated. **b**, Forward titration as in **a** but with cGAS K394E, either pre-treated with 0.1 mM EDTA or not. **c**, Fluorescein (Flc)-labelled 21 bp dsDNA tracer (10 nM) was mixed with cGAS protein (300 nM) pre-treated with EDTA and counter-titrated with unlabelled dsDNA or nucleosome. The data include two technical replicates and all data points are explicitly shown. Affinities are indicated as  $IC_{50}$  values. For (**a**) and (**b**), all data include three technical replicates and are shown as mean  $\pm$  s.d. Affinities are indicated as apparent  $K_d$  ( $K_{app}$ ) values. In (**c**) all replicates are explicitly shown. Adapted from <sup>363</sup>.

#### 4.2.3.5 Disruption of the cGAS-NCP interaction triggers immune activation in cells

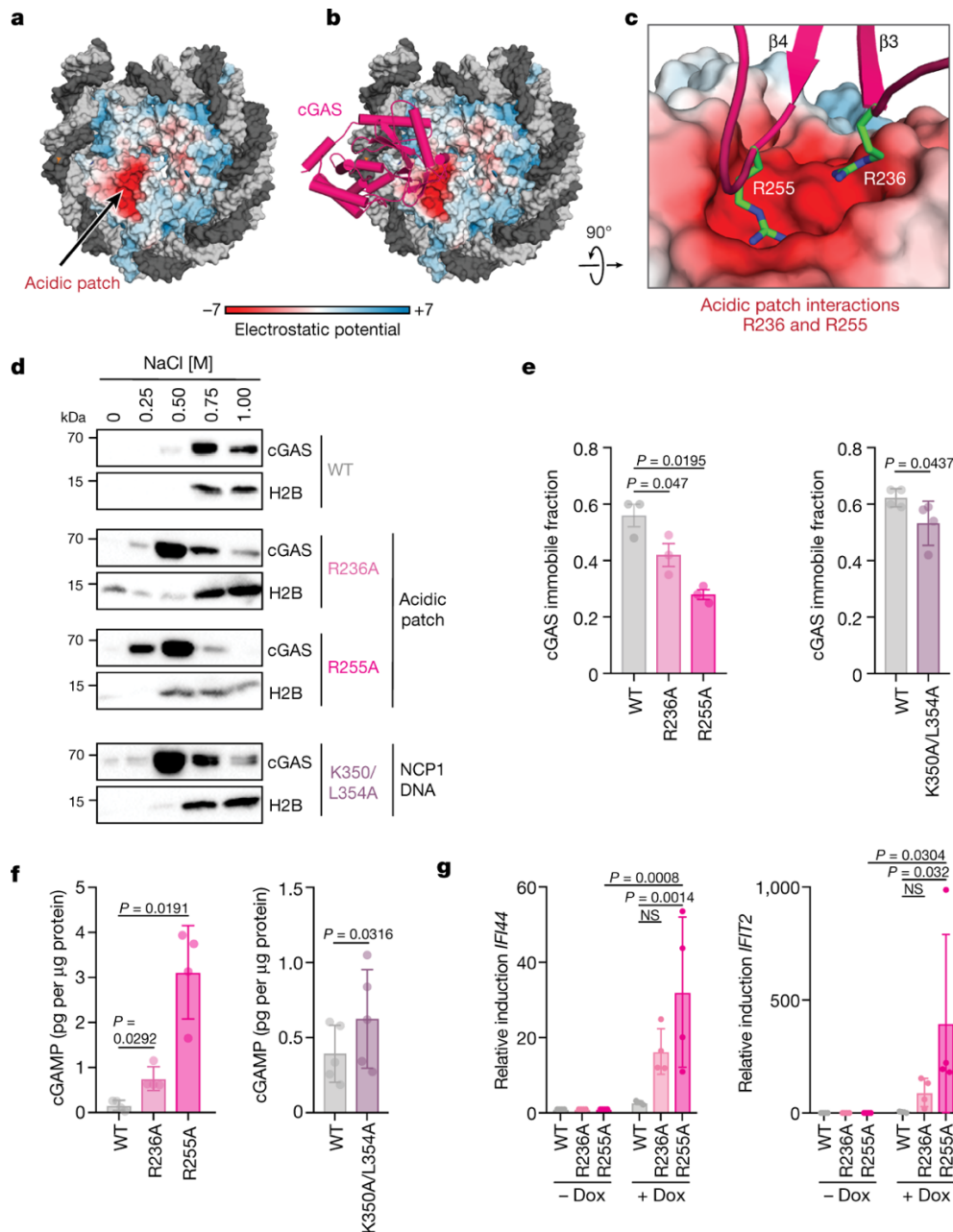
Finally, we probed the interplay between nucleosome binding and cGAS inhibition in cells, focussing on the two key *in vitro* interfaces for cGAS-nucleosome binding, namely residues contacting the acidic patch (R236 and R255; **Figure 4.35a-c**) and those involving in binding the nucleosomal DNA *in cis* (K350/L354). We found that the corresponding alanine mutants were defective in nuclear tethering when expressed in cGAS KO HeLa cells (**Figure 4.35d**). Using fluorescence recovery after photobleaching (FRAP), we observed differential



intranuclear mobility across the mutants, with the degree of dissociation correlating with the cGAMP production level (**Figure 4.35e, f**).

We also explored whether expression of the two most consequential cGAS mutants (R236A and R255A) stimulates a type I IFN response. We observed that the expression of cGAS mutants in HeLa cells induced strong upregulation of interferon-stimulated genes and WT cGAS had no such effect (**Figure 4.35g**). Hence, disrupting the interaction of cGAS with the acidic patch of nucleosomes is sufficient to trigger innate immune activation.

Overall, this study provides an explanation for why nuclear cGAS is not constitutively activated despite its strong binding to chromatin, aiding the understanding of the recognition of self vs non-self DNA by this sensor. It reveals the structural basis of cGAS inhibition by nucleosomes, with the NCP rendering cGAS unable to bind dsDNA and locking it in a catalytically inactive monomeric state. Therefore, cGAS does not recognise foreign dsDNA through any pathogen-specific features but rather it is relying on the suppressive activity of chromatin binding to avoid aberrant activation by genomic DNA. Accordingly, this work demonstrates that mutations on the cGAS-acidic patch interface that abolish cGAS-NCP interactions are sufficient to abrogate the inhibitory effect of nucleosomes *in vitro* and to unbridle cGAS activity on genomic DNA in cells.



**Figure 4.35.** **a, b**, Electrostatic surface representation of the NCP disc surface alone (**a**) or with cGAS (pink ribbon) (**b**). The electrostatic potential is shown from red ( $-7$ ) to blue ( $+7$ ) k T/e. **c**, A magnified view of contacts between cGAS and the acidic patch of the nucleosome. **d**, Differential nuclear salt fractionation probed for cGAS and H2B by immunoblot from HeLa cGAS KO cells reconstituted with doxycycline-inducible WT cGAS or cGAS mutants after 2 days of doxycycline treatment. The experiments in (**d**) were independently repeated at least three times with similar results. **e**, HeLa cGAS KO cells were transfected with WT cGAS–GFP or mutant cGAS–GFP and the immobile fraction of nuclear-localised cGAS was assessed by FRAP. Data are mean  $\pm$  s.d. of  $n = 3$  (left) and  $n = 4$  (right) independent experiments. One-way ANOVA with post-hoc Dunnett multiple comparison test (left) or two-tailed Student’s *t*-test (right). **f**, cGAMP production from HeLa cGAS KO cells reconstituted with doxycycline-inducible WT cGAS or cGAS mutants after 2 days of doxycycline treatment. Data are mean  $\pm$  s.d. of  $n = 4$  (left) and  $n = 5$  (right) independent experiments. One-way ANOVA with post-hoc Dunnett multiple comparison test (left) or two-tailed Student’s *t*-test (right). **g**, HeLa cGAS KO cells reconstituted with doxycycline (Dox)-inducible WT cGAS or cGAS mutants and treated with doxycycline for 24 h were co-cultured with BJ fibroblasts for 24 h. Cells were lysed and mRNA levels of *IFI44* (left) and *IFIT2* (right) were assessed as indicated. Data are presented as fold induction relative to cells without doxycycline and shown are the mean  $\pm$  s.d. of  $n = 4$  independent experiments. Two-way ANOVA with post-hoc Tukey multiple comparison test; NS, not significant. Individual data points are from biological replicates. Reproduced from <sup>363</sup>. *I have not contributed to these results.*

## Chapter 5 : Discussion and perspectives

### 5.1 Cyclin K degraders

Following the elucidation of the mechanism of action of IMiDs and aryl sulphonamides, many other compounds, including prevalent kinase inhibitor drugs<sup>279,368</sup>, have been suspected to have a degradation component to their mode of action. However, a clear molecular link between kinase inhibitors and the degradation machinery has remained elusive<sup>279,368</sup>. The work presented herein identified CR8, a preclinical CDK inhibitor, as a degrader of cyclin K (**section 2.1**)<sup>294</sup>. CR8 was found to be a novel type of a molecular glue degrader that binds the heterodimeric target complex CDK12-cyclin K and recruits the DDB1-CUL4-RBX1 E3 ligase core to ubiquitinate cyclin K. Through structural elucidation, CR8 was shown to bind the active site of CDK12, with a phenylpyridine moiety extending out of the pocket and into the interface, making contacts with several DDB1 residues (**Figure 2.8**)<sup>294</sup>. CDK12 is not a constitutive E3 ligase component, but instead serves as a unique drug-induced substrate receptor, linking DDB1 to the ubiquitination target and bypassing the requirement for a canonical DCAF substrate receptor. Importantly, the DDB1-CDK12 interface is large (~2100 Å<sup>2</sup>) and highly complementary, with a basal affinity of 50 μM measured between DDB1 and CDK12-cyclin K. Hence, CR8 does not *de facto* induce novel protein-protein interactions, but rather facilitates an existing low-affinity interaction and strengthens complex formation by ~1000-fold. Notably, no biological role is known for the kinase-DDB1 interaction.

Another effort to find molecular glue degraders through comparing compound toxicity between hypomodulated and wildtype cells yielded three additional cyclin K degraders, dCeMM2-4 (**section 2.2**)<sup>285</sup>. Soon afterwards, a phenotype-based screen for Nrf2 inhibitors serendipitously uncovered another small-molecular degrader of cyclin K, HQ461<sup>291</sup>. Later in 2021, a patent application from Bayer was disclosed, which encompassed a large derivative series of cyclin K/CDK12 degraders (similar in structure to CR8), although with no description of their mechanism of action<sup>305</sup>. While their mode of action has not been validated through *in vitro* reconstitutions in the original publications, our work confirmed that both HQ461 and several patent-derived compounds operate via a mechanism analogous to CR8 (**Chapter 3**).

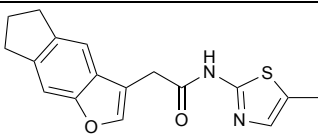
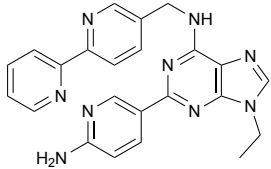
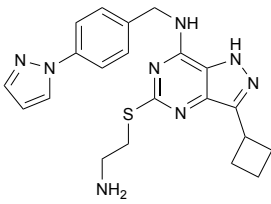
These examples of cyclin K degraders show remarkable chemical diversity, more so than any other class of molecular glues (*c.f.* **Table 1.1**). The further work reported herein evaluated almost 100 compounds and additionally identified multiple novel scaffolds that function as cyclin K degraders, including among published kinase inhibitors, and carefully surveyed the SAR of this novel molecular glue degrader class (**Chapter 3** and **Supplementary Table S3.1**). Crystallographic studies of 29 compound-induced ternary complexes revealed that despite considerable chemical diversity, the overall DDB1-CDK12-cyclin K complex architecture is largely preserved. The small molecules all contact the DDB1 Arg928, predominantly through  $\pi$ -cation interactions but also *via* hydrogen bonding and hydrophobic contacts, with many possible gluing moieties effectively fostering CDK12-DDB1 interactions. The molecular glue compound therefore provides critical binding energy by filling the kinase pocket and bridging over to engage DDB1 residues, tightening this entropy-driven (as determined by ITC, **Figure 2.10**) DDB1-CDK12 interaction over 1000-fold. This, together

with the fact that many distinct scaffolds can engage the kinase pocket and wide spectrum of gluing moieties can effectively bridge this interface, offers a rationale for the unusually broad chemical diversity observed.

While many cyclin K degraders are in fact pleiotropic kinase inhibitors, we did not observe the depletion of other cyclin or kinase targets beyond CDK12/13-cyclin K in this work. Hence, the specificity for the DDB1-CDK12/13 pairing appears to be primed not by the compound, but predominantly by the complementary protein-protein interface ( $\sim 2100 \text{ \AA}^2$ ), which is further underscored by the micromolar binding affinity measured in the absence of a small molecule<sup>294</sup>. Yet, additional compound-mediated interactions are required - not for specificity, but for sufficient stabilisation of the ternary complex for robust cyclin K ubiquitination and degradation. Nonetheless, it remains to be determined if these compounds will be reprogrammable, redirecting DDB1 recruitment within the CDK family, the kinase superfamily, or beyond.

Notably, three additional cyclin K degraders have been reported concomitantly with this work (**Table 5.1**). In late 2021, the compound NCT02 (**Table 5.1**) was identified in a screening for inhibitory compounds of colorectal cancer spheroids<sup>290</sup>. The CDK12-cyclin K complex was identified as the target through a combination of thermal proteome profiling, proteomics, and pathway analysis of transcriptomics data<sup>290</sup>. Neddylation inhibition and DDB1 knockout both inhibited cyclin K degradation upon NCT02 treatment, suggesting that NCT02 has a CR8-like mode of action<sup>290</sup>. Most recently, in 2022, two CR8-like cyclin K degraders were reported by laboratories working on novel CDK inhibitor scaffolds (**Table 5.1**)<sup>296,369</sup>. “Compound 30d” is a 2, 6, 9-trisubstituted purine inhibitor differing from CR8 only in the identity of substituents<sup>296</sup>. The study showed cyclin K depletion upon compound treatment, assessed the relevant SAR, and queried the kinase inhibitory selectivity of these dual inhibitor/degrader small molecules<sup>296</sup>. “Compound 24” is a new 3, 5, 7-trisubstituted pyrazolo[4, 3-*d*]pyrimidine derivative<sup>369</sup>. A crystal structure of this compound with CDK2-cyclin A2 was solved, and both potent pan-CDK inhibition and cyclin K degradation were demonstrated.

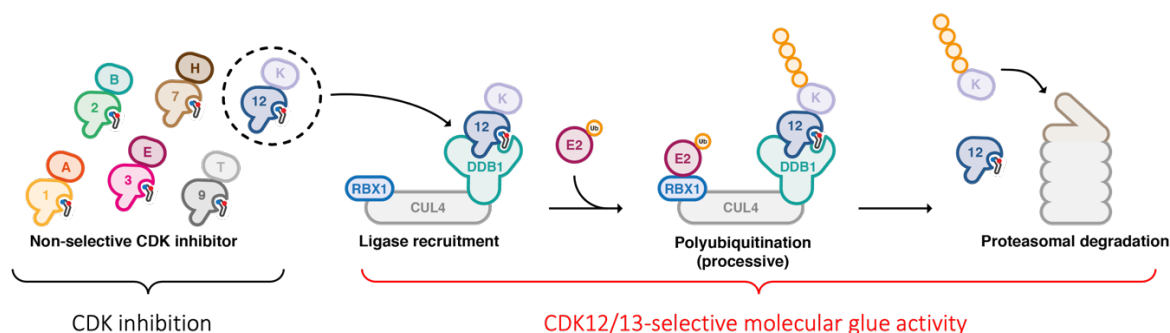
**Table 5.1.** Additional cyclin K degraders published alongside this work.

Name	Structure	Reference
NCT02		290
“Compound 30d”		296
“Compound 24”		369

## 5.2 Therapeutic utility of cyclin K degraders

CDK12-cyclin K are emerging therapeutic targets in breast cancer<sup>183,196,234,242,247</sup>, ovarian cancer<sup>201,205</sup>, Ewing sarcoma<sup>210,223</sup>, hepatocellular carcinoma<sup>254</sup>, anaplastic thyroid carcinoma<sup>203,252</sup>, colorectal cancer<sup>290</sup>, Myc-dependent tumours<sup>201,241</sup>, myotonic dystrophy<sup>212</sup>, and beyond<sup>249,250</sup>. While several CDK12/13 inhibitors have been developed in recent years, most remain multi- or pan-selective<sup>232</sup>. Here, small molecule-induced kinase inactivation that leverages specific protein-protein interactions offers a path towards improved drug selectivity. Therefore, molecular glue degraders of cyclin K are next-generation CDK12-cyclin K targeting compounds that are likely to be of high therapeutic utility. While the degrader mode of action of these compounds inherently gives rise to selectivity superior to traditional inhibitors, our work demonstrates that specificity constraints can be further dialled in by additional ligand interactions with compound-proximal residues unique to CDK12/13.

The work on CR8 provided a structural dissection of how a kinase inhibitor scaffold acquires gain-of-function molecular glue degrader properties leading to robust kinase inactivation<sup>294</sup>. The compound therefore has a unique dual kinase inhibitor/molecular glue degrader mode of action (**Figure 5.1**) – CR8 can bind to multiple CDKs and yet only binding to CDK12-cyclin K, although low-affinity itself ( $K_d = 5 \mu\text{M}$  by ITC), will generate a suitable interface for DDB1 recruitment, leading to high-affinity ternary complex formation ( $K_d < 50 \text{ nM}$  by ITC). This in turn causes cyclin K ubiquitination and its subsequent degradation by the proteasome. Compound derivatisation shows that the inhibitory versus degradative modes of action can be present in variable proportions and these properties of the molecular glue degrader compound can be tuned with chemical modifications. Notably, the broad chemical diversity can also enable the design of cyclin K degraders that differ substantially from traditional kinase inhibitors, such as *de facto* interface stabiliser drugs that show no independent kinase binding that can be developed from scaffolds such as HQ461, Z7, or Z11 (**Chapter 3**).



**Figure 5.1** CR8 has a dual kinase inhibitor/molecular glue degrader mode of action.

Notably, a CDK12/13 inhibitor with high family selectivity, SR-4835, was retrospectively identified to be a potent molecular glue degrader of cyclin K (both through the work presented herein – see **Chapter 3** – and in a recently published report<sup>290</sup>). Hence, the published studies of SR-4835 in mouse xenograft models of TNBC represent a preclinical proof of concept for cyclin K degraders in TNBC, both as single agents and in synergistic combinations with chemotherapeutics or PARP inhibitors<sup>242</sup>.

Importantly, as the transcriptional profiles of cyclin K degraders differ from those of CDK12 PROTAC degraders or CDK12 inhibitors, cyclin K degraders offer unique therapeutic opportunities. Further work, including the dissection of any biological roles of cyclin K independent of CDK12/13, will be required to ascertain under which circumstances cyclin K degradation could give rise to distinct clinical benefits beyond selective CDK12/13-cyclin K inactivation.

### 5.3 Glue-induced interfaces

Molecular glue degrader compounds (**Figure 5.2**) leverage both protein-ligand and protein-protein interfaces to achieve tight complex formation, which is conducive to robust target ubiquitination and degradation. These molecular glue degrader drugs are small (<500 Da), with ligand/neosubstrate interfaces as compact as  $\sim 70 \text{ \AA}^2$  supplemented by more extensive protein-protein contacts (**Table 5.2**). Notably, the size of the protein-protein interface is not necessarily directly proportional to the affinity of the two the proteins without the molecular glue (**Table 5.2**), as the neomorphic interface can comprise attractive as well as repulsive regions.

The learnings from the cyclin K degrader class indicate that the absolute size of the protein-protein interface, as well as the relative contribution of the small molecule to the interface determines both the SAR and the neosubstrate specificity behaviour for the molecular glue. For IMiDs and CRBN, the small protein-protein interface ( $\sim 350\text{-}700 \text{ \AA}^2$ ) to which the compound contributes *ca.* 40 % finds the ligand driving both specificity and affinity with tightly defined SAR (**Table 5.2, Figure 5.2b**)<sup>293,297,298</sup>. In contrast, for DDB1-CDK12-cyclin K, where the large ( $\sim 2100 \text{ \AA}^2$ ) protein-protein interface with a compound contribution of *ca.* 20 % (**Figure 5.2d**) allows for much greater variability in protein-ligand interactions, yielding a relatively flat SAR but limiting the compound's contribution to neosubstrate specificity<sup>293</sup>.

The complementary nature of these protein-protein interfaces suggests that the ligase and neosubstrate are likely to also interact with considerable affinity in the absence of a small molecule. Yet, the original publications evaluated the binding of these protein-protein pairs to find no detectable affinity between CRBN and CK1 $\alpha$  or ZFs in the absence of an IMiD<sup>44</sup>, low affinity ( $\sim 5 \text{ \mu M}$ )<sup>42</sup> or no detectable affinity<sup>137,142</sup> for DCAF15 binding to RBM39, and mid-micromolar affinity ( $\sim 50 \text{ \mu M}$ ) for DDB1 binding to CDK12-cyclin K<sup>294</sup>.

Notably, a study published in 2022 by the Zheng laboratory revisited these measurements for multiple glue examples using biolayer interferometry (BLI) and Alphascreen<sup>323</sup>. This work, focussed on establishing the thermodynamic signature of molecular glue degrader action, found that DDB1-CRBN and IKZF1 interact with a dissociation constant of 223 nM in the absence of an IMiD (shifting to  $K_d = 53 \text{ nM}$  with pomalidomide), while DDB1-CRBN and CK1 $\alpha$  bind with a low micromolar affinity ( $K_d = 2.3 \text{ \mu M}$  in the absence of the compound) and a  $K_d$  of 75 nM with pomalidomide. The reported differences in affinity in the presence and absence of an IMiD are remarkably small and it is unclear where the discrepancies in the biophysical measurements between this work and the original studies stem from. Cao *et al.* further turned to the auxin system (**Figure 5.2a**) and examined the interaction between TIR1 and IAA7 to reveal that TIR1-IAA7 have intrinsic micromolar affinity ( $K_d = 19 \text{ \mu M}$ ), which in the presence of auxin affords a ternary complex with a 30 nM dissociation constant.

The authors ascribe such binding behaviour in the absence of the compound as a hallmark of molecular glue degrader activity<sup>323</sup>. This is in agreement with our ITC measurements for CR8, whereby a low affinity “basal”, compound-independent binding was strengthened over 1000-fold by the molecular glue degrader<sup>294</sup>.

Lastly, the IMiDs have been described to exhibit an apparent affinity threshold, whereby few-fold lower ternary complex affinities translate to no cellular degradation<sup>113</sup>. The initial report of CR8 and the related inhibitors such as roscovitine and flavopiridol promoting DDB1-CDK12-cyclin K interactions yet only CR8 leading to degradation suggested that a similar, almost binary threshold exists for cyclin K degraders<sup>294</sup>. The thorough optimisation of the TR-FRET assay employed to accurately assess *in vitro* binding (**Figure 4.16**), as well as the SAR study spanning almost 100 compounds (**Chapter 3**), allowed us to fine-tune the description of this relationship. The evaluation of this large compound set confirmed that small differences in *in vitro* ternary complex affinity translate to larger disparities in cellular degradation, with a Hill-like equation best describing the relationship between TR-FRET EC<sub>50</sub> values and cyclin K reporter degradation DC<sub>50</sub> values (**Figure 3.13c**, **Figure 3.12c**). Although the origins of this behaviour are not clear, it is conceivable that small differences in ternary complex affinity *in vitro* are amplified *in vivo* as well as across the different stages of neosubstrate processing by the UPS.

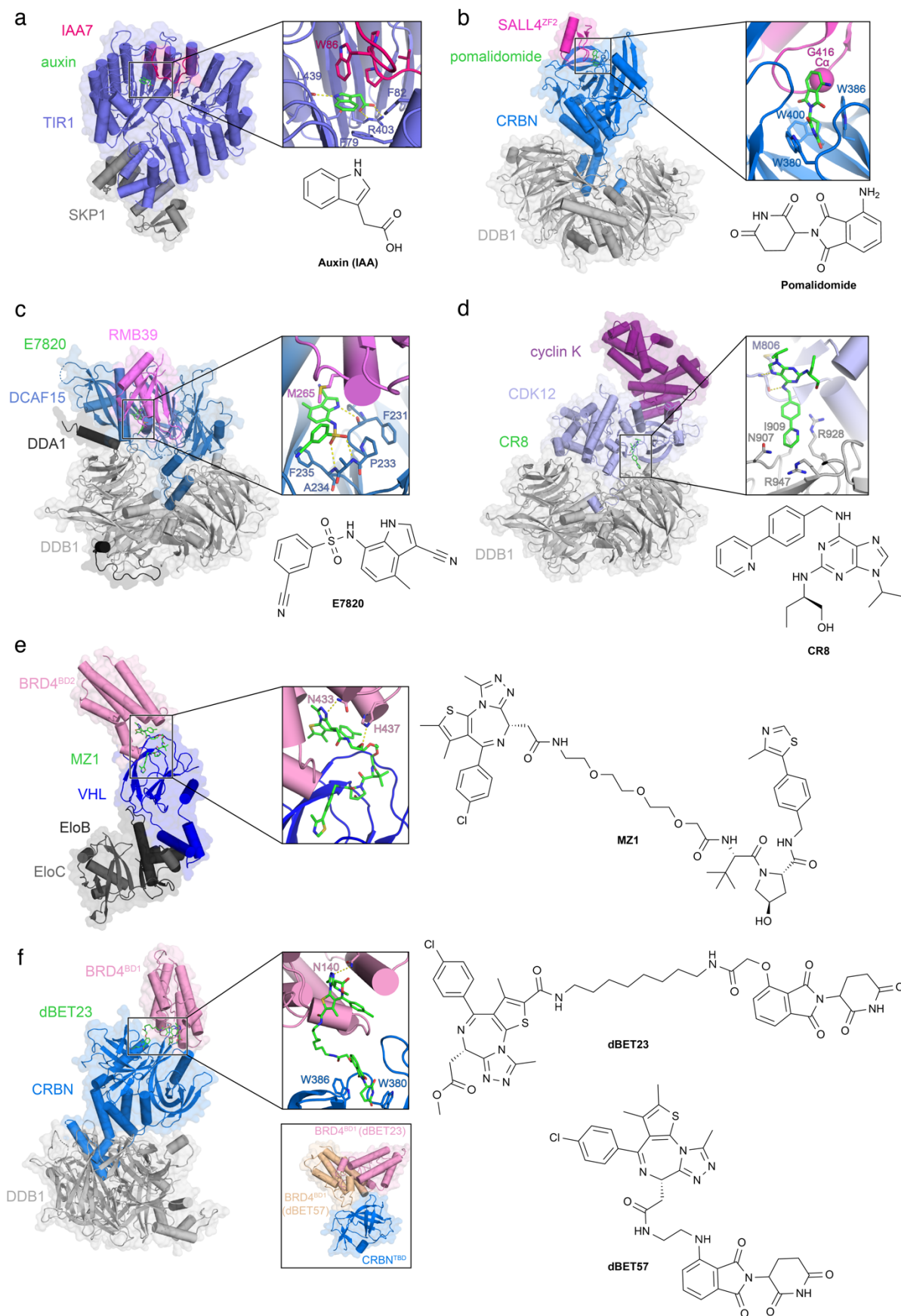
## 5.4 Degradation themes

Molecular glue degraders can be categorised based on their affinity to the two protein partners. IMiD-like ligase-binding glues are advantageous for targeting traditionally undruggable proteins such as ZF TFs, as a ligandable pocket on the target is not strictly required for ternary complex formation (**Figure 5.2b**). Glues that bind on the side of the target, such as the cyclin K degrader CR8 (**Figure 5.2d**), offer the possibility to recruit a ligase that by itself would not be considered ligandable, as is the case for DDB1. Moreover, such target-based degraders can have dual inhibitor/degrader properties potentiating the desired pharmacological response. Glues that bind neither the ligase nor the target with high affinity, such as aryl sulphonamides (**Figure 5.2c**), could efficiently bring together two seemingly non-ligandable partners. Notably, glues that act solely through interface stabilisation – akin to some low molecular weight cyclin K degraders described herein (**Chapter 3**) – could show selectivity profiles that are superior to inhibitor-derived degraders. However, high protein-protein interface complementarity is required for stable complex formation and robust ubiquitination in this case. Finally, PROTACs (**Figure 5.2e, f**) tightly bind both protein partners, which gives rise to the Hook effect but also renders the compounds strongly (albeit not exclusively) dependent on compound-mediated recruitment. This makes the warheads to some extent exchangeable for binders of other targets and ligases but also implies that PROTACs require two independently ligandable pockets. It will be important to see if, by structure-based design leveraging cooperative interactions within the ternary complex, such bifunctional molecules could be converted into molecular glue-type compounds that no longer independently engage either protein.

Conceptually, the recent examples of cyclin K degraders demonstrates that recruitment of a traditionally undruggable neosubstrate (cyclin K) to the ligase via a ligandable protein partner (CDK12) is a viable strategy.

Cyclin K is the primary ubiquitination target, and while CDK12 could become subject to autoubiquitination upon prolonged compound exposure similar to canonical DCAFs<sup>38,106</sup>, data show that the observed destabilisation of CDK12 is most likely due to cyclin deprivation<sup>291</sup>. Thus, it is the overall geometry of the target-ligase complex that governs which components can be efficiently ubiquitinated and degraded. These spatial considerations are also likely at play for PROTACs, as compounds targeting the SMARCA2/4 subunits of the SWI/SNF complex have been shown to only directly degrade their target<sup>370</sup>, while heterobifunctionals binding the EED component of PRC2 have been reported to additionally target other subunits for degradation<sup>371</sup>.





**Figure 5.2. Overview of degrader-induced interfaces.** **a**, SKP1-TIR1-auxin-IAA7 complex (PDB 2P1Q)<sup>85</sup>. **b**, DDB1-CRBN-pomalidomide-SALL4<sup>ZF2</sup> complex (PDB 6UML)<sup>115</sup>. The C<sub>α</sub> of G416 in the β-hairpin loop of SALL4<sup>ZF2</sup> is shown as a sphere. **c**, DDB1-DDA1-DCAF15-E7820-RBM39 complex (PDB 6PAI)<sup>42</sup>. **d**, DDB1-CR8-CDK12-cyclin K complex

(PDB 6TD3)<sup>294</sup>. **e**, EloB/C-VHL-MZ1-BRD4<sup>BD2</sup> complex (PDB 5T35)<sup>90</sup>. **f**, DDB1-CRBN-dBET23-BRD4<sup>BD1</sup> complex (PDB 6BN7)<sup>102</sup>. Bottom panel: overlay of dBET23 and dBET57-induced conformations of BRD4<sup>BD1</sup> (PDB 6BNB) aligned on CRBN thalidomide-binding domain (TBD). **a-f**, Important interface residues are shown as sticks. Key interactions are marked with a dashed yellow line. (**c,f**), The BPB domain of DDB1 was omitted for clarity.

**Table 5.2.** Comparison of structurally characterised drug-induced protein-protein interfaces for small-molecular degraders. Interface areas were calculated with PISA<sup>372</sup>. For indisulam and E7820, average values from the two available structures are shown. For cyclin K degraders, only four representative examples are shown.

Compound	Ligase	Target	PDB	Protein-protein interface [%]	Ligase-ligand interface [%]	Target-ligand interface [%]	Total interface area [Å <sup>2</sup> ]
Auxin	SKP1 <sup>TIR1</sup>	IAA7 peptide	2P1Q <sup>85</sup>	71	22	7	926
Lenalidomide	CRL4 <sup>CRBN</sup>	CK1α	F5QD <sup>44</sup>	68	23	9	1052
Pomalidomide	CRL4 <sup>CRBN</sup>	SALL4 (ZF2)	6UML <sup>115</sup>	52	35	13	681
CC-885	CRL4 <sup>CRBN</sup>	GSPT1	5HXB <sup>112</sup>	52	28	20	1164
CC-90009	CRL4 <sup>CRBN</sup>	GSPT1	6XK9 <sup>122</sup>	54	27	19	1221
Pomalidomide	CRL4 <sup>CRBN</sup>	IKZF1 (ZF2)	6H0F <sup>113</sup>	61	28	12	859
Pomalidomide	CRL4 <sup>CRBN</sup>	ZNF692 (ZF4)	6H0G <sup>113</sup>	61	28	10	880
Thalidomide	CRL4 <sup>CRBN</sup>	SALL4 (ZF2)	7BQU <sup>114</sup>	56	32	13	706
5-Hydroxythalidomide	CRL4 <sup>CRBN</sup>	SALL4 (ZF2)	7BQV <sup>114</sup>	60	29	12	802
E7820	CRL4 <sup>DCAF15</sup>	RBM39	6PAI <sup>42</sup> , 6Q0R <sup>137</sup>	68	22	10	1463
Indisulam	CRL4 <sup>DCAF15</sup>	RBM39	6UD7 <sup>142</sup> , 6Q0W <sup>137</sup>	72	19	9	1665
CR8	CRL4 <sup>Δ</sup>	Cyclin K	6TD3 <sup>294</sup>	79	6	16	2672
21195	CRL4 <sup>Δ</sup>	Cyclin K	this work	80	5	15	2730
919278	CRL4 <sup>Δ</sup>	Cyclin K	this work	84	5	11	2625
HQ461	CRL4 <sup>Δ</sup>	Cyclin K	this work	83	6	11	2646
NRX-103094	SKP1 <sup>β-TrCP</sup>	β-catenin peptide	6M91 <sup>373</sup>	54	25	21	958
MZ1	CRL2 <sup>VHL</sup>	BRD4 (BD2)	5T35 <sup>90</sup>	26	37	38	1338
dBET23	CRL4 <sup>CRBN</sup>	BRD4 (BD1)	6BN7 <sup>102</sup>	39	23	38	1328

## 5.5 Screening for molecular glue degraders

The IMiD, aryl sulphonamide and BI-3802 molecular glues have been identified as ligase binders and/or target degraders by observant research teams. The efforts that followed these initial discoveries focused primarily on screening derivatives of known degraders in the hope of recruiting novel neosubstrates. However, a broad search for binders of other ubiquitin ligases was also pursued in the hope of finding the “new CRBN”, driven on one hand by wanting to expand the repertoire of ligases available for PROTAC design and on the other by the quest for a family of molecular glues as clinically successful as IMiDs. While a VHL-based molecular glue degrader has recently been reported (with details pending disclosure)<sup>152</sup>, it remains to be seen if other ligandable ligases, such as cIAP1, or ones with no described ligands, can recruit and ubiquitinate other proteins upon the binding of monovalent molecular glue degrader compounds.

In addition to these ligase-oriented strategies, two systematic ligase- and target-agnostic approaches for degrader discovery have been reported, and both are described herein. A bioinformatic correlation between drug toxicity and ubiquitin ligase mRNA expression levels across different cancer cell lines allowed for the identification of the CR8-DDB1 drug-ligase pair (**section 2.1**)<sup>294</sup>, while a screen looking at differential cytotoxicity of compounds in wild-type versus hyponeddylated cellular models found one RBM39 degrader and three cyclin K degraders (**section 2.2**)<sup>285</sup>. Both studies used relatively small compound libraries (4518 pre-clinical and clinical compounds and ~2000 cytostatic and cytotoxic small molecules, respectively) and focused solely on cell survival, yielding surprisingly high hit rates. Many more compounds can likely be found by taking advantage of larger existing libraries. Moreover, utilising more complex read-outs beyond cell death would also enable the discovery of molecular glue degraders of proteins that are not essential in the chosen cellular context. Once a molecular glue degrader candidate is identified, a pipeline comprising mass-spectrometry-assisted identification of the degradation target(s), ligase identification using CRISPR screening technology, and biochemical reconstitutions can allow the elucidation of its detailed mechanism of action<sup>146,285,294</sup>.

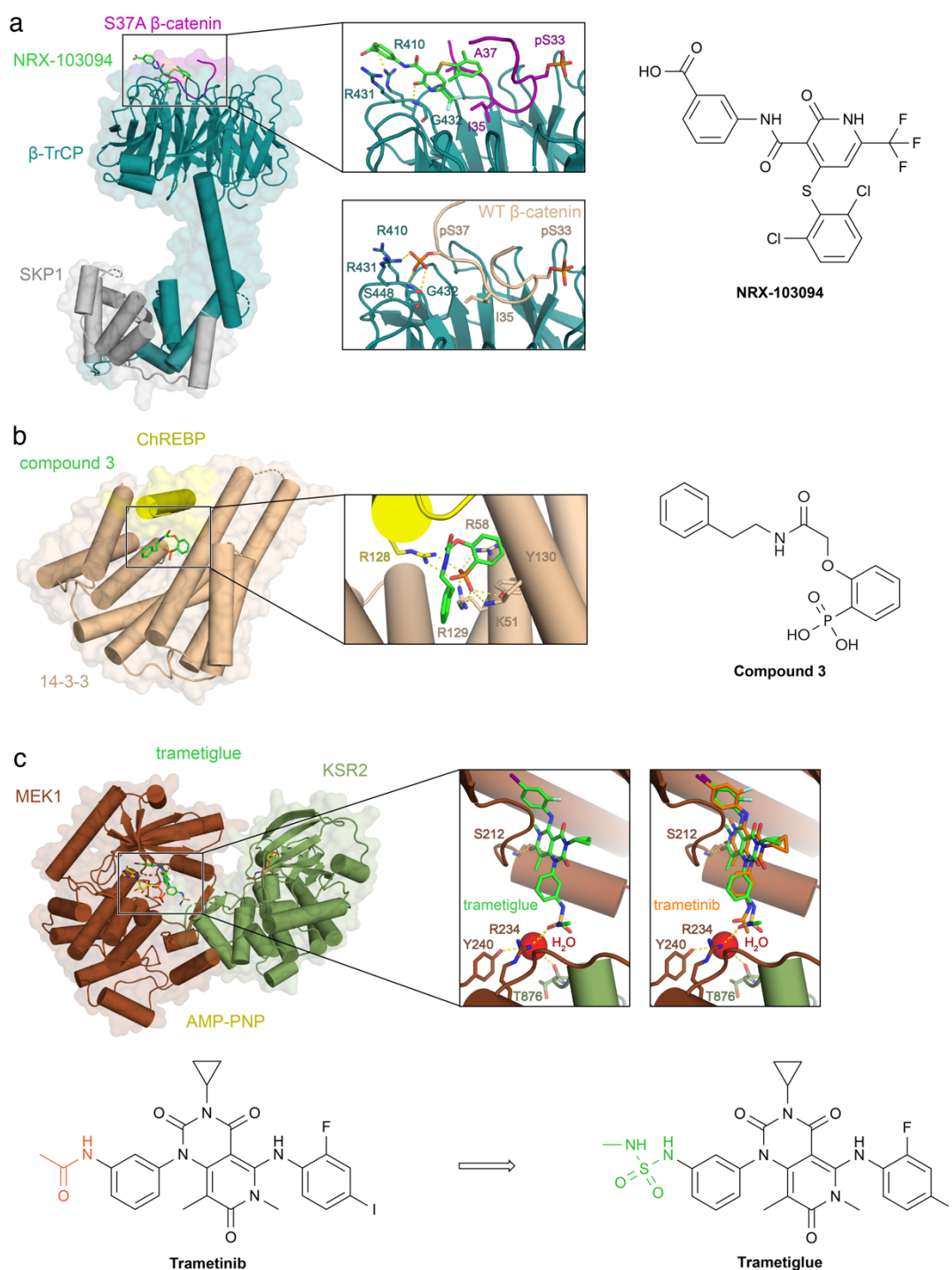
Once identified, glues and their derivatives can be evaluated using various *in vitro* methods – TR-FRET, ITC, pulldowns, ubiquitination assays, structural studies – and cellular approaches, such as reporter stability and viability screens. Our work showed that for cyclin K degraders, the *in vitro* ternary complex affinity is largely predictive of in-cell degradation for the large compound set (**Figure 3.13c**, **Figure 3.12c**). Moreover, this study further demonstrates that screening for compounds’ cytotoxicity in the presence or absence of MLN4924 is a feasible and informative strategy to identify degraders (**Figure 3.12d**). Akin to compound screening in wild-type versus hyponeddylated cellular models<sup>285,294</sup>, such an approach can help infer the contribution of a (CRL-mediated) degradation phenotype to the compound’s mode of action.

## 5.6 Rational design for gluing weak interactors

If both protein-protein and protein-ligand interactions contribute to molecular glue action, molecular glue degraders may be easier to obtain by leveraging existing low-affinity protein-protein interactions. For instance, a protein epitope that has been weakened by mutations and is no longer recognised efficiently by its cognate

ligase may be “repaired” by a molecular glue, as in the case of the mutated oncogenic transcription factor  $\beta$ -catenin, which has lost its affinity for the cognate SCF <sup>$\beta$ -TrCP</sup> E3 ligase<sup>104</sup> (**Figure 5.3a**). Through HTS followed by structure-based design, protein-protein interaction stabilisers were obtained that complemented the mutated interface by providing the binding surface between  $\beta$ -catenin and  $\beta$ -TrCP that is lost as a result of the S37A  $\beta$ -catenin mutation. This compound restored the interaction between SCF <sup>$\beta$ -TrCP</sup> and mutant  $\beta$ -catenin and allowed for efficient ubiquitination *in vitro*<sup>104</sup>.

Although not a molecular glue degrader, a similar strategy was used to develop a glue compound that furthers the interaction between 14-3-3 and its phosphorylation-independent interactor carbohydrate-response element-binding protein (ChREBP), where a weak stabiliser derived from an *in silico* screening campaign was optimised into an effective molecular glue in a structure-guided manner<sup>374</sup> (**Figure 5.3b**). Another example of rational glue design is that of a derivative of the kinase inhibitor trametinib, which targets the mitogen-activated protein kinase (MEK) kinase<sup>375</sup> (**Figure 5.3c**). The inhibitor was found to bind at the interface between the kinase domain of MEK and the kinase suppressor of Ras (KSR) subunit. Given the proximity between the two proteins, a “trametiglue” was designed by substituting an acetamide group for a sulphamide to enhance interfacial binding. This yielded a compound with improved cellular potency, offering a path for selectively targeting sub-populations of MEK. Several other compounds that stabilize weak protein-protein interactions have been reported, such as CC0651 that strengthens the ubiquitin interaction with the E2 ubiquitin-conjugating enzyme CDC34<sup>376</sup>, and the natural product Brefeldin A, which stabilises the ARF-GDP-Sec7 complex<sup>377,378</sup>. Thus, leveraging existing weak interactions, particularly in the context of ubiquitin ligases, offers an exciting route for the rational design of molecular glue degraders.



**Figure 5.3. Rational design of chemical inducers of proximity.** **a**, SKP1-β-TrCP-NRX-103094-β-catenin(S37A) complex (PDB 6M91)<sup>373</sup>. Bottom panel: wild-type (WT) β-catenin peptide bound to β-TrCP (PDB 1P22)<sup>379</sup>. **b**, 14-3-3 monomer bound to compound 3 and ChREBP α2 peptide (PDB 6YGJ)<sup>374</sup>. **c**, MEK1-trametinib-KSR2 complex (PDB 7JUV)<sup>375</sup>. Right panel: overlay of trametinib and the parent inhibitor trametigluce bound at the MEK1-KSR2 interface (PDB 7JUR) through superposition of kinase domains<sup>375</sup>. **a-c**, Important interface residues are shown as sticks. Key hydrogen bonds are indicated with dashed yellow lines.

## 5.7 The “protein pair problem”

While molecular glue degraders may be obtained through leveraging existing low-affinity protein-protein interactions, such suitably predisposed pairs are challenging to identify beyond isolated examples or established degrons. This constitutes what may be referred to as the “protein pair problem”.

Computational methods, such as protein-protein docking<sup>380</sup> or fingerprint-based approaches<sup>381</sup>, could be used to shortlist candidate pairs. However, screening large swathes of the proteome against each other, or even screening a single target of interest against a panel of ubiquitin ligase components is an extremely computationally expensive task. It is further hindered by the lack of structural information on many ubiquitin ligases (an obstacle only partly alleviated by the advent of AlphaFold2<sup>382</sup>), the bias of published ligand-bound or ternary complex structures, and the inherent conformational flexibility<sup>383</sup> of the putative interactors. *In vitro* experimental approaches are also limited, as the detection of protein-protein interactions at scale, especially ones with micro- and millimolar affinities, remains highly challenging. Additionally, the question of whether a given interaction is likely to prove consequential when strengthened in affinity by a small molecule is difficult to address. This will depend on the geometry and kinetics of the interaction but also many factors only at play in the relevant cellular context, such as component localisation, post-translational modifications, or competing interactions of the molecules of interest. These obstacles are largely alleviated when modulating an endogenous ligase-target pair. Actionable native ligase-target pairs, *e.g.* ones with mutationally impaired affinity akin to the above-mentioned  $\beta$ -catenin example<sup>373</sup>, could be identified for instance based on degron loss data<sup>384,385</sup>.

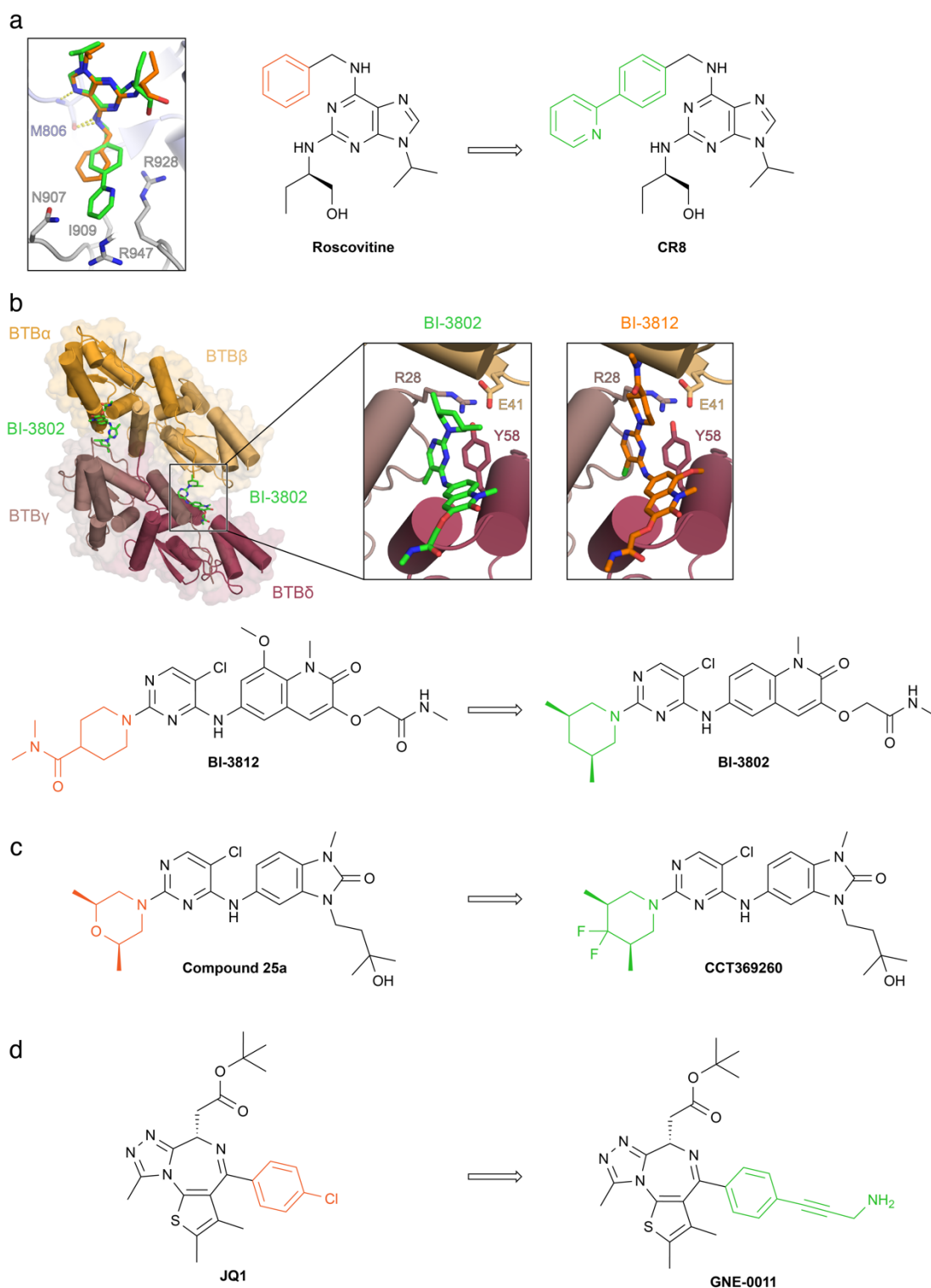
Recently, a preprint from Mikko Taipale’s laboratory described an interesting proteome-scale induced proximity screen aiming to identify actionable effectors<sup>386</sup>. The approach<sup>386</sup> was based on the co-expression of a target and an effector protein in cells with distinct sets of dimerization tags – either PYL1 and ABI1, which interact in the presence of the small molecule abscisic acid, or the vhhGFP nanobody and GFP, which interact constitutively<sup>386</sup>. The study focussed predominantly on the degradation or stabilisation of a mock substrate, a GFP reporter (arrayed screens with 280 E3 ligases or 47 DUBs, respectively), with the effects of tag-induced dimerization monitored by flow cytometry. While this approach, when applied to a target of interest, is likely to be the most relevant for fully neomorphic, PROTAC-induced complexes, important learnings can also be drawn for prospective molecular glue degraders design.

Alternative methods employing unassisted protein-protein dimerization, such as protein-fragment complementation assays (*e.g.* the yeast-two hybrid system<sup>387</sup> or split DHFR-based assays<sup>388</sup>) could also be used in this context. With the target of interest as the bait, such experiments would allow the (indirect) survey of its binding preferences for a library of prey ligase components, yielding a catalogue of preferred pairings. As these complementation assays do not utilise degradation as a readout, they could also be employed for other classes of induced-proximity effectors beyond the UPS. However, the challenge that persists is the effective detection of low-affinity (high micromolar and millimolar) interactions.

## 5.8 Compound-induced stickiness

If the discussed molecular glue degraders indeed induce new interactions with no physiological relevance, how do small molecules give rise to such neomorphic binding? Several recent examples such as CR8 or BI-3802 (**Figure 5.4**) demonstrate that the addition of solvent-facing small chemical groups to a binder or inhibitor can confer gain-of-function molecular glue degrader functionality. Evolutionary studies show that proteins have an intrinsic tendency to form higher-order complexes, and that random single-point mutations give rise to *de novo* interfaces that induce larger assemblies with surprising ease<sup>389</sup>. This is also evident in genetic diseases such as sickle-cell anaemia where a surface mutation in haemoglobin (E6V mutant also referred to as haemoglobin S) triggers fibre formation<sup>282</sup>. The BI-3802 (**Figure 5.4b**, see also **section 1.1.12**) and CR8 examples now argue that small molecules bound to a protein surface can give rise to similar aggregation processes, either as self-polymerisation events (as in BCL6) or by strengthening an existing, albeit weak, interaction of DDB1 with CDK12-cyclin K (**Figure 5.4**). The model formulated by Levy and colleagues stipulates that proteins evolve on the cusp of engaging in macromolecular interactions<sup>389</sup>. The very low affinities are not by themselves directly biologically actionable but such propensities to interact are exploitable by compound-mediated modification of protein surfaces. Across evolution, hydrophobic patches have been suggested to persist as means to favour functionally relevant multimerization<sup>390</sup>. Surface-bound small molecules could thus similarly conjure a sufficiently hydrophobic hotspot to either lead to target aggregation or induce higher-order assemblies. In addition, alterations in the protein surface induced by single nucleotide polymorphisms (SNPs) in the coding region of proteins may already intrinsically affect interactomes of the protein carrying the SNP, and molecular glue compounds could further leverage such sequence variation.

Experimentally, the formation of higher-order assemblies is most easily detectable if ubiquitin ligases are involved, giving rise to target degradation. Many more compounds could alter target interactomes more subtly, for instance through changes in post-translational modifications or localisation of the target protein<sup>280</sup>. Proof-of-principle heterobifunctional compounds that bring together neosubstrates and DUBs, kinases, phosphatases, or acetyltransferases have been reported<sup>391–394</sup>, and it is conceivable that monovalent compounds could similarly redirect the enzymatic activity of other effectors towards new substrates. It is likely, however, that compound-induced aggregation phenomena akin to the behaviour of BI-3802 are even more common and that this behaviour has thus far been largely overlooked.



**Figure 5.4. Modification of protein surface by derivatised binders drives protein-protein interactions.** **a**, Overlay of CR8 (PDB 6TD3) and roscovitine (PDB 2A4L) in the active site of CDK12 in the DDB1-CR8-CDK12-cyclin K complex through superposition of kinase domains<sup>294</sup>. **b**, Structure of two BTB dimers brought together by the degrader BI-3802<sup>146</sup>. Left close-up panel: BI-3802 at the dimer interface. BCL6 residues R28 (BTB $\gamma$ ), E41 (BTB $\beta$ ), Y58 (BTB $\delta$ ) are shown as sticks. Right close-up panel: model of *bona fide* inhibitor BI-3812 at the dimer interface. The inhibitor was docked to the crystal structure of BCL6 BTB (PDB 5MW2) and was modelled at the interface by superposition of BTB domains. The solvent-exposed moiety of BI-3812 clashes with a helix of BTB $\beta$ . **c**, Another inhibitor-degrader pair identified for BCL6<sup>395</sup>. **d**, Inhibitor and degrader of BRD4<sup>396</sup>. The mechanism of action of GNE-0011 is not known.



This hitherto hidden, induced proximity-based mode of action is therefore likely to be more common among existing small molecules than was previously thought. Hence, molecular glue-driven binding events should be taken into consideration when developing drugs and the effect of compound binding on target stability should be routinely evaluated for candidate inhibitors.

Taken one step further, it remains to be seen whether derivatisation of the solvent-facing regions in small molecular binders and inhibitors can be used to systematically induce neomorphic interactions and thereby perturb the target of interest in a forward screening approach. Should this prove to be a viable approach, it could allow for routine target-centric degrader screening, whereby libraries of compound are screened for target destabilisation in a ligase-agnostic manner. This could not only pave the way to the (semi-)rational design of molecular glue degraders for a given target, but also allow for seemingly non-ligandable ligases to be identified as viable effectors for targeted protein degradation.

## 5.9 Compound-induced interfaces

Molecular recognition events between two proteins, or between a protein and a small molecule, are at the heart of every biological process. Through decades of medicinal chemistry efforts, we have attained a good understanding of how to find, design, and optimise protein-ligand interactions<sup>397,398</sup>. Through structural analysis of protein-protein interactions by crystallography and, more recently, cryo-electron microscopy, we have amassed a multitude of examples of protein-protein interfaces to learn from, allowing for recent advances in predicting protein-protein interactions<sup>399-401</sup>. Still lacking, though, is a more systematic understanding of how small molecules can remodel, or repurpose, protein-protein interfaces. In such an understanding lies the key to rationally designing molecular glue-triggered events.

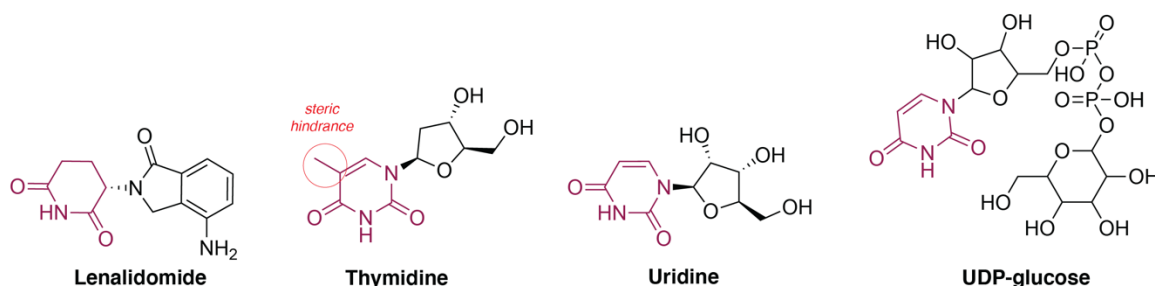
Leveraging compound-induced interactions holds enormous potential for the discovery of novel drugs. However, our thinking of interfaces in drug discovery is largely dominated by high affinity, high-specificity binding events. These lock-and-key interactions, as originally described by Emil Fischer<sup>402</sup>, are highly relevant for small protein-ligand interfaces. However, interfaces that display lower affinity are equally important in induced proximity. The best indication that tight interactions in one part of a complex can prompt protein-protein interactions elsewhere comes from PROTAC-induced complexes. The structure of the MZ1 ternary complex found BRD4 engaged with the VHL E3 ligase receptor through neomorphic protein-protein interactions<sup>90</sup> (**Figure 5.2e**). PROTACs recruiting BRD4 to CRBN showed different BD1 interfaces engaged with CRBN depending on the linker and exit vector used, with the biochemical validation of these binding modes suggesting genuinely different interactions<sup>102</sup> (**Figure 5.2f**). CRBN complexes with different neosubstrates (CK1 $\alpha$  and GSPT1) also utilised distinct regions on the receptor surface<sup>44,112</sup>. This indicates that two proteins, once brought in proximity through a high-affinity interaction elsewhere, will engage in protein-protein interactions, and that these contacts have different levels of complementarity ranging from being strongly attractive to being repulsive<sup>90,100,102</sup>. These interactions are sufficiently common that they can be systematically harnessed using small molecules.

Still, a quantitative and computational framework is currently missing that would allow us to dissect and understand such compound-mediated protein-protein binding events. New examples, such as the large cyclin K degrader class, and a detailed understanding of the interfaces involved considerably further our understanding of glue-dependent events. However, more studies are necessary to establish a rational strategy for routine molecular glue design, and therefore accelerate the development this powerful new class of medicines.

## 5.10 Endogenous molecular glues

Endogenous compounds including nuclear hormones and inositol phosphates have been postulated to act as degraders<sup>403</sup> or inducers of proximity<sup>404–406</sup>, respectively. To my knowledge, however, no endogenous molecular glue degrader compounds akin to plant hormones have yet been conclusively identified in humans. CRBN and its neosubstrates, as well as DCAF15 and RBM39/RBM23 are likely neomorphic substrate-receptor pairs given their biology, mode of action and the limited interface conservation. Yet, especially in the case of the highly conserved tri-tryptophan pocket on CRBN, it is difficult to exclude that endogenous small molecules exist that bind and reprogram the ligase. What is more, considering the recent report of low-micromolar affinity binding between DDB1-CRBN and IKZF1 or CK1 $\alpha$  detected through BLI and AlphaScreen measurements<sup>323</sup>, it cannot be discounted that such compound-independent interactions could become relevant in a specific cellular context.

Uridine, an endogenous and universally present compound, has been previously shown to bind CRBN<sup>407</sup>. Two pyrimidine nucleobases - thymine and uracil - show obvious structural similarity to the IMiD glutarimide moiety (**Figure 5.5**) and NMR experiments demonstrated that uracil and its nucleosides bind CRBN with a low-micromolar affinity, while thymidine only showed binding constants in the millimolar range<sup>407</sup>. In 2022, Cho *et al.* reported that UDP-glucose is a molecular glue degrader able to bring together CRBN and glucokinase and lead to steady-state degradation of glucokinase, a key glycolysis enzyme<sup>408</sup>. While the interaction was not reconstituted *in vitro*, the knockout of the enzyme solely responsible for cellular UDP-glucose production was shown to lead to glucokinase accumulation and UDP-glucose administration caused glucokinase depletion<sup>408</sup>. Further work, spanning both biochemical and structural studies but also the elucidation of biological significance of this interaction will be required.

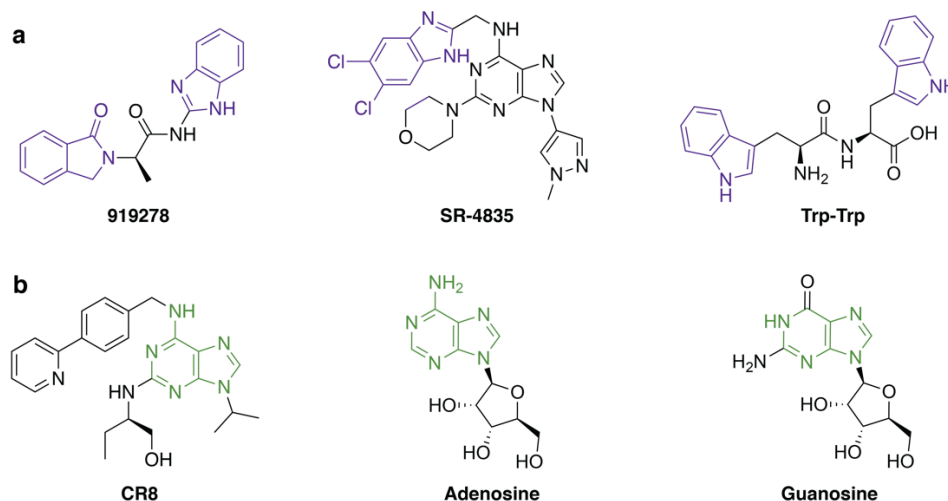


**Figure 5.5.** Uridine and UDP-glucose are endogenous CRBN binders. The glutarimide ring of lenalidomide and similar moieties in the pyrimidine derivatives are shown in purple. Uridine nucleosides bind CRBN with low-micromolar affinities,

while thymidine nucleosides show binding constants in the millimolar range, likely due to steric hindrance of the additional methyl substituent. UDP-glucose was postulated to be a CRBN-based molecular glue degrader of glucokinase.

In case of the DDB1-CDK12/13-cyclin K interaction, it is unclear whether these low-affinity binding events have a physiological role in the absence of an interaction-promoting compound. While data show that cyclin K is not a substrate of DDB1 or CUL4B under steady-state conditions<sup>285,294</sup>, it is conceivable that endogenous ligands could promote this interaction upon a specific stimulus or in a developmentally regulated manner. We showed that native nucleotide cofactors (such as ATP) do not promote this interaction. However, the kinase pocket is in principle able to bind many different ligands and the striking diversity of the identified cyclin K degraders suggests that endogenous small molecules that facilitate this interaction could exist. One intriguing possibility is that of a metabolite able to promote cyclin K degradation.

Metabolites bind to various proteins to regulate a myriad of cellular processes, yet protein-metabolite interactions are understudied. As new methods emerge for capturing such interactions, various previously understudied compound classes emerge as potential regulators<sup>409</sup>. For example, the Ser-Leu dipeptide was recently found to weakly bind phosphoglycerate kinase in *Saccharomyces cerevisiae* and therefore modulate its activity<sup>409</sup>. Multiple dipeptides bear chemical similarity to cyclin K degraders, e.g. 919278 and the gluing moiety of SR-4835 somewhat resemble tryptophan side chains (**Figure 5.6a**), making them an interesting candidate class. However, many other possibilities exist, including modified nucleosides or other nucleobase-containing metabolites. This would echo the purine scaffold seen in CR8, SR-4835 and other cyclin K degraders (**Figure 5.6b**), as well as the role of CDK12 in DNA damage response, where the kinase could additionally serve as a ligand-dependent sensor under certain circumstances.



**Figure 5.6. Broad chemical similarity of cyclin K degraders to various endogenous compound classes. a,** Short peptides, for example the Trp-Trp dipeptide, bear some resemblance to multiple identified degraders, for example the 919278 compound and the gluing moiety of SR-4835. **b,** The core scaffold of CR8 is the nucleobase adenine.

Further work will be required to address this exciting yet exceedingly challenging question. Approaches ranging from virtual screening to *in vitro* experiments using the highly optimised TR-FRET screen developed

herein to cellular assays could be employed. However, a central difficulty will be the access to suitable metabolome libraries, as published databases contain predominantly non-endogenous metabolites (such as metabolites of xenobiotics) while many listed endogenous compounds are based solely on predictions<sup>410</sup>. In the case of looking for an interaction only occurring upon a specific stimulus, stress signal, in development or in disease, the search will certainly have to encompass metabolites outside of the well-annotated, commercially available libraries containing a few hundred compounds.

## 5.11 Significance

The principal aim of this thesis was to further the understanding of how molecular glue degraders reprogram protein-protein interfaces to lead to selective target degradation. Accordingly, the work presented herein uncovered a novel class of molecular glue degraders and established their mechanism of action, visualised their binding modes, and carefully studied the SAR through the evaluation of *ca.* 100 small molecules. This now establishes cyclin K degraders as a third well-described class in the molecular glue degrader repertoire, alongside thalidomide analogues and aryl sulphonamides.

These studies additionally yield important insights into ubiquitin ligase biology, showing that DDB1 can interact with a drug-induced “neosubstrate receptor” CDK12 to degrade cyclin K, while bypassing a canonical DCAF. The SAR dissection reveals that many highly diverse compounds can induce the DDB1-CDK12 interaction, highlighting the importance of the compound’s gluing moiety productively engaging the DDB1 Arg928 residue but also the extensive and omnipresent protein-protein contacts that determine the geometry of this interaction. This work not only informs on the fundamental properties of compound-induced interfaces, but also defines a novel class of CDK12/13-cyclin K targeting molecules, evaluates multiple actionable scaffolds, and conceptualises several optimisation strategies that will enable the clinical pursuit of these important therapeutic targets.

A thorough optimisation of an *in vitro* TR-FRET ternary complex formation assay is also described, with the final experimental setup allowing for the dissection of small differences in the gluing ability of closely related compounds. Such *in vitro* assays are indispensable in molecular glue degrader development and the steps described herein could be employed in the development of similar assay systems for other protein pairs. Our results show that in the case of cyclin K degraders, the *in vitro* ternary complex affinity is largely predictive of in-cell degradation, further supporting the relevance of such *in vitro* assays.

This thesis also describes two novel methods for the prospective, systematic discovery of molecular glue degraders – through bioinformatic correlation of drug toxicity and ubiquitin ligase expression or through differential compound profiling in wild-type versus hypomodulated cell lines. What is more, the learnings from the cyclin K degrader class further indicate that degraders can be obtained through screening for target destabilisation using libraries of suitably modified inhibitors, offering the first semi-rational route towards molecular glue degrader discovery for a given target. The studies described herein therefore constitute an important advance in the understanding of the discovery and design rules for molecular glue degraders.

Finally, this work underscores that molecular glue properties are likely already present in many existing drugs, and the solvent-exposed moieties of such molecules determine their gain-of-function behaviour. It also illustrates that compounds can bring two proteins together with surprising ease and that this phenomenon could be exploited for the development of effective therapeutics.

## Chapter 6 : Materials and methods

### 6.1 General materials and methods

#### 6.1.1 Materials

**Table 6.1.** Cell lines.

Cell line	Organism	Reference
SF-9	<i>Spodoptera frugiperda</i>	Invitrogen™
High-Five™ BTI-TN-5B1-4 (Hi5)	<i>Trichoplusia ni</i>	Invitrogen™
DH5α	<i>Escherichia coli</i>	Thermo Fischer
HEK293T	<i>Homo sapiens</i>	Genetic Perturbation Platform, Broad Institute
MDA-MB-231	<i>Homo sapiens</i>	ATCC

**Table 6.2.** Plasmids.

Plasmid	Description	Reference
pAC8-dsRed	Derived from pBacPAK8, including dsRed as expression level marker	<sup>411</sup>
	Baculovirus expression system: flashBAC™	Oxford Expression Technology

**Table 6.3.** Various reagents

Reagent	Source
Cellfectin™ II (cationic-lipid reagent for insect cell transfection; cat. no. <i>10362100</i> )	Thermo Fisher Scientific
SF-4 Baculo Express ICM (serum-free Insect Culture Medium)	BioConcept
Fetal bovine serum (FBS)	Thermo Fisher
Alexa <sub>488</sub> -C5-maleimide ( <i>A10254</i> )	Thermo Fischer
Alexa <sub>647</sub> -C2-maleimide ( <i>A20347</i> )	Thermo Fischer
Cy5 maleimide ( <i>GEPA15130</i> )	Merck

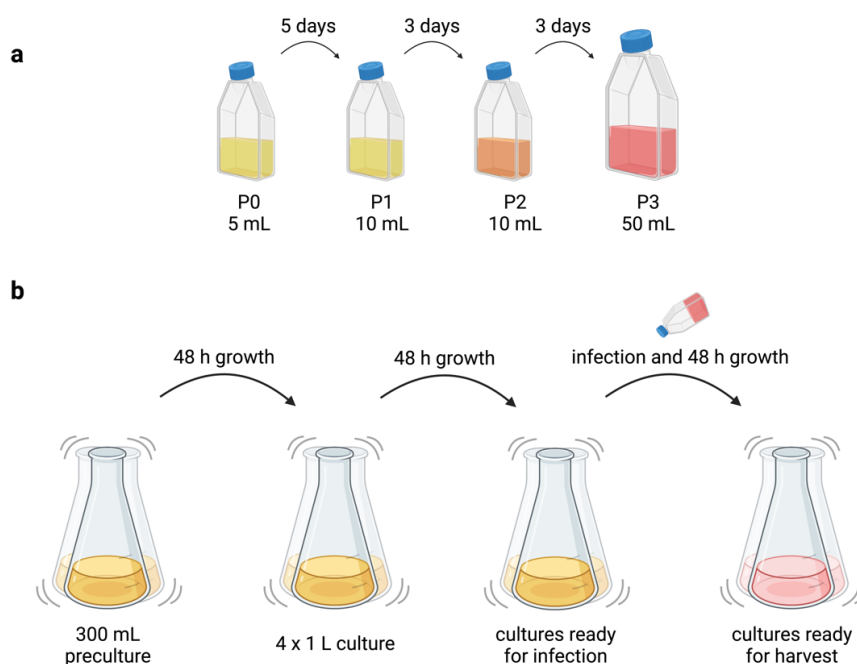
LifeTechnologies Terbium-Streptavidin ( <i>PV3965</i> )	Thermo Fischer
Lanthascreen Europium-Streptavidin ( <i>PV5899</i> )	Thermo Fischer
Biotin anti-His antibody ( <i>PV6089</i> )	Thermo Fischer
LifeTechnologies Kinase Tracer 236 ( <i>PV5592</i> )	Thermo Fischer
Recombinant CDK2-cyclin A protein ( <i>PV3267</i> )	Thermo Fischer
Recombinant CDK9-cyclin T protein ( <i>PV4131</i> )	Thermo Fischer
DMSO Hybri-Max ampules ( <i>D2650</i> )	Sigma-Aldrich
3.5 M Ammonium sulphate ( <i>MD2-250-35</i> )	Molecular Dimensions
1.8M Ammonium citrate tribasic pH 7.0 ( <i>HR2-994-21</i> )	Hampton Research
HEPES pH 6.8 ( <i>HR2-902-01</i> )	Hampton Research
HEPES pH 8.2 ( <i>HR-902-15</i> )	Hampton Research
Dithiothreitol (DTT) ( <i>D0632</i> )	Sigma-Aldrich
TCEP ( <i>P1020</i> )	LucernaChem
BSA, protease and essential globulin-free ( <i>A-7030</i> )	Sigma
QIAprep Spin Miniprep Kit	QIAGEN
SDS running buffer	25 mM Tris pH 8.3 192 mM Glycine 0.1 % SDS
5x loading dye	62.5 mM Tris HCl pH 6.8 2 % SDS 25 % Glycerol 0.01% Bromophenol blue 3 % $\beta$ -mercaptoethanol

### 6.1.2 Insect cell culture

Suspension cultures of Sf9 cells (*Spodoptera frugiperda*; used for virus amplification) and Hi5 cells (High-Five™ BTI-TN-5B1-4, *Trichoplusia ni*; used for protein expression) were cultured in serum-free SF-4 medium (SF4 Baculo Express ICM, BioConcept). Cells were incubated in a shaker incubator (KuhnerShakerX) at 120 rpm and 27 °C. Cell growth and viability were monitored by measuring the cell density (million/mL) and % viability using a cell viability analyser (Vi-CELL™ XR, Beckman Coulter) and confluent cultures (4 million/mL for Hi5 and up to 6 million/mL for Sf9) were regularly diluted to a final density of 0.2-0.5 million/mL for Hi5 and 1 million/mL for Sf9 cells. Some cultures were supplemented with Penicillin (1.6  $\mu$ g/mL) and Streptomycin (2.5  $\mu$ g/mL).

For the generation of passage 0 (P0) viral stock using the pAC8-dsRed (**Table 6.2**) vector system, a transfection mix (8  $\mu$ L Cellfectin™, 0.5  $\mu$ g viral DNA, 2.0  $\mu$ g of the respective plasmid, 150  $\mu$ L SF- 4 medium) was prepared and incubated at RT for 30 min. The transfection mix was added to an Sf9 culture (2.9 million/mL,

850  $\mu$ L) and the cells were incubated for 4 h in a shaker incubator at 120 rpm and 27 °C. After 4 h, SF-4 medium (4 mL) was added to the transfected cells, which were then allowed to grow for 5 days in a shaker incubator at 120 rpm and 27 °C. On day five, the culture was supplemented with 10 % FBS and harvested through centrifugation (1500 rpm, RT, 10 min). The supernatant was transferred to a sterile Falcon™ tube and stored at 4 °C or used directly for further viral amplification (**Figure 6.1a**). For this, the P0 virus for gene of interest was added (1:100) to an Sf9 suspension culture (2 million/mL, 10 mL) and infected cells were incubated for 3 days at 27 °C and 120 rpm. The resulting P1 virus with an increased viral titre was harvested as described above and further used for the generation of P2 viruses. For large-scale protein expression, a P3 virus was generated by infecting larger cell culture volumes (50 mL).



**Figure 6.1. Protein expression in the baculovirus-insect cell system using pAC8-dsRed vectors.** a, Viral amplification. b, Cell culture and infection. Image created with BioRender.com

### 6.1.3 Protein expression and purification

In a general recombinant protein expression workflow (**Figure 6.1b**), Hi5 cells (1-16 L) were grown to confluency, harvested by centrifugation (1500 rpm, RT, 10 min, Sorvall™ BP 16) and resuspended in the appropriate P3 virus (10-25 mL virus per 1 L of cells). The cells were incubated for 30-60 min (27 °C, 120 rpm) and subsequently evenly distributed over fresh, pre-warmed (27 °C) Sf4 medium (2 L per 1 L of cells), usually supplemented with penicillin (1.6  $\mu$ g/mL) and streptomycin (2.5  $\mu$ g/mL). The infected Hi5 cell cultures were incubated for 2 days (27 °C and 120 rpm (KuhnerShakerX)) and harvested by centrifugation (1500 rpm, 4 °C, 10 min, Sorvall™ BP 16). The resulting cell pellet was resuspended in a suitable, ice-cold lysis buffer (75 mL per 1 L of cells) and cells were lysed by sonication (3 min cycles, 10 sec pulse, 20 sec pause, 20 % amplitude, Vibra-Cell™, Sonics) whilst in an ice bath.



The cell debris were then separated from soluble fraction of the lysate by ultracentrifugation (4 °C, 40min, 35000 rpm in a Ti45 rotor, Beckman Coulter). The supernatant was subsequently filtered through Microcloth (Merck) and applied to a suitable affinity column. The elution fraction from the affinity purification step was then subject to ion exchange chromatography and finally size exclusion chromatography to yield the final protein sample.

## **6.1.4 Protein analysis**

### **6.1.4.1 SDS polyacrylamide-gel electrophoresis**

SDS- PAGE (SDS polyacrylamide-gel electrophoresis) was performed to monitor the progress of each protein purification step. For this, precast protein gels (Any kD™ Mini-PROTEAN® TGX™, Bio Rad) were used. Protein samples were diluted (SEC buffer) 1:5 and supplemented with 5x SDS loading dye and pre-stained protein M<sub>w</sub> marker (Precision Plus Protein™ Dual Colour Standards, Bio Rad) was used for reference. The electrophoresis was carried out at 400 mA and 200 V for 30 min and gels were stained in Quick Blue (Lubio Science) and de-stained in water overnight.

### **6.1.4.2 Determination of protein concentration**

Concentration determination for protein samples was performed using two orthogonal methods. Firstly, absorbance at 280 nm was determined using a NanoDrop™ 2000/2000c Spectrophotometer (Thermo Scientific™). Secondly, protein concentration was measured using the Bradford Assay. For this, two different protein dilutions (usually 1:5 and 1:10) and a BSA standard series (1, 2, 4, 6, and 8 mg/mL) were each mixed with Quick Start™ Bradford 1x Dye Reagent (Bio Rad) and incubated at RT for 2 min. Bradford reagent solution was used as reference and the absorbance of the samples was measured at 595 nm using the Ultrospec 3100 Pro UV/Visible Spectrophotometer (GE Healthcare/ Amersham Pharmacia). The standard curve of BSA absorbance was then plotted and used to determine the concentration of each sample.

## **6.2 Discovery of cyclin K degraders (Chapter 2)**

### **6.2.1 Mammalian cell culture**

The human HEK293T cell lines were provided by the Genetic Perturbation Platform, Broad Institute and K562<sub>Cas9</sub>, THP1<sub>Cas9</sub>, P31FUJ<sub>Cas9</sub> cell lines were provided by Zuzana Tothova, Broad Institute and HEK293T<sub>Cas9</sub><sup>113</sup> and MM1S<sub>Cas9</sub><sup>412</sup> were previously published. HEK293T cells were cultured in DMEM (Gibco) and all other cell lines in RPMI (Gibco), with 10% FBS (Invitrogen), glutamine (Invitrogen) and penicillin–streptomycin (Invitrogen) at 37 °C and 5 % CO<sub>2</sub>.

## 6.2.2 Compounds

*R*-CR8 (3605) was obtained from Tocris, *S*-CR8 (ALX-270-509-M005), flavopiridol (ALX-430-161-M005) from Enzo Life Sciences, roscovitine (HY-30237), THZ531 (HY-103618), LDC00067 (HY-15878) from MedChem Express and DRF053 (D6946) from Sigma.

## 6.2.3 Antibodies

The following antibodies were used in this study: anti-cycK (Bethyl Laboratories, A301-939A for full length cycK), anti-cycK (abcam, ab251652, for cycK<sup>1-267</sup>), anti-beta-actin (Cell Signalling, #3700), anti-mouse 800CW (LI-COR Biosciences, 926-32211), anti-rabbit 680LT (LI-COR Biosciences, 925-68021), anti-rabbit IgG antibodies (abcam, ab6721).

## 6.2.4 Reporter vectors

The following reporter were used in this study: Artichoke (SFFV.BsmBICloneSite-17aaRigidLinker-eGFP.IRES.mCherry.cppt.EF1 $\alpha$ .PuroR, Addgene #73320 for genome-wide screen and validation experiments), Cilantro 2 (PGK.BsmBICloneSite-10aaFlexibleLinker-eGFP.IRES.mCherry. cppt.EF1 $\alpha$ .PuroR, Addgene #74450 for degradation kinetics), sgBFP (sgRNA.SFFV.tBFP, for validation of drug-E3 ligase pairs), sgRFP657, (sgRNA.EFS.RFP657 for validation of drug-E3 ligase pairs), sgPuro, (pXPR003, Addgene #52963, for drug sensitivity assays).

## 6.2.5 Oligonucleotides

Detailed lists of all oligonucleotides used in this study can be found in Supplementary Oligo Table 1 in the published manuscript (<https://www.nature.com/articles/s41586-020-2374-x#Sec42>).

## 6.2.6 Bioinformatic screen

We computed Pearson correlations of viability of PRISM repurposing compounds in 8 doses and 578 cell lines<sup>273</sup> with gene expression (GE) and copy number variation (CN) of all detectable protein-coding genes of matched cell lines from The Cancer Cell Line Encyclopaedia (CCLE)<sup>274</sup>. Z score was computed for each pair of compounds, dose viability, and genomic feature (GE or CN) across all cell lines. For each compound-genomic feature pair, the most extreme correlations are ranked from negative to positive. To focus on novel compound-gene relationships, we restricted genes to a curated list of 499 E3 ligase components and compounds that are not one of “EGFR inhibitor”, “RAF inhibitor” or “MDM inhibitor” based on PRISM repurposing annotation<sup>274</sup>. Hit compounds were selected if either the Z score was less than -6 or ranked in the top 15 with Z score less than -4. The resulting list of 158 E3 gene-compound pairs was further curated and shortened manually to 96 E3 gene-compound pairs, which included 95 unique E3 ligases and 85 unique compounds.

### 6.2.7 Cloning and lentiviral packaging of sgRNAs targeting 95 E3 ligases

sgRNAs targeting E3 ligases were selected from the human Brunello CRISPR library<sup>413</sup>. 170 oligo pairs (IDT) targeting 95 E3 ligases were annealed and cloned into the sgRNA.SFFV.tBFP (Guide ID A) or sgRNA.EFS.RFP657 (Guide ID B) fluorescent vectors in a 96-well format using previously published protocols<sup>414</sup>. Briefly, vectors were linearised with BsmBI (New England Biolabs) and gel purified with the Spin Miniprep Kit (Qiagen). Annealed oligos were phosphorylated with T4 Polynucleotide Kinase (New England Biolabs) and ligated into the linearised and purified vector backbones with T4 DNA Ligase (New England Biolabs). Constructs were transformed into XL10-Gold ultracompetent *Escherichia coli* (Stratagene/Agilent Technologies), plasmids were purified using MiniPrep Kit (Qiagen), and guide sequence confirmed by Sanger sequencing. For validation of the primary screen, virus was produced in a 96-well format. Briefly, 11,000 HEK293T cells were seeded per well in 100  $\mu$ L DMEM medium supplemented with 10% FBS and Penicillin-Streptomycin-Glutamine. The next day a Packaging Mix was prepared in a 96-well plate consisting of 500 ng psPAX2, 50 ng pVSV-G and 17 ng sgRNA backbone in 5  $\mu$ L OptiMem (Invitrogen) and incubated for 5 minutes at room temperature. This mix was combined with 0.1  $\mu$ L TRANSIT-LT1 (Mirus) in 5  $\mu$ L of OptiMem, incubated for 30 minutes at room temperature, and then applied to cells. Two days post-transfection, dead cells were removed by centrifugation and lentivirus containing medium was collected and stored at  $-80^{\circ}\text{C}$  prior to use.

### 6.2.8 Validation of drug-E3 ligase pairs from the bioinformatic screen

K562<sub>Cas9</sub>, OVK16<sub>Cas9</sub>, A564<sub>Cas9</sub>, ES2<sub>Cas9</sub> and MOLM13<sub>Cas9</sub> cell lines were individually transduced with 192 sgRNAs targeting 95 E3 ligases in a 96-well plate format. 3000 cells/well were plated in 100  $\mu$ L RPMI supplemented with 10% FCS and Penicillin-Streptomycin-Glutamine and 30 $\mu$ L/well of virus supernatant was added. 24 hours post infection the medium was changed. After three days, the percentage of sgRNA transduced cells was determined by flow cytometry. If more than 60% of cells were transduced, un-transduced cells were added to bring the level below 60%. Eight days post-infection cell density was measured and adjusted to  $1.5 \times 10^5$  cells/mL with RPMI. For treatment, 50  $\mu$ L of sgRNA transduced cells were seeded into each well of a 384 well plate with pre-plated DMSO or cognate drug in three concentrations (0.1  $\mu$ M, 1  $\mu$ M, 10  $\mu$ M) with Agilent BRAVO Automated Liquid Handling Platform. Plates were sealed with White Rayon adhesive sealing tape (Thermo Fisher Scientific) and grown for three days. Adherent cell lines were trypsinised and re-suspended in 50  $\mu$ L of RPMI with Matrix WellMate (Thermo Scientific). Suspension cells were directly subjected to analysis. 10  $\mu$ L of cell suspension was subjected to flow analysis with FACSCanto equipped with High Throughput Sampler (BD Bioscience). The percentage of sgRNA transduced cells in the drug treatment wells was normalised to the DMSO control. Wells with fluorescent drug and samples with less than 120 viable cells events or less than 6% fluorescent cells were removed from analysis. All E3-drug pairs were ranked based on the number of experimental conditions (cell line and drug dose) with more than 50% of sgRNA transduced cells in drug treatment wells in comparison to corresponding DMSO control.

### **6.2.9 Validation of the DDB1 resistance phenotype**

For validation experiments, virus was produced in a 6-well plate format, as described above with the following adjustments:  $2.5 \times 10^5$  HEK293T cells/well in 2 mL DMEM medium, 3  $\mu\text{L}$ /well of TransIT-LT1, 15  $\mu\text{L}$ /well of OPTI-MEM, 500 ng/well of the desired plasmid, 500 ng/well psPAX2, and 50 ng/well pVSV-G in 32.5  $\mu\text{L}$ /well OPTI-MEM. After collecting the virus,  $10 \times 10^3$  HEK293T<sub>Cas9</sub> cells in 100  $\mu\text{L}$  DMEM medium were transduced with 10  $\mu\text{L}$  of virus supernatant. The transduced HEK293T<sub>Cas9</sub> cells were then mixed with untransduced control cells at a 1:9 ratio. Nine days after sgRNA transduction, cells were treated for 3 days with DMSO or 1  $\mu\text{M}$  CR8 and analysed by flow cytometry to determine the percentage of BFP<sup>+</sup> cells. sgRNAs targeting DDB1 provide partial depletion of DDB1 (50% DDB1 alleles modified, reducing DDB1 levels by roughly 50%), which suggests selection towards heterozygous or hypomorphic clones.

### **6.2.10 Whole proteome quantification using tandem mass tag mass spectrometry**

$10 \times 10^6$  Molt-4 cells were treated with DMSO (triplicate) or 1  $\mu\text{M}$  CR8 (single replicate) for 1 or 5 hours and later were harvested by centrifugation. Samples were processed, measured, and analysed as described before<sup>116</sup>. Data are available in the PRIDE repository (PXD016187 and PXD016188).

### **6.2.11 Quantitative PCR**

HEK293T<sub>Cas9</sub> cells were treated with DMSO or 1  $\mu\text{M}$  CR8 for 2 hours, collected by centrifugation, washed with PBS, and snap-frozen at  $-80^\circ\text{C}$ . mRNA was isolated using a QIAGEN RNA kit (Qiagen, 74106). For cDNA synthesis, total RNA was reverse transcribed using a High-Capacity cDNA Reverse Transcription Kit (Thermo Fischer) before qPCR analysis with TaqMan Fast Advanced Master Mix (ThermoFisher Scientific, 4444557) for CCNK (TaqMan, Hs00171095\_m1, Life Technologies) and GAPDH (TaqMan, Hs02758991\_g1). Reactions were run and analysed on a CFX96 Real Time system (Bio-Rad).

### **6.2.12 Immunoblots for whole protein lysate**

Cells were washed with phosphate buffered saline (PBS) and lysed (150 mM NaCl, 50 mM Tris (pH 7.5), 1% NP-40, 1% glycerol, 1x Halt Cocktail protease and phosphatase inhibitors) for 20 minutes on ice. The insoluble fraction was removed by centrifugation, protein concentration was quantified using a BCA protein assay kit (Pierce), and equal amount of lysate was run on SDS-PAGE 4-12% Bis-Tris Protein Gels (NuPAGE, Thermo Fisher) and subsequently transferred to nitrocellulose membrane with Trans-Blot Turbo System (BIO-RAD). Membranes were blocked in Odyssey Blocking Buffer/PBS (LI-COR Biosciences) and incubated with primary antibodies overnight at  $4^\circ\text{C}$ . The membranes were then washed in TBS-T, incubated for 1 hour with secondary IRDye-conjugated antibodies (LI-COR Biosciences), and washed three times in TBS-T for 5 minutes prior to Near-Infrared Western blot detection on an Odyssey Imaging System (LI-COR Biosciences).

### 6.2.13 Cyclin K stability reporter analysis

HEK293T<sub>Cas9</sub> expressing the cycK<sub>eGFP</sub> degradation reporter were transduced with experimental sgRNAs. Nine days after infection the cells were dosed for 2 hours with DMSO or 1  $\mu$ M CR8. Using FLOWJO (flow cytometry analysis software), the geometric mean of eGFP and mCherry fluorescent signal for round and mCherry positive cells was calculated. The ratio of eGFP to mCherry was normalised to the average of three DMSO-treated controls.

### 6.2.14 Genome-wide CRISPR – CR8 resistance screen

5 % (v/v) of the human genome-wide CRISPR-KO Brunello library with 0.4  $\mu$ L Polybrene/mL (stock of 8 mg/mL) was added to  $1.5 \times 10^8$  HEK293T<sub>Cas9</sub> in 75 mL of medium and transduced (2400 rpm, 2 hours, 37°C). 24h after infection sgRNA transduced cells were selected with 2  $\mu$ g/mL of Puromycin for two days. On the ninth day post-infection, cells were treated with either DMSO or 1  $\mu$ M CR8 and cultured for an additional 3 days. Resistant live cells were selected by gently washing away detached dead cells from the medium. Cell pellets were resuspended in multiple direct lysis buffer reactions (1 mM CaCl<sub>2</sub>, 3 mM MgCl<sub>2</sub>, 1 mM EDTA, 1% Triton X-100, Tris pH 7.5 - with freshly supplemented 0.2 mg/mL Proteinase) with  $1 \times 10^6$  cells per 100  $\mu$ L reaction. The sgRNA sequence was amplified in a first PCR reaction with eight staggered forward primers. 20  $\mu$ L of direct lysed cells was mixed with 0.04U Titanium Taq (Takara Bio 639210), 0.5 x Titanium Taq buffer, 800  $\mu$ M dNTP mix, 200 nM SBS3-Stagger-pXPR003 forward primer, 200 nM SBS12-pXPR003 reverse primer in a 50  $\mu$ L reaction (cycles: 5 minutes at 94°C, 15 x (30 sec at 94°C, 15 sec at 58°C, 30 sec at 72°C), 2 minutes at 72°C). 2  $\mu$ L of the first PCR reaction was used as the template for 15 cycles of the second PCR, where Illumina adapters and barcodes were added (0.04U Titanium Taq, 1 x Titanium Taq buffer, 800  $\mu$ M dNTP mix, 200 nM P5-SBS3 forward primer, 200 nM P7-barcode-SBS12 reverse primer). An equal amount of all samples was pooled and subjected to preparative agarose electrophoresis followed by gel purification (Qiagen). Eluted DNA was further purified by NaOAc and isopropanol precipitation. Amplified sgRNAs were quantified using Illumina NextSeq platform. Read counts for all guides targeting the same gene were used to generate p-values. Hits enriched in resistance population with False Discovery Rate (FDR) < 0.05 and enriched > 5-fold, are labelled on the relevant plots<sup>415</sup>.

### 6.2.15 BISON CRISPR – CR8 resistance screen

The BISON CRISPR library targets 713 E1, E2, and E3 ubiquitin ligases, deubiquitinases, and control genes and contains 2,852 guide RNAs. It was cloned into the pXPR003 as previously described<sup>413</sup> by the Broad Institute Genome Perturbation Platform (GPP). The virus for the library was produced in a T-175 flask format, as described above with the following adjustments:  $1.8 \times 10^7$  HEK293T cells in 25 mL complete DMEM medium, 244  $\mu$ L of TransIT-LT1, 5 mL of OPTI-MEM, 32  $\mu$ g of library, 40  $\mu$ g psPAX2, and 4  $\mu$ g pVSV-G in 1 mL OPTI-MEM. 10 % (v/v) of BISON CRISPR library was added to  $6 \times 10^6$  HEK293T<sub>Cas9</sub> cells in triplicates and transduced. Samples were processed as describe above for the genome wide resistance screen.

### 6.2.16 Genome-wide CRISPR – cyclin K stability reporter screen

A single clone of cycK<sub>eGFP</sub> HEK293T<sub>Cas9</sub> was transduced with the genome wide Brunello library as described above with the following modification:  $4.5 \times 10^8$  cycK<sub>eGFP</sub> HEK293T<sub>Cas9</sub> cells in 225 mL of medium. Nine days later cells were treated with CR8 or DMSO for at least 2 hours and the cycK stable population was separated using fluorescence activated cell sorting. Four populations were collected (top 5 %, top 5-15 %, lowest 5-15 % and lowest 5 %) based on the cycK<sub>eGFP</sub> to mCherry mean fluorescent intensity (MFI) ratio. Sorted cells were harvested by centrifugation and subjected to direct lysis as described above. The screen was analysed as described previously by comparing stable populations (top 5 % eGFP/mCherry expression) to unstable populations (lowest 15 % eGFP/mCherry expression). Hits enriched in cycK stable population with FDR < 0.05 are labelled in the relevant figures.

### 6.2.17 Pooled CRISPR screen data analysis

The data analysis pipeline comprised the following steps: (1) Each sample was normalised to the total number of reads. (2) For each guide, the ratio of reads in the stable vs. unstable sorted gate was calculated, and sgRNAs were ranked. (3) The ranks for each guide were summed for all replicates. (3) The gene rank was determined as the median rank of the four guides targeting it. (4) P-values were calculated by simulating a distribution with guide RNAs that had randomly assigned ranks over 100 iterations. R scripts can be found in the Supplementary Information of the published manuscript (<https://www.nature.com/articles/s41586-020-2374-x#Sec42>).

### 6.2.18 DCAF arrayed screen

An arrayed DCAF library (targeting DCAFs substrate receptors, DCAF-like and control genes) was constructed as described above with the appropriate oligos (Supplementary Oligo Table 1). K562<sub>Cas9</sub>, P31FUJ<sub>Cas9</sub>, THP1<sub>Cas9</sub> and MM1S<sub>Cas9</sub> were individually transduced and treated with DMSO or 1  $\mu$ M CR8 (K562<sub>Cas9</sub>, P31FUJ<sub>Cas9</sub>, THP1<sub>Cas9</sub>) or 0.1  $\mu$ M CR8 (MM1S<sub>Cas9</sub>). The analysis was performed as described above for validation of DDB1 resistance phenotype.

### 6.2.19 Protein purification

Human wild-type and mutant versions of DDB1 (Uniprot entry Q16531), CDK12 (Q9NYV4, K965R) and CCNK (O75909) were subcloned into pAC-derived vectors<sup>416</sup> and recombinant proteins were expressed as N-terminal His<sub>6</sub>, His<sub>6</sub>-Spy, StrepII or StrepII-Avi fusions in *Trichoplusia ni* High Five insect cells using the baculovirus expression system (Invitrogen)<sup>417</sup>.

Wild-type or mutant forms of full-length or beta-propeller B domain deletion ( $\Delta$ BPB: aa 396-705 deleted) constructs of His<sub>6</sub>-DDB1 and StrepII-Avi-DDB1 were purified as previously described for DDB1-DCAF complexes<sup>418</sup>. High Five insect cells co-expressing truncated versions of wild-type or mutant His<sub>6</sub>-CDK12 (aa 713-1052 or 713-1032) and His<sub>6</sub>- or His<sub>6</sub>-Spy-tagged cycK (aa 1-267) were lysed by sonication in

50 mM Tris-HCl (pH 8.0), 500 mM NaCl, 10% (v/v) glycerol, 10 mM MgCl<sub>2</sub>, 10 mM imidazole, 0.25 mM tris(2-carboxyethyl)phosphine (TCEP), 0.1% (v/v) Triton X-100, 1 mM phenylmethylsulfonylfluoride (PMSF), and 1 x protease inhibitor cocktail (Sigma). Following ultracentrifugation, the soluble fraction was passed over HIS-Select Ni<sup>2+</sup> affinity resin (Sigma), washed with 50 mM Tris-HCl (pH 8.0), 1 M NaCl, 10% (v/v) glycerol, 0.25 mM TCEP, 10 mM imidazole and eluted in 50 mM Tris-HCl (pH 8.0), 200 mM NaCl, 10% (v/v) glycerol, 0.25 mM TCEP, 250 mM imidazole. When necessary, affinity tags were removed by overnight TEV protease treatment. In case of HIS-Select Ni<sup>2+</sup> affinity purified CDK12-cycK that was not subjected to TEV cleavage, the pH of the eluate was adjusted to 6.8 before ion exchange chromatography. StrepII-tagged versions of CDK12-cycK were affinity purified using Strep-Tactin Sepharose (IBA) omitting imidazole in lysis, wash and elution buffers, supplementing the elution buffer with 2.5 mM desthiobiotin (IBA GmbH), and using 50 mM Tris-HCl (pH 6.8) throughout.

For ion exchange chromatography, affinity purified proteins were diluted in a 1:1 ratio with buffer *A* (50 mM Tris-HCl (pH 6.8), 10 mM NaCl, 2.5% (v/v) glycerol, 0.25 mM TCEP) and passed over an 8 mL Poros 50HQ column. The flow through was again diluted in a 1:1 ratio with buffer *A* and passed over an 8 mL Poros 50HS column. Bound proteins were eluted by a linear salt gradient mixing buffer *A* and buffer *B* (50 mM Tris-HCl (pH 6.8), 1 M NaCl, 2.5% (v/v) glycerol, 0.25 mM TCEP) over 15 column volumes to a final ratio of 80% buffer *B*. Poros 50HS peak fractions containing the CDK12-cycK complex were concentrated and subjected to size exclusion chromatography in 50 mM HEPES (pH 7.4), 200 mM NaCl, 2.5% (v/v) glycerol and 0.25 mM TCEP. The concentrated proteins were flash frozen in liquid nitrogen and stored at -80°C.

### 6.2.20 Co-immunoprecipitation assay

The purified His<sub>6</sub>-CDK12/StrepII-cycK complex was mixed with equimolar concentrations of full-length His<sub>6</sub>-DDB1 or TEV-cleaved DDB1<sup>ΔBPB</sup> (5 μM) in the presence 5 μM *R*-CR8 or DMSO in IP buffer (50 mM HEPES (pH 7.4), 200 mM NaCl, 0.25 mM TCEP, 0.05% (v/v) Tween-20) containing 1 mg/mL bovine serum albumin. The solution was added to Strep-Tactin MacroPrep beads (IBA GmbH) preequilibrated in IP buffer and incubated for 1 hour at 4°C on an end-over-end shaker. The beads were extensively washed with IP buffer, and the bound protein was eluted with IP buffer containing 2.5 mM desthiobiotin for 1 hour at 4°C on an end-over-end shaker. Eluted proteins were separated by SDS-PAGE stained with Coomassie Brilliant Blue.

### 6.2.21 Crystallisation and data collection

The protein solution for crystallization contained 70 μM TEV-cleaved DDB1<sup>ΔBPB</sup>, 80 μM *R*-CR8 and 80 μM TEV-cleaved CDK12-cycK in 50 mM HEPES (pH 7.4), 200 mM NaCl, 0.25 mM TCEP. Crystals were grown by vapour diffusion in drops containing 0.2 μL DDB1<sup>ΔBPB</sup>-*R*-CR8-CDK12<sup>713-1052</sup>-cycK<sup>1-267</sup> complex solution mixed with 0.2 μL of reservoir solution containing 0.9 M ammonium citrate tribasic (pH 7.0) in two-well format sitting drop crystallization plates (Swissci). Plates were incubated at 19°C and crystals appeared 5-13 days after setup. Crystals were flash cooled in liquid nitrogen in reservoir solution supplemented with 25% (v/v) glycerol

as a cryoprotectant prior to data collection. Diffraction data were collected at the Swiss Light Source (beamline PXI) with an Eiger 16M detector (Dectris) at a wavelength of 1 Å and a crystal cooled to 100 K. Data were processed with *DIALS*, scaled with *AIMLESS* supported by other programs of the CCP4 suite<sup>419</sup>, and converted to structure factor amplitudes with *STARANISO*<sup>420</sup>, applying a locally weighted  $CC_{1/2} = 0.3$  resolution cut-off.

### 6.2.22 Structure determination and model building

The DDB1<sup>ΔBPB</sup>-R-CR8-CDK12<sup>713-1052</sup>-cycK<sup>1-267</sup> complex formed crystals belonging space group  $P3_121$ , with three complexes in the crystallographic asymmetric unit (ASU). Their structure was determined using molecular replacement (MR) in *PHASER*<sup>421</sup> with a search model derived from PDB entry 6H0F for DDB1<sup>ΔBPB</sup>, and PDB entry 4NST for CDK12-cycK. The initial model was improved by iterative cycles of building with *COOT*<sup>422</sup>, and refinement using *phenix.refine*<sup>423</sup> or *autoBUSTER*<sup>424</sup>, with ligand restraints generated using *eLBOW* through *phenix.ready\_set*<sup>425</sup>. The final model was produced by refinement with *autoBUSTER*. Analysis with *MOLPROBITY*<sup>426</sup> indicates that 93.9 % of the residues in final model are in favourable regions of the Ramachandran plot, with 0.6 % outliers. Data processing and refinement statistics are presented in **Table 2.1**. Interface analysis was performed using PISA<sup>372</sup>.

### 6.2.23 Biotinylation of DDB1

Purified full-length StrepII-Avi-DDB1 was biotinylated *in vitro* at a concentration of 8 μM by incubation with final concentrations of 2.5 μM BirA enzyme and 0.2 mM D-biotin in 50 mM HEPES (pH 7.4), 200 mM NaCl, 10 mM MgCl<sub>2</sub>, 0.25 mM TCEP and 20 mM ATP. The reaction was incubated for 1 hour at room temperature and stored at 4°C for 14-16 hours. Biotinylated DDB1 (biotinDDB1) was purified by gel filtration chromatography and stored at -80°C (~20 μM).

### 6.2.24 Time-resolved fluorescence resonance energy transfer (TR-FRET)

Increasing concentrations of Alexa488-SpyCatcher-labelled<sup>113</sup> His<sub>6</sub>-Spy-cycK/His<sub>6</sub>-CDK12 (Alexa488CycK-CDK12) were added to a mixture of biotinylated DDB1 (biotinDDB1) at 50 nM, terbium-coupled streptavidin at 4 nM (Invitrogen) and kinase inhibitors at 10 μM (final concentrations) in 384-well microplates (Greiner, 784076) in a buffer containing 50 mM Tris (pH 7.5), 150 mM NaCl, 0.1 % pluronic acid and 0.5% DMSO. CR8 titrations were carried out by adding increasing concentrations CR8 (0-25 μM) into premixed 500 μM Alexa488cycK-CDK12, 50 nM biotinDDB1, and 4 nM terbium-coupled streptavidin. Before TR-FRET measurements, reactions were incubated for 15 minutes at room temperature. After excitation of terbium (Tb) fluorescence at 337 nm, emissions at 490 nm (Tb) and 520 nm (Alexa488) were measured with a 70 μs delay to reduce background fluorescence and the reactions were followed by recording 60 data points of each well over 1 hours using a PHERAstar FS microplate reader (BMG Labtech). The TR-FRET signal of each data point was extracted by calculating the 520:490 nm ratio. Data were analysed with *Prism 7* (GraphPad) assuming equimolar binding of biotinDDB1 to Alexa488cycK-CDK12 using the equations described previously<sup>8</sup>.



Counter titrations with unlabelled proteins were carried out by mixing 500  $\mu\text{M}$  Alexa488cycK-CDK12 with 50 nM biotinDDB1 in the presence of 4 nM terbium-coupled streptavidin and 1  $\mu\text{M}$  compound for DDB1 titrations or 12.5  $\mu\text{M}$  compound for CDK12 titrations. After incubation for 15 minutes at room temperature, increasing amounts of unlabelled cycK-CDK12 or DDB1 (0-10  $\mu\text{M}$ ) were added to the preassembled Alexa488cycK-CDK12/biotinDDB1 complexes in a 1:1 volume ratio and incubated for 15 minutes at room temperature. TR-FRET data were acquired as described above. The 520/490 nm ratios were plotted to calculate the half maximal inhibitory concentrations ( $\text{IC}_{50}$ ) assuming a single binding site using *Prism 7* (GraphPad).  $\text{IC}_{50}$  values were converted to the respective  $K_i$  values as described previously<sup>427</sup>. Three technical replicates were carried out per experiment.

### 6.2.25 DDB1-CUL4-RBX1 reconstitution and *in vitro* CUL4 neddylation

*In vitro* CUL4 reconstitution and CUL4 neddylation were performed as described<sup>44</sup>. His<sub>6</sub>-CUL4A/His<sub>6</sub>-RBX1 at 3.5  $\mu\text{M}$  was incubated with His<sub>6</sub>-DDB1 at 3  $\mu\text{M}$  in a reaction mixture containing 3.8  $\mu\text{M}$  NEDD8, 50 nM NAE1/UBA3 (E1), 30 nM UBC12 (E2), 1 mM ATP, 50 mM Tris (pH 7.5), 100 mM NaCl, 2.5 mM MgCl<sub>2</sub>, 0.5 mM DTT and 5 % (v/v) glycerol for 1.5 hours at room temperature. Neddylated and gel filtration-purified DDB1-CUL4-RBX1 (<sub>N8</sub>DDB1-CUL4-RBX1) was concentrated to 7.6  $\mu\text{M}$ , flash frozen and stored at -80°C.

### 6.2.26 In-vitro ubiquitination assays

*In vitro* ubiquitination was performed by mixing <sub>N8</sub>DDB1-CUL4-RBX1 at 70 nM with a reaction mixture containing kinase inhibitors at 2  $\mu\text{M}$ , CDK12-cycK at 500 nM, E1 (UBA1, BostonBiochem) at 50 nM, E2 (UBCH5a, BostonBiochem) at 1  $\mu\text{M}$ , and ubiquitin at 20  $\mu\text{M}$ . Reactions were carried out in 50 mM Tris (pH 7.5), 150 mM NaCl, 5 mM MgCl<sub>2</sub>, 0.2 mM CaCl<sub>2</sub>, 1 mM ATP, 0.1 % Triton X-100 and 0.1 mg/mL BSA, incubated for 0-30 minutes at 30°C and analysed by western blot using anti-cycK and anti-rabbit IgG antibodies. Blots were scanned on an Amersham 600 CCD-based imaging system (GE Life Sciences).

### 6.2.27 Isothermal titration calorimetry (ITC)

ITC experiments were performed at 25°C on a VP-ITC isothermal titration calorimeter (Microcal Inc.). Purified and TEV-cleaved CDK12-cycK and DDB1<sup>ΔBPB</sup> were exhaustively dialysed in 50 mM HEPES (pH 7.4), 150 mM NaCl, 0.25 mM TCEP, 0.5 % DMSO and loaded into the sample cell at a final concentration of 10-50  $\mu\text{M}$ . Kinase inhibitors (CR8 or roscovitine) were diluted from a 100 mM DMSO stock solution to 100-500  $\mu\text{M}$  in buffer containing 50 mM HEPES (pH 7.4), 150 mM NaCl, 0.25 mM TCEP. The final DMSO concentration was 0.5 %. Titrations with 100-500  $\mu\text{M}$  compound were performed typically through about 30 injections of 6-10  $\mu\text{L}$  at 210 second intervals from a 300  $\mu\text{L}$  syringe rotating at 300 rpm. An initial injection of the ligand (4  $\mu\text{L}$ ) was made and discarded during data analysis. For probing DDB1-CDK12-cycK complex formation, DDB1<sup>ΔBPB</sup> (100  $\mu\text{M}$ , in the syringe) was titrated into the cell containing CDK12-cycK (10  $\mu\text{M}$ ) or CDK12-cycK (10  $\mu\text{M}$ ) pre-incubated with CR8 (30  $\mu\text{M}$ ). The heat change accompanying the titration was recorded as

differential power by the instrument and determined by integration of the peak obtained. Titrations of ligand to buffer only and buffer into protein were performed to allow baseline corrections. The heat change was fitted using nonlinear least-squares minimization to obtain the dissociation constants,  $K_d$ , the enthalpy of binding,  $\Delta H$ , and the stoichiometry,  $n$ . Between one and three replicates were performed per titration.

### 6.2.28 Bioluminescence Resonance Energy Transfer (BRET) analyses

Bioluminescence resonance energy transfer (BRET) experiments were using a NanoBRET PPI starter kit (Promega N1821) according to the manufacturer's instructions and as previously described<sup>428</sup>.

### 6.2.29 Drug sensitivity assays

HEK293T<sub>Cas9</sub> cells were resuspended at  $0.15 \times 10^6$  per mL and plated on a 384 well plate with 50  $\mu$ l per well and MLN4924, MLN7243 or MG132 with or without CR8 serially diluted with D300e Digital Dispenser (Tecan Inc.). HEK293T<sub>Cas9</sub> cells ( $0.625 \times 10^6$  cells/6 well plate format) were seeded the day before transfection. The following day, 2.5  $\mu$ g of pRSF91-GFP or pRSF91-CRBN<sup>109</sup> plasmid DNA was mixed with 250  $\mu$ l OptiMem and 7.5  $\mu$ l TransIT-LT1 (Mirus Bio) according to manufacture protocol. 48 hours post transfection cells were resuspended at  $0.15 \times 10^6$  cells/mL and plated on a 384 well plate with 50  $\mu$ l per well. HEK293T<sub>Cas9</sub> cells were transduced with sgRNAs targeting either DDB1 or Luciferase in pXPR003 backbone (GPP). After nine days of puromycin selection, cells were re-plated into a 96-well format with  $2 \times 10^4$  cells per well and CR8 and Roscovitine were serially diluted with D300e Digital Dispenser (Tecan Inc.). After 3 days of drug exposure, cell viability was assessed using the CellTiter-Glo luminescent assay (Promega, G7572) on an EnVision Multilabel Plate Reader (Perkin Elmer). Cell viabilities were calculated relative to DMSO controls.

### 6.2.30 Cyclin K stability reporter analysis with CRBN overexpression

HEK293T<sub>Cas9</sub> cells expressing the cycK<sub>eGFP</sub> degradation reporter were transiently transfected with pLX307-Luc or pLX307-CRBN (for flow experiment) as described above and 48 hours post infection treated with CR8 for 2 hours and analysed by flow cytometry.

## 6.3 Cyclin K degrader SAR (Chapter 3)

### 6.3.1 Compounds

The majority of compound bearing the "DS" prefix were synthesised in-house based on published synthetic procedures for related small molecules<sup>429-431</sup>. Compounds DS22, DS47, DS74 WX3, WX4-INT were synthesised by WuXi AppTec (Shanghai, China). Other small molecules were obtained from the following commercial vendors:

Compound	Company	Catalogue number
R-CR8	Tocris	3605
GSK2250665A	Tocris	5401
Flavopiridol	Enzo Life Sciences	ALX-430-161-M005
Roscovitine	MedChemExpress	HY-30237
THZ531	MedChemExpress	HY-103618
LDC00067	MedChemExpress	HY-15878
SR-4835	MedChemExpress	HY-130250
Danusertib	MedChemExpress	HY-10179
Dinaciclilb	MedChemExpress	HY-10492
BSJ-4-116	MedChemExpress	HY-139039
21195	Cayman Chemical Company	21195
919278	ProbeChem	PC-35532
DRF053	Sigma	D6946
dCeMM2	LabNetwork	STK119616
dCeMM3	Enamine	Z54609541
dCeMM4	Enamine	Z1126858802
Z7	Enamine	Z200170434
Z11	Enamine	Z27665843
Z12	Enamine	Z225160068
SNS032	MedChemExpress	HY-10008
NVP-2	MedChemExpress	HY-12214A

### 6.3.2 Protein expression and purification

Human wild-type and mutant versions of DDB1 (Uniprot entry Q16531), CDK12 (Q9NYV4, K965R) and CCNK (O75909) were subcloned into pAC-derived vectors<sup>416</sup> and recombinant proteins were expressed as N-terminal His<sub>6</sub>, His<sub>6</sub>-Spy, StrepII or StrepII-Avi fusions in *Trichoplusia ni* High Five insect cells using the baculovirus expression system (Invitrogen)<sup>417</sup>.

#### 6.3.2.1 Expression and purification of DDB1 constructs

Wild-type or mutant forms of full-length or beta-propeller B domain deletion ( $\Delta$ BPB: aa 396-705 deleted) constructs of His<sub>6</sub>-DDB1 $_{\Delta$ BPB, StrepII-Avi-DDB1, or His<sub>6</sub>-Spy-DDB1 were purified as previously described for DDB1-DCAF complexes<sup>418</sup>. Briefly, for His-tagged constructs, High Five insect cells expressing the above-mentioned proteins were lysed by sonication in 50 mM Tris-HCl (pH 8.0), 500 mM NaCl, 10% (v/v) glycerol, 10 mM imidazole, 0.25 mM tris(2-carboxyethyl)phosphine (TCEP), 0.1% (v/v) Triton X-100, 1 mM

phenylmethylsulfonylfluoride (PMSF), and 1 x protease inhibitor cocktail (Sigma). Following ultracentrifugation, the soluble fraction was passed over HIS-Select Ni<sup>2+</sup> affinity resin (Sigma), washed first with 50 mM Tris HCl pH 8.0, 500 mM NaCl, 10 % glycerol, 0.5 mM TCEP, 10 mM imidazole, then with 50 mM Tris HCl pH 8.0, 200 mM NaCl, 10 % glycerol, 0.5 mM TCEP, 10 mM imidazole, and finally eluted in 50 mM Tris-HCl (pH 8.0), 200 mM NaCl, 10% (v/v) glycerol, 0.5 mM TCEP, 250 mM imidazole. For crystallography, affinity tags were removed by overnight TEV protease treatment at a 1:50 (w/w) ratio. StrepII-tagged versions of DDB1 were affinity purified using Strep-Tactin Sepharose (IBA) omitting imidazole in lysis, wash and elution buffers, supplementing the elution buffer with 2.5 mM desthiobiotin (IBA GmbH), and using 50 mM Tris-HCl (pH 6.8) throughout. For ion exchange chromatography, affinity purified proteins were diluted in a 1:1 ratio with buffer *A* (50 mM Tris-HCl (pH 8.0), 10 mM NaCl, 2.5% (v/v) glycerol, 0.5 mM TCEP) and passed over an 8 mL Poros 50HQ column. Bound DDB1 was eluted by a linear salt gradient mixing buffer *A* and buffer *B* (50 mM Tris-HCl (pH 8.0), 1 M NaCl, 2.5% (v/v) glycerol, 0.5 mM TCEP) over 15 column volumes to a final ratio of 60% buffer *B*. DDB1-containing fractions were concentrated and subjected to size exclusion chromatography in 50 mM HEPES (pH 7.4), 200 mM NaCl, 2.5% (v/v) glycerol and 0.5 mM TCEP. Peak fractions were concentrated (to a final concentration of *circa* 200  $\mu$ M for DDB1 $\Delta$ B and 30  $\mu$ M for full-length DDB1), flash frozen in liquid nitrogen and stored at -80°C or directly used in crystallisation trials.

### **6.3.2.2 Expression and purification of CDK12-cyclin K for biophysical assays**

High Five insect cells were infected with CDK12 and cyclin K viruses at a 2:1 ratio to avoid excess expression of cyclin K alone. Cells co-expressing truncated versions of wild-type or mutant His<sub>6</sub>-CDK12 (aa 713-1052 or 713-1032) and His<sub>6</sub>-Spy-tagged or His<sub>6</sub>-Avi-tagged cycK (aa 1-267) were lysed by sonication in 50 mM Tris-HCl (pH 6.8), 500 mM NaCl, 10% (v/v) glycerol, 10 mM MgCl<sub>2</sub>, 10 mM imidazole, 0.25 mM tris(2-carboxyethyl)phosphine (TCEP), 0.1% (v/v) Triton X-100, 1 mM phenylmethylsulfonylfluoride (PMSF), and 1 x protease inhibitor cocktail (Sigma). Following ultracentrifugation, the soluble fraction was passed over HIS-Select Ni<sup>2+</sup> affinity resin (Sigma), washed first with 50 mM Tris-HCl (pH 6.8), 1 M NaCl, 10% (v/v) glycerol, 0.5 mM TCEP, 10 mM imidazole, then with 50 mM Tris HCl pH 6.8, 500 mM NaCl, 10 % (v/v) glycerol, 0.5 mM TCEP, 10 mM imidazole, and eluted in 50 mM Tris-HCl (pH 6.8), 200 mM NaCl, 10% (v/v) glycerol, 0.25 mM TCEP, 250 mM imidazole. For biophysical assays, the eluted fractions were directly concentrated and passed over a gel filtration column in 50 mM HEPES pH 7.4, 200 mM NaCl, 0.5 mM TCEP, 2.5 % (v/v) glycerol. Peak fractions were concentrated to *circa* 20  $\mu$ M, flash frozen in liquid nitrogen and stored at -80°C

### **6.3.2.3 Expression and purification of CDK12-cyclin K for crystallography**

High Five insect cells were infected with CDK12 and cyclin K viruses at a 2:1 ratio as described above. Cells co-expressing truncated versions of wild-type or mutant StrepII-CDK12 (aa 713-1052) and StrepII-cycK (aa 1-267) were lysed by sonication in 50 mM Tris-HCl (pH 6.8), 500 mM NaCl, 10% (v/v) glycerol, 10 mM MgCl<sub>2</sub>, 10 mM imidazole, 0.25 mM tris(2-carboxyethyl)phosphine (TCEP), 0.1% (v/v) Triton X-100, 1 mM

phenylmethylsulfonylfluoride (PMSF), and 1 x protease inhibitor cocktail (Sigma). Following ultracentrifugation, the soluble fraction was passed over Strep-Tactin Sepharose affinity resin (IBA), washed first with 50 mM Tris-HCl (pH 6.8), 1 M NaCl, 10% (v/v) glycerol, 0.5 mM TCEP, then with 50 mM Tris HCl pH 6.8, 500 mM NaCl, 10 % (v/v) glycerol, 0.5 mM TCEP, and eluted in 50 mM Tris-HCl (pH 6.8), 200 mM NaCl, 10% (v/v) glycerol, 0.5 mM TCEP, 2.5 mM desthiobiotin (IBA GmbH). Affinity tags were subsequently removed by overnight incubation with TEV protease at 1:50 (w/w). Before IEX chromatography, the protein solution was diluted in a 1:1 ratio with buffer A (50 mM Tris HCl pH 6.8, 10 mM NaCl, 2.5 % (v/v) glycerol, 0.5 mM TCEP) and passed over an 8 mL Poros 50HQ column to remove contaminants. To flow-through was then again diluted in a 1:1 ratio with buffer A and loaded onto an 8 mL Poros 50HS column. Bound proteins were eluted by a linear salt gradient mixing buffer A and buffer B (50 mM Tris-HCl (pH 6.8), 1 M NaCl, 2.5% (v/v) glycerol, 0.25 mM TCEP) over 15 column volumes to a final ratio of 80% buffer B. Poros 50HS peak fractions containing the CDK12-cycK complex were concentrated and subjected to size exclusion chromatography in 50 mM HEPES (pH 7.4), 200 mM NaCl, 2.5% (v/v) glycerol and 0.25 mM TCEP. The concentrated proteins (at *circa* 200  $\mu$ M) were flash frozen in liquid nitrogen and stored at -80°C or used directly in crystallisation trials.

### 6.3.3 Protein labelling

#### 6.3.3.1 Labelling with fluorophore-conjugated maleimides

The SpyTag/SpyCatcher system was employed as a mean of conjugation of the TR-FRET acceptor (Alexa<sub>488</sub>, Alexa<sub>647</sub>, or Cy5; each in a maleimide form) to the protein of interest as described before<sup>113,294</sup>. For this, mutant (S50C) SpyCatcher protein was purified according to published procedures<sup>432</sup> and was reduced by incubation with dithiothreitol (DTT) (8 mM) at 4°C for 1 hour. DTT was subsequently removed during a gel filtration step in 50 mM HEPES pH 7.4, 200 mM NaCl. Alexa<sub>488</sub>-C5-maleimide (or equivalent) was dissolved in 100% dimethyl sulfoxide (DMSO), mixed with the reduced SpyCatcher at a 4:1 ratio, and incubated for 3 h in a vacuum desiccator. The reaction was subsequently quenched with DTT (8 mM), and labelled SpyCatcher was purified through size exclusion chromatography in 50 mM HEPES pH 7.4, 200 mM NaCl, 0.5 mM TCEP. SpyCatcher<sub>Alexa488</sub> was concentrated to ~ 50  $\mu$ M, flash frozen in liquid nitrogen and stored at -80 °C. Purified Spy-tagged DDB1 or CDK12/cyclin K complex was mixed with an equimolar concentration of labelled SpyCatcher, incubated for 1 hour at room temperature and the labelling efficiency was monitored via SDS-PAGE. Labelled proteins were flash frozen and stored at -80°C (~20  $\mu$ M).

#### 6.3.3.2 Biotinylation

Conjugation of the TR-FRET donor (Strep-Tb or -Eu) to the protein of interest was achieved through enzymatic biotinylation. The biotinylation of purified full-length StrepII-Avi-DDB1 was performed as described before<sup>294</sup> (**section 6.2.23**). For the biotinylation of His<sub>6</sub>-CDK12/StrepII-Avi-cyclin K or His<sub>6</sub>-CDK12 His<sub>6</sub>/cyclin K-Avi, the complex at 20-50  $\mu$ M was incubated with final concentrations of 2.5  $\mu$ M BirA enzyme, 0.2 mM D-biotin,

and 20 mM ATP in 50 mM HEPES (pH 7.4), 200 mM NaCl, 10 mM MgCl<sub>2</sub> and 0.5 mM TCEP. The reaction was incubated for 10-12 hours at 4°C and biotinylated complex was purified by gel filtration chromatography (in 50 mM HEPES pH 7.4, 200 mM NaCl, 0.5 mM TCEP), concentrated to circa ~10-20 μM, flash frozen and stored at -80°C in small aliquots.

### 6.3.4 TR-FRET assay

For TR-FRET-based compound evaluation, a mixture of CDK12–cyclin K<sub>biotin</sub> (50 nM), Alexa488DDB1 (100 nM), terbium-coupled streptavidin (Invitrogen) (4 nM) in the optimised TR-FRET assay buffer (50 mM HEPES pH 7.4, 150 mM NaCl, 0.005% Tween 20, 0.5% DMSO, 0.05% BSA, 1 mM TCEP, 2 mM EDTA) was plated in a white 384-well microplate (Greiner, 784075) at a volume of 6 μL per well. Compounds were then dispensed digitally from DMSO stocks using the D300 dispenser (Tecan) to yield a 13-point dilution series of each small molecule, with the highest final concentration of 20 μM for poor molecular glues and 2 μM for potent compounds.

TR-FRET measurements were carried out using a PHERAstar FS microplate reader (BMG Labtech) equipped with a 337-520-490 filter set and 60 cycles were recorded over 1 h. A 70-μs delay was employed to reduce background fluorescence. The TR-FRET signal was obtained through calculating the ratio of 520 nm to 490 nm fluorescence and Prism 9 (GraphPad) was used for further data analysis. The curves resulting from plotting the TR-FRET against compound concentration were fitted using (**Equation 6.1**) in order to determine the half maximal effective concentrations (EC<sub>50</sub>).

$$Y = \text{Bottom} + \frac{(x^{\text{Hillslope}}) * (\text{Top} - \text{Bottom})}{EC_{50}^{\text{Hillslope}} + x^{\text{Hillslope}}}$$

Equation 6.1. [Agonist] vs. response equation from GraphPad Prism 9. *Top* and *Bottom* refer to the curve plateaus.

For dissociation constant (K<sub>d</sub>) determination, **Equation 6.2** was used:

$$Y = \text{Bottom} + (\text{Top} - \text{Bottom}) \left( \frac{K_d + L + x - \sqrt{((K_d + L + x)^2 - 4Lx)}}{2L} \right)$$

Equation 6.2. A quadratic binding equation used for K<sub>d</sub> determination. *Top* and *Bottom* refer to the curve plateaus. Ligand concentration (L) was constrained to 50 nM.

### 6.3.5 DDB1-CUL4-RBX1 reconstitution and *in vitro* CUL4 neddylation

*In vitro* CUL4 reconstitution and CUL4 neddylation were performed as described in **section 6.2.25**<sup>44</sup>. Briefly, His<sub>6</sub>-CUL4A/His<sub>6</sub>-RBX1 at 3.5 μM was incubated with His<sub>6</sub>-DDB1 at 3 μM in a reaction mixture containing

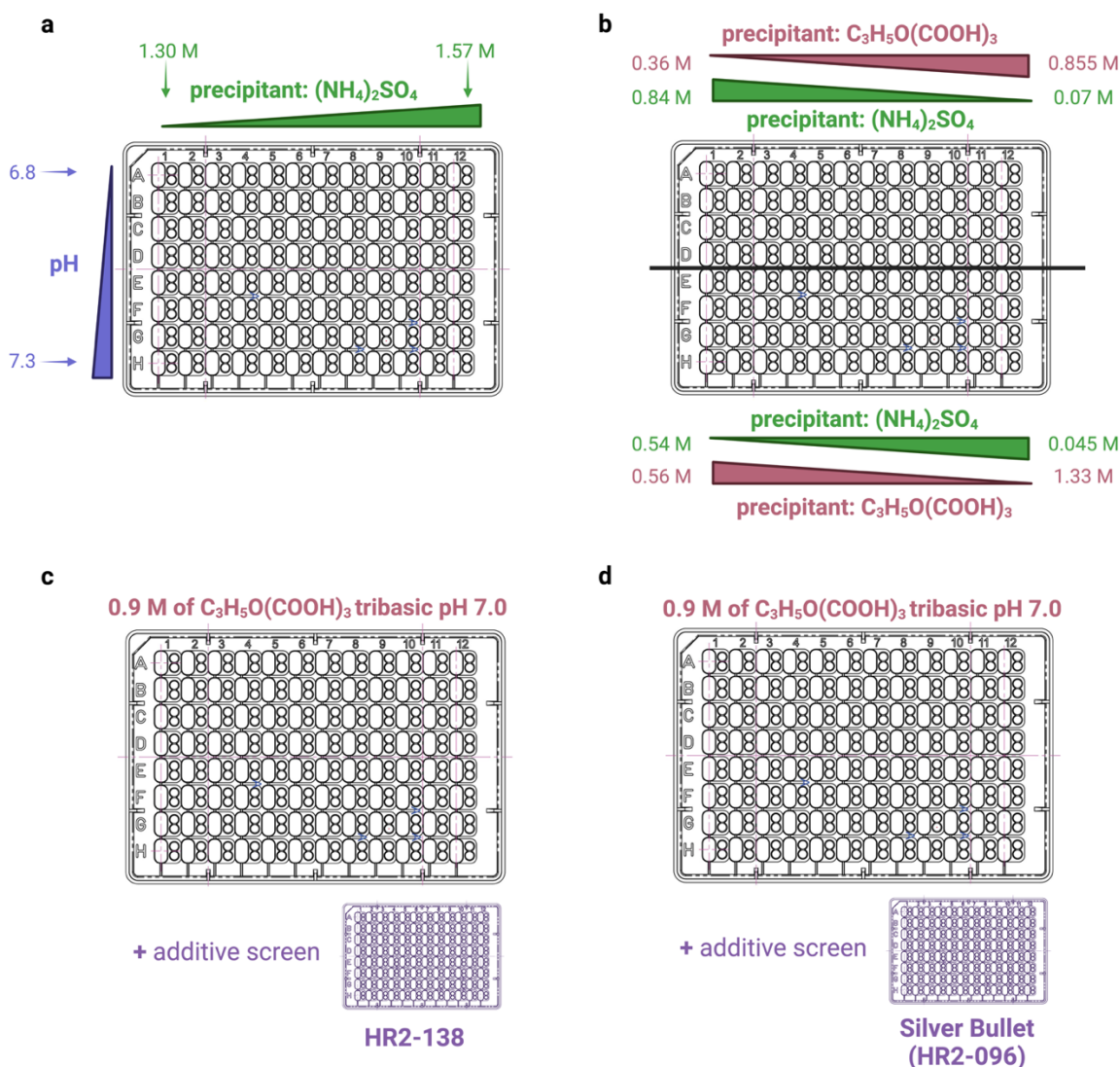
3.8  $\mu\text{M}$  NEDD8, 50 nM NAE1/UBA3 (E1), 30 nM UBC12 (E2), 1 mM ATP, 50 mM Tris (pH 7.5), 100 mM NaCl, 2.5 mM  $\text{MgCl}_2$ , 0.5 mM DTT and 5 % (v/v) glycerol for 1.5 hours at room temperature. Neddylated and gel filtration-purified DDB1-CUL4-RBX1 ( $_{\text{N}_8}$ DDB1-CUL4-RBX1) was concentrated to 5-15  $\mu\text{M}$ , flash frozen and stored at  $-80^\circ\text{C}$ .

### 6.3.6 Crystallographic methods

#### 6.3.6.1 Crystallisation and data collection

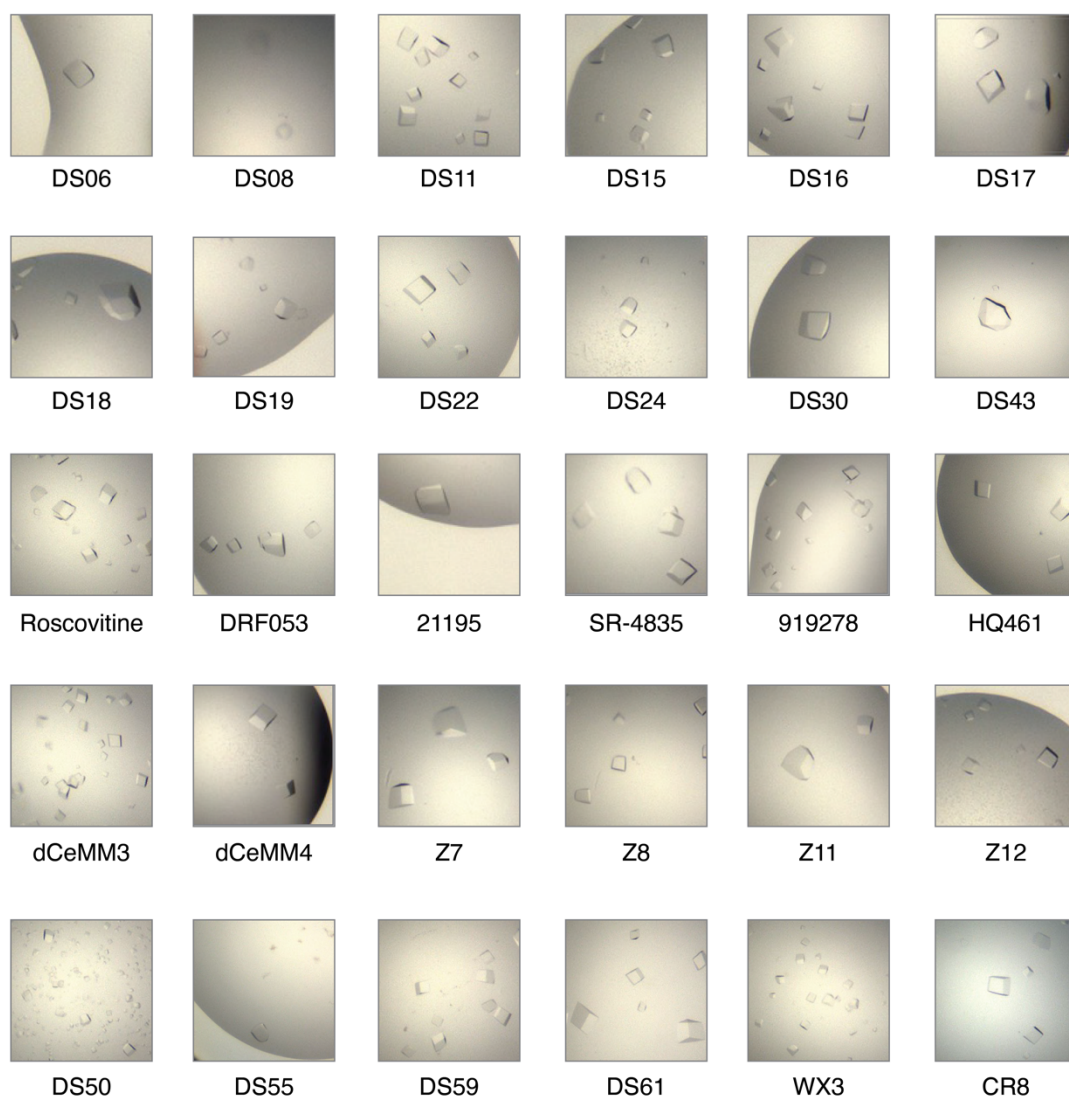
A crystallisation solution of purified and TEV-cleaved DDB1 $\Delta$ B (70  $\mu\text{M}$ ), CDK12-cyclin K (80  $\mu\text{M}$ ) and compound (80-120  $\mu\text{M}$ ) in SEC buffer (50 mM HEPES 7.4, 200 mM NaCl, 0.25 mM TCEP) was set up in two-drop 96-well plates (Swissci) for sitting-drop crystallization. A liquid handling system Phoenix (Art Robbins Instruments, Dunn Labortechnik GmbH) was used to pipette crystallization drops containing 200 nL of the protein complex solution mixed with 200 nL of reservoir solution. The plates were incubated at 20  $^\circ\text{C}$  and initial crystal formation was usually observed within two to five days.

Four different fine screens were formulated around the two conditions in which the DDB1 $\Delta$ B-CDK12-cyclin K-CR8 complex originally crystallised (0.9 M ammonium citrate tribasic pH 7.0 or 1.4 M ammonium sulphate, 70 mM HEPES pH 7.0) (**Figure 6.2**).

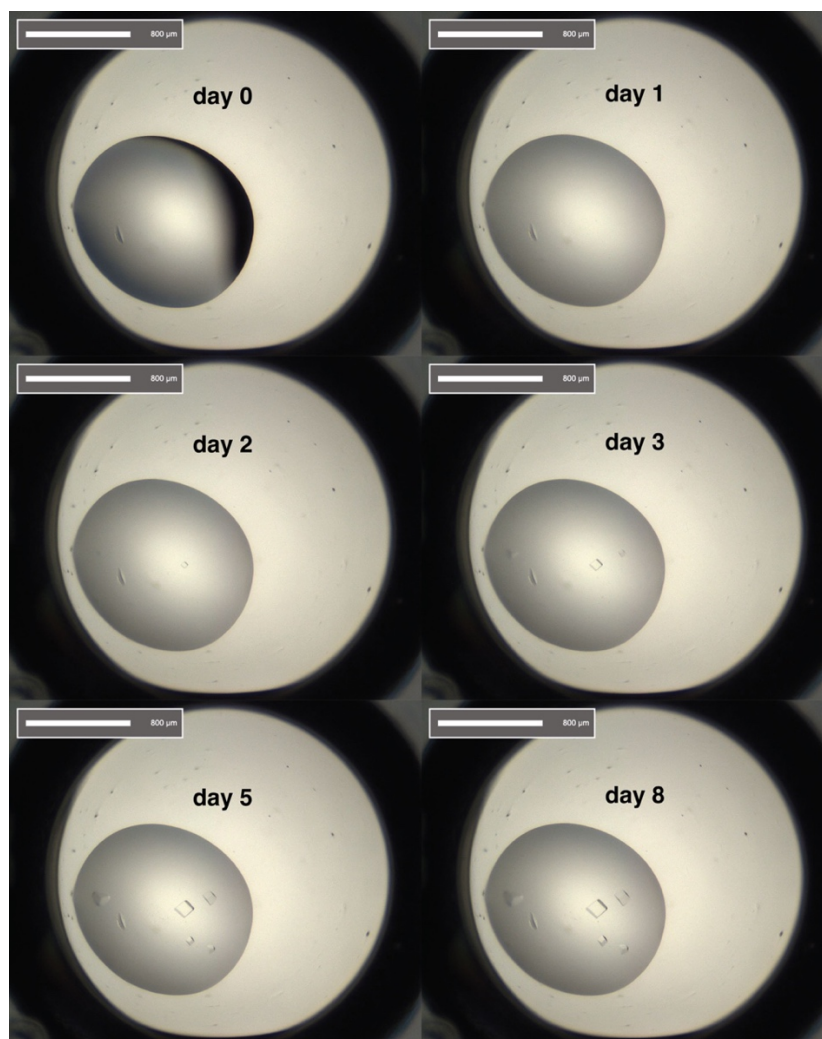


**Figure 6.2. Fine screens employed in crystallization trials.** **a**, Ammonium sulphate pH grid screen, with the precipitant concentration increasing within rows from left to right and the pH increasing within columns from top to bottom. **b**, Dual precipitant (ammonium sulphate - ammonium citrate tribasic) screen. In the upper half of the plate (rows A-D), the concentration of ammonium sulphate decreases from left to right, whereas that of ammonium citrate increases. In the lower half (rows E-H), an analogous strategy was employed but at different precipitant concentrations. **c**, Additive screen with ammonium citrate tribasic (0.9 M, pH 7.0) supplemented with commercially available additives in a 96-well plate format (HR2-138, Hampton Research). **d**, as in (c) but instead supplemented with the Silver Bullet additive screen (HR2-096, Hampton Research)





**Figure 6.3. Crystal morphology.** View of the pre-harvest drop (usually day 8 after plating) for the best-diffracting crystals for each compound.



**Figure 6.4.** Timeline of crystal growth illustrated for an example compound DS22. Crystals were usually harvested on days 9-10. The scale bar corresponds to 800  $\mu\text{m}$  and the drop has a volume of 400 nL.

Crystals designated for data collection were cryoprotected by supplementing the 400 nL drop with 1  $\mu\text{L}$  of reservoir solution containing ethylene glycol (25 % (v/v)) as cryoprotectant. Crystals were flash frozen in liquid nitrogen. Diffraction data collection was performed at the Swiss Light Source (SLS, Paul-Scherrer-Institute, Villigen, Switzerland) beamline PXII (X10SA) equipped with an Eiger2 16M detector (Dectris). The wavelength was set to 1  $\text{\AA}$  and the crystal was cooled to 100 K in a liquid nitrogen cryo-jet.

### 6.3.6.2 Structure determination and model building

Pipedream (GlobalPhasing Ltd.) was used to execute the following steps automatically. Data processing was performed using XDS<sup>433</sup>. Space group determination has been done through POINTLESS<sup>419</sup> and datasets were merged and scaled with AIMLESS<sup>419</sup>. Intensities were converted into structure factor amplitudes with STARANISO (Global Phasing)<sup>277</sup> by applying a weighted  $\text{CC}_{1/2} = 0.3$  resolution cut-off.

The structures of the different DDB1 $\Delta$ B-CDK12-cyclin K-compound complexes were solved by means of molecular replacement with PHASER<sup>421</sup> using a previously solved structure of the complex induced by CR8

(PDB 6TD3<sup>294</sup>; with ligand deleted) as the search model. Reciprocal space refinement was performed using phenix.refine<sup>423</sup>, followed by iterative real space refinement in COOT<sup>422</sup>. Ligand restraints were generated with jligand<sup>434</sup>. In case of ambiguous ligand density, additional simulations were performed with ISOLDE<sup>435</sup> to inform on the preferred ligand conformation. Structure validation was carried out with Molprobit<sup>436</sup>, and visualizations were generated with PyMol (Schrödinger). Interface analysis was performed using PISA<sup>372</sup> and pocket dimensions were evaluated using CASTp<sup>437</sup>.

### 6.3.7 Ligand docking

Ligand docking at the DDB1-CDK12 interface (structure 6TD3<sup>294</sup> minimised using OPLS3e force field, processed for receptor grid generation using default settings, and with the CR8 ligand removed) was performed with Glide (Schrödinger)<sup>438</sup>. For docking of the cyclin K degrader set evaluated in the study, ligands were input as a SMILE strings, prepared using LigPrep (OPLS3e force field), and docked using XP precision<sup>439</sup>.

For the R<sub>2</sub> group screen, CR8 derivatives with a library of 26,736 R<sub>2</sub> groups possessing commercially available precursors was docked at the DDB1-CDK12 interface. Briefly, the “In-Stock Decorators” library was downloaded from chem-space.com and filtered for aliphatic primary amines with the ligfilter module using LigPrep. The r\_group\_enumeration tool was then used to exchange the R<sub>2</sub> aminobutanol moiety in the CR8 ligand for the generated primary amine fragments from the previous LigPrep step. Docking was performed with SP precision.

For the docking of a kinase inhibitor library, a kinase inhibitor catalogue containing commercially available compounds was downloaded from ZINC, prepared using LigPrep (OPLS3e force field), and a database was created using Phase (Schrödinger). The library was docked with HTVS precision.

### 6.3.8 Lanthascreen kinase binding assay

Lanthascreen experiments were performed using commercially available reagents (Thermo Fischer) and according to manufacturer’s instructions. Briefly, a His-tagged kinase (CDK2-cyclin A, CDK9-cyclin T, or CDK12-cyclin K at a final concentration of 5 nM) was mixed with a biotin anti-His antibody (PV6089; final concentration of 2 nM), Lanthascreen Europium-Streptavidin (PV5899; 2 nM), and Tracer 236 (PV5899; final concentration of 30 nM for CDK9, 100 nM for CDK2 and 200 nM for CDK12) in kinase buffer A (PV3189). A serial dilution of a test compound from a DMSO stock was prepared in kinase buffer A. The kinase solution and compound dilution series were mixed in a black 384-well microplate (Greiner, 784076) to a final volume of 12-15 µL. After excitation of europium (Eu) fluorescence at 337 nm, emissions at 665 nm (Eu) and 620 nm (Alexa647) were measured with a 100 µs delay to reduce background fluorescence and the reactions were followed over 60 minutes using a PHERAstar FS microplate reader (BMG Labtech). TR-FRET signal was calculated through the 665:620 nm ratio and data were analysed with Prism 9 (GraphPad) using **Equation 6.3**

$$Y = \text{Bottom} + \frac{(\text{Top} - \text{Bottom})}{1 + \frac{\text{IC}_{50}^{\text{Hillslope}}}{x}}$$

**Equation 6.3.** [Inhibitor] vs. response equation from GraphPad Prism 9. *Top* and *Bottom* refer to the curve plateaus.

### 6.3.9 Mammalian cell culture

The human HEK293T cell lines were provided by the Genetic Perturbation Platform, Broad Institute and MDA-MB-231 cells were purchased from ATCC. HEK293T cell line were mycoplasma-negative and authenticated by STR profiling. HEK293T cells were cultured in Dulbecco's modified Eagle's medium (DMEM (Gibco) with 10% foetal bovine serum (FBS) (Invitrogen), glutamine (Invitrogen) and penicillin–streptomycin (Invitrogen) at 37 °C and 5% CO<sub>2</sub>.

### 6.3.10 Reporter vectors

The following reporters were used in this study: Cilantro 2 (PGK.BsmBICloneSite-10aaFlexibleLinker-eGFP.IRES.mCherry.cppt.EF1α.PuroR, Addgene #74450) for cyclin K degradation assay as previously reported<sup>294</sup>.

### 6.3.11 Cyclin K reporter stability analysis

HEK293T-Cas9 cells expressing the cyclin K - eGFP degradation reporter were resuspended at  $1 \times 10^6$  per mL and 50 μL of cell suspension was seeded in non-cell culture treated 384-well plates. Shortly after, cells were treated with drug or DMSO for 5h. Subsequently, cells were fixed with 50 μL of 4% paraformaldehyde solution (Chem Cruz, sc-281692) and stored at 4°C. The fluorescent signal was quantified by flow cytometry (LSRFortessa flow cytometer, BD Biosciences). Using FlowJo (flow cytometry analysis software, BD Biosciences), the geometric mean of the eGFP and mCherry fluorescent signal for round and mCherry-positive cells was calculated. The ratio of eGFP to mCherry was normalised to the average of three DMSO-treated controls.

### 6.3.12 Drug sensitivity assays

HEK293T-Cas9 cells were resuspended at  $0.15 \times 10^6$  per ml treated with DMSO or 100 nM MLN4924 and plated on 384-well plates. After 2h, the indicated drugs were dispensed with a D300 digital dispenser (Tecan). After three days of drug exposure, cell viability was assessed using the CellTiter-Glo luminescent assay (Promega, G7572) on an CLARIOstar Plus, MARS 3.4 (BMG LabTech). Cell viabilities were calculated relative to DMSO controls. The half-maximum effective concentration (EC<sub>50</sub>) values were derived from standard four-parameter log-logistic curves fitted with the 'dr4pl' R package.

### 6.3.13 Label-free quantitative mass spectrometry

#### 6.3.13.1 Sample preparation

MDA-MB-231 cells were treated with DMSO or a cyclin K degrader compound at 1  $\mu$ M for 5 hrs. Cells were harvested by centrifugation and washed with phosphate buffered saline (PBS) before snap freezing in liquid nitrogen. Cells were lysed by addition of lysis buffer (8 M Urea, 50 mM NaCl, 50 mM 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (EPPS) pH 8.5, Protease and Phosphatase inhibitors) and homogenization by bead beating (BioSpec) for three repeats of 30 seconds at 2400. Bradford assay was used to determine the final protein concentration in the clarified cell lysate. 50  $\mu$ g of protein for each sample was reduced, alkylated and precipitated using methanol/chloroform as previously described<sup>116</sup> and the resulting washed precipitated protein was allowed to air dry. Precipitated protein was resuspended in 4 M Urea, 50 mM HEPES pH 7.4, followed by dilution to 1 M urea with the addition of 200 mM EPPS, pH 8. Proteins were first digested with LysC (1:50; enzyme:protein) for 12 h at RT. The LysC digestion was diluted to 0.5 M Urea with 200 mM EPPS pH 8 followed by digestion with trypsin (1:50; enzyme:protein) for 6 h at 37 °C. Sample digests were acidified with formic acid to a pH of 2-3 prior to desalting using C18 solid phase extraction plates (SOLA, Thermo Fisher Scientific). Desalted peptides were dried in a vacuum-centrifuged and reconstituted in 0.1% formic acid for LC-MS analysis.

Data were collected using a TimsTOF Pro2 (Bruker Daltonics, Bremen, Germany) coupled to a nanoElute LC pump (Bruker Daltonics, Bremen, Germany) via a CaptiveSpray nano-electrospray source. Peptides were separated on a reversed-phase C<sub>18</sub> column (25 cm x 75  $\mu$ m ID, 1.6  $\mu$ M, IonOpticks, Australia) containing an integrated captive spray emitter. Peptides were separated using a 50 min gradient of 2 - 30% buffer B (acetonitrile in 0.1% formic acid) with a flow rate of 250 nL/min and column temperature maintained at 50 °C.

DDA was performed in Parallel Accumulation-Serial Fragmentation (PASEF) mode to determine effective ion mobility windows for downstream diaPASEF data collection<sup>440</sup>. The ddaPASEF parameters included: 100% duty cycle using accumulation and ramp times of 50 ms each, 1 TIMS-MS scan and 10 PASEF ramps per acquisition cycle. The TIMS-MS survey scan was acquired between 100 – 1700  $m/z$  and  $1/k_0$  of 0.7 - 1.3 V.s/cm<sup>2</sup>. Precursors with 1 – 5 charges were selected and those that reached an intensity threshold of 20,000 arbitrary units were actively excluded for 0.4 min. The quadrupole isolation width was set to 2  $m/z$  for  $m/z < 700$  and 3  $m/z$  for  $m/z > 800$ , with the  $m/z$  between 700-800  $m/z$  being interpolated linearly. The TIMS elution voltages were calibrated linearly with three points (Agilent ESI-L Tuning Mix Ions; 622, 922, 1,222  $m/z$ ) to determine the reduced ion mobility coefficients ( $1/K_0$ ). To perform diaPASEF, the precursor distribution in the DDA  $m/z$ -ion mobility plane was used to design an acquisition scheme for DIA data collection which included two windows in each 50 ms diaPASEF scan. Data was acquired using sixteen of these 25 Da precursor double window scans (creating 32 windows) which covered the diagonal scan line for doubly and triply charged precursors, with singly charged precursors able to be excluded by their position in the  $m/z$ -ion mobility plane. These precursor isolation windows were defined between 400 - 1200  $m/z$  and  $1/k_0$  of 0.7 - 1.3 V.s/cm<sup>2</sup>.

### 6.3.13.2 LC-MS data analysis

The diaPASEF raw file processing and controlling peptide and protein level false discovery rates, assembling proteins from peptides, and protein quantification from peptides was performed using directDIA analysis in Spectronaut 14 (Version 15.5.211111.50606, Biognosys). DirectDIA mode includes first extracting the DIA data into a collection of MS2 spectra which are searched using Spectronaut's Pulsar search engine. The search results are then used to generate a spectral library which is then employed for the targeted analysis of the DIA data. MS/MS spectra were searched against a Swissprot human database (January 2021). Database search criteria largely followed the default settings for directDIA including: tryptic with two missed cleavages, fixed carbamidomethylation of cysteine, and variable oxidation of methionine and acetylation of protein N-termini and precursor Q-value (FDR) cut-off of 0.01. Precursor quantification was performed using MS1 areas, cross run normalization was set to localised and imputation strategy was set to no imputation. Proteins with poor quality data were excluded from further analysis (summed abundance across channels of <100 and mean number of precursors used for quantification <2). Protein abundances were scaled using in-house scripts in the R framework<sup>441</sup> and statistical analysis was carried out using the limma package within the R framework<sup>442</sup>.

### 6.3.14 RNA sequencing

#### 6.3.14.1 Library preparation

300,000 MDA-MB-231 cells (ATCC) per technical replicate were treated with either 1 or 10  $\mu\text{M}$  of compound for 6 h. Pre-treatment of relevant conditions with 0.1  $\mu\text{M}$  MLN4924 (MedChemExpress) was performed for 2 h prior to compound treatment. RNA extraction was performed using TRIzol (Life Technologies Company) and quantified using Qubit (Thermo Fisher Scientific). RNA was then treated with DNase I, quantified using Qubit, and evaluated using a High Sensitivity RNA kit on a TapeStation (Agilent). ERCC spike-in (Thermo Fisher Scientific) was then added. Total RNA prep, cDNA synthesis, and library preparation was performed based on Illumina Stranded Total RNA prep, Ligation with Ribo-Zero protocol (Illumina). RNA was rRNA depleted and bead purified using RNAClean XP beads (Beckman Coulter). RNA was then fragmented followed by cDNA synthesis and bead-purification using AMPure XP beads (Beckman Coulter, New Jersey, USA). cDNA was then dual indexed, amplified, bead purified, evaluated using a DNA 1000 kit (Agilent) on a TapeStation, and quantified using a Qubit. Finally, libraries were pooled and sequenced on a NovaSeqS4 (Illumina), using 150bp paired-end reads.

#### 6.3.14.2 RNA-seq analysis

Sequencing reads were aligned to the human genome (*H.sapiens*-ENSEMBL-GRCh38.r91) and quantified by BrowserGenome (<https://www.nature.com/articles/nmeth.3615>). Differential gene expression analysis was performed in R (v4.1.1). In detail, raw counts were imported using the DESeqDataSetFromMatrix function implemented in the DESeq2 package (v.1.32.0,

<https://genomebiology.biomedcentral.com/articles/10.1186/s13059-014-0550-8>). After vst transformation and normalization using default parameters, normalised transcripts with an average expression lower than 10 across all replicates as well as a total overall expression lower than 200 across all samples were filtered out. Principal component analysis was performed using all transcripts as input. Differential expressed (DE) genes were computed comparing each sample against the DMSO control separately using a significance cut-off of  $p = 0.05$ , a log<sub>2</sub> fold change threshold of 0 and independent hypothesis weighting. Log<sub>2</sub> fold change shrunken DE genes passing the filtering criteria corresponding to the figure descriptions were displayed using EnhancedVolcano (v.1.10.0).

## 6.4 Fluorescence-based methods (Chapter 4)

### 6.4.1 TR-FRET

TR-FRET experiments were performed in three different titration modes: compound titrations (**Chapter 3**), forward titrations, and counter titrations. The details of buffer composition, protein labels, filter set, plate colour, and compound dispensing method employed are specified in individual figure legends.

#### 6.4.1.1 Compound titrations

For compound titrations into a mixture of labelled protein partners, a two-fold serial dilution of every compound was prepared in a chosen TR-FRET assay buffer (variable composition, specified in individual figure legends). This dilution series was then added to a mix of labelled CDK12–cyclin K (50 nM), labelled DDB1 (100 nM), terbium-coupled streptavidin (Invitrogen) (4 nM) and TR-FRET assay buffer in a 384-well microplate (Greiner, 784075/6). In every titration mode, 4  $\mu$ L of the respective dilution series was added to 4  $\mu$ L of the master mix unless otherwise stated.

TR-FRET measurements were carried out using a PHERAstar FS microplate reader (BMG Labtech) providing the possibility to record 60 data points of each well over 1 h, by first exciting donor fluorescence at 337 nm and then measuring donor and acceptor emissions at the appropriate wavelengths with a 70- $\mu$ s delay to reduce background fluorescence. TR-FRET signal was obtained through calculating the ratio of donor to acceptor fluorescence and Prism 9 (GraphPad) was used for further data analysis. For compound titrations, the resulting curves were fitted using an [Agonist] vs. response–equation (**Equation 6.1**) to determine the half maximal effective concentrations ( $EC_{50}$ ).

#### 6.4.1.2 Forward titrations

In the forward titration mode, a serial dilution of labelled DDB1 (2–2.5  $\mu$ M highest concentration), was prepared in TR-FRET assay buffer. The dilution series was titrated into a mix of labelled CDK12–cyclin K (50 nM), a test compound (10  $\mu$ M) and terbium-coupled streptavidin (Invitrogen) (4 nM) in TR-FRET assay buffer in a 384-well microplate (Greiner, 784075/6). In the case of forward titrations, the data were also fitted with **Equation 6.1** to determine  $EC_{50}$  values.

### 6.4.1.3 Counter titrations

Counter titrations with unlabelled proteins were carried out by mixing labelled CDK12–cyclin K (50 nM), DDB1 (100 nM), the compounds to be tested (10  $\mu$ M) and terbium-coupled streptavidin (Invitrogen) (4 nM) in TR-FRET assay buffer. After incubation for 15 minutes at room temperature, increasing amounts of unlabelled DDB1 (0-20  $\mu$ M) were added. TR-FRET data were acquired as described above and the 520/490 nm ratios were plotted to calculate the half maximal inhibitory concentrations ( $IC_{50}$ ) using **Equation 6.3** from *Prism 9* (GraphPad).

## 6.4.2 Interactions of Oct4-Sox2 with DNA and nucleosomes

### 6.4.2.1 Human octamer histones expression, purification, and reconstitution

Human histones were expressed and purified as described previously<sup>443</sup>. Lyophilised histones were mixed at equimolar ratios in 20 mM Tris-HCl (pH 7.5) buffer, containing 7 M guanidine hydrochloride and 20 mM 2-mercaptoethanol. Samples were dialyzed against 10 mM Tris-HCl (pH 7.5) buffer, containing 2 M NaCl, 1 mM EDTA, and 2 mM 2-mercaptoethanol. The resulting histone complexes were purified by size exclusion chromatography (Superdex 200; GE Healthcare).

### 6.4.2.2 DNA preparation

DNA for medium to large scale individual nucleosome purifications was generated by Phusion (Thermo Fisher) PCR amplification (on average 2 x 96 well plates) or large-scale plasmid purification (GigaPrep, Invitrogen) followed by EcoRV-HF (New England Biosciences) blunt-end restriction enzyme cleavage. The resulting DNA fragment was purified by a monoQ column (GE Healthcare). The unmodified 601 Widom sequence was purified with a large-scale plasmid purification using a high copy plasmid containing 32 copies of the 601 sequence previously cloned into pUC19 vector<sup>444,445</sup>. All purified DNA was concentrated and stored at -20°C in 10 mM Tris-HCl pH 7.5 until use.

### 6.4.2.3 Nucleosome assembly

The DNA and the histone octamer complex were mixed in a 1:1.5 molar ratio in the presence of 2 M KCl. The samples were dialyzed against refolding buffer (RB) high (10 mM Tris-HCl (pH 7.5), 2 M KCl, 1 mM EDTA, and 1 mM DTT). The KCl concentration was gradually reduced from 2 M to 0.25 M using a peristaltic pump with RB low (10 mM Tris-HCl (pH 7.5), 250 mM KCl, 1 mM EDTA, and 1 mM DTT) at 4°C. Samples were further dialyzed against RB low buffer at 4°C overnight. After dialysis, nucleosomes were incubated at 55°C for 2 hours. Large scale assemblies of individual nucleosomes were purified by on a monoQ 5/50 ion exchange gradient (GE Healthcare) and desalted using a Zeba spin column (Thermo Fisher) into 20mM Tris-HCl pH 7.5 and 500  $\mu$ M TCEP and stored at 4°C.



#### 6.4.2.4 Protein expression and purification of OCT4 and SOX2

Human full-length OCT4 (residues 1 – 360), OCT4 DNA binding domain (residues 134-290) or human SOX2 DNA binding domain (residues 37 – 118) were subcloned into pAC-derived vectors<sup>411</sup> containing an N-terminal StrepII tag. An additional N-terminal EGFP tag and C-terminal sortase-6XHIS tag (LPETGGHHHHHH) were fused in frame with OCT4 full-length to improve purification and was also used for cryo-EM sample preparation. Recombinant proteins were expressed in 2- 4 L cultures of *Trichoplusia ni* High Five cells using the Bac-to-Bac system (Thermo Fisher). Cells were cultured at 27°C, harvested 2 days after infection, resuspended in lysis buffer (50 mM Tris-HCl pH 8.0, 1M NaCl, 100 μM phenylmethylsulfonyl fluoride, 1 × protease inhibitor cocktail (Sigma), 250 μM TCEP), and lysed by sonication. The supernatant was harvested, and the proteins were purified by Streptactin affinity chromatography (IBA), and then purified by Heparin ion exchange chromatography (GE Healthcare). All proteins were further purified by size exclusion chromatography (Superdex 200; GE Healthcare) in GF buffer (20 mM HEPES pH 7.4, 150 mM NaCl, 5% glycerol, 500 μM TCEP) as a last purification step. The purified proteins were concentrated and stored at -80°C.

#### 6.4.2.5 Determination of OCT4 and/or SOX2 binding affinity for DNA by FP forward titrations

Fic-labelled DNA containing the canonical OCT4-SOX2 motif (5'-Fic-GACCTTTGTTATGCAAATTAA-3') was used as a fluorescent tracer. Increasing amounts of OCT4 or SOX2 (0.3-2500 nM) were mixed with tracer (10 nM final concentration) in a 384-well microplate (Greiner, 784076) and incubated for 15 min at room temperature. The interaction was measured in a buffer containing 15 mM HEPES pH 7.4, 250 μM TCEP, 75 mM NaCl, 10 mM KCl, 1 mM MgCl<sub>2</sub>, 0.1% (v/v) pluronic acid. Changes in fluorescence polarization were monitored by a PHERAstar FS microplate reader (BMG Labtech) equipped with a fluorescence polarization filter unit. The polarization units were converted to fraction bound as described previously<sup>446</sup>. The fraction bound was plotted versus OCT4 or SOX2 concentration and fitted assuming a one-to-one binding model to determine the dissociation constant (K<sub>d</sub>) using Prism 7 (GraphPad). Since the oligonucleotide that was used contained a fluorescent label, we refer to these as apparent K<sub>d</sub> (K<sub>d</sub>(app)). All measurements were performed in triplicates.

#### 6.4.2.6 Determination of OCT4 and/or SOX2 binding affinity for nucleosomes by FP counter titrations

For the competitive titration assays, the OCT4 or SOX2 bound to the fluorescent oligo tracer was backtitrated with unlabelled oligo or nucleosomes containing the canonical motif at different sites. The competitive titration experiments were carried out by mixing tracer (10 nM), OCT4 (300 nM) or SOX2 (150 nM) and increasing concentration of different nucleosomes or DNA (0 - 3.2μM). The fraction bound vs. the nucleosome or DNA concentration were fitted with a nonlinear regression curve to obtain the IC<sub>50</sub> values in Prism 7 (GraphPad) (**Equation 6.3**). Two to three technical replicates were measured for each reaction. We note that the assay does not allow us to differentiate between a specific and nonspecific contribution to the binding.

### 6.4.3 Interactions of cGAS with DNA and nucleosomes

#### 6.4.3.1 Determination of cGAS binding affinity for DNA by fluorescence polarisation forward titrations

Fic-labelled 21-bp dsDNA oligonucleotide (5'-Fic-GACCTTTGTTATGCAACCTAA-3') was used as a fluorescent tracer. Increasing amounts of WT or K394E cGAS (0.3–2500 nM) were mixed with tracer (10 nM final concentration) in a 384-well microplate (784076, Greiner) at room temperature. The interaction was measured in a buffer containing 20 mM HEPES pH 7.4, 500  $\mu$ M TCEP, 40 mM NaCl, 10 mM KCl and 0.1 % (v/v) pluronic acid. PHERAstar FS microplate reader (BMG Labtech) equipped with a fluorescence polarization filter unit was used to determine the changes in fluorescence polarization. The polarization units were converted to fraction bound as described previously<sup>446</sup>. The fraction bound was plotted versus cGAS concentration and fitted assuming a 1:1 binding model to determine the dissociation constant ( $K_d$ ) using Prism 7 (GraphPad). Since the oligonucleotide that was used contained a fluorescent label, we refer to these as apparent  $K_d$  ( $K_{app}$ ). All measurements were performed in triplicates.

#### 6.4.3.2 Determination of cGAS binding affinity for nucleosomes by fluorescence polarisation counter-titrations

For the competitive titration assays, the cGAS bound to the fluorescent oligo tracer was backtitrated with unlabelled dsDNA (21, 147, 167 and 227 bp) or nucleosomes (146, 167 and 227 bp). These counter-titration experiments were carried out by mixing tracer (10 nM) and cGAS (300 nM), and titrating increasing concentration of the unlabelled competitor (0–2.5  $\mu$ M). The fraction bound was plotted versus competitor concentration and the data were fitted with a non-linear regression curve to obtain the  $IC_{50}$  values in Prism 7 (GraphPad) (**Equation 6.3**). At least two technical replicates were performed per experiment.

## References

1. The top 15 best-selling cancer drugs in 2022 | Fierce Pharma. <https://www.fiercepharma.com/special-report/special-report-top-15-best-selling-cancer-drugs-2022>.
2. Hershko, A. & Ciechanover, A. Mechanisms of Intracellular Protein Breakdown. <http://dx.doi.org/10.1146/annurev.bi.51.070182.002003> **51**, 335–364 (2003).
3. Ciechanover, A., Heller, H., Elias, S., Haas, A. L. & Hershko, A. ATP-dependent conjugation of reticulocyte proteins with the polypeptide required for protein degradation. *Proc Natl Acad Sci U S A* **77**, 1365–1368 (1980).
4. Ciechanover, A., Hod, Y. & Hershko, A. A heat-stable polypeptide component of an ATP-dependent proteolytic system from reticulocytes. *Biochem Biophys Res Commun* **81**, 1100–1105 (1978).
5. Goldknopf, I. L. & Busch, H. Isopeptide linkage between nonhistone and histone 2A polypeptides of chromosomal conjugate-protein A24. *Proc Natl Acad Sci U S A* **74**, 864–868 (1977).
6. Hershko, A. & Ciechanover, A. The ubiquitin system. *Annu Rev Biochem* **67**, 425–479 (1998).
7. Varshavsky, A. The ubiquitin system, an Immense Realm. *Annu Rev Biochem* **81**, 167–176 (2012).
8. Hershko, A., Heller, H., Elias, S. & Ciechanover, A. Components of ubiquitin-protein ligase system. Resolution, affinity purification, and role in protein breakdown. *Journal of Biological Chemistry* **258**, 8206–8214 (1983).
9. Haas, A. L. & Rose, I. A. The mechanism of ubiquitin activating enzyme. A kinetic and equilibrium analysis. *Journal of Biological Chemistry* **257**, 10329–10337 (1982).
10. Pickart, C. M. & Rose, I. A. Functional heterogeneity of ubiquitin carrier proteins. *Journal of Biological Chemistry* **260**, 1573–1581 (1985).
11. Huang, D. T. *et al.* Basis for a ubiquitin-like protein thioester switch toggling E1-E2 affinity. *Nature* **445**, 394–398 (2007).
12. Komander, D. & Rape, M. The ubiquitin code. *Annu Rev Biochem* **81**, 203–229 (2012).
13. Mevissen, T. E. T. & Komander, D. Mechanisms of Deubiquitinase Specificity and Regulation. *Annu Rev Biochem* **86**, 159–192 (2017).
14. George, A. J., Hoffiz, Y. C., Charles, A. J., Zhu, Y. & Mabb, A. M. A comprehensive atlas of E3 ubiquitin ligase mutations in neurological disorders. *Front Genet* **9**, 29 (2018).
15. Stewart, M. D., Ritterhoff, T., Klevit, R. E. & Brzovic, P. S. E2 enzymes: more than just middle men. *Cell Res* **26**, 423 (2016).
16. Borden, K. L. B. & Freemont, P. S. The RING finger domain: a recent example of a sequence-structure family. *Curr Opin Struct Biol* **6**, 395–401 (1996).
17. Scheffner, M., Nuber, U. & Huibregtse, J. M. Protein ubiquitination involving an E1–E2–E3 enzyme ubiquitin thioester cascade. *Nature* *1995* **373:6509** **373**, 81–83 (1995).

18. Soucy, T. A. *et al.* An inhibitor of NEDD8-activating enzyme as a new approach to treat cancer. *Nature* **458**, 732–736 (2009).
19. Scrima, A. Structural Basis of UV DNA-Damage Recognition by the DDB1–DDB2 Complex Andrea. *Cell* **135**, 1213–1223 (2008).
20. Duda, D. M. *et al.* Structural regulation of cullin-RING ubiquitin ligase complexes. *Current Opinion in Structural Biology* vol. 21 257–264 Preprint at <https://doi.org/10.1016/j.sbi.2011.01.003> (2011).
21. Zimmerman, E. S., Schulman, B. A. & Zheng, N. Structural assembly of cullin-RING ubiquitin ligase complexes. *Current Opinion in Structural Biology* vol. 20 714–721 Preprint at <https://doi.org/10.1016/j.sbi.2010.08.010> (2010).
22. Pan, Z. Q., Kentsis, A., Dias, D. C., Yamoah, K. & Wu, K. Nedd8 on cullin: building an expressway to protein destruction. *Oncogene* **23**, 1985–1997 (2004).
23. Baek, K. *et al.* NEDD8 nucleates a multivalent cullin-RING-UBE2D ubiquitin ligation assembly. *Nature* **578**, 461–466 (2020).
24. Duda, D. M. *et al.* Structural insights into NEDD8 activation of cullin-RING ligases: conformational control of conjugation. *Cell* **134**, 995–1006 (2008).
25. Furukawa, M., Zhang, Y., McCarville, J., Ohta, T. & Xiong, Y. The CUL1 C-terminal sequence and ROC1 are required for efficient nuclear accumulation, NEDD8 modification, and ubiquitin ligase activity of CUL1. *Mol Cell Biol* **20**, 8185–8197 (2000).
26. Gong, L. & Yeh, E. T. H. Identification of the activating and conjugating enzymes of the NEDD8 conjugation pathway. *J Biol Chem* **274**, 12036–12042 (1999).
27. Schwechheimer, C. *et al.* Interactions of the COP9 signalosome with the E3 ubiquitin ligase SCFTIR1 in mediating auxin response. *Science* **292**, 1379–1382 (2001).
28. Wei, N. & Deng, X. W. The COP9 signalosome. *Annu Rev Cell Dev Biol* **19**, 261–286 (2003).
29. Lingaraju, G. M. *et al.* Crystal structure of the human COP9 signalosome. *Nature* **512**, 161–165 (2014).
30. Zheng, J. *et al.* CAND1 binds to unneddylated CUL1 and regulates the formation of SCF ubiquitin E3 ligase complex. *Mol Cell* **10**, 1519–1526 (2002).
31. Reichermeier, K. M. *et al.* PIKES Analysis Reveals Response to Degradation and Key Regulatory Mechanisms of the CRL4 Network. *Mol Cell* **77**, 1092–1106.e9 (2020).
32. Angers, S. *et al.* Molecular architecture and assembly of the DDB1-CUL4A ubiquitin ligase machinery. *Nature* **443**, 590–593 (2006).
33. Hannah, J. & Zhou, P. Distinct and overlapping functions of the cullin E3 ligase scaffolding proteins CUL4A and CUL4B. *Gene* **573**, 33–45 (2015).
34. He, Y. J., McCall, C. M., Hu, J., Zeng, Y. & Xiong, Y. DDB1 functions as a linker to recruit receptor WD40 proteins to CUL4-ROC1 ubiquitin ligases. *Genes Dev* **20**, 2949–2954 (2006).
35. Jin, J., Arias, E. E., Chen, J., Harper, J. W. & Walter, J. C. A Family of Diverse Cul4-Ddb1-Interacting Proteins Includes Cdt2, which Is Required for S Phase Destruction of the Replication Factor Cdt1. *Mol Cell* **23**, 709–721 (2006).

36. Bennett, E. J., Rush, J., Gygi, S. P. & Harper, J. W. Dynamics of Cullin-RING Ubiquitin Ligase Network Revealed by Systematic Quantitative Proteomics. *Cell* **143**, 951–965 (2010).
37. Li, T., Chen, X., Garbutt, K. C., Zhou, P. & Zheng, N. Structure of DDB1 in Complex with a Paramyxovirus V Protein: Viral Hijack of a Propeller Cluster in Ubiquitin Ligase. *Cell* **124**, 105–117 (2006).
38. Fischer, E. S. *et al.* The molecular basis of CRL4 DDB2/CSA ubiquitin ligase architecture, targeting, and activation. *Cell* **147**, 1024–1039 (2011).
39. Fischer, E. S. *et al.* Structure of the DDB1-CRBN E3 ubiquitin ligase in complex with thalidomide. *Nature* **512**, 49–53 (2014).
40. Li, T., Robert, E. I., van Breugel, P. C., Strubin, M. & Zheng, N. A promiscuous  $\alpha$ -helical motif anchors viral hijackers and substrate receptors to the CUL4–DDB1 ubiquitin ligase machinery. *Nat Struct Mol Biol* **17**, 105–111 (2010).
41. Wu, Y. *et al.* The DDB1–DCAF1–Vpr–UNG2 crystal structure reveals how HIV-1 Vpr steers human UNG2 toward destruction. *Nature Structural & Molecular Biology* 2016 23:10 **23**, 933–940 (2016).
42. Du, X. *et al.* Structural Basis and Kinetic Pathway of RBM39 Recruitment to DCAF15 by a Sulfonamide Molecular Glue E7820. *Structure* **27**, 1625-1633.e3 (2019).
43. Wu, Y. *et al.* The DDB1–DCAF1–Vpr–UNG2 crystal structure reveals how HIV-1 Vpr steers human UNG2 toward destruction. *Nat Struct Mol Biol* **23**, 933–939 (2016).
44. Petzold, G., Fischer, E. S. & Thomä, N. H. Structural basis of lenalidomide-induced CK1 $\alpha$  degradation by the CRL4 CRBN ubiquitin ligase. *Nature* **532**, 127–130 (2016).
45. Sugasawa, K. *et al.* UV-induced ubiquitylation of XPC protein mediated by UV-DDB-ubiquitin ligase complex. *Cell* **121**, 387–400 (2005).
46. Montagut, A. M. *et al.* Recent advances in the pharmacological targeting of ubiquitin-regulating enzymes in cancer. *Semin Cell Dev Biol* (2022) doi:10.1016/J.SEMCDB.2022.02.007.
47. Adams, J. *et al.* Potent and selective inhibitors of the proteasome: Dipeptidyl boronic acids. *Bioorg Med Chem Lett* **8**, 333–338 (1998).
48. Richardson, P. G. *et al.* A Phase 2 Study of Bortezomib in Relapsed, Refractory Myeloma. *n engl j med* **26**, 2609–2626 (2003).
49. Adams, J. & Kauffman, M. Development of the proteasome inhibitor Velcade (Bortezomib). *Cancer Invest* **22**, 304–311 (2004).
50. Manasanch, E. E. & Orłowski, R. Z. Proteasome inhibitors in cancer therapy. *Nature Reviews Clinical Oncology* 2017 14:7 **14**, 417–433 (2017).
51. Wertz, I. E. & Wang, X. From Discovery to Bedside: Targeting the Ubiquitin System. *Cell Chem Biol* (2018) doi:10.1016/j.chembiol.2018.10.022.
52. Hyer, M. L. *et al.* A small-molecule inhibitor of the ubiquitin activating enzyme for cancer treatment. *Nature Medicine* 2018 24:2 **24**, 186–193 (2018).

53. Soucy, T. A. *et al.* An inhibitor of NEDD8-activating enzyme as a new approach to treat cancer. *Nature* 2009 458:7239 **458**, 732–736 (2009).
54. Kanda, H. & Yamawaki, K. Bardoxolone methyl: drug development for diabetic kidney disease. *Clin Exp Nephrol* **24**, 857–864 (2020).
55. Vassilev, L. T. *et al.* In Vivo Activation of the p53 Pathway by Small-Molecule Antagonists of MDM2. *Science (1979)* **303**, 844–848 (2004).
56. Wang, S. & Chen, F. E. Small-molecule MDM2 inhibitors in clinical trials for cancer therapy. *Eur J Med Chem* **236**, 114334 (2022).
57. Diehl, C. J. & Ciulli, A. Discovery of small molecule ligands for the von Hippel-Lindau (VHL) E3 ligase and their use as inhibitors and PROTAC degraders. *Chem Soc Rev* (2022) doi:10.1039/D2CS00387B.
58. Amaravadi, R. K. *et al.* A Phase I study of the SMAC-mimetic birinapant in adults with refractory solid tumors or lymphoma. *Mol Cancer Ther* **14**, 2569–2575 (2015).
59. Flygare, J. A. *et al.* Discovery of a Potent Small-Molecule Antagonist of Inhibitor of Apoptosis (IAP) Proteins and Clinical Candidate for the Treatment of Cancer (GDC-0152). (2012) doi:10.1021/jm300060k.
60. Wang, X. *et al.* The proteasome deubiquitinase inhibitor VLX1570 shows selectivity for ubiquitin-specific protease-14 and induces apoptosis of multiple myeloma cells. *Scientific Reports* 2016 6:1 **6**, 1–15 (2016).
61. Schauer, N. J. *et al.* Selective USP7 inhibition elicits cancer cell killing through a p53-dependent mechanism. *Scientific Reports* 2020 10:1 **10**, 1–15 (2020).
62. Futran, A. *et al.* Abstract 1338: Discovery of novel, potent USP7 inhibitors that upregulate p53 leading to anti-proliferative effects in cancer cells. *Cancer Res* **81**, 1338–1338 (2021).
63. Schlierf, A. *et al.* Targeted inhibition of the COP9 signalosome for treatment of cancer. *Nat Commun* **7**, (2016).
64. Zhou, H. J. *et al.* Discovery of a First-in-Class, Potent, Selective, and Orally Bioavailable Inhibitor of the p97 AAA ATPase (CB-5083). *J Med Chem* **58**, 9480–9497 (2015).
65. Scott, D. E., Bayly, A. R., Abell, C. & Skidmore, J. Small molecules, big targets: Drug discovery faces the protein-protein interaction challenge. *Nat Rev Drug Discov* **15**, 533–550 (2016).
66. Giordanetto, F., Schäfer, A. & Ottmann, C. Stabilization of protein-protein interactions by small molecules. *Drug Discov Today* **19**, 1812–1821 (2014).
67. Sabatini, D. M., Erdjument-Bromage, H., Lui, M., Tempst, P. & Snyder, S. H. RAFT1: A mammalian protein that binds to FKBP12 in a rapamycin-dependent fashion and is homologous to yeast TORs. *Cell* **78**, 35–43 (1994).
68. Schreiber, S. L. & Crabtree, G. R. The mechanism of action of cyclosporin A and FK506. *Immunology Today* vol. 13 136–142 Preprint at [https://doi.org/10.1016/0167-5699\(92\)90111-J](https://doi.org/10.1016/0167-5699(92)90111-J) (1992).
69. Sabers, C. J. *et al.* Isolation of a protein target of the FKBP12-rapamycin complex in mammalian cells. *Journal of Biological Chemistry* **270**, 815–822 (1995).

70. Liu, J. *et al.* Calcineurin is a common target of cyclophilin-cyclosporin A and FKBP-FK506 complexes. *Cell* **66**, 807–815 (1991).
71. Gerry, C. J. & Schreiber, S. L. Unifying principles of bifunctional, proximity-inducing small molecules. *Nat Chem Biol* **16**, 369–378 (2020).
72. Schreiber, S. L. The Rise of Molecular Glues. *Cell* **184**, 3–9 (2021).
73. Schreiber, S. L. Chemistry and biology of the immunophilins and their immunosuppressive ligands. *Science* **251**, 283–287 (1991).
74. Liu, J. *et al.* Calcineurin is a common target of cyclophilin-cyclosporin A and FKBP-FK506 complexes. *Cell* **66**, 807–815 (1991).
75. Brown, E. J. *et al.* A mammalian protein targeted by G1-arresting rapamycin-receptor complex. *Nature* **369**, 756–758 (1994).
76. Sabatini, D. M., Erdjument-Bromage, H., Lui, M., Tempst, P. & Snyder, S. H. RAFT1: a mammalian protein that binds to FKBP12 in a rapamycin-dependent fashion and is homologous to yeast TORs. *Cell* **78**, 35–43 (1994).
77. Choi, J., Chen, J., Schreiber, S. L. & Clardy, J. Structure of the FKBP12-Rapamycin Complex Interacting with Binding Domain of Human FRAP. *Science (1979)* **273**, 239–242 (1996).
78. Juvvadi, P. R. *et al.* Harnessing calcineurin-FK506-FKBP12 crystal structures from invasive fungal pathogens to develop antifungal agents. *Nature Communications* *2019 10:1* **10**, 1–18 (2019).
79. Jin, L. & Harrison, S. C. Crystal structure of human calcineurin complexed with cyclosporin A and human cyclophilin. *Proc Natl Acad Sci U S A* **99**, 13522–13526 (2002).
80. Mahon, C., Krogan, N., Craik, C. & Pick, E. Cullin E3 Ligases and Their Rewiring by Viral Factors. *Biomolecules* **4**, 897–930 (2014).
81. Ahn, J. *et al.* HIV-1 Vpr loads uracil DNA glycosylase-2 onto DCAF1, a substrate recognition subunit of a cullin 4A-RING E3 ubiquitin ligase for proteasome-dependent degradation. *Journal of Biological Chemistry* **285**, 37333–37341 (2010).
82. Zemke, N. *et al.* Adenovirus E1A Binding to DCAF10 Targets Degradation of AAA+ ATPases Required for Quaternary Assembly of Multiprotein Machines and Innate Immunity. *bioRxiv* (2020) doi:10.1101/2020.12.16.423151.
83. Glotzer, M., Murray, A. W. & Kirschner, M. W. Cyclin is degraded by the ubiquitin pathway. *Nature* **349**, 132–138 (1991).
84. Yamano, H., Tsurumi, C., Gannon, J. & Hunt, T. The role of the destruction box and its neighbouring lysine residues in cyclin B for anaphase ubiquitin-dependent proteolysis in fission yeast: Defining the D-box receptor. *EMBO Journal* **17**, 5670–5678 (1998).
85. Tan, X. *et al.* Mechanism of auxin perception by the TIR1 ubiquitin ligase. *Nature* **446**, 640–645 (2007).
86. Sheard, L. B. *et al.* Jasmonate perception by inositol-phosphate-potentiated COI1-JAZ co-receptor. *Nature* **468**, 400–407 (2010).

87. Tal, L., Gil, M. X. A., Guercio, A. M. & Shabek, N. Structural aspects of plant hormone signal perception and regulation by ubiquitin ligases. *Plant Physiol* **182**, 1537–1554 (2020).
88. Nishimura, K., Fukagawa, T., Takisawa, H., Kakimoto, T. & Kanemaki, M. An auxin-based degron system for the rapid depletion of proteins in nonplant cells. *Nat Methods* **6**, 917–922 (2009).
89. Sakamoto, K. M. *et al.* Protacs: chimeric molecules that target proteins to the Skp1-Cullin-F box complex for ubiquitination and degradation. *Proc Natl Acad Sci U S A* **98**, 8554–8559 (2001).
90. Gadd, M. S. *et al.* Structural basis of PROTAC cooperative recognition for selective protein degradation. *Nat Chem Biol* **13**, 514–521 (2017).
91. Winter, G. E. *et al.* Phthalimide conjugation as a strategy for in vivo target protein degradation. *Science (1979)* **348**, 1376–1381 (2015).
92. Bondeson, D. P. *et al.* Catalytic in vivo protein knockdown by small-molecule PROTACs. *Nat Chem Biol* **11**, 611–617 (2015).
93. Burslem, G. M. & Crews, C. M. Proteolysis-Targeting Chimeras as Therapeutics and Tools for Biological Discovery. *Cell* **0**, (2020).
94. Paiva, S. L. & Crews, C. M. Targeted protein degradation: elements of PROTAC design. *Current Opinion in Chemical Biology* vol. 50 111–119 Preprint at <https://doi.org/10.1016/j.cbpa.2019.02.022> (2019).
95. Qin, C. *et al.* Discovery of QCA570 as an Exceptionally Potent and Efficacious Proteolysis Targeting Chimera (PROTAC) Degradation of the Bromodomain and Extra-Terminal (BET) Proteins Capable of Inducing Complete and Durable Tumor Regression. *J Med Chem* **61**, 6685–6704 (2018).
96. Lu, J. *et al.* Hijacking the E3 Ubiquitin Ligase Cereblon to Efficiently Target BRD4. *Chem Biol* **22**, 755 (2015).
97. Mares, A. *et al.* Extended pharmacodynamic responses observed upon PROTAC-mediated degradation of RIPK2. *Commun Biol* **3**, (2020).
98. Riching, K. M. *et al.* CDK Family PROTAC Profiling Reveals Distinct Kinetic Responses and Cell Cycle-Dependent Degradation of CDK2. *SLAS Discovery* **26**, 560–569 (2021).
99. Alabi, S. *et al.* Mutant-selective degradation by BRAF-targeting PROTACs. *Nat Commun* **12**, (2021).
100. Donovan, K. A. *et al.* Mapping the Degradable Kinome Provides a Resource for Expedited Degradation Development. *Cell* **183**, 1714–1731.e10 (2020).
101. Békés, M., Langley, D. R. & Crews, C. M. PROTAC targeted protein degraders: the past is prologue. *Nature Reviews Drug Discovery* 2022 21:3 **21**, 181–200 (2022).
102. Nowak, R. P. *et al.* Plasticity in binding confers selectivity in ligand-induced protein degradation article. *Nat Chem Biol* **14**, 706–714 (2018).
103. Chamberlain, P. Linkers for protein degradation. *Nat Chem Biol* **14**, 639–640 (2018).
104. Simonetta, K. R. *et al.* Prospective discovery of small molecule enhancers of an E3 ligase-substrate interaction. *Nat Commun* **10**, 1–12 (2019).



105. Vargesson, N. Thalidomide-induced teratogenesis: History and mechanisms. *Birth Defects Res C Embryo Today* **105**, 140–156 (2015).
106. Ito, T. *et al.* Identification of a primary target of thalidomide teratogenicity. *Science (1979)* **327**, 1345–1350 (2010).
107. Chamberlain, P. P. *et al.* Structure of the human Cereblon-DDB1-lenalidomide complex reveals basis for responsiveness to thalidomide analogs. *Nat Struct Mol Biol* **21**, 803–809 (2014).
108. Lu, G. *et al.* The myeloma drug lenalidomide promotes the cereblon-dependent destruction of ikaros proteins. *Science (1979)* **343**, 305–309 (2014).
109. Krönke, J. *et al.* Lenalidomide causes selective degradation of IKZF1 and IKZF3 in multiple myeloma cells. *Science (1979)* **343**, 301–305 (2014).
110. Gandhi, A. K. *et al.* Immunomodulatory agents lenalidomide and pomalidomide co-stimulate T cells by inducing degradation of T cell repressors Ikaros and Aiolos via modulation of the E3 ubiquitin ligase complex CRL4 CRBN. *Br J Haematol* (2014) doi:10.1111/bjh.12708.
111. Krönke, J. *et al.* Lenalidomide induces ubiquitination and degradation of CK1 $\alpha$  in del(5q) MDS. *Nature* (2015) doi:10.1038/nature14610.
112. Matyskiela, M. E. *et al.* A novel cereblon modulator recruits GSPT1 to the CRL4 CRBN ubiquitin ligase. *Nature* **535**, 252–257 (2016).
113. Sievers, Q. L. *et al.* Defining the human C2H2 zinc finger degrome targeted by thalidomide analogs through CRBN. *Science (1979)* **362**, (2018).
114. Furihata, H. *et al.* Structural bases of IMiD selectivity that emerges by 5-hydroxythalidomide. *Nat Commun* **11**, 1–11 (2020).
115. Matyskiela, M. E. *et al.* Crystal structure of the SALL4–pomalidomide–cereblon–DDB1 complex. *Nat Struct Mol Biol* 1–4 (2020) doi:10.1038/s41594-020-0405-9.
116. Donovan, K. A. *et al.* Thalidomide promotes degradation of SALL4, a transcription factor implicated in Duane radial ray syndrome. *Elife* **7**, (2018).
117. Matyskiela, M. E. *et al.* Cereblon Modulators Target ZBTB16 and Its Oncogenic Fusion Partners for Degradation via Distinct Structural Degrons. *ACS Chem Biol* acschembio.0c00674 (2020) doi:10.1021/acschembio.0c00674.
118. An, J. *et al.* pSILAC mass spectrometry reveals ZFP91 as IMiD-dependent substrate of the CRL4 CRBN ubiquitin ligase. *Nat Commun* **8**, 1–11 (2017).
119. Bjorklund, C. C. *et al.* Iberdomide (CC-220) is a potent cereblon E3 ligase modulator with antitumor and immunostimulatory activities in lenalidomide- and pomalidomide-resistant multiple myeloma cells with dysregulated CRBN. *Leukemia* **34**, 1197–1201 (2020).
120. Schafer, P. H. *et al.* Cereblon modulator iberdomide induces degradation of the transcription factors Ikaros and Aiolos: Immunomodulation in healthy volunteers and relevance to systemic lupus erythematosus. *Ann Rheum Dis* **77**, 1516–1523 (2018).

121. Lopez-Girona, A. *et al.* CC-92480 Is a Novel Cereblon E3 Ligase Modulator with Enhanced Tumoricidal and Immunomodulatory Activity Against Sensitive and Resistant Multiple Myeloma Cells. *Blood* **134**, 1812–1812 (2019).
122. Surka, C. *et al.* CC-90009, a novel cereblon E3 ligase modulator targets acute myeloid leukemia blasts and leukemia stem cells. *Blood* (2020) doi:10.1182/blood.2020008676.
123. Powell, C. E. *et al.* Selective Degradation of GSPT1 by Cereblon Modulators Identified via a Focused Combinatorial Library. *ACS Chem Biol* **15**, 2722–2730 (2020).
124. Wang, E. S. *et al.* Acute pharmacological degradation of Helios destabilizes regulatory T cells. *Nature Chemical Biology* 2021 17:6 **17**, 711–717 (2021).
125. Matyskiela, M. E. *et al.* A Cereblon Modulator (CC-220) with Improved Degradation of Ikaros and Aiolos. *J Med Chem* **61**, 535–542 (2018).
126. Asatsuma-Okumura, T. *et al.* p63 is a cereblon substrate involved in thalidomide teratogenicity. *Nat Chem Biol* **15**, 1077–1084 (2019).
127. Yamanaka, S. *et al.* PLZF is a new substrate of CRBN with thalidomide and 5-hydroxythalidomide. *bioRxiv* (2020) doi:10.1101/2020.02.28.969071.
128. Krönke, J. *et al.* Lenalidomide causes selective degradation of IKZF1 and IKZF3 in multiple myeloma cells. *Science (1979)* **343**, 301–305 (2014).
129. Yu, H. H. *et al.* Single subunit degradation of WIZ, a lenalidomide- And pomalidomide-dependent substrate of E3 ubiquitin ligase CRL4CRBN. *bioRxiv* 595389 Preprint at <https://doi.org/10.1101/595389> (2019).
130. Hagner, P. R. *et al.* CC-122, a pleiotropic pathway modifier, mimics an interferon response and has antitumor activity in DLBCL. *Blood* **126**, 779–789 (2015).
131. Gemechu, Y. *et al.* Humanized cereblon mice revealed two distinct therapeutic pathways of immunomodulatory drugs. *Proc Natl Acad Sci U S A* **115**, 11802–11807 (2018).
132. Li, L. *et al.* A Cereblon Modulator CC-885 Induces CRBN- and p97-Dependent PLK1 Degradation and Synergizes with Volasertib to Suppress Lung Cancer. *Mol Ther Oncolytics* **18**, 215–225 (2020).
133. Solomon, J. *et al.* Targeted degradation of IKZF2 for cancer immunotherapy. *Research Square (preprint)* (2022) doi:10.21203/RS.3.RS-1531006/V1.
134. Han, T. *et al.* Anticancer sulfonamides target splicing by inducing RBM39 degradation via recruitment to DCAF15. *Science* **356**, 3755 (2017).
135. Ting, T. C. *et al.* Aryl Sulfonamides Degrade RBM39 and RBM23 by Recruitment to CRL4-DCAF15. *Cell Rep* **29**, 1499-1510.e6 (2019).
136. Uehara, T. *et al.* Selective degradation of splicing factor CAPER $\alpha$  by anticancer sulfonamides. *Nat Chem Biol* **13**, 675–680 (2017).
137. Faust, T. B. *et al.* Structural complementarity facilitates E7820-mediated degradation of RBM39 by DCAF15. *Nat Chem Biol* **16**, 7–14 (2020).

138. Cavadini, S. *et al.* Cullin-RING ubiquitin E3 ligase regulation by the COP9 signalosome. *Nature* (2016) doi:10.1038/nature17416.
139. Enchev, R. I. *et al.* Structural Basis for a Reciprocal Regulation between SCF and CSN. *Cell Rep* **2**, 616–627 (2012).
140. Yamanaka, K., Sasagawa, Y. & Ogura, T. Recent advances in p97/VCP/Cdc48 cellular functions. *Biochim Biophys Acta* **1823**, 130–137 (2012).
141. Owa, T. *et al.* Discovery of novel antitumor sulfonamides targeting G1 phase of the cell cycle. *J Med Chem* (1999) doi:10.1021/jm9902638.
142. Bussiere, D. E. *et al.* Structural basis of indisulam-mediated RBM39 recruitment to DCAF15 E3 ligase complex. *Nat Chem Biol* **16**, 15–23 (2020).
143. Baek, K. & Schulman, B. A. Molecular glue concept solidifies. *Nat Chem Biol* **16**, 2–3 (2020).
144. Li, L. *et al.* In vivo target protein degradation induced by PROTACs based on E3 ligase DCAF15. *Signal Transduct Target Ther* **5**, 4–6 (2020).
145. Kerres, N. *et al.* Chemically Induced Degradation of the Oncogenic Transcription Factor BCL6. *Cell Rep* **20**, 2860–2875 (2017).
146. Słabicki, M. *et al.* Small-molecule-induced polymerization triggers degradation of BCL6. *Nature* **588**, (2020).
147. Dong, B. *et al.* Targeting therapy-resistant prostate cancer via a direct inhibitor of the human heat shock transcription factor 1. *Sci Transl Med* **12**, 5647 (2020).
148. Cohen, S. *et al.* Single UM171 Expanded Cord Blood Transplant is Feasible, Safe, and Permits Transplantation of Better HLA Matched Cords with Very Low Transplant Related Mortality. *Biology of Blood and Marrow Transplantation* **24**, S190–S191 (2018).
149. Chagraoui, J. *et al.* UM171 Preserves Epigenetic Marks that Are Reduced in Ex Vivo Culture of Human HSCs via Potentiation of the CLR3-KBTBD4 Complex. *Cell Stem Cell* **28**, 48–62.e6 (2021).
150. Subramaniam, A. *et al.* Lysine-specific demethylase 1A restricts ex vivo propagation of human HSCs and is a target of UM171. *Blood* **136**, 2151–2161 (2020).
151. Song, K. W. *et al.* RTK-Dependent Inducible Degradation of Mutant PI3K $\alpha$  Drives GDC-0077 (Inavolisib) Efficacy O. *Cancer Discov* **12**, 204–219 (2022).
152. Ma, X. Discovery and structural characterization of a VHL-mediated molecular glue degrader targeting cysteine dioxygenase. *Acta Crystallogr A Found Adv* **77**, a245–a245 (2021).
153. Malumbres, M. Cyclin-dependent kinases. *Genome Biol* **15**, 1–10 (2014).
154. Asghar, U., Witkiewicz, A. K., Turner, N. C. & Knudsen, E. S. The history and future of targeting cyclin-dependent kinases in cancer therapy. *Nature Reviews Drug Discovery* **14**, 130–146 (2015).
155. Malumbres, M. & Barbacid, M. To cycle or not to cycle: a critical decision in cancer. *Nature Reviews Cancer* **1**, 222–231 (2001).
156. Lundberg, A. S. & Weinberg, R. A. Functional inactivation of the retinoblastoma protein requires sequential modification by at least two distinct cyclin-cdk complexes. *Mol Cell Biol* **18**, 753–761 (1998).

157. Harbour, J. W., Luo, R. X., Dei Santi, A., Postigo, A. A. & Dean, D. C. Cdk phosphorylation triggers sequential intramolecular interactions that progressively block Rb functions as cells move through G1. *Cell* **98**, 859–869 (1999).
158. Malumbres, M. & Barbacid, M. Mammalian cyclin-dependent kinases. *Trends Biochem Sci* **30**, 630–641 (2005).
159. Tin Su, T. Cell cycle: How, when and why cells get rid of cyclin A. *Current Biology* **11**, R467–R469 (2001).
160. Yang, V. W. The Cell Cycle. *Physiology of the Gastrointestinal Tract: Sixth Edition* **1–2**, 197–219 (2018).
161. Sansó, M. & Fisher, R. P. Pause, play, repeat: CDKs push RNAP II's buttons. *Transcription* **4**, 146–152 (2013).
162. Buratowski, S. Progression through the RNA polymerase II CTD cycle. *Mol Cell* **36**, 541–546 (2009).
163. Peterlin, B. M. & Price, D. H. Controlling the elongation phase of transcription with P-TEFb. *Mol Cell* **23**, 297–305 (2006).
164. Bartkowiak, B. *et al.* CDK12 is a transcription elongation-associated CTD kinase, the metazoan ortholog of yeast Ctk1. *Genes Dev* **24**, 2303–2316 (2010).
165. Wood, D. J. & Endicott, J. A. Structural insights into the functional diversity of the CDK–cyclin family. *Open Biol* **8**, (2018).
166. Echaliier, A., Endicott, J. A. & Noble, M. E. M. Recent developments in cyclin-dependent kinase biochemical and structural studies. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics* **1804**, 511–519 (2010).
167. Pavletich, N. P. Mechanisms of cyclin-dependent kinase regulation: structures of Cdks, their cyclin activators, and Cip and INK4 inhibitors. *J Mol Biol* **287**, 821–828 (1999).
168. de Bondt, H. L. *et al.* Crystal structure of cyclin-dependent kinase 2. (1993).
169. Russo, A. A., Jeffrey, P. D. & Pavletich, N. P. Structural basis of cyclin-dependent kinase activation by phosphorylation. (1996).
170. Pilarova, K., Herudek, J. & Blazek, D. CDK12: cellular functions and therapeutic potential of versatile player in cancer. *NAR Cancer* **2**, (2020).
171. Greenleaf, A. L. Human CDK12 and CDK13, multi-tasking CTD kinases for the new millenium. <https://doi.org/10.1080/21541264.2018.1535211> **10**, 91–110 (2018).
172. Greifenberg, A. K. *et al.* Structural and Functional Analysis of the Cdk13/Cyclin K Complex. *Cell Rep* **14**, 320–331 (2016).
173. Choi, S. H., Kim, S. & Jones, K. A. Gene expression regulation by CDK12: a versatile kinase in cancer with functions beyond CTD phosphorylation. *Experimental & Molecular Medicine* **2020 52:5** **52**, 762–771 (2020).
174. Lei, T. *et al.* Cyclin K regulates prereplicative complex assembly to promote mammalian cell proliferation. *Nat Commun* **9**, 1–15 (2018).

175. Tsherniak, A. *et al.* Defining a Cancer Dependency Map. *Cell* **170**, 564-576.e16 (2017).
176. Blazek, D. *et al.* The Cyclin K/Cdk12 complex maintains genomic stability via regulation of expression of DNA damage response genes. *Genes Dev* **25**, 2158–2172 (2011).
177. Xiang, X. *et al.* A Distinct Expression Pattern of Cyclin K in Mammalian Testes Suggests a Functional Role in Spermatogenesis. *PLoS One* **9**, e101539 (2014).
178. Dai, Q. *et al.* Cyclin K-containing Kinase Complexes Maintain Self-renewal in Murine Embryonic Stem Cells. *Journal of Biological Chemistry* **287**, 25344–25352 (2012).
179. Qie, S. & Diehl, J. A. Cyclin D degradation by E3 ligases in cancer progression and treatment. *Semin Cancer Biol* **67**, 159–170 (2020).
180. Fu, X.-D. The superfamily of arginine/serine-rich splicing factors. *RNA* **1**, 663–680 (1995).
181. Kay, B. K., Williamson, M. P. & Sudol, M. The importance of being proline: the interaction of proline-rich motifs in signaling proteins with their cognate domains. *The FASEB Journal* **14**, 231–241 (2000).
182. Liang, K. *et al.* Characterization of Human Cyclin-Dependent Kinase 12 (CDK12) and CDK13 Complexes in C-Terminal Domain Phosphorylation, Gene Transcription, and RNA Processing. *Mol Cell Biol* **35**, 928–938 (2015).
183. Tien, J. F. *et al.* CDK12 regulates alternative last exon mRNA splicing and promotes breast cancer cell invasion. *Nucleic Acids Res* **45**, 6698–6716 (2017).
184. Davidson, L., Muniz, L. & West, S. 3' end formation of pre-mRNA and phosphorylation of Ser2 on the RNA polymerase II CTD are reciprocally coupled in human cells. *Genes Dev* **28**, 342–356 (2014).
185. Eifler, T. T. *et al.* Cyclin-dependent kinase 12 increases 3' end processing of growth factor-induced c-FOS transcripts. *Mol Cell Biol* **35**, 468–478 (2015).
186. Bartkowiak, B. & Greenleaf, A. L. Expression, purification, and identification of associated proteins of the full-length hCDK12/CyclinK complex. *J Biol Chem* **290**, 1786–1795 (2015).
187. Ko, T. K., Kelly, E. & Pines, J. CrkRS: a novel conserved Cdc2-related protein kinase that colocalises with SC35 speckles. *J Cell Sci* **114**, 2591–2603 (2001).
188. Choi, S. H. *et al.* CDK12 phosphorylates 4E-BP1 to enable mTORC1-dependent translation and mitotic genome stability The nuclear cyclin-dependent kinases CDK7, CDK8, CDK9, and CDK12 function to integrate transcription with changes in phosphorylation of the RNA polymerase II (RNAPII) C-terminal domain (CTD) heptad repeats. (2019) doi:10.1101/gad.322339.118.
189. Paul, A. *et al.* CDK12 controls G1/S progression by regulating RNAPII processivity at core DNA replication genes. *EMBO Rep* **20**, e47592 (2019).
190. Liu, H., Liu, K. & Dong, Z. Targeting CDK12 for Cancer Therapy: Function, Mechanism, and Drug Discovery. *Cancer Res* **81**, 18–26 (2021).
191. Bell, D. *et al.* Integrated Genomic Analyses of Ovarian Carcinoma. *Nature* **474**, 609 (2011).
192. Chilà, R., Guffanti, F. & Damia, G. Role and therapeutic potential of CDK12 in human cancers. *Cancer Treat Rev* **50**, 83–88 (2016).

193. Nguyen, B. *et al.* Pan-cancer Analysis of CDK12 Alterations Identifies a Subset of Prostate Cancers with Distinct Genomic and Clinical Characteristics. *Eur Urol* **78**, 671–679 (2020).
194. Reimers, M. A. *et al.* Clinical Outcomes in Cyclin-dependent Kinase 12 Mutant Advanced Prostate Cancer. *Eur Urol* **77**, 333–341 (2020).
195. Pan, E. *et al.* Analysis of CDK12 alterations in a pan-cancer database. *Cancer Med* **11**, 753–763 (2022).
196. Choi, H.-J. *et al.* CDK12 drives breast tumor initiation and trastuzumab resistance via WNT and IRS1-ErbB-PI3K signaling. *EMBO Rep* **20**, e48058 (2019).
197. Yang, B., Chen, J. & Teng, Y. CDK12 Promotes Cervical Cancer Progression through Enhancing Macrophage Infiltration. *J Immunol Res* **2021**, (2021).
198. Zhang, X. *et al.* Quantitative mass spectrometry to interrogate proteomic heterogeneity in metastatic lung adenocarcinoma and validate a novel somatic mutation CDK12-G879V. *Molecular and Cellular Proteomics* **18**, 622–641 (2019).
199. Liu, H. *et al.* CDK12 and PAK2 as novel therapeutic targets for human gastric cancer. *Theranostics* **10**, 6201 (2020).
200. Sokol, E. S. *et al.* Pan-Cancer Analysis of CDK12 Loss-of-Function Alterations and Their Association with the Focal Tandem-Duplicator Phenotype. *Oncologist* **24**, 1526–1533 (2019).
201. Zeng, M. *et al.* Targeting MYC dependency in ovarian cancer through inhibition of CDK7 and CDK12/13. *Elife* **7**, (2018).
202. Henry, K. L. *et al.* CDK12-mediated transcriptional regulation of noncanonical NF- $\kappa$ B components is essential for signaling. <http://stke.sciencemag.org/> (2018).
203. Bai, N. *et al.* CDK12 promotes papillary thyroid cancer progression through regulating the c-myc/ $\beta$ -catenin pathway. *J Cancer* **11**, 4308–4315 (2020).
204. Bajrami, I. *et al.* Genome-wide profiling of genetic synthetic lethality identifies CDK12 as a novel determinant of PARP1/2 inhibitor sensitivity. *Cancer Res* **74**, 287–297 (2014).
205. Joshi, P. M., Sutor, S. L., Huntoon, C. J. & Karnitz, L. M. Ovarian Cancer-associated Mutations Disable Catalytic Activity of CDK12, a Kinase That Promotes Homologous Recombination Repair and Resistance to Cisplatin and Poly(ADP-ribose) Polymerase Inhibitors. *Journal of Biological Chemistry* **289**, 9247–9253 (2014).
206. Paculová, H. *et al.* BRCA1 or CDK12 loss sensitizes cells to CHK1 inhibitors. *Tumor Biology* **39**, 1–11 (2017).
207. Elliott, A. *et al.* Predicted Immunogenicity of CDK12 Biallelic Loss-of-Function Tumors Varies across Cancer Types. *The Journal of Molecular Diagnostics* **23**, 1761–1773 (2021).
208. Chipumuro, E. *et al.* CDK7 Inhibition Suppresses Super-Enhancer-Linked Oncogenic Transcription in MYCN-Driven Cancer. *Cell* **159**, 1126–1139 (2014).
209. Toyoshima, M. *et al.* Functional genomics identifies therapeutic targets for MYC-driven cancer. *Proc Natl Acad Sci U S A* **109**, 9545–9550 (2012).

210. Iniguez, A. B. *et al.* EWS/FLI Confers Tumor Cell Synthetic Lethality to CDK12 Inhibition in Ewing Sarcoma. *Cancer Cell* **33**, 202-216.e6 (2018).
211. Davis, B. M., Mccurrach, M. E., Taneja, K. L., Singer, R. H. & Housman, D. E. Expansion of a CUG trinucleotide repeat in the 3' untranslated region of myotonic dystrophy protein kinase transcripts results in nuclear retention of transcripts. *Proc Natl Acad Sci U S A* **94**, 7388–7393 (1997).
212. Ketley, A. *et al.* CDK12 inhibition reduces abnormalities in cells from patients with myotonic dystrophy and in a mouse model. *Sci Transl Med* **12**, eaaz2415 (2020).
213. Yao, G. *et al.* Cyclin K interacts with  $\beta$ -catenin to induce Cyclin D1 expression and facilitates tumorigenesis and radioresistance in lung cancer. *Theranostics* **10**, 11144–11158 (2020).
214. Fu, T. J., Peng, J., Lee, G., Price, D. H. & Flores, O. Cyclin K functions as a CDK9 regulatory subunit and participates in RNA polymerase II transcription. *J Biol Chem* **274**, 34527–34530 (1999).
215. Fan, Y. *et al.* De Novo Mutations of CCNK Cause a Syndromic Neurodevelopmental Disorder with Distinctive Facial Dysmorphism. *Am J Hum Genet* **103**, 448–455 (2018).
216. Hoshii, T. *et al.* A Non-catalytic Function of SETD1A Regulates Cyclin K and the DNA Damage Response. *Cell* **172**, 1007-1021.e17 (2018).
217. Singh, T. *et al.* Rare loss-of-function variants in SETD1A are associated with schizophrenia and developmental disorders. *Nat Neurosci* **19**, 571–577 (2016).
218. Böskén, C. A. *et al.* The structure and substrate specificity of human Cdk12/Cyclin K. *Nat Commun* **5**, (2014).
219. Dixon-Clarke, S. E., Elkins, J. M., Cheng, S. W. G., Morin, G. B. & Bullock, A. N. Structures of the CDK12/CycK complex with AMP-PNP reveal a flexible C-terminal kinase extension important for ATP binding. *Sci Rep* **5**, (2015).
220. Zhang, T. *et al.* Covalent targeting of remote cysteine residues to develop CDK12 and CDK13 inhibitors. *Nat Chem Biol* **12**, 876–884 (2016).
221. Johannes, J. W. *et al.* Structure-Based Design of Selective Noncovalent CDK12 Inhibitors. *ChemMedChem* (2018) doi:10.1002/cmdc.201700695.
222. Ito, M. *et al.* Discovery of 3-Benzyl-1-(trans-4-((5-cyanopyridin-2-yl)amino)cyclohexyl)-1-arylurea Derivatives as Novel and Selective Cyclin-Dependent Kinase 12 (CDK12) Inhibitors. *J Med Chem* (2018) doi:10.1021/acs.jmedchem.8b00683.
223. Jiang, B. *et al.* Structure-activity relationship study of THZ531 derivatives enables the discovery of BSJ-01-175 as a dual CDK12/13 covalent inhibitor with efficacy in Ewing sarcoma. *Eur J Med Chem* **221**, 113481 (2021).
224. Baek, K., Brown, R. S., Birrane, G. & Ladas, J. A. A. Crystal Structure of Human Cyclin K, a Positive Regulator of Cyclin-dependent Kinase 9. *J Mol Biol* **366**, 563–573 (2007).
225. Ferguson, F. M. & Gray, N. S. Kinase inhibitors: The road ahead. *Nature Reviews Drug Discovery* Preprint at <https://doi.org/10.1038/nrd.2018.21> (2018).

226. Wang, B. *et al.* An overview of kinase downregulators and recent advances in discovery approaches. *Signal Transduction and Targeted Therapy* 2021 6:1 **6**, 1–19 (2021).
227. Xie, Z. *et al.* Lessons Learned from Past Cyclin-Dependent Kinase Drug Discovery Efforts. *J Med Chem* (2021) doi:10.1021/ACS.JMEDCHEM.1C02190/SUPPL\_FILE/JM1C02190\_SI\_001.PDF.
228. Kwiatkowski, N. *et al.* Targeting transcription regulation in cancer with a covalent CDK7 inhibitor. *Nature* 2014 511:7511 **511**, 616–620 (2014).
229. Bradley, M. *et al.* Abstract 1143: Targeting the transcriptional kinases CDK12 and CDK13 in breast and ovarian cancer. *Cancer Res* **77**, 1143–1143 (2017).
230. Gao, Y. *et al.* Overcoming Resistance to the THZ Series of Covalent Transcriptional CDK Inhibitors. *Cell Chem Biol* **25**, 135-142.e5 (2018).
231. Liu, Y. *et al.* Discovery of MFH290: A Potent and Highly Selective Covalent Inhibitor for Cyclin-Dependent Kinase 12/13. *J Med Chem* **63**, 6708–6726 (2020).
232. Bhurta, D. & Bharate, S. B. Analyzing the scaffold diversity of cyclin-dependent kinase inhibitors and revisiting the clinical and preclinical pipeline. *Med Res Rev* **42**, 654–709 (2022).
233. Schneider, E. v. *et al.* The Structure of CDK8/CycC Implicates Specificity in the CDK/Cyclin Family and Reveals Interaction with a Deep Pocket Binder. *J Mol Biol* **412**, 251–266 (2011).
234. Johnson, S. F. *et al.* CDK12 Inhibition Reverses De Novo and Acquired PARP Inhibitor Resistance in BRCA Wild-Type and Mutated Models of Triple-Negative Breast Cancer. *Cell Rep* (2016) doi:10.1016/j.celrep.2016.10.077.
235. Bibian, M. *et al.* Development of highly selective casein kinase 1 $\delta$ /1 $\epsilon$  (CK1 $\delta$ / $\epsilon$ ) inhibitors with potent antiproliferative properties. *Bioorg Med Chem Lett* **23**, 4374–4380 (2013).
236. Bettayeb, K. *et al.* CR8, a potent and selective, roscovitine-derived inhibitor of cyclin-dependent kinases. *Oncogene* **27**, 5797–5807 (2008).
237. Bettayeb, K. *et al.* N-&-N, a new class of cell death-inducing kinase inhibitors derived from the purine roscovitine. *Mol Cancer Ther* **7**, 2713–2724 (2008).
238. Meijer, L. *et al.* Biochemical and Cellular Effects of Roscovitine, a Potent and Selective Inhibitor of the Cyclin-Dependent Kinases cdc2, cdk2 and cdk5. *Eur J Biochem* **243**, 527–536 (1997).
239. Oumata, N. *et al.* Roscovitine-Derived, Dual-Specificity Inhibitors of Cyclin-Dependent Kinases and Casein Kinases 1. *J Med Chem* **51**, 5229–5242 (2008).
240. Bettayeb, K. *et al.* CDK Inhibitors Roscovitine and CR8 Trigger Mcl-1 Down-Regulation and Apoptotic Cell Death in Neuroblastoma Cells. *Genes Cancer* **1**, 369–380 (2010).
241. Delehouz , C. *et al.* CDK/CK1 inhibitors roscovitine and CR8 downregulate amplified MYCN in neuroblastoma cells. *Oncogene* **33**, 5675–5687 (2014).
242. Quereda, V. *et al.* Therapeutic Targeting of CDK12/CDK13 in Triple-Negative Breast Cancer. *Cancer Cell* **36**, 545-558.e7 (2019).
243. Jiang, B. *et al.* Discovery and resistance mechanism of a selective CDK12 degrader. *Nature Chemical Biology* 2021 17:6 **17**, 675–683 (2021).



244. Mayor-Ruiz, C. *et al.* Plasticity of the Cullin-RING Ligase Repertoire Shapes Sensitivity to Ligand-Induced Protein Degradation. *Mol Cell* **75**, 849-858.e8 (2019).
245. Niu, T. *et al.* Noncovalent CDK12/13 dual inhibitors-based PROTACs degrade CDK12-Cyclin K complex and induce synthetic lethality with PARP inhibitor. *Eur J Med Chem* **228**, 114012 (2022).
246. Lim, E., Johnson, S. F., Geyer, M., Serra, V. & Shapiro, G. I. Sensitizing HR-proficient cancers to PARP inhibitors. *Mol Cell Oncol* **4**, (2017).
247. Li, Y. *et al.* CDK12/13 inhibition induces immunogenic cell death and enhances anti-PD-1 anticancer activity in breast cancer. *Cancer Lett* **495**, 12–21 (2020).
248. Md Sakib Hossain, D. *et al.* Dinaciclib induces immunogenic cell death and enhances anti-PD1-mediated tumor suppression. *J Clin Invest* **128**, 644–654 (2018).
249. Lei, P. *et al.* Current progress and novel strategies that target CDK12 for drug discovery. *Eur J Med Chem* 114603 (2022) doi:10.1016/J.EJMECH.2022.114603.
250. Emadi, F., Teo, T., Rahaman, M. H. & Wang, S. CDK12: a potential therapeutic target in cancer. *Drug Discov Today* **25**, 2257–2267 (2020).
251. Shyamsunder, P. *et al.* THZ531 Induces a State of BRCAness in Multiple Myeloma Cells: Synthetic Lethality with Combination Treatment of THZ 531 with DNA Repair Inhibitors. *International Journal of Molecular Sciences* 2022, Vol. 23, Page 1207 **23**, 1207 (2022).
252. Geng, M. *et al.* Targeting CDK12-mediated transcription regulation in anaplastic thyroid carcinoma. *Biochem Biophys Res Commun* **520**, 544–550 (2019).
253. Gao, J. *et al.* Response and resistance to CDK12 inhibition in aggressive B-cell lymphomas. *Haematologica* **107**, 1119 (2022).
254. Wang, C. *et al.* CDK12 inhibition mediates DNA damage and is synergistic with sorafenib treatment in hepatocellular carcinoma. *Gut* **69**, 727–736 (2020).
255. Alagpulinsa, D. A., Ayyadevara, S., Yaccoby, S. & Reis, R. J. S. A Cyclin-dependent kinase inhibitor, dinaciclib, impairs homologous recombination and sensitizes multiple myeloma cells to PARP inhibition. *Mol Cancer Ther* **15**, 241–250 (2016).
256. Chen, X. X. *et al.* Cyclin-dependent kinase inhibitor dinaciclib potently synergizes with cisplatin in preclinical models of ovarian cancer. *Oncotarget* **6**, 14926 (2015).
257. Choi, H.-J. *et al.* CDK12 drives breast tumor initiation and trastuzumab resistance via WNT and IRS1-ErbB-PI3K signaling. *EMBO Rep* **20**, e48058 (2019).
258. Chen, Z. *et al.* Multiple CDK inhibitor dinaciclib suppresses neuroblastoma growth via inhibiting CDK2 and CDK9 activity. *Scientific Reports* 2016 6:1 **6**, 1–12 (2016).
259. Xu, X. *et al.* Rational combination therapy for melanoma with dinaciclib by targeting BAK-dependent cell death. *Mol Cancer Ther* **19**, 627–636 (2020).
260. Feldmann, G. *et al.* Cyclin-dependent kinase inhibitor Dinaciclib (SCH727965) inhibits pancreatic cancer growth and progression in murine xenograft models. <http://dx.doi.org/10.4161/cbt.12.7.16475> **12**, 598–609 (2011).

261. Murphy, A. G. *et al.* A Phase I Study of Dinaciclib in Combination With MK-2206 in Patients With Advanced Pancreatic Cancer. *Clin Transl Sci* **13**, 1178–1188 (2020).
262. Hu, C. *et al.* Combined inhibition of cyclin-dependent kinases (dinaciclib) and AKT(MK-2206) blocks pancreatic tumor growth and metastases in patient-derived xenograft models. *Mol Cancer Ther* **14**, 1532–1539 (2015).
263. L Lui, G. Y., Grandori, C. & Kemp, C. J. CDK12: an emerging therapeutic target for cancer. *J Clin Pathol* **71**, 957–962 (2018).
264. Stanton, B. Z., Chory, E. J. & Crabtree, G. R. Chemically induced proximity in biology and medicine. *Science* **359**, (2018).
265. Salami, J. & Crews, C. M. Waste disposal - An attractive strategy for cancer therapy. *Science* vol. 355 1163–1167 Preprint at <https://doi.org/10.1126/science.aam7340> (2017).
266. Chopra, R., Sadok, A. & Collins, I. A critical evaluation of the approaches to targeted protein degradation for drug discovery. *Drug Discov Today Technol* **31**, 5–13 (2019).
267. Fink, E. C. & Ebert, B. L. The novel mechanism of lenalidomide activity. *Blood* **126**, 2366–2369 (2015).
268. Lu, G. *et al.* The myeloma drug lenalidomide promotes the cereblon-dependent destruction of ikaros proteins. *Science (1979)* **343**, 305–309 (2014).
269. Krönke, J. *et al.* Lenalidomide causes selective degradation of IKZF1 and IKZF3 in multiple myeloma cells. *Science (1979)* **343**, 301–305 (2014).
270. Matyskiela, M. E. *et al.* A novel cereblon modulator recruits GSPT1 to the CRL4 CRBN ubiquitin ligase. *Nature* **535**, 252–257 (2016).
271. Faust, T. B. *et al.* Structural complementarity facilitates E7820-mediated degradation of RBM39 by DCAF15. *Nat Chem Biol* (2019) doi:10.1038/s41589-019-0378-3.
272. Yu, C. *et al.* High-throughput identification of genotype-specific cancer vulnerabilities in mixtures of barcoded tumor cell lines. *Nat Biotechnol* **34**, 419–423 (2016).
273. Corsello, S. M. *et al.* Non-oncology drugs are a source of previously unappreciated anti-cancer activity. *Biorxiv* (2019) doi:10.1101/730119.
274. Ghandi, M. *et al.* Next-generation characterization of the Cancer Cell Line Encyclopedia. *Nature* **569**, 503–508 (2019).
275. Angers, S. *et al.* Molecular architecture and assembly of the DDB1-CUL4A ubiquitin ligase machinery. *Nature* **443**, 590–593 (2006).
276. Lee, J. & Zhou, P. DCAFs, the Missing Link of the CUL4-DDB1 Ubiquitin Ligase. *Mol Cell* **26**, 775–780 (2007).
277. Tickle, I.J., Flensburg, C., Keller, P., Paciorek, W., Sharff, A., Vornrhein, C., Bricogne, G. STARANISO. Preprint at (2018).
278. Sedlacek, H. H. *et al.* Flavopiridol (L86 8275; NSC 649890), a new kinase inhibitor for tumor therapy. *Int J Oncol* **9**, 1143–1168 (1996).
279. Jones, L. H. Small-Molecule Kinase Downregulators. *Cell Chem Biol* **25**, 30–35 (2018).

280. Schreiber, S. L. A chemical biology view of bioactive small molecules and a binder-based approach to connect biology to precision medicines. *Israel Journal of Chemistry* vol. 59 52–59 Preprint at <https://doi.org/10.1002/ijch.201800113> (2019).
281. Fischer, E. S. *et al.* The molecular basis of CRL4DDB2/CSAubiquitin ligase architecture, targeting, and activation. *Cell* **147**, 1024–1039 (2011).
282. Dykes, G. W., Crepeau, R. H. & Edelstein, S. J. Three-dimensional reconstruction of the 14-filament fibers of hemoglobin S. *J Mol Biol* **130**, 451–72 (1979).
283. Levy, E. D. A simple definition of structural regions in proteins and its use in analyzing interface evolution. *J Mol Biol* **403**, 660–670 (2010).
284. garcia-seiseded, H., Empereur-Mot, C., Elad, N. & Levy, E. D. Proteins evolve on the edge of supramolecular self-assembly. *Nature Publishing Group* **548**, (2017).
285. Mayor-Ruiz, C. *et al.* Rational discovery of molecular glue degraders via scalable chemical profiling. *Nat Chem Biol* (2020) doi:10.1038/s41589-020-0594-x.
286. Branon, T. C. *et al.* Efficient proximity labeling in living cells and organisms with TurboID. *Nat Biotechnol* **36**, 880–898 (2018).
287. Cowan, A. D. & Ciulli, A. Driving E3 Ligase Substrate Specificity for Targeted Protein Degradation: Lessons from Nature and the Laboratory. <https://doi.org/10.1146/annurev-biochem-032620-104421> **91**, (2022).
288. Mullard, A. Targeted protein degraders crowd into the clinic. *Nat Rev Drug Discov* **20**, 247–250 (2021).
289. Geiger, T. M., Schäfer, S. C., Dreizler, J. K., Walz, M. & Hausch, F. Clues to molecular glues. *Current Research in Chemical Biology* **2**, 100018 (2022).
290. Dieter, S. M. *et al.* Degradation of CCNK/CDK12 is a druggable vulnerability of colorectal cancer. *Cell Rep* **36**, 109394 (2021).
291. Lv, L. *et al.* Discovery of a molecular glue promoting CDK12-DDB1 interaction to trigger Cyclin K degradation. *Elife* **9**, (2020).
292. Hanan, E. J. *et al.* Monomeric Targeted Protein Degraders. *J Med Chem* **63**, 11330–11361 (2020).
293. Kozicka, Z. & Thomä, N. H. Haven't got a glue: Protein surface variation for the design of molecular glue degraders. *Cell Chem Biol* **28**, 1032–1047 (2021).
294. Słabicki, M. *et al.* The CDK inhibitor CR8 acts as a molecular glue degrader that depletes cyclin K. *Nature* (2020) doi:10.1038/s41586-020-2374-x.
295. Nekardová, M. *et al.* Structural Basis of the Interaction of Cyclin-Dependent Kinase 2 with Roscovitine and Its Analogues Having Bioisosteric Central Heterocycles. *ChemPhysChem* **18**, 785–795 (2017).
296. Kuchukulla, R. R. *et al.* Novel 2,6,9-Trisubstituted Purines as Potent CDK Inhibitors Alleviating Trastuzumab-Resistance of HER2-Positive Breast Cancers. *Pharmaceuticals* 2022, Vol. 15, Page 1041 **15**, 1041 (2022).
297. Leissing, T. M., Luh, L. M. & Cromm, P. M. Structure driven compound optimization in targeted protein degradation. *Drug Discov Today Technol* (2020) doi:10.1016/J.DDTEC.2020.11.005.

298. Powell, C. E. *et al.* Selective Degradation of GSPT1 by Cereblon Modulators Identified via a Focused Combinatorial Library. *ACS Chem Biol* **15**, 2722–2730 (2020).
299. Cortopassi, W. A., Kumar, K. & Paton, R. S. Cation- $\pi$  interactions in CREBBP bromodomain inhibition: an electrostatic model for small-molecule binding affinity and selectivity. *Org Biomol Chem* **14**, 10926–10938 (2016).
300. Sörme, P. *et al.* Structural and thermodynamic studies on cation- $\pi$  interactions in lectin-ligand complexes: High-affinity galectin-3 inhibitors through fine-tuning of an arginine-arene interaction. *J Am Chem Soc* **127**, 1737–1743 (2005).
301. Parry, D. *et al.* Dinaciclib (SCH 727965), a novel and potent cyclin-dependent kinase inhibitor. *Mol Cancer Ther* **9**, 2344–2353 (2010).
302. SRpatent\_WO2019217421-PAMPH-20191114-0350.
303. el Hage, K. *et al.* A Simple Isomerization of the Purine Scaffold of a Kinase Inhibitor, Roscovitine, Affords a Four- to Seven-Fold Enhancement of Its Affinity for Four CDKs. Could This Be Traced Back to Conjugation-Induced Stiffenings/Loosenings of Rotational Barriers? *ACS Omega* **2**, 3467–3474 (2017).
304. Popowycz, F. *et al.* Pyrazolo[1,5-a]-1,3,5-triazine as a purine bioisostere: access to potent cyclin-dependent kinase inhibitor (R)-roscovitine analogue. *J Med Chem* **52**, 655–663 (2009).
305. Thede, K. *et al.* PYRAZOLOTRIAZINES. Patent WO2021116178. (2021).
306. Sterling, T. & Irwin, J. J. ZINC 15 - Ligand Discovery for Everyone. *J Chem Inf Model* **55**, 2324–2337 (2015).
307. Caligiuri, M. *et al.* A proteome-wide CDK/CRK-specific kinase inhibitor promotes tumor cell death in the absence of cell cycle progression. *Chem Biol* **12**, 1103–1115 (2005).
308. Schmidpeter, P. A. M. & Nimigeon, C. M. Fluorescence Titrations to Determine the Binding Affinity of Cyclic Nucleotides to SthK Ion Channels. *Iss* **8**, (2018).
309. Lenci, E. & Trabocchi, A. Peptidomimetic toolbox for drug discovery. *Chem Soc Rev* **49**, 3262–3277 (2020).
310. Armstrong, C. T., Mason, P. E., Anderson, J. L. R. & Dempsey, C. E. Arginine side chain interactions and the role of arginine as a gating charge carrier in voltage sensitive ion channels. *Scientific Reports* **2016 6:1** **6**, 1–10 (2016).
311. Gallivan, J. P. & Dougherty, D. A. Cation- $\pi$  interactions in structural biology. *Proceedings of the National Academy of Sciences* **96**, 9459–9464 (1999).
312. Kumar, K. *et al.* Cation- $\pi$  interactions in protein–ligand binding: theory and data-mining reveal different roles for lysine and arginine. *Chem Sci* **9**, 2655–2665 (2018).
313. Tian, W., Chen, C., Lei, X., Zhao, J. & Liang, J. CASTp 3.0: computed atlas of surface topography of proteins. *Nucleic Acids Res* **46**, W363–W367 (2018).
314. Bissantz, C., Kuhn, B. & Stahl, M. A Medicinal Chemist’s Guide to Molecular Interactions. *J Med Chem* **53**, 5061–5084 (2010).

315. Förster, T. Zwischenmolekulare Energiewanderung und Fluoreszenz. *Ann Phys* **437**, 55–75 (1948).
316. Cho, U. & Chen, J. K. Lanthanide-Based Optical Probes of Biological Systems. *Cell Chem Biol* **27**, 921–936 (2020).
317. Ullman, E. F. *et al.* Luminescent oxygen channeling immunoassay: measurement of particle binding kinetics by chemiluminescence. *Proc Natl Acad Sci U S A* **91**, 5426 (1994).
318. Imbert, P. E. *et al.* Recommendations for the Reduction of Compound Artifacts in Time-Resolved Fluorescence Resonance Energy Transfer Assays. <https://home.liebertpub.com/adt> **5**, 363–372 (2007).
319. Liu, X. *et al.* Assays and technologies for developing proteolysis targeting chimera degraders. *Future Med Chem* **12**, 1155–1179 (2020).
320. Lin, W. & Chen, T. General Stepwise Approach to Optimize a TR-FRET Assay for Characterizing the BRD/PROTAC/CRBN Ternary Complex. *ACS Pharmacol Transl Sci* **4**, 941 (2021).
321. Payne, N. C., Maksoud, S., Tannous, B. A. & Mazitschek, R. A direct high-throughput protein quantification strategy facilitates discovery and characterization of a celastrol-derived BRD4 degrader. *Cell Chem Biol* **29**, 1333-1340.e5 (2022).
322. Liu, X. *et al.* Assays and technologies for developing proteolysis targeting chimera degraders. *Future Med Chem* **12**, 1155–1179 (2020).
323. Cao, S. *et al.* Defining molecular glues with a dual-nanobody cannabidiol sensor. *Nature Communications* **2022 13:1** **13**, 1–14 (2022).
324. Ekins, S., Olechno, J. & Williams, A. J. Dispensing Processes Impact Apparent Biological Activity as Determined by Computational and Statistical Analyses. *PLoS One* **8**, e62325 (2013).
325. Degorce, F. *et al.* HTRF: A Technology Tailored for Drug Discovery –A Review of Theoretical Aspects and Recent Applications. *Curr Chem Genomics* **3**, 22 (2009).
326. Zhang, J. H., Chung, T. D. Y. & Oldenburg, K. R. A simple statistical parameter for use in evaluation and validation of high throughput screening assays. *J Biomol Screen* **4**, 67–73 (1999).
327. Perrin, F. Polarisation de la lumière de fluorescence. Vie moyenne des molécules dans l'état excité. *Journal de Physique et le Radium* **7**, 390–401 (1926).
328. Jameson, D. & Croney, J. Fluorescence polarization: past, present and future. *Comb Chem High Throughput Screen* **6**, 167–176 (2003).
329. Lundblad, J. R., Laurance, M. & Goodman, R. H. Fluorescence polarization analysis of protein-DNA and protein-protein interactions. *Molecular Endocrinology* **10**, 607–612 (1996).
330. Lea, W. A. & Simeonov, A. Fluorescence Polarization Assays in Small Molecule Screening. *Expert Opin Drug Discov* **6**, 17 (2011).
331. Leopoldo, M., Lacivita, E., Berardi, F. & Perrone, R. Developments in fluorescent probes for receptor research. *Drug Discov Today* **14**, 706–712 (2009).
332. Parker, G. J., Law, T. L., Lenocho, F. J. & Bolger, R. E. Development of high throughput screening assays using fluorescence polarization: nuclear receptor-ligand-binding and kinase/phosphatase assays. *J Biomol Screen* **5**, 77–88 (2000).

333. Rodda, D. J. *et al.* Transcriptional regulation of nanog by OCT4 and SOX2. *J Biol Chem* **280**, 24731–24737 (2005).
334. Soufi, A. *et al.* Pioneer transcription factors target partial DNA motifs on nucleosomes to initiate reprogramming. *Cell* **161**, 555–568 (2015).
335. McGinty, R. K. & Tan, S. Nucleosome structure and function. *Chem Rev* **115**, 2255–2273 (2015).
336. Matsumoto, S. *et al.* DNA damage detection in nucleosomes involves DNA register shifting. *Nature* (2019) doi:10.1038/s41586-019-1259-3.
337. Li, G., Levitus, M., Bustamante, C. & Widom, J. Rapid spontaneous accessibility of nucleosomal DNA. *Nat Struct Mol Biol* **12**, 46–53 (2005).
338. Fierz, B. & Poirier, M. G. Biophysics of Chromatin Dynamics. *Annu Rev Biophys* **48**, 321–345 (2019).
339. Zhu, F. *et al.* The interaction landscape between transcription factors and the nucleosome. *Nature* **562**, 76–81 (2018).
340. Cirillo, L. A. *et al.* Binding of the winged-helix transcription factor HNF3 to a linker histone site on the nucleosome. *EMBO J* **17**, 244–254 (1998).
341. Lowary, P. T. & Widom, J. New DNA sequence rules for high affinity binding to histone octamer and sequence-directed nucleosome positioning. *J Mol Biol* **276**, 19–42 (1998).
342. Michael, A. K. *et al.* Mechanisms of OCT4-SOX2 motif readout on nucleosomes. *Science* (1979) eabb0074 (2020) doi:10.1126/science.abb0074.
343. Malik, V. *et al.* Pluripotency reprogramming by competent and incompetent POU factors uncovers temporal dependency for Oct4 and Sox2. *Nat Commun* **10**, (2019).
344. Luger, K., Mäder, A. W., Richmond, R. K., Sargent, D. F. & Richmond, T. J. Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature* **389**, 251–260 (1997).
345. Cirillo, L. A. & Zaret, K. S. An early developmental transcription factor complex that is more stable on nucleosome core particles than on free DNA. *Mol Cell* **4**, 961–969 (1999).
346. Esch, D. *et al.* A unique Oct4 interface is crucial for reprogramming to pluripotency. *Nat Cell Biol* **15**, 295–301 (2013).
347. Reményi, A. *et al.* Crystal structure of a POU/HMG/DNA ternary complex suggests differential assembly of Oct4 and Sox2 on two enhancers. *Genes Dev* **17**, 2048–2059 (2003).
348. Chronis, C. *et al.* Cooperative Binding of Transcription Factors Orchestrates Reprogramming. *Cell* **168**, 442–459.e20 (2017).
349. Liu, Z. & Kraus, W. L. Catalytic-Independent Functions of PARP-1 Determine Sox2 Pioneer Activity at Intractable Genomic Loci. *Mol Cell* **65**, 589–603.e9 (2017).
350. Ablasser, A. & Chen, Z. J. cGAS in action: Expanding roles in immunity and inflammation. *Science* **363**, (2019).
351. Sun, L., Wu, J., Du, F., Chen, X. & Chen, Z. J. Cyclic GMP-AMP synthase is a cytosolic DNA sensor that activates the type I interferon pathway. *Science* **339**, 786–791 (2013).

352. Wu, J. *et al.* Cyclic GMP-AMP is an endogenous second messenger in innate immune signaling by cytosolic DNA. *Science* **339**, 826–830 (2013).
353. Ablasser, A. *et al.* cGAS produces a 2'-5'-linked cyclic dinucleotide second messenger that activates STING. *Nature* **498**, 380–384 (2013).
354. Gao, P. *et al.* Cyclic [G(2',5')pA(3',5')p] is the metazoan second messenger produced by DNA-activated cyclic GMP-AMP synthase. *Cell* **153**, 1094–1107 (2013).
355. Diner, E. J. *et al.* The innate immune DNA sensor cGAS produces a noncanonical cyclic dinucleotide that activates human STING. *Cell Rep* **3**, 1355–1361 (2013).
356. Zhang, X. *et al.* Cyclic GMP-AMP containing mixed phosphodiester linkages is an endogenous high-affinity ligand for STING. *Mol Cell* **51**, 226–235 (2013).
357. Gentili, M. *et al.* The N-Terminal Domain of cGAS Determines Preferential Association with Centromeric DNA and Innate Immune Activation in the Nucleus. *Cell Rep* **26**, 2377–2393.e13 (2019).
358. Volkman, H. E., Cambier, S., Gray, E. E. & Stetson, D. B. Tight nuclear tethering of cGAS is essential for preventing autoreactivity. *Elife* **8**, (2019).
359. Orzalli, M. H. *et al.* cGAS-mediated stabilization of IFI16 promotes innate signaling during herpes simplex virus infection. *Proc Natl Acad Sci U S A* **112**, E1773–E1781 (2015).
360. Zierhut, C. *et al.* The Cytoplasmic DNA Sensor cGAS Promotes Mitotic Cell Death. *Cell* **178**, 302–315.e23 (2019).
361. Lahaye, X. *et al.* NONO Detects the Nuclear HIV Capsid to Promote cGAS-Mediated Innate Immune Activation. *Cell* **175**, 488–501.e22 (2018).
362. Stark, H. GraFix: stabilization of fragile macromolecular complexes for single particle cryo-EM. *Methods Enzymol* **481**, 109–126 (2010).
363. Pathare, G. R. *et al.* Structural mechanism of cGAS inhibition by the nucleosome. *Nature* **2020** 587:7835 **587**, 668–672 (2020).
364. Li, X. *et al.* Cyclic GMP-AMP synthase is activated by double-stranded DNA-induced oligomerization. *Immunity* **39**, 1019–1031 (2013).
365. Andreeva, L. *et al.* cGAS senses long and HMGB/TFAM-bound U-turn DNA by forming protein-DNA ladders. *Nature* **549**, 394–398 (2017).
366. Zhou, W. *et al.* Structure of the Human cGAS-DNA Complex Reveals Enhanced Control of Immune Surveillance. *Cell* **174**, 300–311.e11 (2018).
367. Zhang, X. *et al.* The cytosolic DNA sensor cGAS forms an oligomeric complex with DNA and undergoes switch-like conformational changes in the activation loop. *Cell Rep* **6**, 421–430 (2014).
368. Hanan, E. J. *et al.* Monomeric Targeted Protein Degraders. *J Med Chem* **63**, 11330–11361 (2020).
369. Jorda, R. *et al.* 3,5,7-Substituted Pyrazolo[4,3-d]Pyrimidine Inhibitors of Cyclin-Dependent Kinases and Cyclin K Degraders. *J Med Chem* (2022) doi:10.1021/ACS.JMEDCHEM.1C02184.
370. Farnaby, W. *et al.* BAF complex vulnerabilities in cancer demonstrated via structure-based PROTAC design. *Nat Chem Biol* **15**, 672–680 (2019).

371. Potjewyd, F. *et al.* Degradation of Polycomb Repressive Complex 2 with an EED-Targeted Bivalent Chemical Degradator. *Cell Chem Biol* **27**, 47-56.e15 (2020).
372. Krissinel, E. & Henrick, K. Inference of Macromolecular Assemblies from Crystalline State. *J Mol Biol* **372**, 774–797 (2007).
373. Simonetta, K. R. *et al.* Prospective discovery of small molecule enhancers of an E3 ligase-substrate interaction. *Nat Commun* **10**, 1–12 (2019).
374. Sijbesma, E. *et al.* Structure-based evolution of a promiscuous inhibitor to a selective stabilizer of protein–protein interactions. *Nat Commun* **11**, 1–9 (2020).
375. Khan, Z. M. *et al.* Structural basis for the action of the drug trametinib at KSR-bound MEK. *Nature* 1–9 (2020) doi:10.1038/s41586-020-2760-4.
376. Huang, H. *et al.* E2 enzyme inhibition by stabilization of a low-affinity interface with ubiquitin. *Nat Chem Biol* **10**, 156–163 (2014).
377. Mossessova, E., Corpina, R. A. & Goldberg, J. Crystal Structure of ARF1•Sec7 Complexed with Brefeldin A and Its Implications for the Guanine Nucleotide Exchange Mechanism. *Mol Cell* **12**, 1403–1411 (2003).
378. Renault, L., Guibert, B. & Cherfils, J. Structural snapshots of the mechanism and inhibition of a guanine nucleotide exchange factor. *Nature* **426**, 525–530 (2003).
379. Wu, G. *et al.* Structure of a  $\beta$ -TrCP1-Skp1- $\beta$ -catenin complex: Destruction motif binding and lysine specificity of the SCF $\beta$ -TrCP1 ubiquitin ligase. *Mol Cell* **11**, 1445–1456 (2003).
380. Vakser, I. A. Protein-Protein Docking: From Interaction to Interactome. *Biophys J* **107**, 1785 (2014).
381. Gainza, P. *et al.* Deciphering interaction fingerprints from protein molecular surfaces using geometric deep learning. *Nature Methods* 2019 17:2 **17**, 184–192 (2019).
382. Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* 2021 596:7873 **596**, 583–589 (2021).
383. Watson, E. R. *et al.* Molecular glue CELMoD compounds are allosteric regulators of cereblon conformation. *bioRxiv* 2022.07.02.498551 (2022) doi:10.1101/2022.07.02.498551.
384. Liu, J. *et al.* Genetic fusions favor tumorigenesis through degron loss in oncogenes. *Nature Communications* 2021 12:1 **12**, 1–15 (2021).
385. Tokheim, C. *et al.* Systematic characterization of mutations altering protein degradation in human cancers. *Mol Cell* **81**, 1292-1308.e11 (2021).
386. Poirson, J. *et al.* Proteome-scale induced proximity screens reveal highly potent protein degraders and stabilizers. doi:10.1101/2022.08.15.503206.
387. Fields, S. & Song, O. K. A novel genetic system to detect protein-protein interactions. *Nature* **340**, 245–246 (1989).
388. Diss, G. & Lehner, B. The genetic landscape of a physical interaction. *Elife* **7**, (2018).
389. Garcia-Seisdedos, H., Empereur-Mot, C., Elad, N. & Levy, E. D. Proteins evolve on the edge of supramolecular self-assembly. *Nature* **548**, 244–247 (2017).



390. Hochberg, G. K. A. *et al.* A hydrophobic ratchet entrenches molecular complexes. *Nature* **588**, (2020).
391. Yamazoe, S. *et al.* Heterobifunctional Molecules Induce Dephosphorylation of Kinases—A Proof of Concept Study. *J Med Chem* (2020) doi:10.1021/acs.jmedchem.9b01167.
392. Wang, W. W. *et al.* Targeted Protein Acetylation in Cells Using Heterobifunctional Molecules. *J Am Chem Soc* **143**, 16700–16708 (2021).
393. Siriwardena, S. U. *et al.* Phosphorylation-Inducing Chimeric Small Molecules. *J Am Chem Soc* **142**, 14052–14057 (2020).
394. Modell, A. E., Lai, S., Nguyen, T. M. & Choudhary, A. Bifunctional modalities for repurposing protein function. *Cell Chem Biol* **28**, 1081–1089 (2021).
395. Bellenie, B. R. *et al.* Achieving in Vivo Target Depletion through the Discovery and Optimization of Benzimidazolone BCL6 Degradable. *J Med Chem* **63**, 4047–4068 (2020).
396. Blake, R. A. Abstract 4452: GNE-0011, a novel monovalent BRD4 degrader. in *Cancer Research* vol. 79 4452–4452 (American Association for Cancer Research (AACR), 2019).
397. Du, X. *et al.* Insights into protein–ligand interactions: Mechanisms, models, and methods. *International Journal of Molecular Sciences* vol. 17 Preprint at <https://doi.org/10.3390/ijms17020144> (2016).
398. Lin, X., Li, X. & Lin, X. A review on applications of computational methods in drug screening and design. *Molecules* **25**, 1–17 (2020).
399. Cong, Q., Anishchenko, I., Ovchinnikov, S. & Baker, D. Protein interaction networks revealed by proteome coevolution. *Science (1979)* **365**, 185–189 (2019).
400. Koukos, P. I. & Bonvin, A. M. J. J. Integrative Modelling of Biomolecular Complexes. *J Mol Biol* **432**, 2861–2881 (2020).
401. Savojardo, C., Martelli, P. L. & Casadio, R. Protein–Protein Interaction Methods and Protein Phase Separation. *Annu Rev Biomed Data Sci* **3**, 89–112 (2020).
402. Fischer, E. Einfluss der Configuration auf die Wirkung der Enzyme. *Berichte der deutschen chemischen Gesellschaft* **27**, 2985–2993 (1894).
403. Wu, Y. L. *et al.* Structural basis for an unexpected mode of SERM-Mediated ER antagonism. *Mol Cell* **18**, 413–424 (2005).
404. Lin, H. *et al.* Basis for metabolite-dependent Cullin-RING ligase deneddylation by the COP9 signalosome. *Proc Natl Acad Sci U S A* **117**, 4117–4124 (2020).
405. Blind, R. D. Structural analyses of inositol phosphate second messengers bound to signaling effector proteins. *Adv Biol Regul* **75**, 100667 (2020).
406. Scherer, P. C. *et al.* Inositol hexakisphosphate (IP6) generated by IP5K mediates cullin-COP9 signalosome interactions and CRL function. *Proc Natl Acad Sci U S A* **113**, 3503–3508 (2016).
407. Hartmann, M. D. *et al.* Thalidomide mimics uridine binding to an aromatic cage in cereblon. *J Struct Biol* **188**, 225–232 (2014).
408. Cho, J. *et al.* UDP-Glucose: A Cereblon-Dependent Glucokinase Protein Degradable. *International Journal of Molecular Sciences* 2022, Vol. 23, Page 9094 **23**, 9094 (2022).

409. Luzarowski, M. *et al.* Global mapping of protein–metabolite interactions in *Saccharomyces cerevisiae* reveals that Ser-Leu dipeptide regulates phosphoglycerate kinase activity. *Communications Biology* **2021 4:1** **4**, 1–15 (2021).
410. Wishart, D. S. *et al.* HMDB 5.0: the Human Metabolome Database for 2022. *Nucleic Acids Res* **50**, D622–D631 (2022).
411. Abdulrahman, W. *et al.* A set of baculovirus transfer vectors for screening of affinity tags and parallel expression strategies. *Anal Biochem* **385**, 383–385 (2009).
412. Sievers, Q. L., Gasser, J. A., Cowley, G. S., Fischer, E. S. & Ebert, B. L. Genome-wide screen identifies cullin-RING ligase machinery required for lenalidomide-dependent CRL4CRBN activity. *Blood* **132**, 1293–1303 (2018).
413. Doench, J. G. *et al.* Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. *Nat Biotechnol* **34**, 184–191 (2016).
414. Sanjana, N. E., Shalem, O. & Zhang, F. Improved vectors and genome-wide libraries for CRISPR screening. *Nat Methods* **11**, 783–784 (2014).
415. Benjamini, Y. & Hochberg, Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society. Series B (Methodological)* **57**, 289–300 (1995).
416. Abdulrahman, W. *et al.* A set of baculovirus transfer vectors for screening of affinity tags and parallel expression strategies. *Anal Biochem* **385**, 383–5 (2009).
417. Li, T., Pavletich, N. P., Schulman, B. A. & Zheng, N. High-level expression and purification of recombinant SCF ubiquitin ligases. *Methods in Enzymology* vol. 398 125–142 Preprint at [https://doi.org/10.1016/S0076-6879\(05\)98012-9](https://doi.org/10.1016/S0076-6879(05)98012-9) (2005).
418. Fischer, E. S. *et al.* Structure of the DDB1-CRBN E3 ubiquitin ligase in complex with thalidomide. *Nature* **512**, 49–53 (2014).
419. Winn, M. D. *et al.* Overview of the CCP4 suite and current developments. *Acta Crystallogr D Biol Crystallogr* **67**, 235–42 (2011).
420. Tickle, I.J., Flensburg, C., Keller, P., Paciorek, W., Sharff, A., Vonrhein, C., Bricogne, G. STARANISO. Preprint at (2018).
421. McCoy, A. J. *et al.* Phaser crystallographic software. *J Appl Crystallogr* **40**, 658–674 (2007).
422. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* **66**, 486–501 (2010).
423. Afonine, P. v *et al.* Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr D Biol Crystallogr* **68**, 352–67 (2012).
424. Bricogne, G. B. E. BUSTER. Preprint at (2011).
425. Moriarty, N. W., Grosse-Kunstleve, R. W. & Adams, P. D. electronic Ligand Builder and Optimization Workbench (eLBOW): a tool for ligand coordinate and restraint generation. *Acta Crystallogr D Biol Crystallogr* **65**, 1074–80 (2009).

426. Chen, V. B. *et al.* MolProbity: All-atom structure validation for macromolecular crystallography. *Acta Crystallogr D Biol Crystallogr* **66**, 12–21 (2010).
427. Cer, R. Z., Mudunuri, U., Stephens, R. & Lebeda, F. J. IC50-to-Ki: a web-based tool for converting IC50 to Ki values for inhibitors of enzyme activity and ligand binding. *Nucleic Acids Res* **37**, 441–5 (2009).
428. Sperling, A. S. *et al.* Patterns of substrate affinity, competition, and degradation kinetics underlie biological activity of thalidomide analogs. *Blood* **134**, 160–170 (2019).
429. Oumata, N., Ferandin, Y., Meijer, L. & Galons, H. Practical synthesis of roscovitine and CR8. *Org Process Res Dev* **13**, 641–644 (2009).
430. Monastyrskyi, A. *et al.* Development of dual casein kinase 1 $\delta$ /1 $\epsilon$  (CK1 $\delta$ / $\epsilon$ ) inhibitors for treatment of breast cancer. *Bioorg Med Chem* **26**, 590–602 (2018).
431. Larsen, A. F. & Ulven, T. Direct N9-arylation of purines with aryl halides. *Chemical Communications* **50**, 4997–4999 (2014).
432. Zakeri, B. *et al.* Peptide tag forming a rapid covalent bond to a protein, through engineering a bacterial adhesin. *Proc Natl Acad Sci U S A* **109**, (2012).
433. Kabsch, W. Integration, scaling, space-group assignment and post-refinement. *Acta Crystallogr D Biol Crystallogr* **66**, 133–144 (2010).
434. Lebedev, A. A. *et al.* JLigand: a graphical tool for the CCP4 template-restraint library. *Acta Crystallogr D Biol Crystallogr* **68**, 431 (2012).
435. Croll, T. I. ISOLDE: A physically realistic environment for model building into low-resolution electron-density maps. *Acta Crystallogr D Struct Biol* **74**, 519–530 (2018).
436. Williams, C. J. *et al.* MolProbity: More and better reference data for improved all-atom structure validation. *Protein Science* **27**, 293–315 (2018).
437. Tian, W., Chen, C., Lei, X., Zhao, J. & Liang, J. CASTp 3.0: computed atlas of surface topography of proteins. *Nucleic Acids Res* **46**, W363–W367 (2018).
438. Friesner, R. A. *et al.* Glide: A New Approach for Rapid, Accurate Docking and Scoring. 1. Method and Assessment of Docking Accuracy. *J Med Chem* **47**, 1739–1749 (2004).
439. Friesner, R. A. *et al.* Extra precision glide: Docking and scoring incorporating a model of hydrophobic enclosure for protein-ligand complexes. *J Med Chem* **49**, 6177–6196 (2006).
440. Meier, F. *et al.* diaPASEF: parallel accumulation–serial fragmentation combined with data-independent acquisition. *Nat Methods* **17**, 1229–1236 (2020).
441. R Foundation for Statistical Computing. R Development Core Team. *R: A Language and Environment for Statistical Computing*. *Journal of Computer and Communications* (R Foundation for Statistical Computing, 2014). doi:10.4236/JCC.2016.413003.
442. Ritchie, M. E. *et al.* limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* **43**, e47 (2015).
443. Osakabe, A. *et al.* Structural basis of pyrimidine-pyrimidone (6–4) photoproduct recognition by UV-DDB in the nucleosome. *Scientific Reports 2015 5:1* **5**, 1–9 (2015).

444. Ruthenburg, A. J. *et al.* Recognition of a Mononucleosomal Histone Modification Pattern by BPTF via Multivalent Interactions. *Cell* **145**, 692–706 (2011).
445. Fierz, B. *et al.* Histone H2B ubiquitylation disrupts local and higher-order chromatin compaction. *Nature Chemical Biology* *2011* 7:2 **7**, 113–119 (2011).
446. Marks, B. D. *et al.* Multiparameter Analysis of a Screen for Progesterone Receptor Ligands: Comparing Fluorescence Lifetime and Fluorescence Polarization Measurements. <https://home.liebertpub.com/adt> **3**, 613–622 (2006).

## Abbreviations

aa	amino acid
ADP	adenoside diphosphate
AMP	adenoside monophosphate
ANOVA	analysis of variance
APC	Anaphase Promoting Complex
ATP	adenoside triphosphate
BET	Bromodomain and Extra-Terminal
BFP	blue fluorescent protein
BLI	biolayer interferometry
BPA/B/C	B-propeller A/B/C
BRCA	BRest CAncer gene
BRET	Bioluminescence Resonance Energy Transfer
CDK	cyclin-dependent kinase
cGAMP	2'3'-Cyclic GMP-AMP
CK1 $\alpha$	casein kinase 1 $\alpha$
CN	copy number
CRL	cullin-RING ligase
CSN	COP9 signalosome
CTD	C-terminal domain
cycK	cyclin K
DBD	DNA-binding domain
DC <sub>50</sub>	concentration at which 50% of the maximal degradation is observed
DCAF	DDB1-CUL4 associated factor
DDB1/2	Damage Specific DNA Binding Protein 1/2
DDR	DNA damage response
DM1	myotonic dystrophy
DMEM	Dulbecco's Modified Eagle Medium
DMSO	dimethyl sulfoxide
dsDNA	double-stranded deoxyribonucleic acid
DTT	dithiothreitol
DUB	deubiquitinating enzyme
EC <sub>50</sub>	half maximal effective concentration
EMSA	electron mobility shift assays
FDA	Food and Drug Administration
FDR	false discovery rate
FP	fluorescence polarisation

FRAP	fluorescence recovery after photobleaching
FTD	focal tandem duplications
GPP	Genome Perturbation Platform
HBA	hydrogen bond acceptor
HLH	helix-loop-helix
HR	homologous recombination
HSC	haematopoietic stem cells
HTS	high-throughput screening
IC <sub>50</sub>	half maximal inhibitory concentration
IEX	ion exchange
IFN	interferon
IMiD	immunomodulatory drug
ITC	isothermal titration calorimetry
KSR	kinase suppressor of Ras
MDM2	mouse double minute 2
MR	molecular replacement
NCP	nucleosome core particle
NGS	next-generation sequencing
NTC	non-targeting control
PARP	poly (ADP-ribose) polymerase
PBS	phosphate buffered saline
PCA	principal component analysis
PCR	polymerase chain reaction
PDB	Protein Data Bank
PEG	polyethylene glycol
PR	proline-rich motif
PROTAC	proteolysis targeting chimera
RBM23/39	RNA Binding Motif Protein 23/39
RING	Really Interesting New Gene
RNAPII	RNA polymerase II
SAR	structure-activity relationship
SEC	size exclusion chromatography
SeEN-seq	Selected Engagement on Nucleosome sequencing
SLS	Swiss Light Source
SNP	single nucleotide polymorphism
SPR	surface plasmon resonance
TAD	transactivation domain
TBD	thalidomide-binding domain
TF	transcription factor
TNBC	triple-negative breast cancer

TR-FRET	time-resolved Förster resonance energy transfer
UPS	ubiquitin-proteasome system
VHL	von Hippel-Lindau
WT	wild type
ZF	zinc finger

## Acknowledgements

First and foremost, I am grateful to Nico Thomä for entrusting me with many exciting and ambitious projects during my time in his laboratory and for his guidance and infectious enthusiasm throughout. I would also like to acknowledge my thesis committee members Angela Koehler, Prisca Liberali, and Dirk Schübeler for their ongoing support and helpful advice.

I extend my gratitude to all the former and current members of the Thomä laboratory: Jacob Aguirre, Amparo Andres-Pons, Anja Basters, Richard Bunker, Manuel Carminati, Simone Cavadini, Deyasini Chakraborty, Seemon Coomar, David Domjan, Vivian Focht, Pius Galli, Susanne Kassube, Georg Kempf, Syota Matsumoto, Alicia Michael, Ganesh Pathare, Georg Petzold, Carlos Pla Prats, Pooja Sadana, Colby Sandate, Andreas Schenk, Mira Schütz, Lisa Stoos, Dakota Suchyta, Julius Rabl, Luca Vecchia, and Joscha Weiss for creating an enjoyable working atmosphere. I would like to give special thanks to Georg Petzold, not only for his guidance and unwavering scientific rigour, but also for his excellent sense of humour and many great experiences during our shared travels; to Alicia Michael for her friendship and pottery sessions that always kept my spirits up; Vivian Focht for great teamwork and graciously putting up with me as her Master thesis supervisor; to Mira Schütz for keeping everything in order; to Georg Kempf for always coming to the rescue in case of X-ray crystallography or modelling trouble; and to Jake Aguirre and Carlos Pla Prats for being excellent office mates.

I have been very fortunate to be involved in many fruitful collaborations during my doctoral work. I am grateful to the Ebert laboratory for the many exciting joint projects, especially to Ben Ebert for his brilliant insights and helpful advice throughout and to Mikołaj Słabicki for excellent over-the-pond teamwork and for his ongoing friendship. I also extend my gratitude to Georg Winter for a great collaboration and to the Fischer laboratory, in particular Katherine Donovan, for help with proteomics experiments during multiple projects. I am thankful also to Ilya Lukonin, Prisca Liberali, and Laurent Gelman for embarking with us on an unusual high-content screening project, as well as to Martin Rénatus, Uli Hassiepen, and Sandra Romero-Lopez for their help with conducting experiments at Novartis. Additionally, I am grateful to the UbiCODE Marie-Curie Skłodowska International Training Network for funding my thesis work and I would like to acknowledge the principal investigators and SAB members involved (in particular Ivan Dikič and Andrea Pichler) for their helpful insights and the student cohort for many fun pre-pandemic excursions.

Finally, I would like to thank my parents, Dorota Kozicka and Marek Kozicki, for their continued love and encouragement, my loving husband, Michał Kępa, for his unfaltering support, and my wonderful friends for providing many a welcome distraction during the four years of my PhD.



## Supplementary Information

The Supplementary Information includes the following data pertaining to **Chapter 3**:

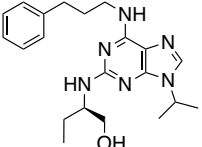
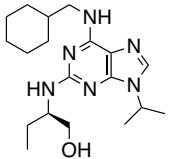
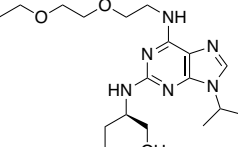
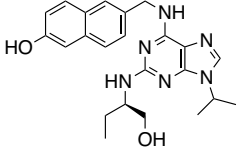
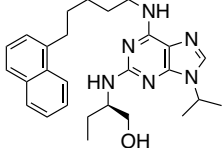
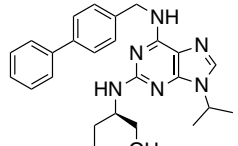
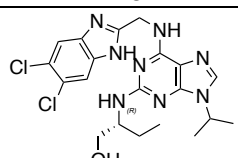
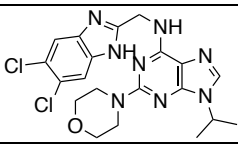
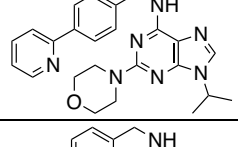
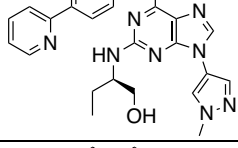
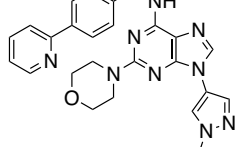
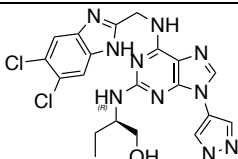
**Table S3.1:** Table summarising the key assay data for all putative cyclin K molecular glue degrader compounds evaluated.

**Figure S3.1:** Supplementary crystallographic data (density).

**Figure S3.2:** Supplementary crystallographic data (interaction distances).

**Figure S3.3:** Supplementary kinase inhibitory selectivity characterisation data (Lanthascreen).



DS11		56.77	n.d.	n.d.	n.d.	0 %						✓	This work
DS12		(774)	n.d.	n.d.	n.d.	0 %							This work
DS13		n.d.	n.d.	n.d.	n.d.	0 %							This work
DS14		n.d.	n.d.	n.d.	n.d.	0 %							This work
DS15		182	n.d.	n.d.	n.d.	0 %						✓	This work
DS16		16.68	2423	629	n.d.	1490 %						✓	This work
DS17		13.19	12	93	142	53 %	117	10260	3061	✓	✓	✓	This work
DS18		8.771	13	101	191	89 %	289	5729	4482		✓	✓	This work
DS19		14	62	203	813	300 %	163	239	143		✓	✓	This work
DS20		1837	n.d.	n.d.	n.d.	0 %	n.d.	n.d.	3195				This work
DS21		299	3124	4940	n.d.	102 %	2862	n.d.	953				This work
DS22		19.33	109	935	1549	66 %	1316	n.d.	n.d.			✓	This work



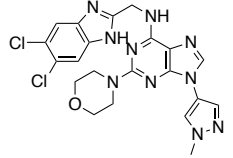
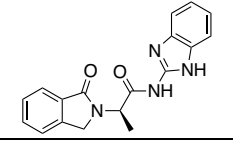
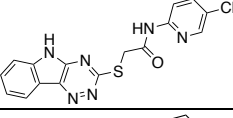
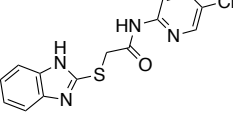
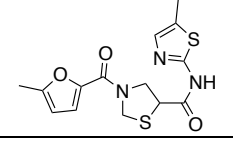
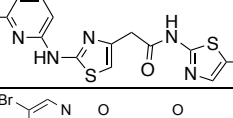
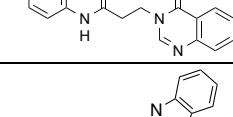
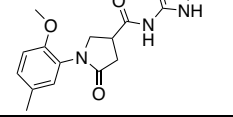
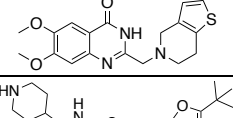
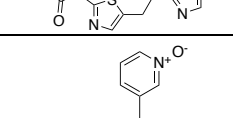
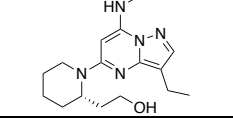
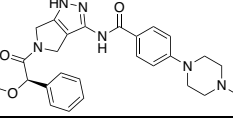
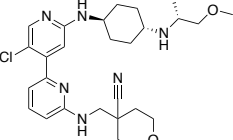
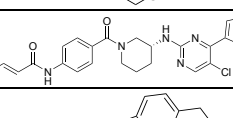
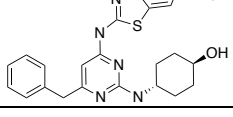


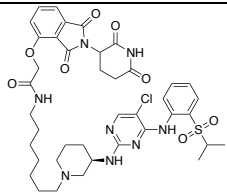
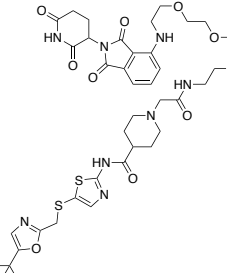




<b>DS70</b>		17.56	*	265	345	30 %	1	31	5				This work
<b>DS71 (P133)</b>		27.6	4	10	71	610 %	134	n.d.	684				2
<b>DS72 (P25)</b>		12.24	6	6	38	533 %	1	1056	104				2
<b>DS73 (P342)</b>		15.32	1	2	19	850 %	25	6116	557				2
<b>DS74 (P419)</b>		19.77	51	104	1057	916 %	25	n.d.	2583				2
<b>WX3</b>		21.36	23	179	1887	954 %						✓	This work
<b>WX4-INT</b>		78.53	365	1175	6251	432 %							This work
<b>Roscovitine</b>		702.6	n.d.	n.d.	n.d.	0 %					✓	✓	3
<b>DRF053</b>		703.6	n.d.	1310	6383	387 %						✓	4
<b>Flavopiridol</b>		51.77	n.d.	395	400	1 %							5
<b>21195</b>		12.03	26	39	127	226 %	31	3	1	✓	✓	✓	6



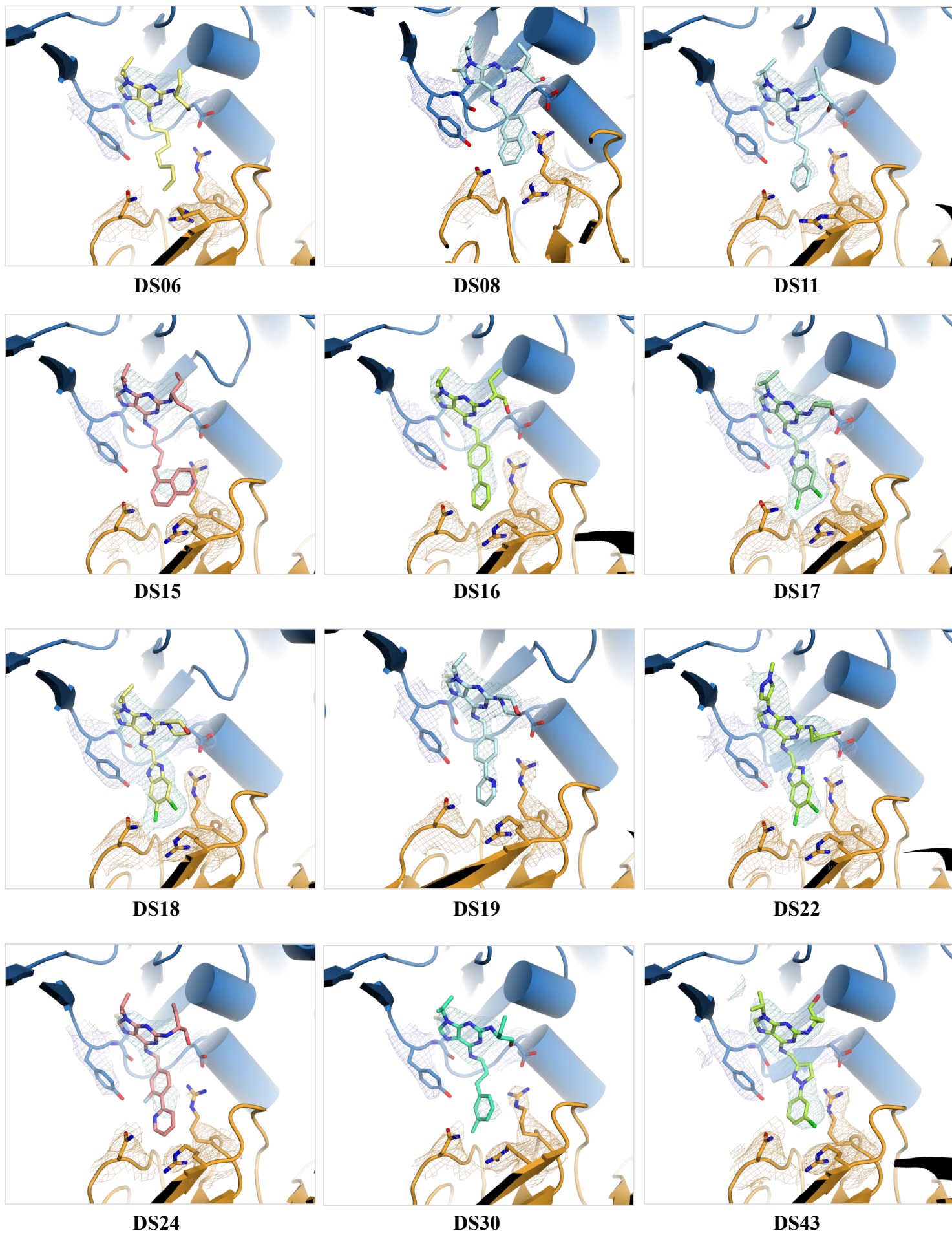
SR-4835		15.9	14	42	313	645 %	27	n.d.	3664	✓	✓	✓	7
919278		37.8	176	559	4748	749 %	1335	n.d.	120	✓	✓	✓	8
dCeMM2		83.22	633	1423	7084	398 %	n.d.	n.d.	1113				9
dCeMM3		263.4	n.d.	n.d.	n.d.	0 %	2666	n.d.	3020			✓	9
dCeMM4		278.8	1210	2626	n.d.	281 %	6547	n.d.	2459			✓	9
HQ461		42.7	493	1296	3358	159 %	n.d.	n.d.	n.d.	✓	✓	✓	10
Z7		302.6	n.d.	n.d.	n.d.	0 %	n.d.	n.d.	873			✓	This work
Z11		177.9	436	1066	n.d.	838 %	3218	n.d.	n.d.			✓	This work
Z12		833	3492	1366	2036	49 %	1944	n.d.	2233			✓	This work
SNS-032		n.d.	n.d.	265	207	-22 %							11
Dinacliclib		n.d.	n.d.	17	15	-12 %					✓		12
Danusertib		n.d.	n.d.	n.d.	n.d.	0 %							13
NVP-2		n.d.	n.d.	23	18	-22 %							14
THZ531		n.d.	n.d.	271	212	-22 %					✓		15
GSK2250665A		n.d.	n.d.	?	?	?							16

<b>BSJ-4-116</b>		n.d.	n.d.	155	511	230 %					✓	17
<b>THAL-SNS032</b>		n.d.	<50	192	705	267 %					✓	14

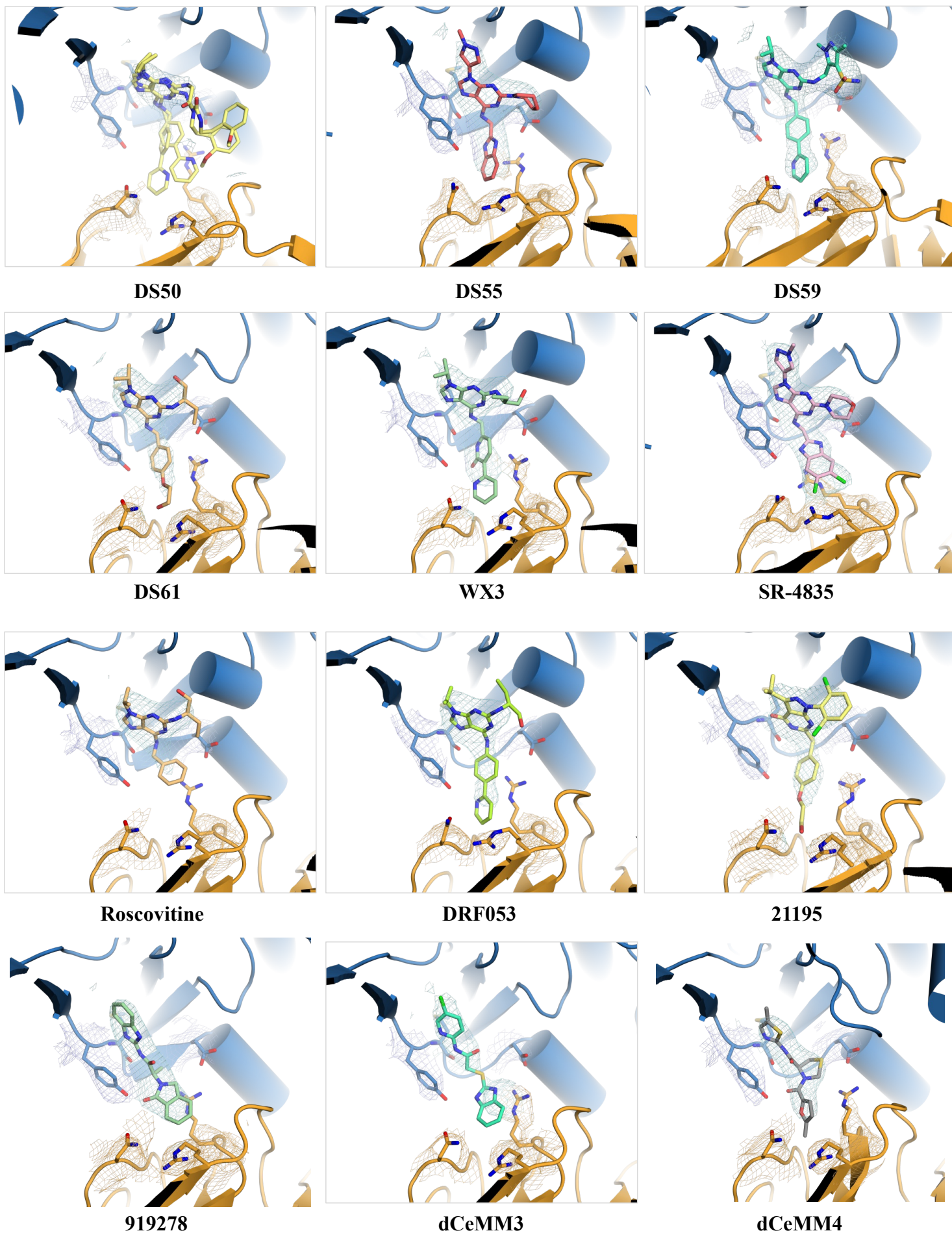
n.d. refers to measurements where  $EC_{50}/DC_{50}/IC_{50}$  value is higher than the highest compound concentration tested (10  $\mu$ M). For MLN4924-mediated cytotoxicity rescue calculation, the value of 10  $\mu$ M is used in such cases. Asterisk (\*) denotes cases where the reporter degradation curve could not be fitted due to compound autofluorescence. Question mark (?) indicates cases where the fit was ambiguous for other reasons.

## References

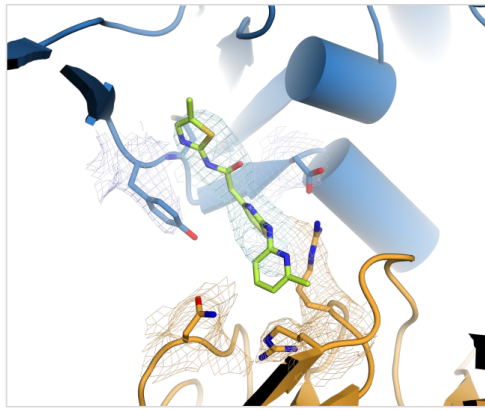
- Bettayeb, K. *et al.* CR8, a potent and selective, roscovitine-derived inhibitor of cyclin-dependent kinases. *Oncogene* **27**, 5797–5807 (2008).
- Thede, K. *et al.* PYRAZOLOTRIAZINES. Patent WO2021116178. (2021).
- Meijer, L. *et al.* Biochemical and Cellular Effects of Roscovitine, a Potent and Selective Inhibitor of the Cyclin-Dependent Kinases cdc2, cdk2 and cdk5. *Eur J Biochem* **243**, 527–536 (1997).
- Oumata, N. *et al.* Roscovitine-derived, dual-specificity inhibitors of cyclin-dependent kinases and casein kinases 1. *J Med Chem* **51**, 5229–5242 (2008).
- Sedlacek, H. H. *et al.* Flavopiridol (L86 8275; NSC 649890), a new kinase inhibitor for tumor therapy. *Int J Oncol* **9**, 1143–1168 (1996).
- Caligiuri, M. *et al.* A proteome-wide CDK/CRK-specific kinase inhibitor promotes tumor cell death in the absence of cell cycle progression. *Chem Biol* **12**, 1103–1115 (2005).
- Quereda, V. *et al.* Therapeutic Targeting of CDK12/CDK13 in Triple-Negative Breast Cancer. *Cancer Cell* **36**, 545-558.e7 (2019).
- Henry, K. L. *et al.* CDK12-mediated transcriptional regulation of noncanonical NF- $\kappa$ B components is essential for signaling. <http://stke.sciencemag.org/> (2018).
- Mayor-Ruiz, C. *et al.* Rational discovery of molecular glue degraders via scalable chemical profiling. *Nat Chem Biol* (2020) doi:10.1038/s41589-020-0594-x.
- Lv, L. *et al.* Discovery of a molecular glue promoting CDK12-DDB1 interaction to trigger Cyclin K degradation. *Elife* **9**, (2020).
- SNS-032 is a potent and selective inhibitor of CDK2, 7 and 9 and induces cell death by inhibiting cell cycle progression and the expression of antiapoptotic proteins | Cancer Research | American Association for Cancer Research. [https://aacrjournals.org/cancerres/article/66/8\\_Supplement/491/528216/SNS-032-is-a-potent-and-selective-inhibitor-of](https://aacrjournals.org/cancerres/article/66/8_Supplement/491/528216/SNS-032-is-a-potent-and-selective-inhibitor-of).
- Parry, D. *et al.* Dinacliclib (SCH 727965), a novel and potent cyclin-dependent kinase inhibitor. *Mol Cancer Ther* **9**, 2344–2353 (2010).
- Carpinelli, P. *et al.* PHA-739358, a potent inhibitor of Aurora kinases with a selective target inhibition profile relevant to cancer. *Mol Cancer Ther* **6**, 3158–3168 (2007).
- Olson, C. M. *et al.* Pharmacological perturbation of CDK9 using selective CDK9 inhibition or degradation. *Nat Chem Biol* **14**, 163–170 (2018).
- Zhang, T. *et al.* Covalent targeting of remote cysteine residues to develop CDK12 and CDK13 inhibitors. *Nat Chem Biol* **12**, 876–884 (2016).
- Alder, C. M. *et al.* Identification of a Novel and Selective Series of Itk Inhibitors via a Template-Hopping Strategy. *ACS Med Chem Lett* **4**, 948–952 (2013).
- Jiang, B. *et al.* Discovery and resistance mechanism of a selective CDK12 degrader. *Nature Chemical Biology* 2021 17:6 **17**, 675–683 (2021).



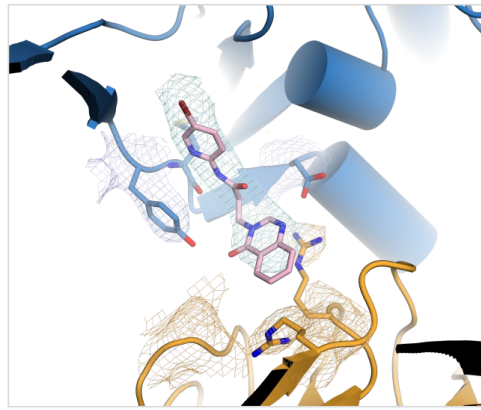
**Supplementary Figure S3.1.** Density map (at the contour level of  $1\sigma$ ) for all ternary complex structures. Density for the small molecule, DDB1 residues R907, R928, R947, and CDK12 residues M816 and Y815 is shown.



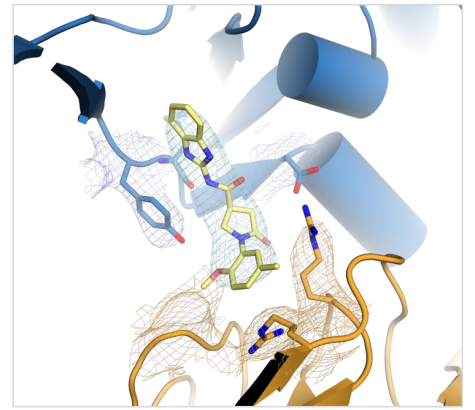
**Supplementary Figure S3.1 (cont.).** Density map (at the contour level of  $1\sigma$ ) for all ternary complex structures. Density for the small molecule, DDB1 residues R907, R928, R947, and CDK12 residues M816 and Y815 is shown.



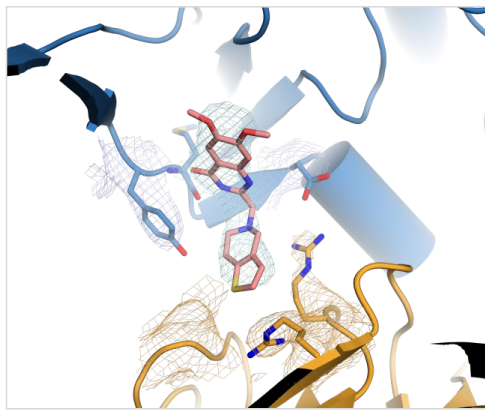
**HQ461**



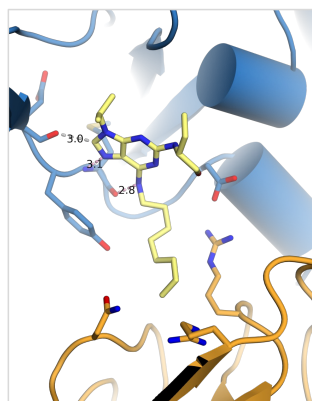
**Z7**



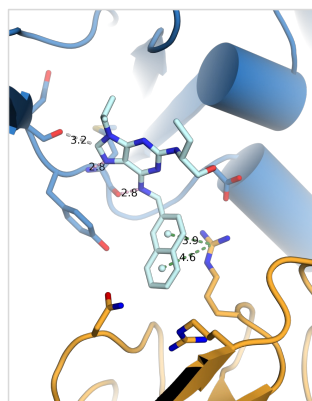
**Z11**



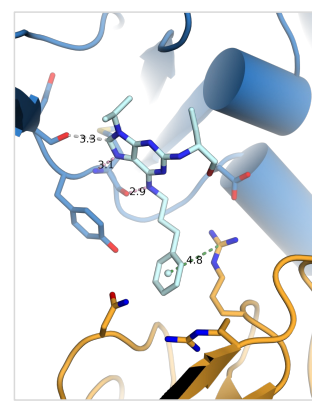
**Z12**



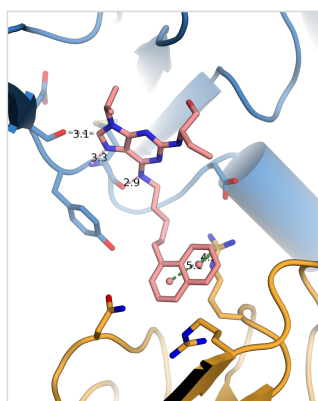
**DS06**



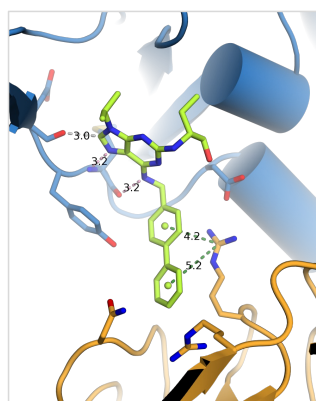
**DS08**



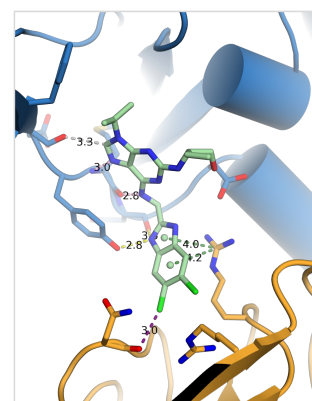
**DS11**



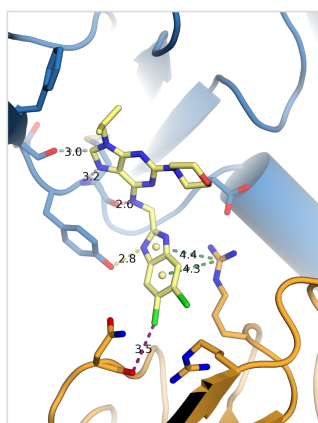
**DS15**



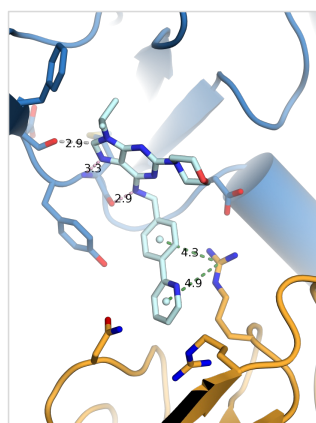
**DS16**



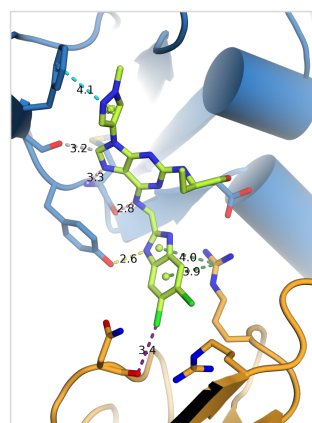
**DS17**



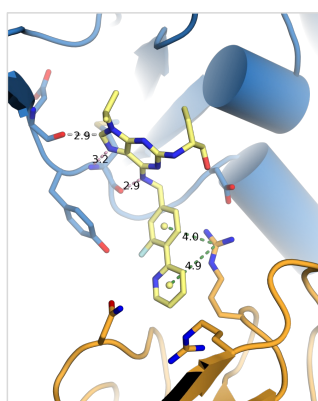
**DS18**



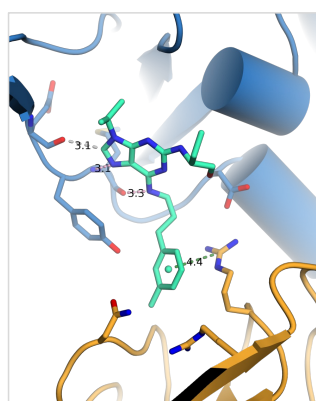
**DS19**



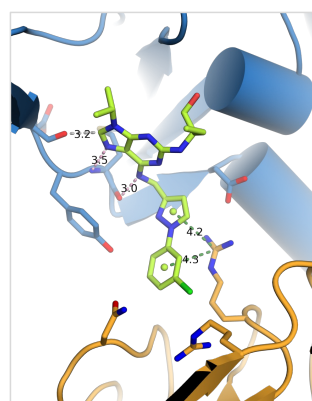
**DS22**



**DS24**

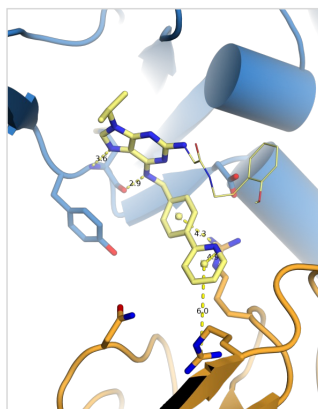


**DS30**

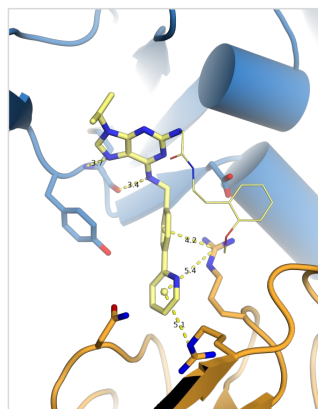


**DS43**

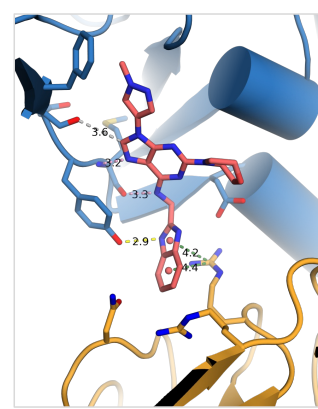
**Supplementary Figure S3.2.** Key protein-ligand interactions for all ternary complex structures. Distances to other proximal residues are also shown where appropriate. Spheres indicate the centre of mass of each aromatic ring in the gluing moiety. These centre points were used for the relevant measurements.



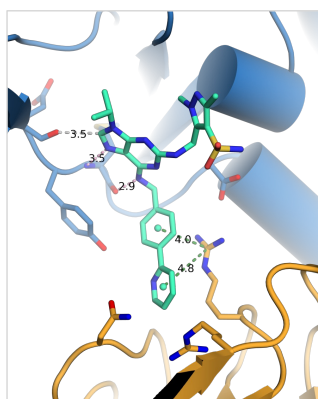
**DS50 (1)**



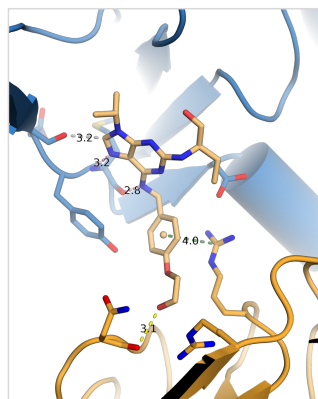
**DS50 (2)**



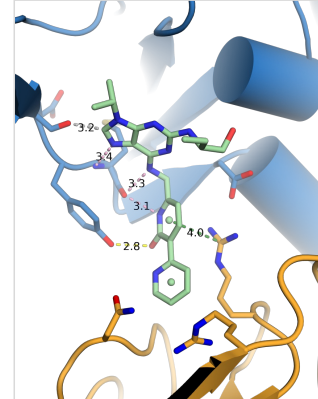
**DS55**



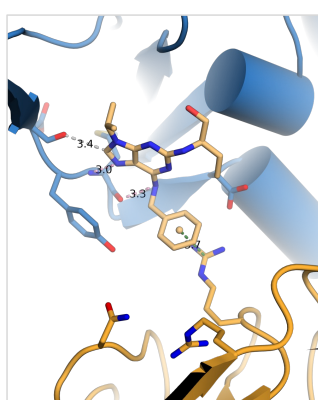
**DS59**



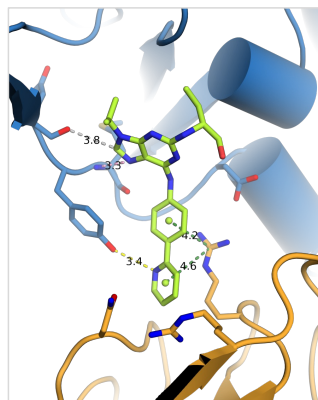
**DS61**



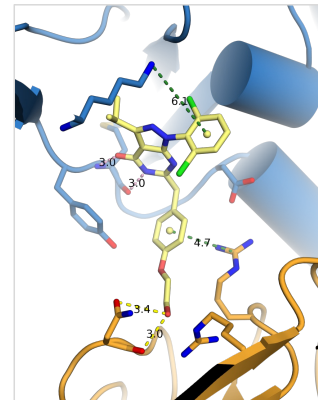
**WX3**



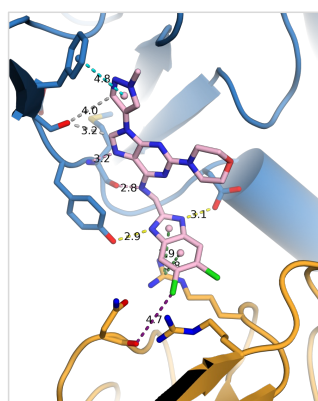
**Roscovitine**



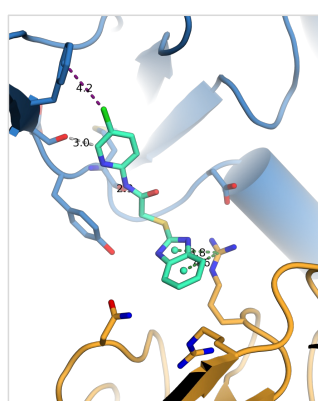
**DRF053**



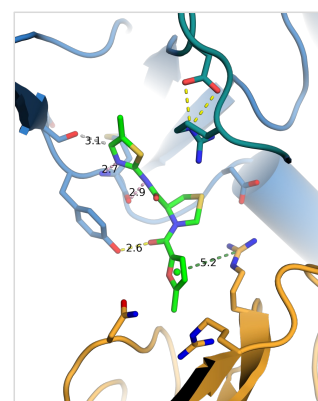
**21195**



**SR-4835**

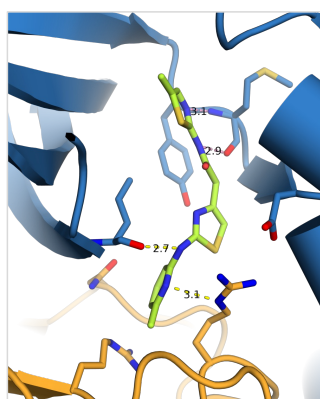


**dCeMM3**

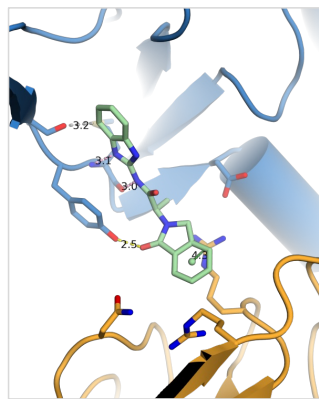


**dCeMM4**

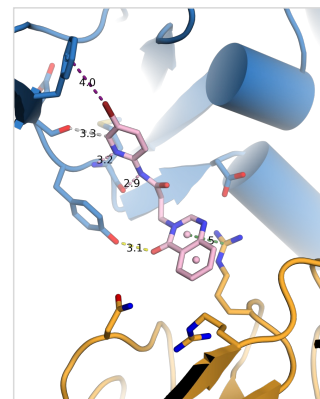
**Supplementary Figure S3.2 (cont.).** Key protein-ligand interactions for all ternary complex structures. Distances to other proximal residues are also shown where appropriate. Spheres indicate the centre of mass of each aromatic ring in the gluing moiety. These centre points were used for the relevant measurements.



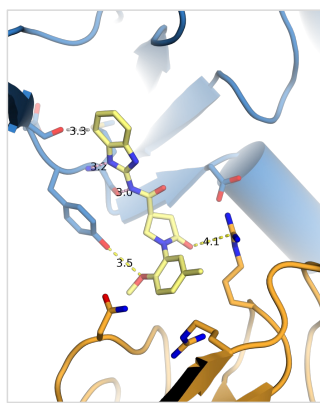
**HQ461**



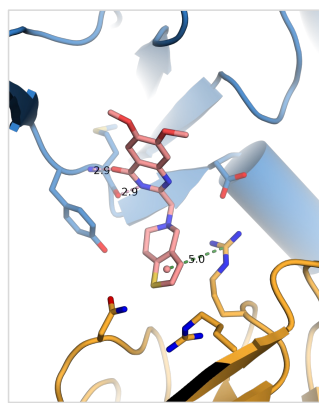
**919278**



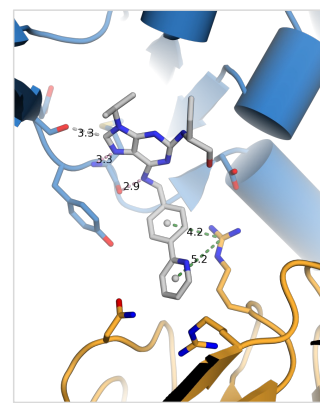
**Z7**



**Z11**



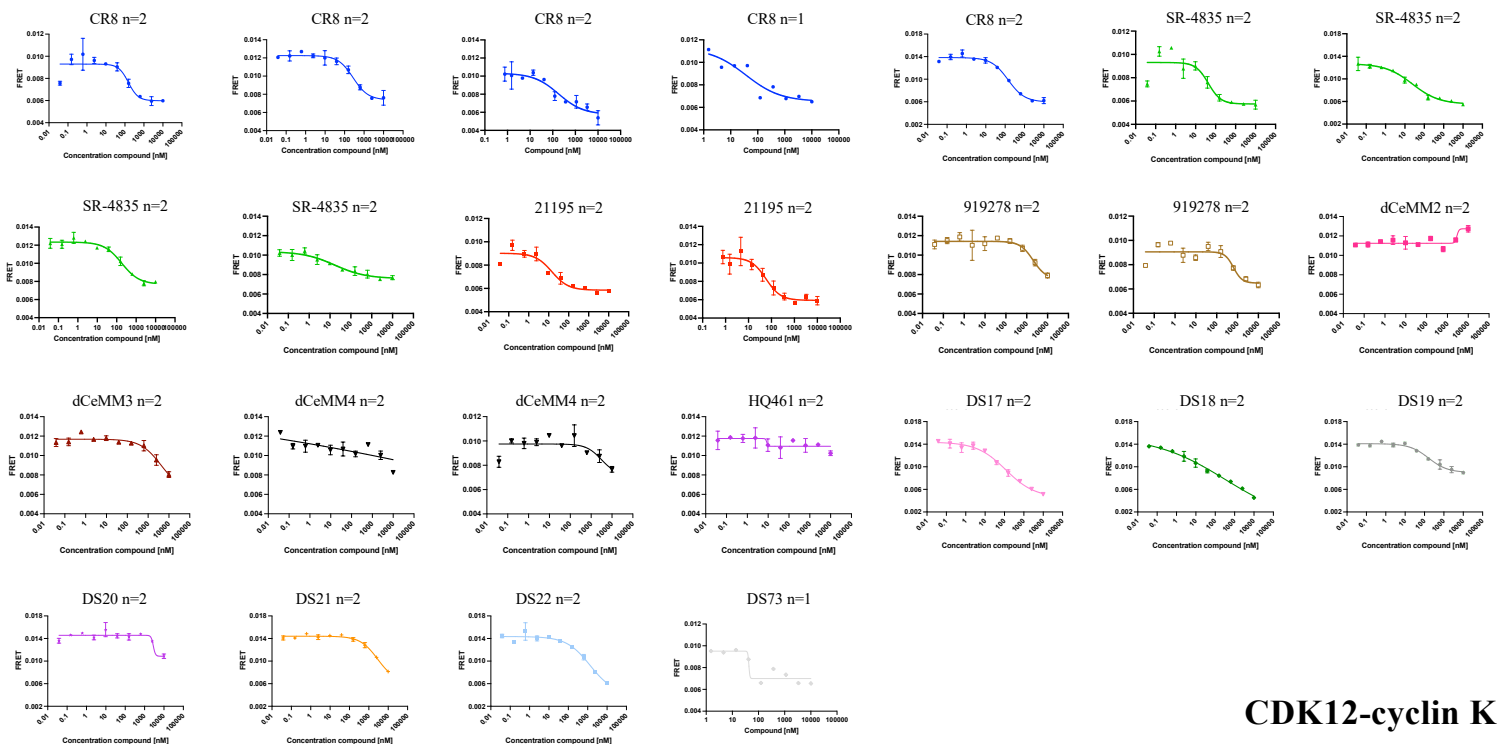
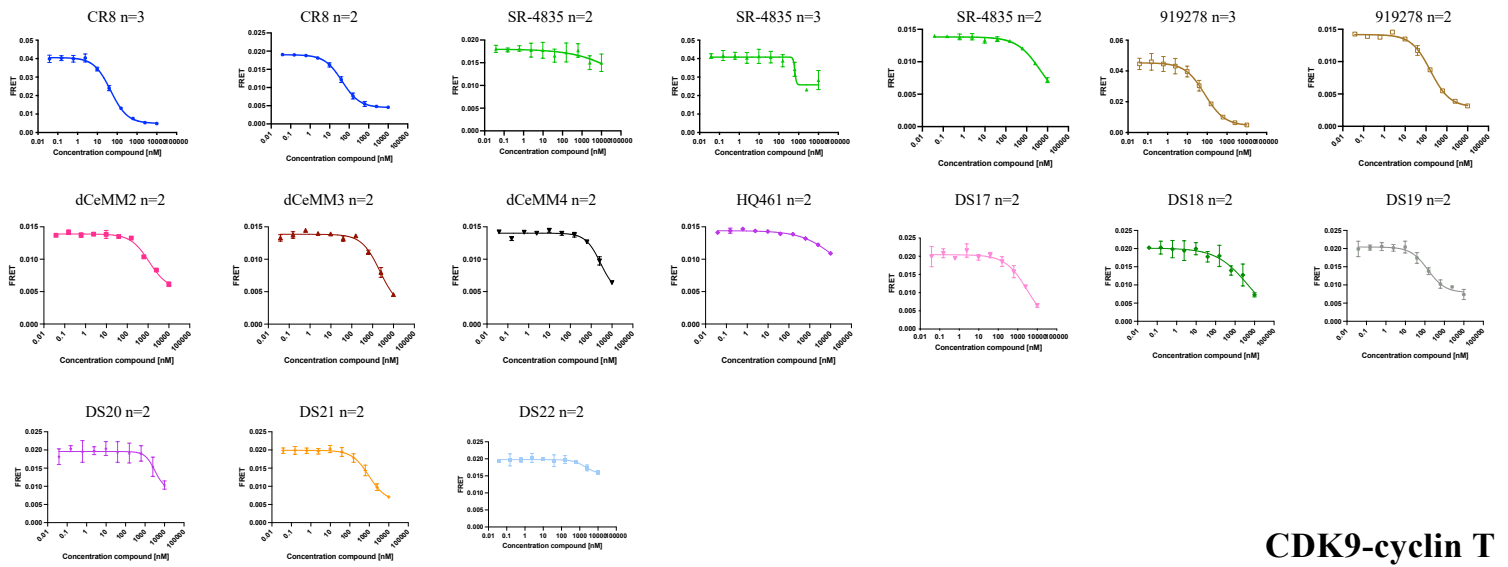
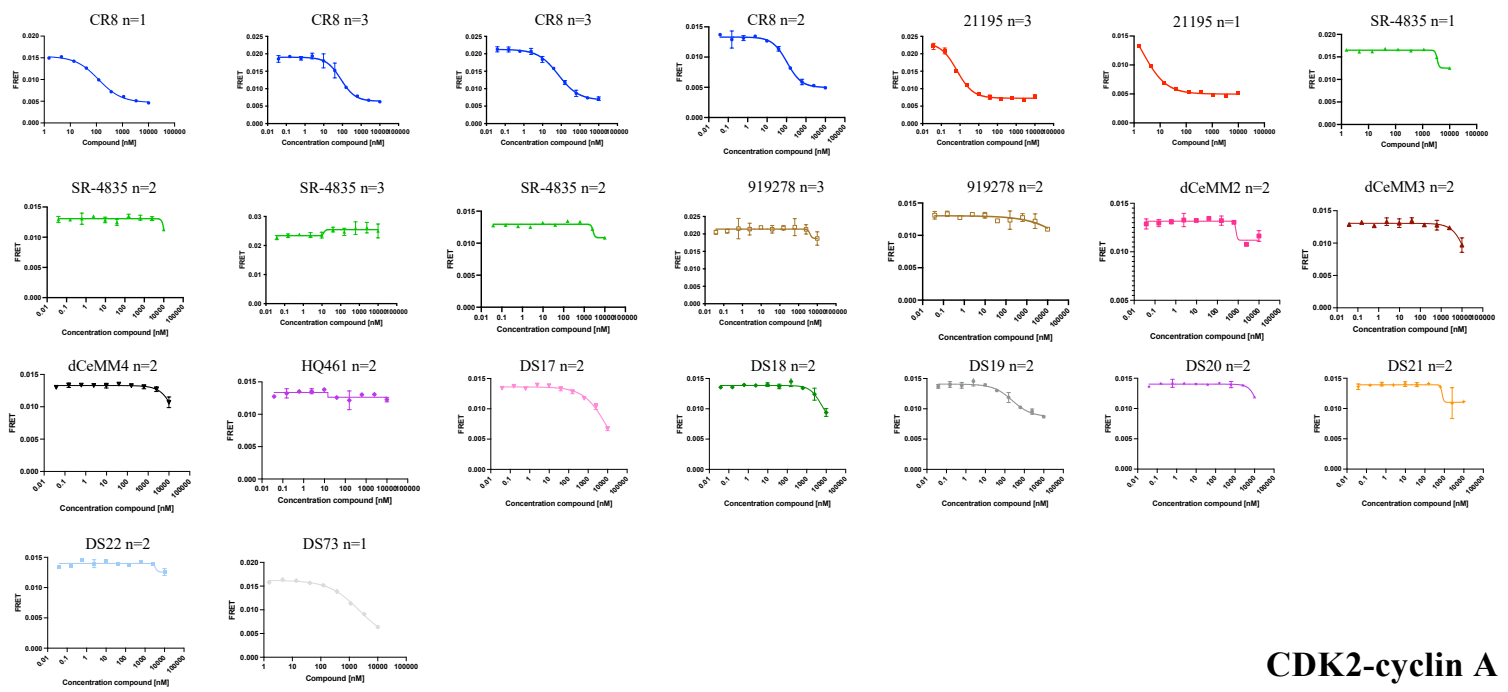
**Z12**



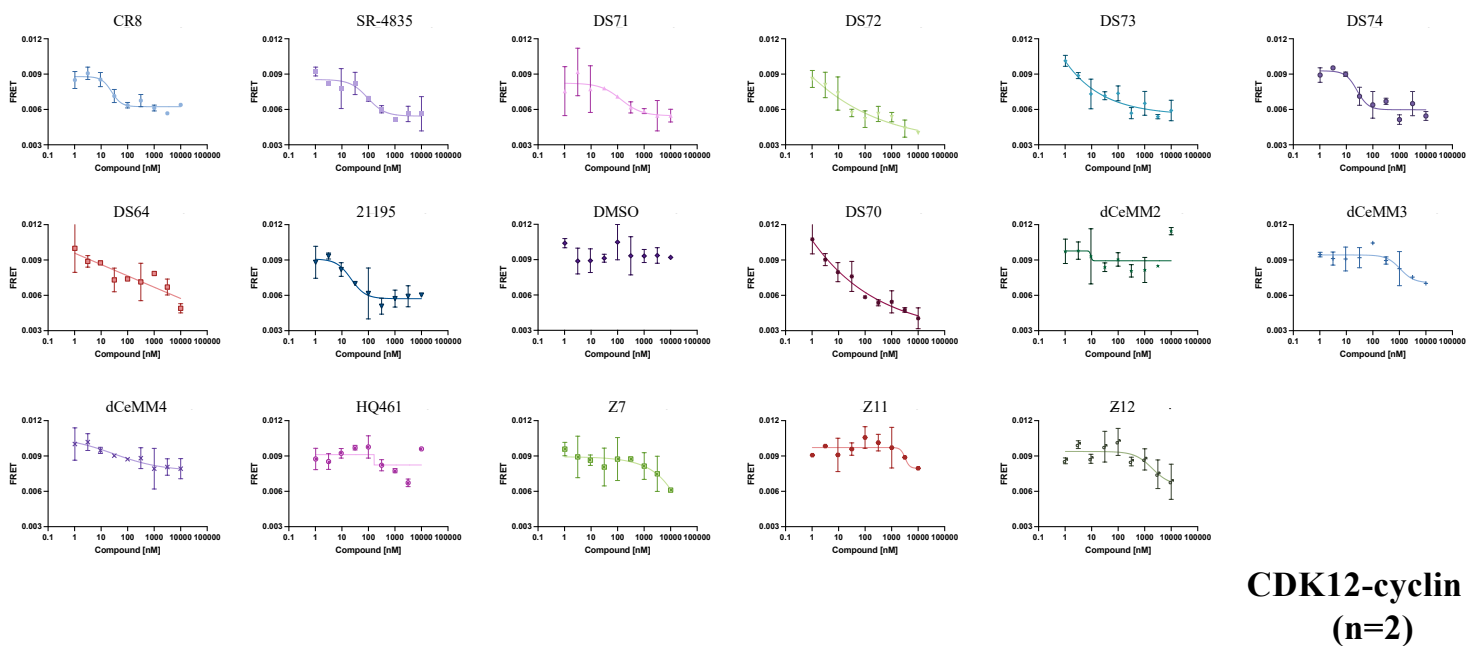
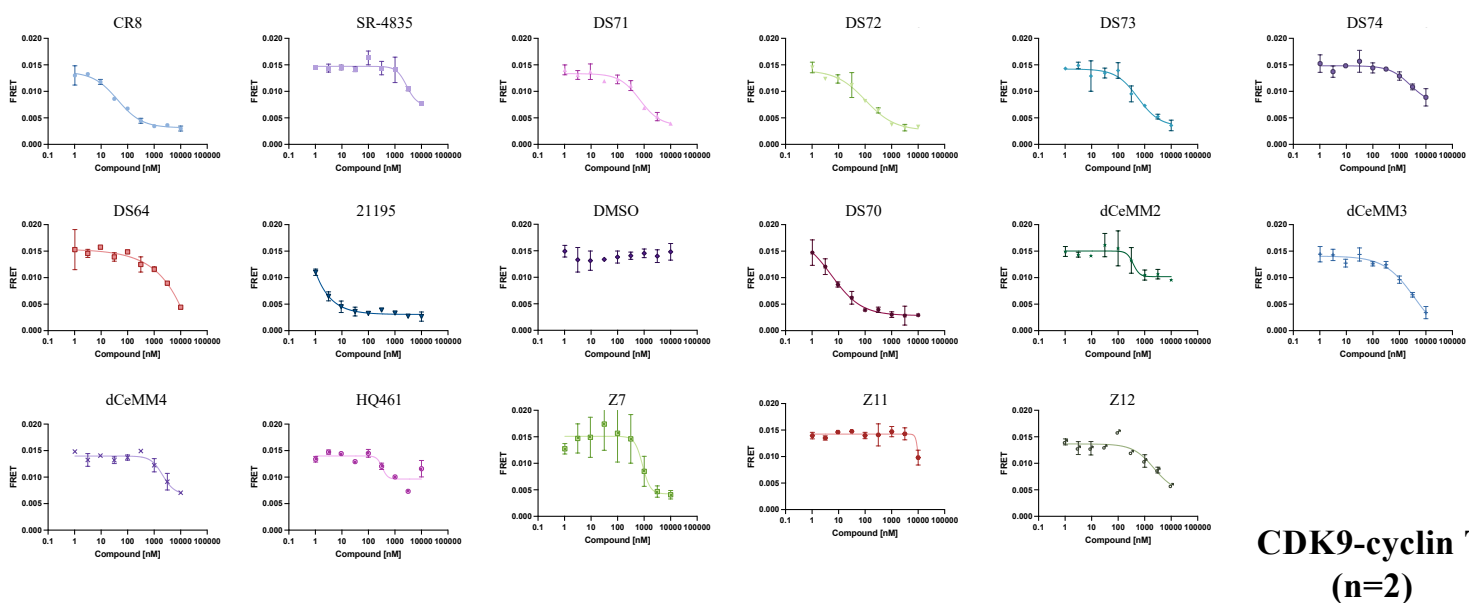
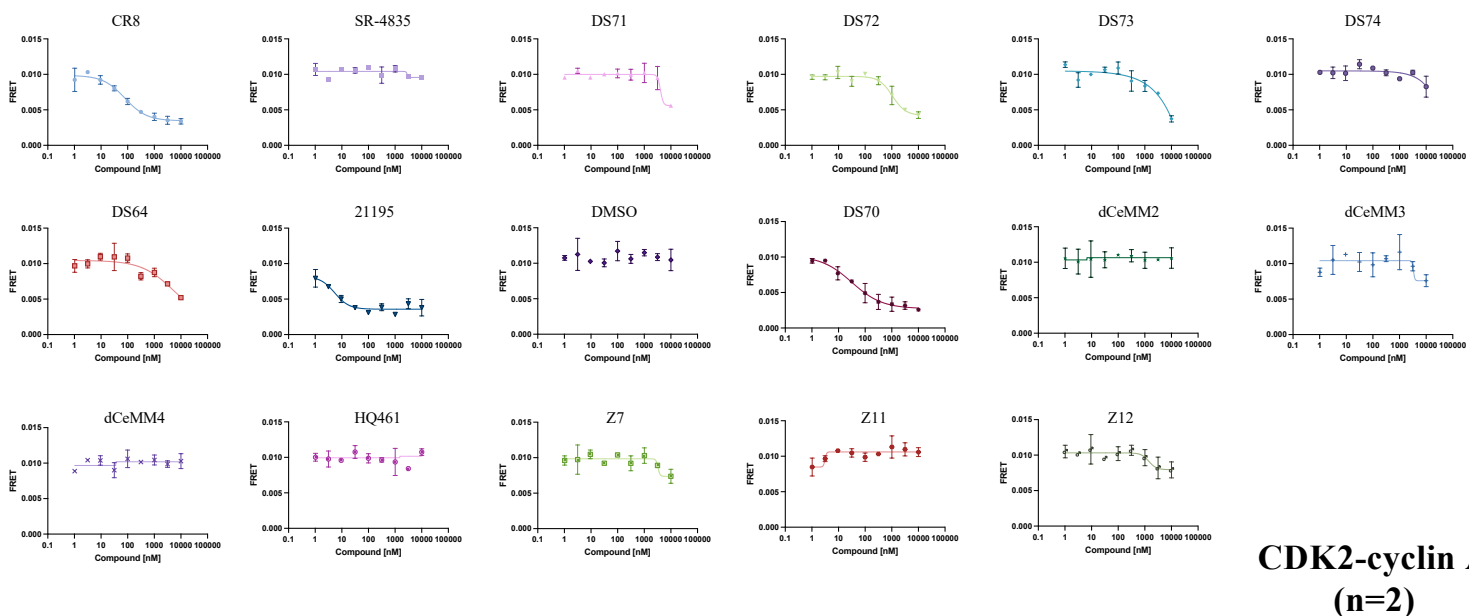
**CR8**

**Supplementary Figure S3.2 (cont.).** Key protein-ligand interactions for all ternary complex structures. Distances to other proximal residues are also shown where appropriate. Spheres indicate the centre of mass of each aromatic ring in the gluing moiety. These centre points were used for the relevant measurements.





Supplementary Figure S3.3. Lanthascreen data for the indicated compound and kinase. Compound dilutions were performed manually.



**Supplementary Figure S3.3 (cont.).** Lanthascreen data for the indicated compound and kinase. Compound dilutions were performed using a D300 digital dispenser (Tecan).

## Appendix

The Appendix includes the main text and figures of the publications listed below. The extended manuscript files can be accessed via the doi links provided.

*Haven't got a glue: Protein surface variation for the design of molecular glue degraders*

**Z. Kozička**, N. Thomä

*Cell Chemical Biology* (2021), <https://doi.org/10.1016/j.chembiol.2021.04.009>

*The CDK inhibitor CR8 acts as a molecular glue degrader that depletes cyclin K*

M. Słabicki\*, **Z. Kozička**\*, G. Petzold\*, Y.-D. Li, M. Manojkumar, R. Bunker, K. Donovan, Q. Sievers, J. Koepfel, D. Suchyta, A. Sperling, E. Fink, J. Gasser, L. Wang, S. Corsello, R. Sellar, M. Jan, D. Gillingham, C. Scholl, S. Fröhling, T. Golub, E. Fischer, N. Thomä<sup>†</sup>, B. Ebert<sup>†</sup>.

*Nature* (2020), <https://doi.org/10.1038/s41586-020-2374-x> (\*equal contribution)

*Rational discovery of molecular glue degraders via scalable chemical profiling*

C. Mayor-Ruiz, S. Bauer, M. Brand, **Z. Kozička**, M. Siklos, H. Imrichova, I. Kalthener, E. Hahn, K. Seiler, A. Koren, G. Petzold, M. Fellner, C. Bock, A. Müller, J. Zuber, M. Geyer, N. Thomä, S. Kubicek, G. Winter.

*Nature Chemical Biology* (2020), <https://doi.org/10.1038/s41589-020-0594-x>

*Taking the brakes off targeted protein degradation*

**Z. Kozička**, G. Petzold, N. Thomä

*Cell Chemical Biology* (2020), <https://doi.org/10.1016/j.chembiol.2019.12.010>

*Mechanisms of OCT4-SOX2 motif readout on nucleosomes*

A. Michael\*, R. Grant\*, L. Isbel\*, S. Cavadini, **Z. Kozička**, R. Bunker, A. Schenk, A. Graff-Meyer, G. Pathare, J. Weiss, S. Matsumoto, L. Burger, D. Schübeler<sup>†</sup>, N. Thomä<sup>†</sup>

*Science* (2020); <https://doi.org/10.1126/science.abb0074>

*Structural mechanism of cGAS inhibition by the nucleosome*

G. Pathare, A. Decout, S. Glück, S. Cavadini, K. Makasheva, R. Hovius, G. Kempf, J. Weiss, **Z. Kozička**, B. Guey, P. Melenec, B. Fierz, N. Thomä<sup>†</sup>, A. Ablasser<sup>†</sup>

*Nature* (2020), <https://doi.org/10.1038/s41586-020-2750-6>

The following preprint is not included and can be accessed online:

*Readout of histone methylation by Trim24 locally restricts chromatin opening by p53*

L. Isbel, I. Murat, S. Durdu, R. S. Grand, J. Weiss, E. Pfeiffer, **Z. Kozička**, A. K. Michael, L. Burger, N. H. Thomä, D. Schübeler

Biorxiv (2022), <https://doi.org/10.1101/2022.08.23.504916>

## Review

# Haven't got a glue: Protein surface variation for the design of molecular glue degraders

Zuzanna Kozicka<sup>1,2</sup> and Nicolas Holger Thomä<sup>1,\*</sup><sup>1</sup>Friedrich Miescher Institute for Biomedical Research, Basel, Switzerland<sup>2</sup>University of Basel, Basel, Switzerland\*Correspondence: [nicolas.thoma@fmi.ch](mailto:nicolas.thoma@fmi.ch)<https://doi.org/10.1016/j.chembiol.2021.04.009>**SUMMARY**

Molecular glue degraders are small, drug-like compounds that induce interactions between an E3 ubiquitin ligase and a target, which result in ubiquitination and subsequent degradation of the recruited protein. In recent years, serendipitous discoveries revealed that some preclinical and clinical compounds already work as molecular glue degraders, with many more postulated to destabilize their targets through indirect or yet unresolved mechanisms. Here we review strategies by which E3 ubiquitin ligases can be reprogrammed by monovalent degraders, with a focus on molecular glues hijacking cullin-RING ubiquitin ligases. We argue that such drugs exploit the intrinsic property of proteins to form higher-order assemblies, a phenomenon previously seen with disease-causing sequence variations. Modifications of the protein surface by a bound small molecule can change the interactome of the target protein. By inducing interactions between a ligase and a substrate, molecular glue degraders offer an exciting path for the development of novel therapeutics.

**INTRODUCTION**

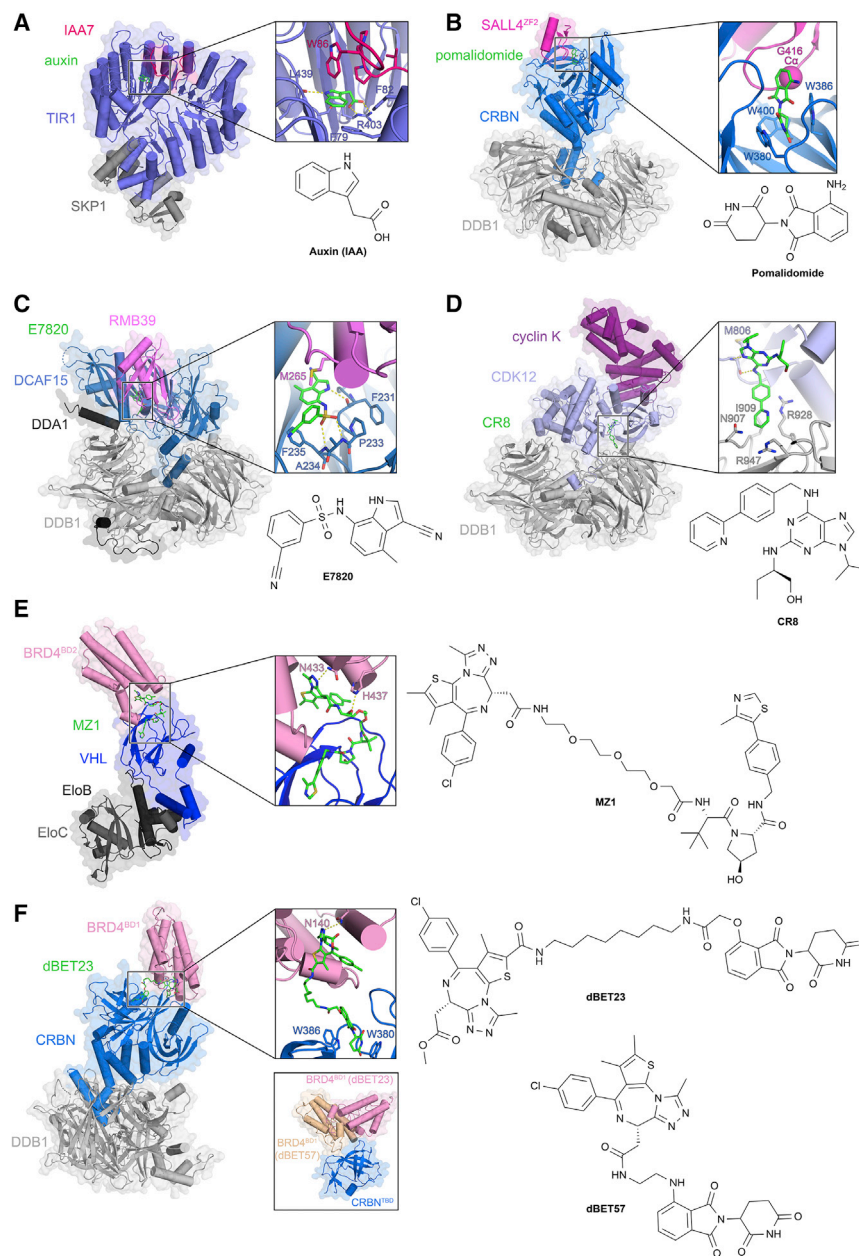
Molecular recognition events between two proteins, or between a protein and a small molecule, are at the heart of every biological process. Through decades of medicinal chemistry efforts, we have attained a good understanding of how to find, design, and optimize protein-ligand interactions (Du et al., 2016; Lin et al., 2020). Through structural analysis of protein-protein interactions by crystallography and, more recently, cryoelectron microscopy (cryo-EM), we have amassed a multitude of examples of protein-protein interfaces to learn from, allowing for recent advances in predicting protein-protein interactions (Cong et al., 2019; Koukos and Bonvin, 2020; Savojardo et al., 2020). Still lacking, though, is a more systematic understanding of how small molecules can remodel, or repurpose, protein-protein interfaces. In drug-discovery efforts, inhibiting protein-protein interactions with small molecules has turned out to be arduous but not infeasible (Scott et al., 2016). Inducing them, on the other hand, was long considered impossible (Giordanetto et al., 2014), but this perception shifted over the last years with the discovery of molecular glues. The elucidation of the proximity-inducing properties of macrolides FK506, cyclosporin, and rapamycin (Liu et al., 1991; Sabatini et al., 1994; Sabers et al., 1995; Schreiber and Crabtree, 1992) paved the way for the discovery of numerous compounds with a similar mode of action (Gerry and Schreiber, 2020). In recent years, molecular glue degrader drugs that cause target protein degradation via induced proximity with a ubiquitin ligase have attracted particular attention,

and will be the focus of this review. Inducing novel interactions between two proteins with small molecules has already yielded Revlimid (lenalidomide), one of the most successful anticancer drugs of our time. While some consider it an isolated happenstance, we argue that such interactions can be triggered much more easily than previously anticipated. We suspect that many possible low-affinity protein-protein interactions are already hardwired into protein surfaces and the appropriate small molecule might tip the balance toward a more stable ternary complex. We reason that understanding and leveraging compound-induced interactions holds enormous potential for the discovery of novel drugs.

**HIJACKING OF THE UBIQUITIN SYSTEM**

Small-molecule-induced proximity is particularly consequential for E3 ubiquitin ligases. The cullin-RING ligase (CRL) family is composed of the CRL 1–5 and 7 clades, which together with ~200 substrate receptors are responsible for ~20% of proteasome-mediated protein degradation in the cell (Soucy et al., 2009). CRLs operate through a dedicated substrate receptor domain that engages the substrate and connects it directly, or with the help of an adaptor protein, to the N terminus of the cullin ligase arm (Duda et al., 2011; Zimmerman et al., 2010). Through its C terminus, cullin engages the RBX1/RBX2 subunit, which serves as the recruitment site for the ubiquitin-loaded E2, allowing for direct ubiquitin transfer onto the substrate. CRLs ubiquitinate diverse substrate proteins when brought into direct proximity to





**Figure 1. Overview of degrader-induced interfaces**

(A) SKP1-TIR1-auxin-IAA7 complex (PDB: 2P1Q) (Tan et al., 2007). (B) DDB1-CRBN-pomalidomide-SALL4<sup>ZF2</sup> complex (PDB: 6UJML) (Matyskiela et al., 2020a). The G416 C<sub>α</sub> in the β-hairpin loop of SALL4<sup>ZF2</sup> is shown as a sphere. (C) DDB1-DDA1-DCAF15-E7820-RBM39 complex (PDB: 6PAI) (Du et al., 2019). (D) DDB1-CR8-CDK12-cyclin K complex (PDB: 6TD3) (Stabicki et al., 2020a). (E) ELoB/C-VHL-MZ1-BRD4<sup>BD2</sup> complex (PDB: 5T35) (Gadd et al., 2017). (F) DDB1-CRBN-dBET23-BRD4<sup>BD1</sup> complex (PDB: 6BN7) (Nowak et al., 2018). Bottom panel: overlay of dBET23- and dBET57-induced conformations of BRD4<sup>BD1</sup> (PDB: 6BNB) aligned on CRBN thalidomide-binding domain (TBD). In (A) to (F), important interface residues are shown as sticks. Key interactions are marked with dashed yellow lines. In (C) and (F), The BPB domain of DDB1 was omitted for clarity.

early gene e1a hijacking CRL4<sup>DCAF1</sup> to degrade RUVBL1/2 (Zemke et al., 2020). The CRL4 clade appears to be particularly amenable to reprogramming by viruses and, as detailed below, by drugs. The flexible attachment of the CUL4 ligase arm to the DDB1 adaptor protein renders the ligase able to rotate by up to 150° by thermal motion. This gives rise to a larger search radius and establishes a ubiquitination zone of about 340 Å × 110 Å × 30 Å from the receptor (Angers et al., 2006; Fischer et al., 2014), facilitating the search for lysine residues in non-native substrates. Ubiquitination by proximity was also observed for other ubiquitin ligases such as the anaphase-promoting complex (APC), where grafting the destruction box of cyclin B onto non-substrate proteins rendered these subject to APC-mediated ubiquitination and degradation (Glotzer et al., 1991; Yamano et al., 1998).

The first biological indication that a small molecule can induce protein degradation

came from plant hormones (Tan et al., 2007). The phytohormones auxin and methyl jasmonate were found to bind the plant F-box CRL receptors (TIR1 and CO11, respectively) and facilitate their interaction with target proteins, two TFs, resulting in enhanced degradation of these targets (Sheard et al., 2010; Tan et al., 2007). Structures revealed that the compact small-molecule hormones, through a small protein-ligand interface, are able to facilitate nanomolar target-ligase interactions (Figure 1A). The molecular mechanism of these compounds differs substantially from that of large macrolides. Phytohormones rely predominantly on large and complementary protein-protein interfaces, with the relatively small compound/neosubstrate interface conferring sufficient additional binding energy for tight complex formation (Tal et al., 2020).

the ligase: CRL4<sup>DDB2</sup>, for example, is recruited to sites of DNA damage, where it ubiquitinates not the DNA lesion itself but various proteins in the immediate proximity of the lesion (Sugasawa et al., 2005). This property is exploited by several virus families, with viral proteins binding and thereby redirecting ligases to ubiquitinate neosubstrates to benefit viral replication. Examples include HBx, which directly binds the CRL4 adaptor DDB1 and leads to the degradation of the HBV restriction factors SMC5/6 (Li et al., 2010), and S5V, which brings about degradation of host transcription factors (TFs) STAT in an analogous manner (Mahon et al., 2014). A similar mechanism is observed for substrate receptor hijacking, with the HIV-1 protein Vpr binding the CRL4 substrate receptor DCAF1 and directing the ligase to ubiquitinate several host proteins including UNG2 (Ahn et al., 2010), and the adenovirus

## BIFUNCTIONAL DEGRADERS AND MOLECULAR GLUES

Following these discoveries, the auxin-inducible degron system was developed for controlled cellular degradation of non-physiological substrate proteins fused to an IAA7 degron (Nishimura et al., 2009). The observation that viruses and plants can reprogram ubiquitin ligases through induced proximity further suggested that targeted protein degradation does not require protein engineering and could instead be achieved with appropriately designed compounds. The man-made proof of concept came from broad successes with rationally designed heterobifunctional degrader molecules (PROTACs) (Bondeson et al., 2015; Gadd et al., 2017; Sakamoto et al., 2001; Winter et al., 2015). These compounds, comprising a ligase-binding ligand and a target-engaging warhead connected by a linker, position the target in proximity to the ubiquitin ligase, leading to target ubiquitination and its proteasomal degradation. As targeted protein degradation offers multiple advantages over traditional inhibition (Burslem and Crews, 2020), PROTACs have attracted keen interest from both academia and industry. Their key advantage is the fact that bifunctionality allows exchange of the target-binding warhead in a plug-and-play fashion to inactivate other proteins. This modularity, at the same time, also gives rise to a key shortcoming, as it renders the compounds comparatively large, making it more challenging to progress these molecules into the clinic.

Molecular glue degraders, on the other hand, also facilitate dimerization of a target protein and an E3 ubiquitin ligase but they are single linker-less scaffolds that induce proximity by relying on extensive and direct ligase-target protein-protein interactions. These drugs alter the interactome of the ligase or the target, strengthening or enabling *de novo* protein-protein interactions, which in the context of a ubiquitin ligase can result in degradation. In some regards, the difference between molecular glues and bifunctional degraders is one in degree, not kind, as PROTACs can also induce extensive direct protein-protein interactions (Chamberlain, 2018; Gadd et al., 2017; Nowak et al., 2018) (Figure 1). Differences in size and monovalency versus bivalency remain to distinguish the classes. A significant disadvantage of molecular glues is that they are difficult to find and all have thus far been discovered serendipitously for a given target, with a notable exception previously reported (Simonetta et al., 2019). In the following sections we review the known molecular glue degrader compounds, focusing on examples that have been structurally characterized to date, and suggest strategies for how to more rationally develop such compounds.

## MOLECULAR GLUES FOR CRL4<sup>CRBN</sup>

Thalidomide was introduced in West Germany in 1957 as a sedative and antiemetic, and was found to be highly teratogenic, affecting more than 10,000 patients before the drug was withdrawn from the market in 1962 (Vargesson, 2015). Through a series of unexpected discoveries, thalidomide and its analogs re-emerged as powerful therapeutics in cancer and hematological malignancies. The long-sought-after understanding of their mode of action came with the identification of CRBN as the efficacy target of thalidomide by the Handa laboratory (Ito et al., 2010) and subsequent structural studies showing that CRBN plays the role of the substrate receptor of the CRL4<sup>CRBN</sup> E3 ubiquitin ligase (Chamber-

lain et al., 2014; Fischer et al., 2014). Thalidomide and derivatives (termed IMiDs for IMMunomodulatory imide Drugs) were eventually identified to be molecular glue degraders that bind a highly conserved tri-tryptophan pocket on CRBN, with the composite CRBN-ligand interface recruiting several different neosubstrates for ubiquitination. Degradation of TFs Ikaros/Aiolos (IKZF1/3) has been linked to the clinical efficacy of thalidomide analogs in multiple myeloma, while casein kinase 1 $\alpha$  (CK1 $\alpha$ ) is the likely efficacy target of lenalidomide in 5q-myelodysplastic syndrome (Chamberlain et al., 2014; Fischer et al., 2014; Gandhi et al., 2014; Ito et al., 2010; Krönke et al., 2014, 2015; Lu et al., 2014).

Structural studies of glue-induced ternary complexes of CK1 $\alpha$  (Petzold et al., 2016), GSPT1 (Matyskiela et al., 2016), Ikaros (Sievers et al., 2018), ZNF692 (Sievers et al., 2018), and SALL4 (Furihata et al., 2020; Matyskiela et al., 2020a) all revealed a structural degron in the neosubstrate comprising a characteristic  $\beta$ -hairpin loop and a key glycine residue that engage the otherwise solvent-exposed phthalimide moiety of IMiDs, and contact residues lining the IMiD pocket on CRBN (Figure 1B) (Petzold et al., 2016; Sievers et al., 2018). With the help of a relatively small protein-protein interface ( $\sim$ 350–700 Å<sup>2</sup>) (Table 1), the compound-bound CRBN complex provides a nanomolar binding platform for neosubstrates.

Arguably the biggest structural class of neosubstrates for CRBN is zinc-finger (ZF) TFs. Structures of thalidomide derivatives with IKZF1, ZNF692, and SALL4 showed that binding of the small and distinct ZF fold is similarly driven by a  $\beta$ -hairpin-containing ZF, while the surrounding ZFs appear to facilitate binding *in vitro*. In cells, only those ZF-containing proteins that carry an additional ZF close by were proficient in degradation of the corresponding full-length protein (Donovan et al., 2018; Sievers et al., 2018), but the role of the nearby ZFs in specificity and degradation proficiency requires further investigation. Several other Cys<sub>2</sub>His<sub>2</sub> (C2H2) ZF TFs have been shown to be degraded by thalidomide and derivatives (Table 2). For example, candidate acute promyelocytic leukemia therapeutics CC-647 and CC-3060 have recently been reported to promote ZBTB16 degradation, but each engage different ZF domains within this transcription factor (Matyskiela et al., 2020b). Multiple additional ZF proteins were found to bind CRBN *in vitro*, with the engagement being insufficient to drive cellular degradation (An et al., 2017; Sievers et al., 2018).

Many new thalidomide derivatives have been developed in recent years, and several compounds with improved efficacy have entered clinical trials (Bjorklund et al., 2020; Lopez-Girona et al., 2019; Schafer et al., 2018) (Table 2). Following the clinical success of early thalidomide analogs, IMiDs have also been successfully modified to recruit novel substrates. Still, even if apparent specificity is observed in given cell types, many of these compounds appear to be intrinsically polypharmacological (Sievers et al., 2018). For instance, growing out the isoindolinone scaffold of lenalidomide led to the pleiotropic ZF/GSPT1 degrader CC-885 (Matyskiela et al., 2016), and later yielded the more GSPT1-selective compounds CC-90009 (Surka et al., 2020) and ZXH-1-161 (Powell et al., 2020). Small changes in compound structure strongly alter neosubstrate degradation specificity (Figure 2), with a recent report on 5-hydroxythalidomide, a thalidomide metabolite, illustrating how the presence of a single hydroxyl leads to a striking change in degradation selectivity from IKZF1 to SALL4, possibly explaining the

**Table 1. Comparison of structurally characterized drug-induced protein-protein interfaces for small-molecular degraders**

Compound	Ligase	Target	PDB ID	Protein-protein interface (%)	Ligase-ligand interface (%)	Target-ligand interface (%)	Total interface area (Å <sup>2</sup> )
Auxin	SKP1 <sup>TIR1</sup>	IAA7 peptide	2P1Q <sup>a</sup>	71	22	7	926
Lenalidomide	CRL4 <sup>CRBN</sup>	CK1α	F5QD <sup>b</sup>	68	23	9	1,052
Pomalidomide	CRL4 <sup>CRBN</sup>	SALL4 (ZF2)	6UML <sup>c</sup>	52	35	13	681
CC-885	CRL4 <sup>CRBN</sup>	GSPT1	5HXB <sup>d</sup>	52	28	20	1,164
CC-90009	CRL4 <sup>CRBN</sup>	GSPT1	6XK9 <sup>e</sup>	54	27	19	1,221
Pomalidomide	CRL4 <sup>CRBN</sup>	IKZF1 (ZF2)	6H0F <sup>f</sup>	61	28	12	859
Pomalidomide	CRL4 <sup>CRBN</sup>	ZNF692 (ZF4)	6H0G <sup>f</sup>	61	28	10	880
Thalidomide	CRL4 <sup>CRBN</sup>	SALL4 (ZF2)	7BQU <sup>g</sup>	56	32	13	706
5-Hydroxythalidomide	CRL4 <sup>CRBN</sup>	SALL4 (ZF2)	7BQV <sup>g</sup>	60	29	12	802
E7820	CRL4 <sup>DCAF15</sup>	RBM39	6PAI <sup>h</sup> , 6Q0R <sup>i</sup>	68	22	10	1,463
Indisulam	CRL4 <sup>DCAF15</sup>	RBM39	6UD7 <sup>j</sup> , 6Q0W <sup>i</sup>	72	19	9	1,665
CR8	CRL4 <sup>Δ</sup>	cyclin K	6TD3 <sup>k</sup>	79	6	16	2,672
NRX-103094	SKP1 <sup>β-TrCP</sup>	β-catenin peptide	6M91 <sup>l</sup>	54	25	21	958
MZ1	CRL2 <sup>VHL</sup>	BRD4 (BD2)	5T35 <sup>m</sup>	26	37	38	1,338
dBET23	CRL4 <sup>CRBN</sup>	BRD4 (BD1)	6BN7 <sup>n</sup>	39	23	38	1,328

Interface areas were calculated with PISA (Krissinel and Henrick, 2007). For indisulam and E7820, average values from the two available structures are shown.

<sup>a</sup>Tan et al., 2007.

<sup>b</sup>Petzold et al., 2016.

<sup>c</sup>Matyskiela et al., 2020a.

<sup>d</sup>Matyskiela et al., 2016.

<sup>e</sup>Surka et al., 2020.

<sup>f</sup>Sievers et al., 2018.

<sup>g</sup>Furihata et al., 2020.

<sup>h</sup>Du et al., 2019.

<sup>i</sup>Faust et al., 2020.

<sup>j</sup>Bussiere et al., 2020.

<sup>k</sup>Stabicki et al., 2020a.

<sup>l</sup>Simonetta et al., 2019.

<sup>m</sup>Gadd et al., 2017.

<sup>n</sup>Nowak et al., 2018.

teratogenicity of thalidomide (Donovan et al., 2018; Furihata et al., 2020). However, it remains highly challenging to rationally tune IMiD selectivity or to predict which compound will trigger degradation of which subset of β-hairpin-containing proteins, and hence to rationally decode the relationship between IMiD used and target degraded. Future efforts are required to map which β-hairpin-containing proteins in the genome can be recruited to CRL4<sup>CRBN</sup> and investigate whether the spectrum of structural CRL4<sup>CRBN</sup> degrons can be further extended, possibly to also target non-glycine containing protein loops.

The structure-activity relationship (SAR) for IMiD-induced complexes is somewhat unusual in that mutating the CRBN/neosubstrate interface, or using less efficient molecular glue compounds that give rise to 2- to 5-fold weaker neosubstrate recruitment *in vitro*, can trigger steep transitions between degradation and non-degradation *in vivo*, in what appears to be an all-or-nothing manner. Relatively small changes in binding mode convert already engaged substrates into degraded targets while at the same time rendering other tightly bound neosubstrates less tightly engaged and not degraded.

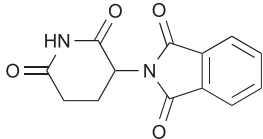
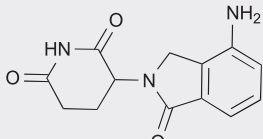
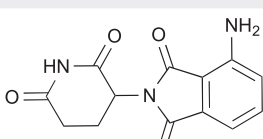
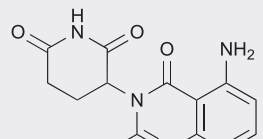
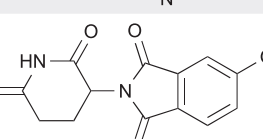
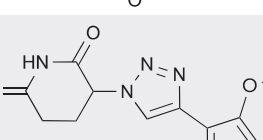
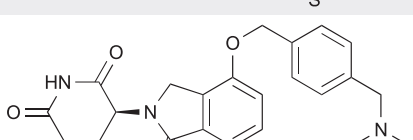

All CRLs, including the CRL4 clade, are regulated by the COP9 signalosome, an isopeptidase that removes the ubiquitin-like acti-

vator NEDD8 from the cullin using a steric gating mechanism. Those CRLs that are engaged with substrates remain neddylated, while idle ligases are deneddylated (Cavadini et al., 2016). Thalidomide-induced substrate complexes have been shown to be sufficiently substrate-like to maintain the ligase in a neddylated and active state (Cavadini et al., 2016). All glue-induced neosubstrates for cullin ubiquitin ligases must pass this substrate criterion, and it is unclear to what extent very small domains, or linear peptides, are sufficiently good mimics to support the ubiquitination process (Cavadini et al., 2016; Enchev et al., 2012; Fischer et al., 2011). Analogously, other components of the ubiquitin-proteasome system such as p97 and the proteasome also require the neosubstrate to be recognized as sufficiently substrate-like to enable efficient degradation (Donovan et al., 2020).

### MOLECULAR GLUES FOR CRL4<sup>DCAF15</sup>

Following the discovery of thalidomide analogs and their mechanism of action, aryl sulfonamides such as indisulam and E7820 were found to drive degradation of the essential RNA-binding protein RBM39 (RNA-Binding Motif Protein 39) and RBM23 by the CRL4<sup>DCAF15</sup> E3 ubiquitin ligase (Han et al., 2017; Owa

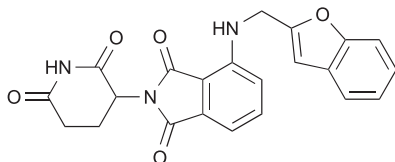
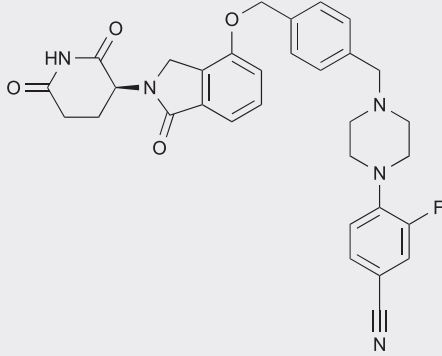
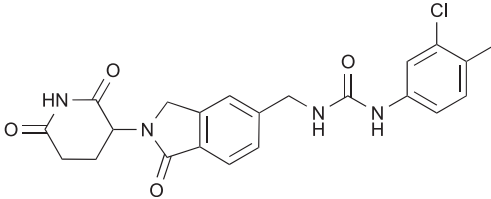
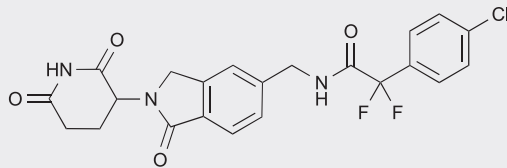
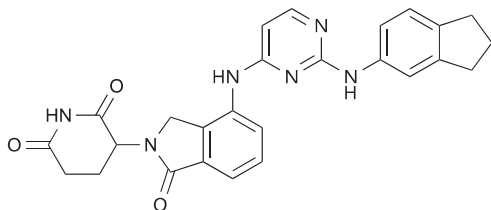
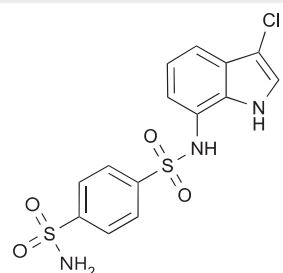
**Table 2. Chemical diversity and neosubstrate selectivity among known molecular glue degrader classes**

Compound	Chemical structure	Reported degradation substrates	References
<b>Molecular glues for CRL4<sup>CRBN</sup></b>			
Thalidomide		<u>IKZF1</u> , <u>IKZF3</u> , <u>ZNF692</u> , <u>ZNF276</u> , <u>SALL4</u> , <u>RNF166</u> , <u>ZBTB16</u> , FAM83F, p63	(Asatsuma-Okumura et al., 2019; Donovan et al., 2018; Sievers et al., 2018; Yamanaka et al., 2020)
Lenalidomide		<u>IKZF1</u> , <u>IKZF3</u> , <u>ZFP91</u> , <u>ZFP692</u> , <u>ZNF276</u> , <u>ZNF653</u> , <u>ZNF827</u> , <u>SALL4</u> , <u>RNF166</u> , <u>WIZ1</u> , CK1 $\alpha$ , FAM83F, RAB28	(Donovan et al., 2018; Krönke et al., 2014, 2015; Lu et al., 2014; Sievers et al., 2018; Yu et al., 2019)
Pomalidomide		<u>IKZF1</u> , <u>IKZF3</u> , <u>ZFP91</u> , <u>ZFP692</u> , <u>ZNF276</u> , <u>ZNF653</u> , <u>ZNF827</u> , <u>SALL4</u> , <u>RNF166</u> , <u>GZF1</u> , <u>ZBTB39</u> , <u>ZNF98</u> , <u>WIZ1</u> , <u>ZBTB16</u> , FAM83F, RAB28, DTWD1	(An et al., 2017; Donovan et al., 2018; Krönke et al., 2014; Lu et al., 2014; Matyskiela et al., 2020b; Sievers et al., 2018; Yu et al., 2019)
Avadomide (CC-122)		<u>IKZF1</u> , <u>IKZF3</u> , <u>ZFP91</u>	(Hagner et al., 2015)
5-hydroxythalidomide		<u>SALL4</u> , <u>ZBTB16</u>	(Furihata et al., 2020; Yamanaka et al., 2020)
FPFT-2216		<u>IKZF1</u> , CK1 $\alpha$	(Gemetchu et al., 2018)
Iberdomide (CC-220)		<u>IKZF1</u> , <u>IKZF3</u> , <u>ZFP91</u> , <u>ZNF98</u>	(Donovan et al., 2018; Matyskiela et al., 2018)
CC-647		<u>ZBTB16</u>	(Matyskiela et al., 2020b)

(Continued on next page)

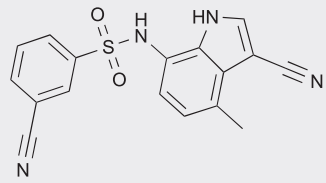
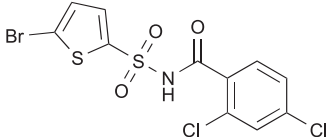
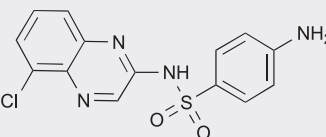
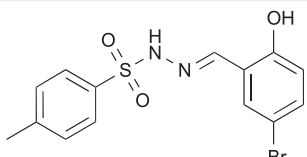
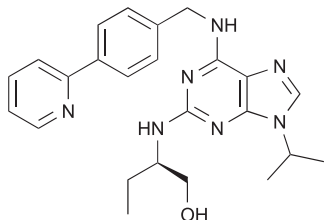
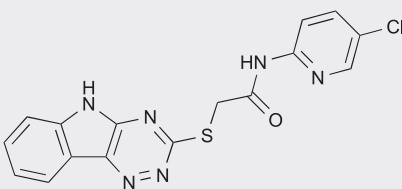
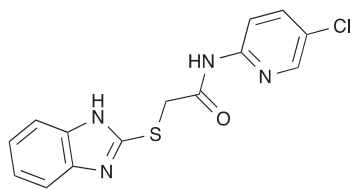
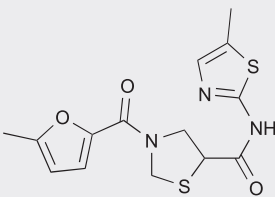


Table 2. Continued

Compound	Chemical structure	Reported degradation substrates	References
CC-3060		ZBTB16, IKZF1, ZFP91, ZNF276	(Matyskiela et al., 2020b)
CC-92480		IKZF1, IKZF3	(Lopez-Girona et al., 2019)
CC-885		IKZF1, IKZF3, GSPT1, CK1 $\alpha$ , PLK1, HBS1L	(Li et al., 2020b; Matyskiela et al., 2016; Surka et al., 2020)
CC-90009		GSPT1	(Surka et al., 2020)
ZXH-1-161		GSPT1, GSPT2	(Powell et al., 2020)
<b>Molecular glues for CRL4<sup>DCAF15</sup></b>			
Indisulam		RBM39, RBM23	(Han et al., 2017; Ting et al., 2019; Uehara et al., 2017)

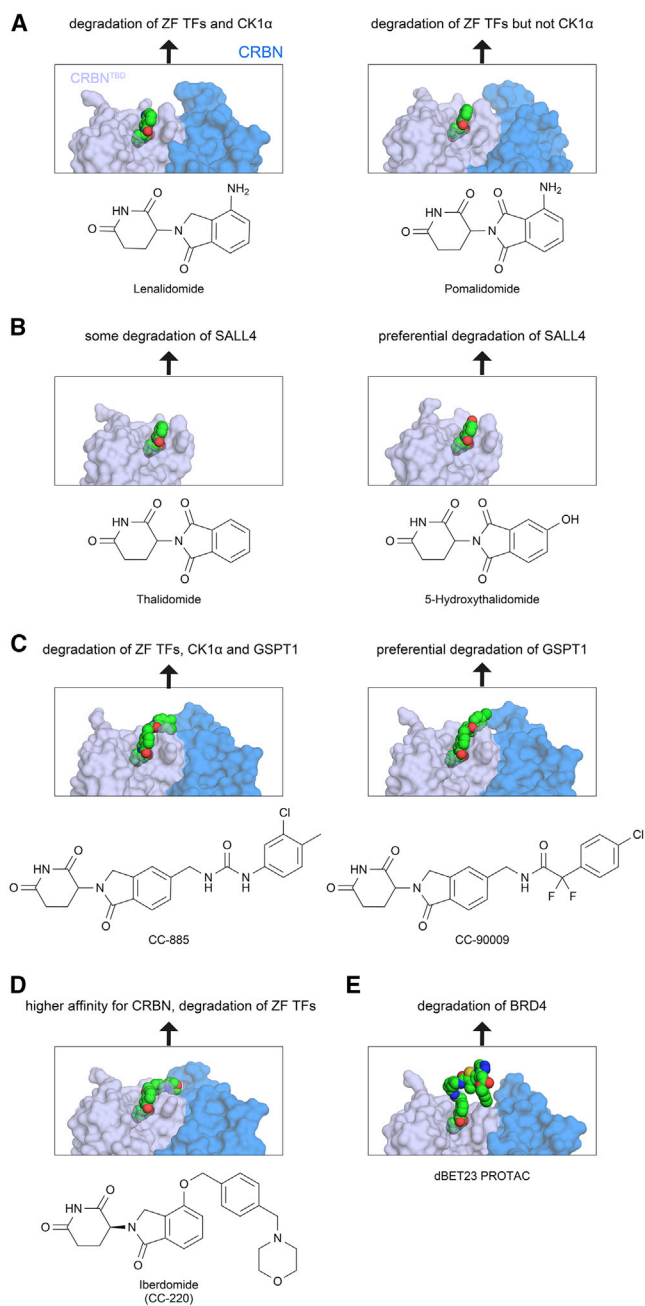
(Continued on next page)

Table 2. Continued

Compound	Chemical structure	Reported degradation substrates	References
E7820		RBM39, RBM23	(Faust et al., 2020; Ting et al., 2019; Uehara et al., 2017)
Tasisulam		RBM39, RBM23	(Han et al., 2017; Ting et al., 2019)
CQS (chloroquinoxaline sulfonamide)		RBM39, RBM23	(Han et al., 2017; Ting et al., 2019; Uehara et al., 2017)
dCeMM1		RBM39	(Mayor-Ruiz et al., 2020)
<b>Cyclin K degraders</b>			
CR8		cyclin K	(Stabicki et al., 2020a)
dCeMM2		cyclin K	(Mayor-Ruiz et al., 2020)
dCeMM3		cyclin K	(Mayor-Ruiz et al., 2020)
dCeMM4		cyclin K	(Mayor-Ruiz et al., 2020)

(Continued on next page)





**Figure 2. Reprogramming of the CRBN surface**

(A) CRBN-bound lenalidomide (PDB: 5FQD) (Petzold et al., 2016) and pomalidomide (PDB: 6UML) (Matyskiela et al., 2020a). (B) CRBN<sup>TBD</sup>-bound thalidomide (PDB: 7BQU) and 5-hydroxythalidomide (PDB: 7BQV) (Furihata et al., 2020). (C) CC-885 (PDB: 5HXB) (Matyskiela et al., 2016) and CC-90009 (PDB: 6XK9) (Surka et al., 2020) bound to CRBN. (D) CRBN-bound iberdomide (PDB: 5V3O) (Matyskiela et al., 2018). (E) CRBN-bound dBET23 PROTAC (PDB: 6BN7) (Nowak et al., 2018). CRBN is shown in a surface representation with 40% transparency. The compounds are shown as spheres (scale 1.0) and their chemical structures are given under each panel (for dBET23 the chemical structure can be found in Figure 1F). Broad neosubstrate degradation selectivity is indicated for each compound. More details can be found in Table 2. TBD, thalidomide-binding domain.

remarkable chemical diversity, more so than any aforementioned class of molecular glues (Table 2). Even more remarkable is that while most are promiscuous kinase binders, they all exclusively degrade cyclin K. In addition to the five reported degraders, a number of additional kinase inhibitors (roscovitine, DRF053, flavopiridol) have been shown to facilitate the DDB1-CDK12-cyclin K interaction *in vitro*, albeit with no robust ubiquitination and no cyclin K degradation in cells (Stabicki et al., 2020a).

Like IMiDs, cyclin K degraders exhibit an apparent affinity threshold, whereby a few-fold lower ternary complex affinity translates to no cellular degradation. We suspect this to be a common SAR principle for CRL4 ligase reprogramming and, although the molecular origins of this behavior are not clear, the small differences in ternary complex affinity *in vitro* are likely amplified *in vivo* as well as across the different stages of neosubstrate processing.

The example of CR8 also demonstrates that recruitment of a traditionally undruggable neosubstrate (cyclin K) to the ligase via a ligandable protein partner (CDK12) is a viable strategy. Cyclin K is the primary ubiquitination target, and while CDK12 could become subject to autoubiquitination upon prolonged compound exposure similar to canonical DCAFs (Fischer et al., 2011; Ito et al., 2010), data show that the observed destabilization of CDK12 is most likely due to cyclin deprivation (Lv et al., 2020). Thus, it is the overall geometry of the target-ligase complex that governs which components can be efficiently ubiquitinated and degraded. These spatial considerations are also likely at play for PROTACs, as compounds targeting the SMARCA2/4 subunits of the SWI/SNF complex have been shown to only directly degrade their target (Farnaby et al., 2019), while heterobifunctionals binding the EED component of PRC2 have been reported to additionally target other subunits for degradation (Potjeywyd et al., 2020).

As kinase inhibitors often show poor target selectivity, small-molecule-induced kinase inactivation that leverages specific protein-protein interactions offers a path toward improved drug selectivity and may facilitate the pursuit of CDK12/CDK13 and cyclin K as emerging therapeutic targets (Lui et al., 2018; Querada et al., 2019). The work on CR8 provided the structural dissection of how a kinase inhibitor scaffold acquires gain-of-function molecular glue degrader properties leading to robust kinase inactivation (Stabicki et al., 2020a), and hence provided the first example of a compound that binds the target complex and is capable of *de novo* ligase recruitment. How such diverse compounds triggered CDK12-cyclin K recruitment and in some cases cyclin K degradation remains unclear at this time. It further remains to be determined whether these compounds will be reprogrammable, redirecting DDB1 recruitment within the CDK family, the kinase superfamily, or beyond.

## BCL6 DEGRADERS

BI-3802 is a small molecule that binds the BTB domain of the oncogenic TF BCL6 and results in its proteasomal degradation (Kerres et al., 2017). A recent report revealed that the compound induces polymerization, sequestration into cellular foci, and subsequent degradation of BCL6 (Stabicki et al., 2020b). The TF was reported to be a substrate for the E3 ubiquitin ligase SIAH1, with the BI-3802-induced formation of BCL6 filaments strongly facilitating BCL6 ubiquitination by SIAH1 and its subsequent degradation. Cryo-EM studies showed that the solvent-exposed dimethylpiperidine

moiety of BI-3802 drives polymerization through direct hydrophobic interactions with distal amino acids of an adjacent BCL6 dimer. Further modeling explained the lack of polymerization with BI-3812, a closely related bona fide BCL6 inhibitor, by revealing a steric clash between the extended carboxamide group and residues of the adjacent BTB dimer. BI-3802 is thus a second example of a target-binding compound that, upon modification of its solvent-facing moiety, shows gain-of-function molecular glue activity leading to enhanced target inactivation. Surface-exposed inhibitors thus appear to be able to strongly facilitate novel protein-protein interactions. It will be important to investigate whether other BTB domains as well as other protein folds share a small-molecule-induced polymerization-dependent mechanism of degradation, and to uncover what governs these reprogramming processes.

### MOLECULAR GLUE THEMES

The discussed diverse molecular glue degrader compounds (Figure 1) leverage both protein-ligand and protein-protein interfaces to achieve tight complex formation, which is conducive to robust target ubiquitination and degradation. These molecular glue degrader drugs are small (<500 Da), with ligand/neosubstrate interfaces as compact as  $\sim 70 \text{ \AA}^2$  supplemented by more extensive protein-protein contacts (Table 1). In all these cases, the affinity between the ligase and neosubstrate increases by orders of magnitude in the presence of the small molecule. The complementary nature of these protein-protein interfaces suggests that the two proteins likely also interact in the absence of a compound. Biophysical assays measuring binding of these protein-protein pairs found no detectable affinity between CRBN and CK1a or ZFs in the absence of an IMiD (Petzold et al., 2016); low affinity ( $\sim 5 \text{ \mu M}$ ) (Du et al., 2019), or no detectable affinity (Bussiere et al., 2020; Faust et al., 2020) for DCAF15 binding to RBM39, and mid-micromolar affinity ( $\sim 50 \text{ \mu M}$ ) for DDB1 binding to CDK12-cyclin K (Słabicki et al., 2020a). Interestingly, the size of the protein-protein interface is not necessarily directly proportional to the affinity of the two proteins without the molecular glue (Table 1), as the neomorphic interface can comprise attractive as well as repulsive regions.

Molecular glue degraders can be further categorized based on their affinity to the two protein partners. IMiD-like ligase-binding glues are advantageous for targeting traditionally undruggable proteins such as ZF TFs, as a ligandable pocket on the target is not strictly required for ternary complex formation (Figure 1B). Glues that bind on the side of the target, such as CR8, offer the possibility to recruit a ligase that by itself would not be considered ligandable, like in the case of DDB1. Moreover, such target-based degraders can have dual inhibitor/degrader properties potentiating the desired pharmacological response. Glues that bind neither the ligase nor the target with high affinity, such as aryl sulfonamides, could efficiently bring together two seemingly non-ligandable partners. However, high protein-protein interface complementarity is required for stable complex formation and robust ubiquitination in this case. Finally, PROTACs tightly bind both protein partners, which gives rise to the undesirable Hook effect but also renders the compounds strongly (albeit not exclusively) dependent on compound-mediated recruitment. This makes the warheads to some extent exchangeable for binders of other targets and ligases but also implies that PROTACs require

two independently ligandable pockets. It will be important to find out whether, by structure-based design leveraging cooperative interactions within the ternary complex, such bifunctional molecules could be converted into molecular glue-type compounds that no longer independently engage either protein.

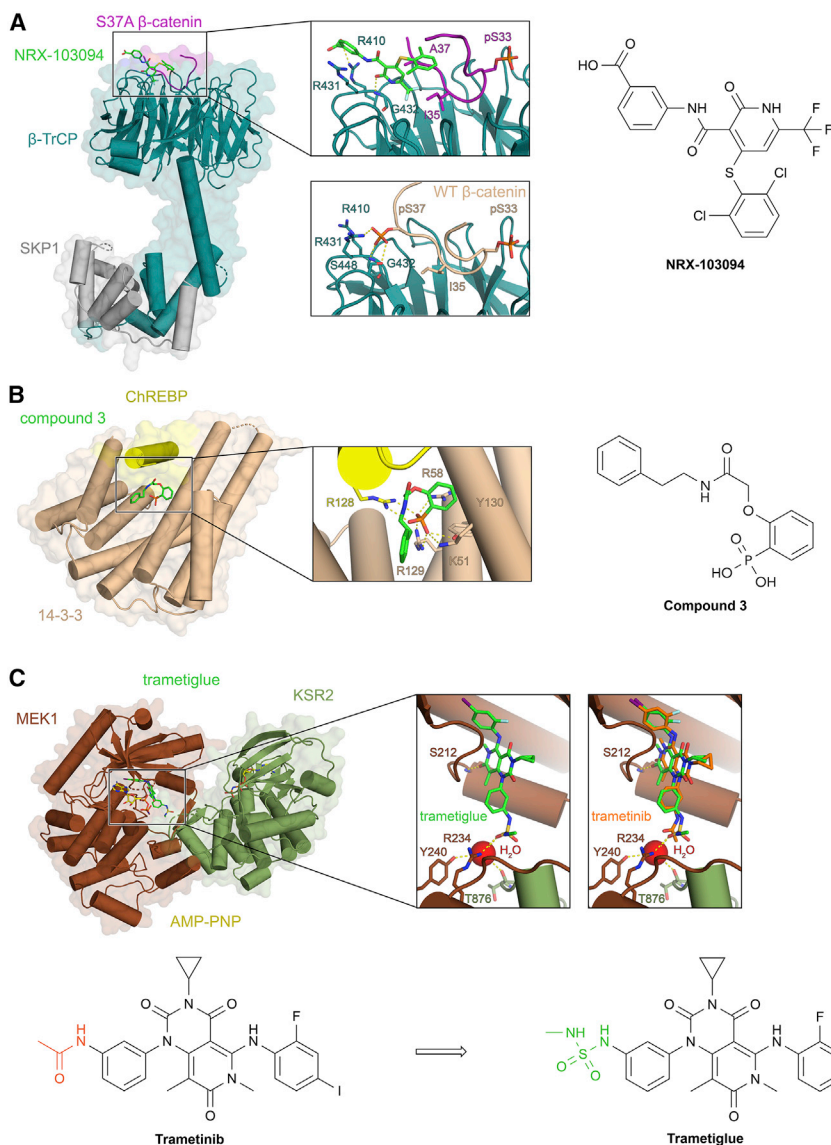
### SCREENING FOR MOLECULAR GLUE DEGRADERS

The IMiD, aryl sulfonamide, and BI-3802 molecular glues have been picked up as ligase binders and/or target degraders by observant research teams. The efforts that followed these initial discoveries focused primarily on screening derivatives of known degraders in the hope of recruiting novel neosubstrates. However, a broad search for binders of other ubiquitin ligases was also pursued in the hope of finding the “new CRBN,” driven on one hand by wanting to expand the repertoire of ligases accessible to PROTACs and on the other by the quest for a family of molecular glues as clinically successful as IMiDs. It remains to be seen whether other ligandable ligases, such as VHL or cIAP1, can recruit and ubiquitinate other proteins upon the binding of monovalent molecular glue degrader compounds.

In addition to these ligase-oriented strategies, two systematic ligase- and target-agnostic approaches for degrader discovery have recently been reported. A bioinformatic correlation between drug toxicity and ubiquitin ligase mRNA expression levels across different cancer cell lines allowed identification of the CR8-DDB1 drug-ligase pair (Słabicki et al., 2020a); and a screen looking at differential cytotoxicity of compounds in wild-type versus hyponed-dylated cellular models found one RBM39 degrader and three cyclin K degraders (Mayor-Ruiz et al., 2020). Both studies used relatively small compound libraries (4,518 preclinical and clinical compounds and  $\sim 2,000$  cytostatic and cytotoxic small molecules, respectively) and focused solely on cell survival, yielding surprisingly high hit rates. Many more compounds can likely be found by taking advantage of larger existing libraries with complex readouts beyond cell death. Once a molecular glue degrader candidate is identified, a pipeline composed of mass spectrometry-assisted identification of the degradation target(s), ligase identification using CRISPR screening technology, and biochemical reconstitutions allows the full elucidation of its mechanism of action (Mayor-Ruiz et al., 2020; Słabicki et al., 2020a, 2020b).

### GLUING WEAK INTERACTORS

If both protein-protein and protein-ligand interactions contribute to molecular glue action, molecular glue degraders may be easier to obtain by leveraging existing low-affinity protein-protein interactions. For instance, a protein epitope that has been weakened by mutations and is no longer recognized efficiently by its cognate ligase may thus be “repaired” by a molecular glue, as in the case of the mutated oncogenic TF  $\beta$ -catenin, which lost its affinity for its cognate SCF <sup>$\beta$ -TrCP</sup> E3 ligase (Simonetta et al., 2019) (Figure 3A). Through high-throughput screening followed by structure-based design, protein-protein interaction stabilizers were obtained that complemented the mutated interface by providing the binding surface between  $\beta$ -catenin and  $\beta$ -TrCP that is lost as a result of the S37A  $\beta$ -catenin mutation. The molecular glue restored



### Figure 3. Rational design of chemical inducers of proximity

(A) SKP1- $\beta$ -TrCP-NRX-103094- $\beta$ -catenin(S37A) complex (PDB: 6M91) (Simonetta et al., 2019). Bottom panel: wild-type (WT)  $\beta$ -catenin peptide bound to  $\beta$ -TrCP (PDB: 1P22) (Wu et al., 2003). (B) 14-3-3 monomer bound to compound **3** and ChREBP  $\alpha$ 2 peptide (PDB: 6YGJ) (Sijbesma et al., 2020). (C) MEK1-trametinib-KSR2 complex (PDB: 7JUR) (Khan et al., 2020). Right panel: trametinib and the parent inhibitor trametinib bound at the MEK1-KSR2 interface (PDB: 7JUR) overlaid through superposition of kinase domains (Khan et al., 2020). Important interface residues are shown as sticks. Key hydrogen bonds are indicated by dashed yellow lines.

ing a path for selectively targeting subpopulations of MEK. A number of other compounds that stabilize weak protein-protein interactions have been reported, such as CC0651 that strengthens the ubiquitin interaction with the E2 ubiquitin-conjugating enzyme CDC34 (Huang et al., 2014), and the natural product brefeldin A, which stabilizes the ARF-GDP-Sec7 complex (Mossessova et al., 2003; Renault et al., 2003). Thus, leveraging existing weak interactions, particularly in the context of ubiquitin ligases, offers an exciting route for the rational design of molecular glue degraders.

### COMPOUND-INDUCED STICKINESS

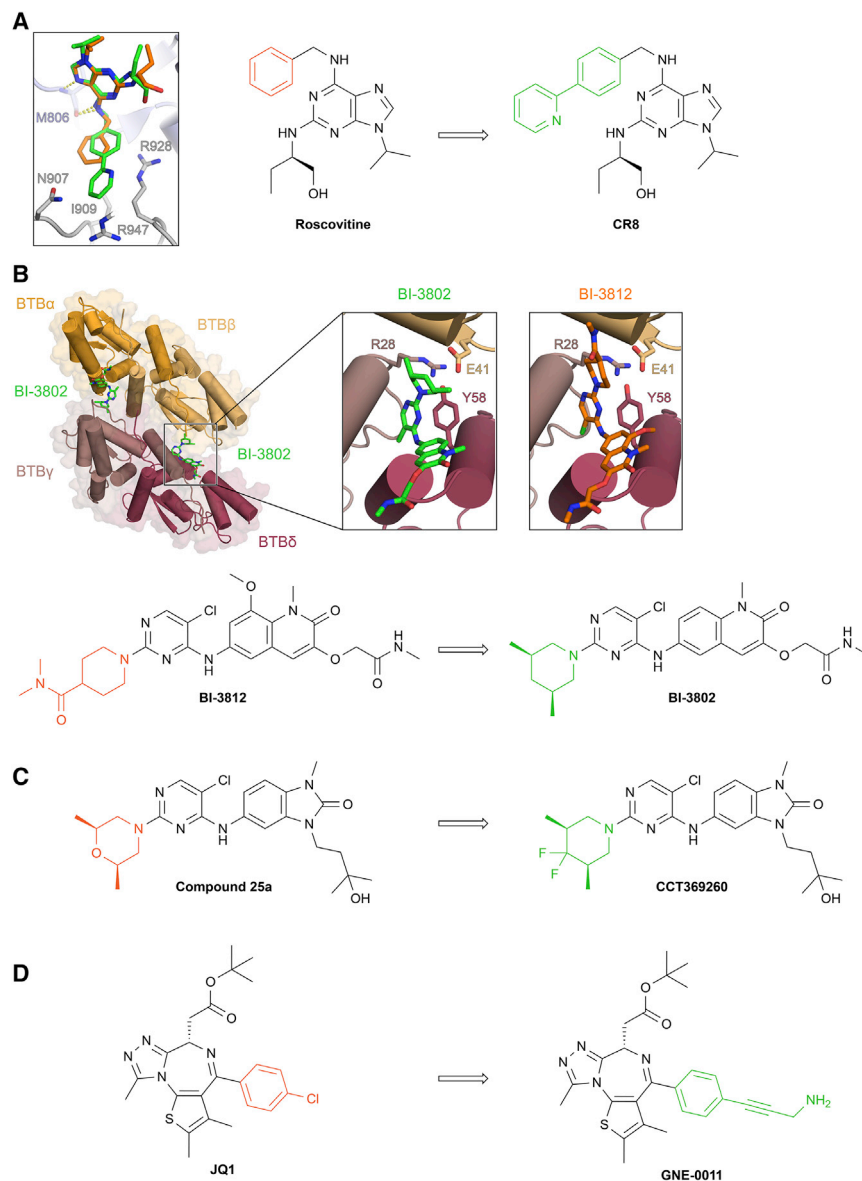
Endogenous compounds including nuclear hormones and inositol phosphates have been postulated to act as degraders (Wu et al., 2005) or inducers of proximity (Scherer et al., 2016), respectively. To our knowledge, however, no endogenous

the interaction between SCF $^{\beta$ -TrCP and mutant  $\beta$ -catenin and allowed for efficient ubiquitination *in vitro*.

Although not a molecular glue degrader, a similar strategy was used to develop a glue compound that furthers the interaction between 14-3-3 and its phosphorylation-independent interactor carbohydrate-response element-binding protein (ChREBP). A weak stabilizer derived from an *in silico* screening campaign was optimized into an effective molecular glue in a structure-guided manner (Sijbesma et al., 2020) (Figure 3B). Another example of rational glue design is that of a derivative of the kinase inhibitor trametinib, which targets the mitogen-activated protein kinase (MEK) kinase (MEK) kinase (Khan et al., 2020) (Figure 3C). The inhibitor was found to bind at the interface between the kinase domain of MEK and the kinase suppressor of Ras (KSR) subunit. Given the proximity between the two proteins, a “trametinib glue” was designed by substituting an acetamide group for a sulfonamide to enhance interfacial binding. This yielded a compound with improved cellular potency, offer-

ing a path for selectively targeting subpopulations of MEK. A number of other compounds that stabilize weak protein-protein interactions have been reported, such as CC0651 that strengthens the ubiquitin interaction with the E2 ubiquitin-conjugating enzyme CDC34 (Huang et al., 2014), and the natural product brefeldin A, which stabilizes the ARF-GDP-Sec7 complex (Mossessova et al., 2003; Renault et al., 2003). Thus, leveraging existing weak interactions, particularly in the context of ubiquitin ligases, offers an exciting route for the rational design of molecular glue degraders.

Assuming that the discussed examples (Figures 1B–1F) are drug-induced binding events and not mere variations of physiological processes, how do small molecules give rise to such neomorphic binding? A number of recent examples (Figure 4) demonstrate that the addition of solvent-facing small chemical groups to a binder or inhibitor can confer gain-of-function molecular glue degrader functionality. Evolutionary studies show that



**Figure 4. Modification of protein surface by derivatized binders drives protein-protein interactions**

(A) Overlay of CR8 (PDB: 6TD3) and roscovitine (PDB: 2A4L) in the active site of CDK12 in the DDB1-CR8-CDK12-cyclin K complex through superposition of kinase domains (Stabicki et al., 2020a).

(B) Structure of two BTB dimers brought together by BI-3802 (Stabicki et al., 2020b). Right panel: model of BI-3812 at the dimer interface. The inhibitor was docked to the crystal structure of BCL6 BTB (PDB: 5MW2) and modeled at the interface by superposition of BTB domains. The solvent-exposed moiety of BI-3812 clashes with a helix of BTBβ.

(C) Another inhibitor-degrader pair identified for BCL6 (Bellenie et al., 2020).

(D) Inhibitor and degrader of BRD4 (Blake, 2019). The mechanism of action of GNE-0011 has not been reported to date.

patches have been suggested to persist as means to favor functionally relevant multimerization (Hochberg et al., 2020). Surface-bound small molecules could thus similarly conjure a sufficiently hydrophobic hotspot to either lead to target aggregation or induce higher-order assemblies. In addition, alterations in the protein surface induced by single-nucleotide polymorphisms (SNPs) in the coding region of proteins may already intrinsically affect interactomes of the protein carrying the SNP, and molecular glue compounds could further leverage such sequence variation. Experimentally, such higher-order assemblies are most easily detectable if ubiquitin ligases are involved, giving rise to target degradation. Many more compounds could alter target interactomes more subtly, for instance through changes in post-translational modifications or localization of the target protein

proteins have an intrinsic tendency to form higher-order complexes and that random single-point mutations, with surprising ease, give rise to *de novo* interfaces that induce larger assemblies (Garcia-Seisdedos et al., 2017). This has been evident in genetic disease such as sickle-cell anemia where a surface mutation in hemoglobin (E6V; hemoglobin S) triggers fiber formation (Dykes et al., 1979). The BI-3802 and CR8 examples now argue that small molecules bound to a protein surface can give rise to similar aggregation processes, either as self-polymerization events (as in BCL6) or by strengthening an existing, albeit weak, interaction of DDB1 with CDK12-cyclin K (Figure 4). The model formulated by Levy and colleagues stipulates that proteins evolve on the cusp of engaging in macromolecular interactions (Garcia-Seisdedos et al., 2017). The very low affinities are not by themselves directly biologically actionable, but such propensities to interact are exploitable by compound-mediated modification of protein surfaces. Across evolution, hydrophobic

(Schreiber, 2019). It is likely, however, that compound-induced aggregation phenomena akin to the behavior of BI-3802 are even more common and that this behavior has thus far been largely overlooked.

We thus expect this hitherto hidden mode of action to be more common among existing small molecules than was previously thought. Hence, molecular glue-driven binding events should be taken into consideration when developing drugs, and the effect of compound binding on target stability should be routinely evaluated for candidate inhibitors. Taken one step further, it remains to be seen whether derivatization of the solvent-facing regions in small-molecular binders and inhibitors can be used to systematically induce neomorphic interactions in a forward screening approach. Should this prove to be a viable approach, it could allow for routine target-centric degrader screening, whereby libraries of compound are screened for target destabilization in a ligase-agnostic manner. This could not only pave the

way to the (semi-)rational design of molecular glue degraders for a given target, but also allow for seemingly non-ligandable ligases to be identified as viable for targeted protein degradation.

### A NEED FOR A BETTER UNDERSTANDING OF PROTEIN-PROTEIN INTERFACES

The key to rationally designing molecular glue events lies in understanding protein-protein interfaces and the ways in which these can be affected by small molecules. Our thinking of interfaces in drug discovery is largely dominated by high-affinity, high-specificity binding events. These lock-and-key interactions, as originally described by Emil Fischer (1894), are particularly relevant for small protein-ligand interfaces. But do other types of interfaces exist? For example, those that are more greasy, with lower affinity and possibly less specificity? The best indication that tight interactions in one part of a complex can prompt protein-protein interactions elsewhere comes from structures of PROTAC-induced complexes. The structure of the MZ1 ternary complex found BRD4 engaged with the VHL E3 ligase receptor through neomorphic protein-protein interactions (Gadd et al., 2017) (Figure 1E). PROTACs recruiting BRD4 to CRBN showed different BD1 interfaces engaged with CRBN depending on the linker and exit vector used, with the biochemical validation of these binding modes suggesting genuinely different interactions (Nowak et al., 2018) (Figure 1F). CRBN complexes with different neosubstrates (CK1a and GSPT1) also utilized distinct regions on the receptor surface (Matyskiela et al., 2016; Petzold et al., 2016). We interpret this as an indication that two proteins, once brought into close proximity through a high-affinity interaction elsewhere, engage in protein-protein interactions, and that these contacts have different levels of complementarity ranging from being strongly attractive to being repulsive (Donovan et al., 2020; Gadd et al., 2017; Nowak et al., 2018). These interactions are sufficiently common that they can be efficiently redirected by the compound.

A quantitative and computational framework that would allow us to dissect and understand such compound-mediated protein-protein binding events is currently missing. More examples of molecular glue events and a detailed understanding of the interfaces involved are necessary to design novel molecular glues and develop them into a powerful new class of medicines.

### SIGNIFICANCE

**With the recent advances in understanding molecular glue degraders that reprogram the ubiquitin system to achieve (neo)substrate degradation, there are now exciting possibilities at hand to systematically develop this novel class of therapeutics. Molecular degrader properties are likely already present in many existing drugs, and the solvent-exposed handle of such molecules is what determines their glue properties. Molecular glue degraders offer important lessons about compound-induced interfaces, as well as the principles of what makes two proteins bind. The theme that emerges is that compounds can bring two proteins together with surprising ease and that this property can be exploited for the development of efficient therapeutics.**

### ACKNOWLEDGMENTS

We thank Dennis Gillingham, Georg Petzold, Guillaume Diss and Colby Sandate for discussion and helpful comments on the manuscript. We are grateful to Hojong Yoon for sharing the coordinates of BI-3812 docked to BCL6. This work was supported by funding from the European Research Council under the European Union's Horizon 2020 Research and Innovation Program grant agreement no. 666068, the Swiss National Science Foundation under the Sinergia grant agreement no. 186230, and Oncosuisse (KFS-4980-02-2020). Z.K. was supported by the European Union's Horizon 2020 Research and Innovation Program under the Marie Skłodowska-Curie grant agreement no. 765445.

### DECLARATION OF INTERESTS

N.H.T. receives funding from the Novartis Research Foundation and is a member of the scientific advisory board of Monte Rosa Therapeutics.

### REFERENCES

- Ahn, J., Vu, T., Novince, Z., Guerrero-Santoro, J., Rapic-Otrin, V., and Gronenborn, A.M. (2010). HIV-1 Vpr loads uracil DNA glycosylase-2 onto DCAF1, a substrate recognition subunit of a cullin 4A-RING E3 ubiquitin ligase for proteasome-dependent degradation. *J. Biol. Chem.* *285*, 37333–37341.
- An, J., Ponthier, C.M., Sack, R., Seebacher, J., Stadler, M.B., Donovan, K.A., and Fischer, E.S. (2017). pSILAC mass spectrometry reveals ZFP91 as IMiD-dependent substrate of the CRL4 CRBN ubiquitin ligase. *Nat. Commun.* *8*, 15398. <https://doi.org/10.1038/ncomms15398>.
- Angers, S., Li, T., Yi, X., MacCoss, M.J., Moon, R.T., and Zheng, N. (2006). Molecular architecture and assembly of the DDB1-CUL4A ubiquitin ligase machinery. *Nature* *443*, 590–593.
- Asatsuma-Okumura, T., Ando, H., De Simone, M., Yamamoto, J., Sato, T., Shimizu, N., Asakawa, K., Yamaguchi, Y., Ito, T., Guerrini, L., et al. (2019). p63 is a cereblon substrate involved in thalidomide teratogenicity. *Nat. Chem. Biol.* *15*, 1077–1084.
- Baek, K., and Schulman, B.A. (2020). Molecular glue concept solidifies. *Nat. Chem. Biol.* *16*, 2–3.
- Bellenie, B.R., Cheung, K.M.J., Varela, A., Pierrat, O.A., Collie, G.W., Box, G.M., Bright, M.D., Gowan, S., Hayes, A., Rodrigues, M.J., et al. (2020). Achieving in vivo target depletion through the discovery and optimization of benzimidazolone BCL6 degraders. *J. Med. Chem.* *63*, 4047–4068.
- Bjorklund, C.C., Kang, J., Amatangelo, M., Polonskaia, A., Katz, M., Chiu, H., Couto, S., Wang, M., Ren, Y., Ortiz, M., et al. (2020). Iberdomide (CC-220) is a potent cereblon E3 ligase modulator with antitumor and immunostimulatory activities in lenalidomide- and pomalidomide-resistant multiple myeloma cells with dysregulated CRBN. *Leukemia* *34*, 1197–1201.
- Blake, R.A. (2019). Abstract 4452: GNE-0011, a novel monovalent BRD4 degrader. *Cancer Res.* *49*, <https://doi.org/10.1158/1538-7445.AM2019-4452>.
- Bondeson, D.P., Mares, A., Smith, I.E.D., Ko, E., Campos, S., Miah, A.H., Mulholland, K.E., Routly, N., Buckley, D.L., Gustafson, J.L., et al. (2015). Catalytic in vivo protein knockdown by small-molecule PROTACs. *Nat. Chem. Biol.* *11*, 611–617.
- Burslem, G.M., and Crews, C.M. (2020). Proteolysis-targeting chimeras as therapeutics and tools for biological discovery. *Cell* *181*, 102–114.
- Bussiere, D.E., Xie, L., Srinivas, H., Shu, W., Burke, A., Be, C., Zhao, J., Godbole, A., King, D., Karki, R.G., et al. (2020). Structural basis of indisulam-mediated RBM39 recruitment to DCAF15 E3 ligase complex. *Nat. Chem. Biol.* *16*, 15–23.
- Cavadini, S., Fischer, E.S., Bunker, R.D., Potenza, A., Lingaraju, G.M., Goldie, K.N., Mohamed, W.I., Faty, M., Petzold, G., Beckwith, R.E.J., et al. (2016). Cullin-RING ubiquitin E3 ligase regulation by the COP9 signalosome. *Nature* *531*, 598–603.
- Chamberlain, P. (2018). Linkers for protein degradation. *Nat. Chem. Biol.* *14*, 639–640.
- Chamberlain, P.P., Lopez-Girona, A., Miller, K., Carmel, G., Pagarigan, B., Chie-Leon, B., Rychak, E., Corral, L.G., Ren, Y.J., Wang, M., et al. (2014).



- Structure of the human Cereblon-DDB1-lenalidomide complex reveals basis for responsiveness to thalidomide analogs. *Nat. Struct. Mol. Biol.* **21**, 803–809.
- Cong, Q., Anishchenko, I., Ovchinnikov, S., and Baker, D. (2019). Protein interaction networks revealed by proteome coevolution. *Science* **365**, 185–189.
- Donovan, K.A., An, J., Nowak, R.P., Yuan, J.C., Fink, E.C., Berry, B.C., Ebert, B.L., and Fischer, E.S. (2018). Thalidomide promotes degradation of SALL4, a transcription factor implicated in Duane radial ray syndrome. *eLife* **7**, e38430.
- Donovan, K.A., Ferguson, F.M., Bushman, J.W., Eleuteri, N.A., Bhunia, D., Ryu, S., Tan, L., Shi, K., Yue, H., Liu, X., et al. (2020). Mapping the degradable kinome provides a resource for expedited degrader development. *Cell* **183**, 1714–1731.e10.
- Du, X., Li, Y., Xia, Y.L., Ai, S.M., Liang, J., Sang, P., Ji, X.L., and Liu, S.Q. (2016). Insights into protein–ligand interactions: mechanisms, models, and methods. *Int. J. Mol. Sci.* **17**, 144.
- Du, X., Volkov, O.A., Czerwinski, R.M., Tan, H.L., Huerta, C., Morton, E.R., Rizzi, J.P., Wehn, P.M., Xu, R., Nijhawan, D., et al. (2019). Structural basis and kinetic pathway of RBM39 recruitment to DCAF15 by a sulfonamide molecular glue E7820. *Structure* **27**, 1625–1633.e3.
- Duda, D.M., Scott, D.C., Calabrese, M.F., Zimmerman, E.S., Zheng, N., and Schulman, B.A. (2011). Structural regulation of cullin-RING ubiquitin ligase complexes. *Curr. Opin. Struct. Biol.* **21**, 257–264.
- Dykes, G.W., Crepeau, R.H., and Edelstein, S.J. (1979). Three-dimensional reconstruction of the 14-filament fibers of hemoglobin S. *J. Mol. Biol.* **130**, 451–472.
- Enchev, R.I., Scott, D.C., da Fonseca, P.C.A., Schreiber, A., Monda, J.K., Schulman, B.A., Peter, M., and Morris, E.P. (2012). Structural basis for a reciprocal regulation between SCF and CSN. *Cell Rep.* **2**, 616–627.
- Farnaby, W., Koegl, M., Roy, M.J., Whitworth, C., Diers, E., Trainor, N., Zollman, D., Steurer, S., Karolyi-Oezguer, J., Riedmueller, C., et al. (2019). BAF complex vulnerabilities in cancer demonstrated via structure-based PROTAC design. *Nat. Chem. Biol.* **15**, 672–680.
- Faust, T.B., Yoon, H., Nowak, R.P., Donovan, K.A., Li, Z., Cai, Q., Eleuteri, N.A., Zhang, T., Gray, N.S., and Fischer, E.S. (2020). Structural complementarity facilitates E7820-mediated degradation of RBM39 by DCAF15. *Nat. Chem. Biol.* **16**, 7–14.
- Fischer, E. (1894). Einfluss der Configuration auf die Wirkung der Enzyme. *Berichte Der Dtsch. Chem. Gesellschaft* **27**, 2985–2993.
- Fischer, E.S., Scrima, A., Böhm, K., Matsumoto, S., Lingaraju, G.M., Faty, M., Yasuda, T., Cavadini, S., Wakasugi, M., Hanaoka, F., et al. (2011). The molecular basis of CRL4 DDB2/CSA ubiquitin ligase architecture, targeting, and activation. *Cell* **147**, 1024–1039.
- Fischer, E.S., Böhm, K., Lydeard, J.R., Yang, H., Stadler, M.B., Cavadini, S., Nagel, J., Serluca, F., Acker, V., Lingaraju, G.M., et al. (2014). Structure of the DDB1-CRBN E3 ubiquitin ligase in complex with thalidomide. *Nature* **512**, 49–53.
- Furihata, H., Yamanaka, S., Honda, T., Miyauchi, Y., Asano, A., Shibata, N., Tanokura, M., Sawasaki, T., and Miyakawa, T. (2020). Structural bases of IMiD selectivity that emerges by 5-hydroxythalidomide. *Nat. Commun.* **11**, 4578. <https://doi.org/10.1038/s41467-020-18488-4>.
- Gadd, M.S., Testa, A., Lucas, X., Chan, K.H., Chen, W., Lamont, D.J., Zengerle, M., and Ciulli, A. (2017). Structural basis of PROTAC cooperative recognition for selective protein degradation. *Nat. Chem. Biol.* **13**, 514–521.
- Gandhi, A.K., Kang, J., Havens, C.G., Conklin, T., Ning, Y., Wu, L., Ito, T., Ando, H., Waldman, M.F., Thakurta, A., et al. (2014). Immunomodulatory agents lenalidomide and pomalidomide co-stimulate T cells by inducing degradation of T cell repressors Ikaros and Aiolos via modulation of the E3 ubiquitin ligase complex CRL4 CRBN. *Br. J. Haematol.* **164**, 811–821.
- Garcia-Seisdedos, H., Empereur-Mot, C., Elad, N., and Levy, E.D. (2017). Proteins evolve on the edge of supramolecular self-assembly. *Nature* **548**, 244–247.
- Gemechu, Y., Millrine, D., Hashimoto, S., Prakash, J., Sanchenkova, K., Metwally, H., Gyanu, P., Kang, S., and Kishimoto, T. (2018). Humanized cereblon mice revealed two distinct therapeutic pathways of immunomodulatory drugs. *Proc. Natl. Acad. Sci. U S A* **115**, 11802–11807.
- Gerry, C.J., and Schreiber, S.L. (2020). Unifying principles of bifunctional, proximity-inducing small molecules. *Nat. Chem. Biol.* **16**, 369–378.
- Giordanetto, F., Schäfer, A., and Ottmann, C. (2014). Stabilization of protein-protein interactions by small molecules. *Drug Discov. Today* **19**, 1812–1821.
- Glotzer, M., Murray, A.W., and Kirschner, M.W. (1991). Cyclin is degraded by the ubiquitin pathway. *Nature* **349**, 132–138.
- Hagner, P.R., Man, H.W., Fontanillo, C., Wang, M., Couto, S., Breider, M., Bjorklund, C., Havens, C.G., Lu, G., Rychak, E., et al. (2015). CC-122, a pleiotropic pathway modifier, mimics an interferon response and has antitumor activity in DLBCL. *Blood* **126**, 779–789.
- Han, T., Goralski, M., Gaskill, N., Capota, E., Kim, J., Ting, T.C., Xie, Y., Williams, N.S., and Nijhawan, D. (2017). Anticancer sulfonamides target splicing by inducing RBM39 degradation via recruitment to DCAF15. *Science* **356**, 3755.
- Hanan, E.J., Liang, J., Wang, X., Blake, R.A., Blaquiére, N., and Staben, S.T. (2020). Monomeric targeted protein degraders. *J. Med. Chem.* **63**, 11330–11361.
- Hochberg, G.K.A., Liu, Y., Marklund, E.G., Metzger, B.P.H., Laganowsky, A., and Thornton, J.W. (2020). A hydrophobic ratchet entrenches molecular complexes. *Nature* **588**, 503–508.
- Huang, H., Ceccarelli, D.F., Orlicky, S., St-Cyr, D.J., Ziemba, A., Garg, P., Plamondon, S., Auer, M., Sidhu, S., Marinier, A., et al. (2014). E2 enzyme inhibition by stabilization of a low-affinity interface with ubiquitin. *Nat. Chem. Biol.* **10**, 156–163.
- Ito, T., Ando, H., Suzuki, T., Ogura, T., Hotta, K., Imamura, Y., Yamaguchi, Y., and Handa, H. (2010). Identification of a primary target of thalidomide teratogenicity. *Science* **327**, 1345–1350.
- Jones, L.H. (2018). Small-molecule kinase downregulators. *Cell Chem. Biol.* **25**, 30–35.
- Kerres, N., Steurer, S., Schlager, S., Bader, G., Berger, H., Caligiuri, M., Dank, C., Engen, J.R., Ettmayer, P., Fischerauer, B., et al. (2017). Chemically induced degradation of the oncogenic transcription factor BCL6. *Cell Rep.* **20**, 2860–2875.
- Khan, Z.M., Real, A.M., Marsiglia, W.M., Chow, A., Duffy, M.E., Yerabolu, J.R., Scopton, A.P., and Dar, A.C. (2020). Structural basis for the action of the drug trametinib at KSR-bound MEK. *Nature* **588**, 509–514.
- Koukos, P.I., and Bonvin, A.M.J.J. (2020). Integrative modelling of biomolecular complexes. *J. Mol. Biol.* **432**, 2861–2881.
- Krissinel, E., and Henrick, K. (2007). Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.* **372**, 774–797.
- Krönke, J., Udeshi, N.D., Narla, A., Grauman, P., Hurst, S.N., McConkey, M., Svinikina, T., Heckl, D., Comer, E., Li, X., et al. (2014). Lenalidomide causes selective degradation of IKZF1 and IKZF3 in multiple myeloma cells. *Science* **343**, 301–305.
- Krönke, J., Fink, E.C., Hollenbach, P.W., MacBeth, K.J., Hurst, S.N., Udeshi, N.D., Chamberlain, P.P., Mani, D.R., Man, H.W., Gandhi, A.K., et al. (2015). Lenalidomide induces ubiquitination and degradation of CK1 $\alpha$  in del(5q) MDS. *Nature* **523**, 183–188.
- Lui, G.Y.L., Grandori, C., and Kemp, C.J. (2018). CDK12: an emerging therapeutic target for cancer. *J. Clin. Pathol.* **71**, 957–962.
- Li, L., Mi, D., Pei, H., Duan, Q., Wang, X., Zhou, W., Jin, J., Li, D., Liu, M., and Chen, Y. (2020a). In vivo target protein degradation induced by PROTACs based on E3 ligase DCAF15. *Signal Transduct. Target. Ther.* **5**, 4–6.
- Li, L., Xue, W., Shen, Z., Liu, J., Hu, M., Cheng, Z., Wang, Y., Chen, Y., Chang, H., Liu, Y., et al. (2020b). A cereblon modulator CC-885 induces CRBN- and p97-dependent PLK1 degradation and synergizes with volasertib to suppress lung cancer. *Mol. Ther. Oncol.* **18**, 215–225.
- Li, T., Robert, E.I., van Breugel, P.C., Strubin, M., and Zheng, N. (2010). A promiscuous  $\alpha$ -helical motif anchors viral hijackers and substrate receptors to the CUL4–DDB1 ubiquitin ligase machinery. *Nat. Struct. Mol. Biol.* **17**, 105–111.
- Lin, X., Li, X., and Lin, X. (2020). A review on applications of computational methods in drug screening and design. *Molecules* **25**, 1375. <https://doi.org/10.3390/molecules25061375>.

- Liu, J., Farmer, J.D., Lane, W.S., Friedman, J., Weissman, I., and Schreiber, S.L. (1991). Calcineurin is a common target of cyclophilin-cyclosporin A and FKBP-FK506 complexes. *Cell* 66, 807–815.
- Lopez-Girona, A., Havens, C.G., Lu, G., Rychak, E., Mendy, D., Gaffney, B., Surka, C., Lu, C.-C., Matyskiela, M., Khambatta, G., et al. (2019). CC-92480 is a novel cereblon E3 ligase modulator with enhanced tumoricidal and immunomodulatory activity against sensitive and resistant multiple myeloma cells. *Blood* 134, 1812.
- Lu, G., Middleton, R.E., Sun, H., Naniong, M.V., Ott, C.J., Mitsiades, C.S., Wong, K.K., Bradner, J.E., and Kaelin, W.G. (2014). The myeloma drug lenalidomide promotes the cereblon-dependent destruction of ikaros proteins. *Science* 343, 305–309.
- Lv, L., Chen, P., Cao, L., Li, Y., Zeng, Z., Cui, Y., Wu, Q., Li, J., Wang, J.-H., Dong, M.-Q., et al. (2020). Discovery of a molecular glue promoting CDK12-DDB1 interaction to trigger Cyclin K degradation. *eLife* 9, e59994.
- Mahon, C., Krogan, N., Craik, C., and Pick, E. (2014). Cullin E3 ligases and their rewiring by viral factors. *Biomolecules* 4, 897–930.
- Matyskiela, M.E., Lu, G., Ito, T., Pagarigan, B., Lu, C.C., Miller, K., Fang, W., Wang, N.Y., Nguyen, D., Houston, J., et al. (2016). A novel cereblon modulator recruits GSPT1 to the CRL4 CRBN ubiquitin ligase. *Nature* 535, 252–257.
- Matyskiela, M.E., Zhang, W., Man, H.W., Muller, G., Khambatta, G., Baculi, F., Hickman, M., Lebrun, L., Pagarigan, B., Carmel, G., et al. (2018). A cereblon modulator (CC-220) with improved degradation of Ikaros and Aiolos. *J. Med. Chem.* 61, 535–542.
- Matyskiela, M.E., Clayton, T., Zheng, X., Mayne, C., Tran, E., Carpenter, A., Pagarigan, B., McDonald, J., Rolfe, M., Hamann, L.G., et al. (2020a). Crystal structure of the SALL4-pomalidomide-cereblon-DDB1 complex. *Nat. Struct. Mol. Biol.* 27, 319–322.
- Matyskiela, M.E., Zhu, J., Baughman, J.M., Clayton, T., Slade, M., Wong, H.K., Danga, K., Zheng, X., Labow, M., LeBrun, L., et al. (2020b). Cereblon modulators target ZBTB16 and its oncogenic fusion partners for degradation via distinct structural degrons. *ACS Chem. Biol.* 15, 3149–3158.
- Mayor-Ruiz, C., Bauer, S., Brand, M., Kozicka, Z., Siklos, M., Imrichova, H., Kaltheuner, I.H., Hahn, E., Seiler, K., Koren, A., et al. (2020). Rational discovery of molecular glue degraders via scalable chemical profiling. *Nat. Chem. Biol.* 16, 1199–1207.
- Mossessova, E., Corpina, R.A., and Goldberg, J. (2003). Crystal structure of ARF1<sup>Sec7</sup> complexed with Brefeldin A and its implications for the guanine nucleotide exchange mechanism. *Mol. Cell* 12, 1403–1411.
- Nishimura, K., Fukagawa, T., Takisawa, H., Kakimoto, T., and Kanemaki, M. (2009). An auxin-based degron system for the rapid depletion of proteins in nonplant cells. *Nat. Methods* 6, 917–922.
- Nowak, R.P., Deangelo, S.L., Buckley, D., He, Z., Donovan, K.A., An, J., Saface, N., Jedrychowski, M.P., Ponthier, C.M., Ishoey, M., et al. (2018). Plasticity in binding confers selectivity in ligand-induced protein degradation article. *Nat. Chem. Biol.* 14, 706–714.
- Owa, T., Yoshino, H., Okauchi, T., Yoshimatsu, K., Ozawa, Y., Sugi, N.H., Nagasu, T., Koyanagi, N., and Kitoh, K. (1999). Discovery of novel antitumor sulfonamides targeting G1 phase of the cell cycle. *J. Med. Chem.* 42, 3789–3799.
- Petzold, G., Fischer, E.S., and Thomä, N.H. (2016). Structural basis of lenalidomide-induced CK1 $\alpha$  degradation by the CRL4 CRBN ubiquitin ligase. *Nature* 532, 127–130.
- Potjewyd, F., Turner, A.M.W., Beri, J., Rectenwald, J.M., Norris-Drouin, J.L., Cholensky, S.H., Margolis, D.M., Pearce, K.H., Herring, L.E., and James, L.I. (2020). Degradation of polycomb repressive complex 2 with an EED-targeted bivalent chemical degrader. *Cell Chem. Biol.* 27, 47–56.e15.
- Powell, C.E., Du, G., Che, J., He, Z., Donovan, K.A., Yue, H., Wang, E.S., Nowak, R.P., Zhang, T., Fischer, E.S., et al. (2020). Selective degradation of GSPT1 by cereblon modulators identified via a focused combinatorial library. *ACS Chem. Biol.* 15, 2722–2730.
- Quereda, V., Bayle, S., Vena, F., Frydman, S.M., Monastyrskiy, A., Roush, W.R., and Duckett, D.R. (2019). Therapeutic targeting of CDK12/CDK13 in triple-negative breast cancer. *Cancer Cell* 36, 545–558.e7.
- Renault, L., Guibert, B., and Cherfils, J. (2003). Structural snapshots of the mechanism and inhibition of a guanine nucleotide exchange factor. *Nature* 426, 525–530.
- Sabatini, D.M., Erdjument-Bromage, H., Lui, M., Tempst, P., and Snyder, S.H. (1994). RAFT1: a mammalian protein that binds to FKBP12 in a rapamycin-dependent fashion and is homologous to yeast TORs. *Cell* 78, 35–43.
- Sabers, C.J., Martin, M.M., Brunn, G.J., Williams, J.M., Dumont, F.J., Wiederrecht, G., and Abraham, R.T. (1995). Isolation of a protein target of the FKBP12-rapamycin complex in mammalian cells. *J. Biol. Chem.* 270, 815–822.
- Sakamoto, K.M., Kim, K.B., Kumagai, A., Mercurio, F., Crews, C.M., and DeShaies, R.J. (2001). PROTACs: chimeric molecules that target proteins to the Skp1-Cullin-F box complex for ubiquitination and degradation. *Proc. Natl. Acad. Sci. U S A* 98, 8554–8559.
- Savojarjo, C., Martelli, P.L., and Casadio, R. (2020). Protein-protein interaction methods and protein phase separation. *Annu. Rev. Biomed. Data Sci.* 3, 89–112.
- Schafer, P.H., Ye, Y., Wu, L., Kosek, J., Ringheim, G., Yang, Z., Liu, L., Thomas, M., Palmisano, M., and Chopra, R. (2018). Cereblon modulator iberdomide induces degradation of the transcription factors Ikaros and Aiolos: immunomodulation in healthy volunteers and relevance to systemic lupus erythematosus. *Ann. Rheum. Dis.* 77, 1516–1523.
- Scherer, P.C., Ding, Y., Liu, Z., Xu, J., Mao, H., Barrow, J.C., Wei, N., Zheng, N., Snyder, S.H., and Rao, F. (2016). Inositol hexakisphosphate (IP6) generated by IP5K mediates cullin-COP9 signalosome interactions and CRL function. *Proc. Natl. Acad. Sci. U S A* 113, 3503–3508.
- Schreiber, S.L. (2019). A chemical biology view of bioactive small molecules and a binder-based approach to connect biology to precision medicines. *Isr. J. Chem.* 59, 52–59.
- Schreiber, S.L., and Crabtree, G.R. (1992). The mechanism of action of cyclosporin A and FK506. *Immunol. Today* 13, 136–142.
- Scott, D.E., Bayly, A.R., Abell, C., and Skidmore, J. (2016). Small molecules, big targets: drug discovery faces the protein-protein interaction challenge. *Nat. Rev. Drug Discov.* 15, 533–550.
- Scrima, A., Fischer, E.S., Lingaraju, G.M., Böhm, K., Cavadini, S., and Thomä, N.H. (2011). Detecting UV-lesions in the genome: the modular CRL4 ubiquitin ligase does it best!. *FEBS Lett.* 585, 2818–2825.
- Sheard, L.B., Tan, X., Mao, H., Withers, J., Ben-Nissan, G., Hinds, T.R., Kobayashi, Y., Hsu, F.F., Sharon, M., Browne, J., et al. (2010). Jasmonate perception by inositol-phosphate-potentiated COI1-JAZ co-receptor. *Nature* 468, 400–407.
- Sievers, Q.L., Petzold, G., Bunker, R.D., Renneville, A., Stabicki, M., Liddicoat, B.J., Abdulrahman, W., Mikkelsen, T., Ebert, B.L., and Thomä, N.H. (2018). Defining the human C2H2 zinc finger degrome targeted by thalidomide analogs through CRBN. *Science* 362, eaat0572.
- Sijbesma, E., Visser, E., Plitzko, K., Thiel, P., Milroy, L.G., Kaiser, M., Brunsfeld, L., and Ottmann, C. (2020). Structure-based evolution of a promiscuous inhibitor to a selective stabilizer of protein-protein interactions. *Nat. Commun.* 11, 3954. <https://doi.org/10.1038/s41467-020-17741-0>.
- Simonetta, K.R., Taygerly, J., Boyle, K., Basham, S.E., Padovani, C., Lou, Y., Cummins, T.J., Yung, S.L., von Soly, S.K., Kayser, F., et al. (2019). Prospective discovery of small molecule enhancers of an E3 ligase-substrate interaction. *Nat. Commun.* 10, 1402. <https://doi.org/10.1038/s41467-019-09358-9>.
- Stabicki, M., Kozicka, Z., Petzold, G., Li, Y.-D., Manojkumar, M., Bunker, R.D., Donovan, K.A., Sievers, Q.L., Koepfel, J., Suchyta, D., et al. (2020a). The CDK inhibitor CR8 acts as a molecular glue degrader that depletes cyclin K. *Nature* 585, 293–297.
- Stabicki, M., Yoon, H., Koepfel, J., Nitsch, L., Roy Burman, S.S., Di Genua, C., Donovan, K.A., Sperling, A.S., Hunkeler, M., Tsai, J.M., et al. (2020b). Small-molecule-induced polymerization triggers degradation of BCL6. *Nature* 588, 164–168.
- 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., et al. (2009). An inhibitor of NEDD8-activating enzyme as a new approach to treat cancer. *Nature* 458, 732–736.

- Sugasawa, K., Okuda, Y., Saijo, M., Nishi, R., Matsuda, N., Chu, G., Mori, T., Iwai, S., Tanaka, K., Tanaka, K., et al. (2005). UV-induced ubiquitylation of XPC protein mediated by UV-DDB-ubiquitin ligase complex. *Cell* **121**, 387–400.
- Surka, C., Jin, L., Mbong, N., Lu, C.-C., Jang, I.S., Rychak, E., Mendy, D., Clayton, T., Tindall, E.A., Hsu, C., et al. (2020). CC-90009, a novel cereblon E3 ligase modulator targets acute myeloid leukemia blasts and leukemia stem cells. *Blood* **137**, 661–677.
- Tal, L., Gil, M.X.A., Guercio, A.M., and Shabek, N. (2020). Structural aspects of plant hormone signal perception and regulation by ubiquitin ligases. *Plant Physiol.* **182**, 1537–1554.
- Tan, X., Calderon-Villalobos, L.I.A., Sharon, M., Zheng, C., Robinson, C.V., Estelle, M., and Zheng, N. (2007). Mechanism of auxin perception by the TIR1 ubiquitin ligase. *Nature* **446**, 640–645.
- Ting, T.C., Goralski, M., Klein, K., Wang, B., Kim, J., Xie, Y., and Nijhawan, D. (2019). Aryl sulfonamides degrade RBM39 and RBM23 by recruitment to CRL4-DCAF15. *Cell Rep.* **29**, 1499–1510.e6.
- Uehara, T., Minoshima, Y., Sagane, K., Sugi, N.H., Mitsuhashi, K.O., Yamamoto, N., Kamiyama, H., Takahashi, K., Kotake, Y., Uesugi, M., et al. (2017). Selective degradation of splicing factor CAPER $\alpha$  by anticancer sulfonamides. *Nat. Chem. Biol.* **13**, 675–680.
- Vargesson, N. (2015). Thalidomide-induced teratogenesis: history and mechanisms. *Birth Defects Res. C Embryo Today Rev.* **105**, 140–156.
- Winter, G.E., Buckley, D.L., Paulk, J., Roberts, J.M., Souza, A., Dhe-Paganon, S., and Bradner, J.E. (2015). Phthalimide conjugation as a strategy for in vivo target protein degradation. *Science* **348**, 1376–1381.
- Wu, G., Xu, G., Schulman, B.A., Jeffrey, P.D., Harper, J.W., and Pavletich, N.P. (2003). Structure of a  $\beta$ -TrCP1-Skp1- $\beta$ -catenin complex: destruction motif binding and lysine specificity of the SCF $\beta$ -TrCP1 ubiquitin ligase. *Mol. Cell* **11**, 1445–1456.
- Wu, Y.L., Yang, X., Ren, Z., McDonnell, D.P., Norris, J.D., Willson, T.M., and Greene, G.L. (2005). Structural basis for an unexpected mode of SERM-mediated ER antagonism. *Mol. Cell* **18**, 413–424.
- Yamanaka, S., Murai, H., Saito, D., Abe, G., Tokunaga, E., Iwasaki, T., Takahashi, H., Takeda, H., Suzuki, T., Shibata, N., et al. (2020). PLZF is a new substrate of CRBN with thalidomide and 5-hydroxythalidomide. *BioRxiv*. <https://doi.org/10.1101/2020.02.28.969071>.
- Yamano, H., Tsurumi, C., Gannon, J., and Hunt, T. (1998). The role of the destruction box and its neighbouring lysine residues in cyclin B for anaphase ubiquitin-dependent proteolysis in fission yeast: defining the D-box receptor. *EMBO J.* **17**, 5670–5678.
- Yu, H.H., Reitsma, J.M., Sweredoski, M.J., Moradian, A., Hess, S., and Deshaies, R.J. (2019). Single subunit degradation of WIZ, a lenalidomide- and pomalidomide-dependent substrate of E3 ubiquitin ligase CRL4CRBN. *BioRxiv*. <https://doi.org/10.1101/595389>.
- Zemke, N., Barshop, W., Sha, J., Hsu, E., Wohlschlegel, J., and Berk, A. (2020). Adenovirus E1A binding to DCAF10 targets degradation of AAA+ ATPases required for quaternary assembly of multiprotein machines and innate immunity. *BioRxiv*. <https://doi.org/10.1101/2020.12.16.423151>.
- Zimmerman, E.S., Schulman, B.A., and Zheng, N. (2010). Structural assembly of cullin-RING ubiquitin ligase complexes. *Curr. Opin. Struct. Biol.* **20**, 714–721.

# The CDK inhibitor CR8 acts as a molecular glue degrader that depletes cyclin K

<https://doi.org/10.1038/s41586-020-2374-x>

Received: 9 November 2019

Accepted: 29 April 2020

Published online: 3 June 2020

 Check for updates

Mikołaj Stabicki<sup>1,2,3,15</sup>, Zuzanna Kozicka<sup>4,5,15</sup>, Georg Petzold<sup>4,15</sup>, Yen-Der Li<sup>1,2,6</sup>, Manisha Manojkumar<sup>1,2,3</sup>, Richard D. Bunker<sup>4,14</sup>, Katherine A. Donovan<sup>7,8</sup>, Quinlan L. Sievers<sup>1,2</sup>, Jonas Koeppel<sup>1,2,3</sup>, Dakota Suchyta<sup>4,5</sup>, Adam S. Sperling<sup>1,2</sup>, Emma C. Fink<sup>1,2</sup>, Jessica A. Gasser<sup>1,2</sup>, Li R. Wang<sup>1</sup>, Steven M. Corsello<sup>1,2</sup>, Rob S. Sellar<sup>1,2,9</sup>, Max Jan<sup>1,2</sup>, Dennis Gillingham<sup>5</sup>, Claudia Scholl<sup>10</sup>, Stefan Fröhling<sup>3,11</sup>, Todd R. Golub<sup>1,12,13</sup>, Eric S. Fischer<sup>7,8</sup>, Nicolas H. Thomä<sup>4</sup>✉ & Benjamin L. Ebert<sup>1,2,13</sup>✉

Molecular glue compounds induce protein–protein interactions that, in the context of a ubiquitin ligase, lead to protein degradation<sup>1</sup>. Unlike traditional enzyme inhibitors, these molecular glue degraders act substoichiometrically to catalyse the rapid depletion of previously inaccessible targets<sup>2</sup>. They are clinically effective and highly sought-after, but have thus far only been discovered serendipitously. Here, through systematically mining databases for correlations between the cytotoxicity of 4,518 clinical and preclinical small molecules and the expression levels of E3 ligase components across hundreds of human cancer cell lines<sup>3–5</sup>, we identify CR8—a cyclin-dependent kinase (CDK) inhibitor<sup>6</sup>—as a compound that acts as a molecular glue degrader. The CDK-bound form of CR8 has a solvent-exposed pyridyl moiety that induces the formation of a complex between CDK12–cyclin K and the CUL4 adaptor protein DDB1, bypassing the requirement for a substrate receptor and presenting cyclin K for ubiquitination and degradation. Our studies demonstrate that chemical alteration of surface-exposed moieties can confer gain-of-function glue properties to an inhibitor, and we propose this as a broader strategy through which target-binding molecules could be converted into molecular glues.

Molecular glues are a class of small-molecule drugs that induce or stabilize interactions between proteins<sup>1</sup>. In the context of a ubiquitin ligase, drug-induced interactions can lead to protein degradation, and this is an emerging strategy for the inactivation of therapeutic targets that are intractable by conventional pharmacological means<sup>2</sup>. Known molecular glue degraders bind to the substrate receptors of E3 ubiquitin ligases and recruit target proteins for their ubiquitination and subsequent degradation by the proteasome.

Thalidomide analogues and aryl sulfonamides are two classes of drugs that act as molecular glue degraders. Widely used in the clinic, thalidomide analogues are an effective treatment for multiple myeloma, other B cell malignancies and myelodysplastic syndrome with a deletion in chromosome 5q<sup>7</sup>. Thalidomide analogues recruit zinc-finger transcription factors and other target proteins to cereblon (CRBN)<sup>8–11</sup>, the substrate receptor of the cullin-RING E3 ubiquitin ligase CUL4A/B–RBX1–DDB1–CRBN (CRL4<sup>CRBN</sup>)<sup>12</sup>. Similarly, aryl sulfonamides degrade the essential RNA-binding protein RBM39 by engaging DCAF15, the substrate receptor of the CRL4<sup>DCAF15</sup> E3 ubiquitin ligase<sup>13–15</sup>. In these

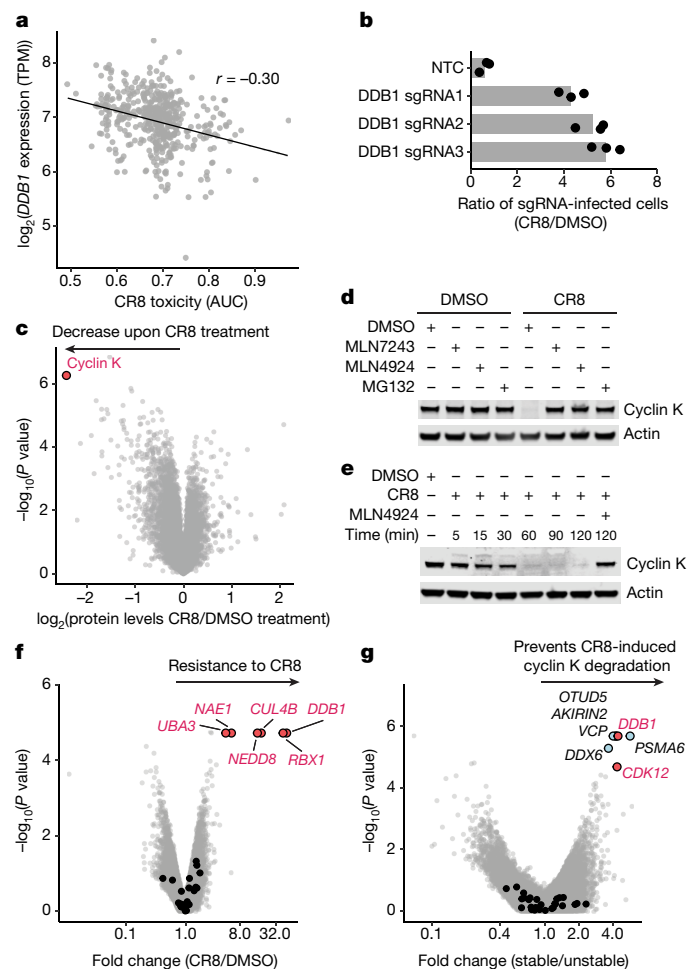
examples, the degraders are not dependent on a ligandable pocket on the target protein, but instead exploit complementary protein–protein interfaces between the receptor and the target. By reprogramming the selectivity of the ubiquitin ligase, these molecules divert the ligase to drive multiple rounds of target ubiquitination in a catalytic manner<sup>16</sup>. Such compounds can thus circumvent the limitations of classical inhibitors and expand the repertoire of ‘druggable’ proteins. Although highly desirable, molecular glue degraders have only been found serendipitously, and the strategies available for identifying or designing these compounds are limited.

## CR8 induces proteasomal cyclin K degradation

To identify small molecules that mediate protein degradation through an E3 ubiquitin ligase, we correlated drug-sensitivity data for 4,518 clinical and preclinical drugs that have been tested against 578 cancer cell lines<sup>3,4</sup> with the respective mRNA levels of 499 E3 ligase components<sup>5</sup> (Extended Data Fig. 1a). Expression of *DCAF15* correlated with

<sup>1</sup>Broad Institute of MIT and Harvard, Cambridge, MA, USA. <sup>2</sup>Department of Medical Oncology, Dana-Farber Cancer Institute, Boston, MA, USA. <sup>3</sup>Division of Translational Medical Oncology, German Cancer Research Center (DKFZ) and National Center for Tumor Diseases (NCT), Heidelberg, Germany. <sup>4</sup>Friedrich Miescher Institute for Biomedical Research, Basel, Switzerland.

<sup>5</sup>Faculty of Science, University of Basel, Basel, Switzerland. <sup>6</sup>Department of Molecular and Cellular Biology, Harvard University, Cambridge, MA, USA. <sup>7</sup>Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, MA, USA. <sup>8</sup>Department of Cancer Biology, Dana-Farber Cancer Institute, Boston, MA, USA. <sup>9</sup>Department of Haematology, UCL Cancer Institute, University College London, London, UK. <sup>10</sup>Division of Applied Functional Genomics, German Cancer Research Center (DKFZ) and National Center for Tumor Diseases (NCT), Heidelberg, Germany. <sup>11</sup>German Cancer Consortium, Heidelberg, Germany. <sup>12</sup>Department of Pediatric Oncology, Dana-Farber Cancer Institute, Boston, MA, USA. <sup>13</sup>Howard Hughes Medical Institute, Boston, MA, USA. <sup>14</sup>Present address: Monte Rosa Therapeutics, Basel, Switzerland. <sup>15</sup>These authors contributed equally: Mikołaj Stabicki, Zuzanna Kozicka, Georg Petzold. ✉e-mail: nicolas.thoma@fmi.ch; benjamin.ebert@dfci.harvard.edu



**Fig. 1 | CR8-induced degradation of cyclin K depends on DDB1 and CDK12.** **a**, Pearson correlation between CR8 toxicity and *DDB1* mRNA levels. Dots represent cancer cell lines ( $n = 471$ ). A lower value for the area under the curve (AUC; y axis) corresponds to higher drug toxicity. The Pearson correlation coefficient ( $r$ ) is shown. TPM, transcripts per million. **b**, Flow cytometry analysis of HEK293T-Cas9 cells expressing three different sgRNAs against *DDB1* and a blue fluorescent protein (BFP) marker after a three-day treatment with 1  $\mu$ M CR8 (bars represent mean,  $n = 3$ ). NTC, non-targeting control. **c**, Whole-proteome quantification of MOLT-4 cells treated with 1  $\mu$ M CR8 ( $n = 1$ ) or DMSO ( $n = 3$ ) for 5 h (two-sided moderated  $t$ -test,  $n = 3$ ). **d**, Immunoblots of cyclin K degradation in HEK293T-Cas9 cells that were pretreated with 0.5  $\mu$ M MLN7243, 1  $\mu$ M MLN4924 or 10  $\mu$ M MG132 for 4 h and then treated with 1  $\mu$ M CR8 for 2 h ( $n = 3$ ). **e**, Immunoblots of the time course of cyclin K degradation in HEK293T-Cas9 cells treated with 1  $\mu$ M CR8 ( $n = 3$ ). **f**, Genome-wide CRISPR-Cas9 resistance screen for CR8 resistance in HEK293T-Cas9 cells. **g**, Genome-wide CRISPR-Cas9 reporter screen for cyclin K-eGFP stability after treatment with 1  $\mu$ M CR8 in HEK293T-Cas9 cells. In **f**, **g**, guide counts were collapsed to gene level ( $n = 4$  guides per gene; two-sided empirical rank-sum test statistics). Black dots denote DCAF substrate receptors (**f**, **g**).

indisulam and tasisulam toxicity, consistent with the known function of these drugs as degraders of the essential protein RBM39 through the CRL4<sup>DCAF15</sup> E3 ubiquitin ligase and thus demonstrating the potential of the approach (Extended Data Fig. 1b, c). We sought to validate the correlations between ligase expression and drug toxicity that scored most highly by examining whether CRISPR-mediated inactivation of the identified E3 ligase component would rescue the respective drug-induced toxicity (Extended Data Fig. 1d). These experiments confirmed that single-guide RNAs (sgRNAs) that target *DCAF15* confer resistance to indisulam and tasisulam. In addition, we observed a correlation between the cytotoxicity of (*R*)-CR8, a CDK inhibitor<sup>6</sup>, and the

mRNA levels of the CUL4 adaptor protein DDB1 (Fig. 1a, Extended Data Fig. 1e). Consistently, sgRNAs targeting *DDB1* conferred resistance to (*R*)-CR8 (Fig. 1b).

As the DDB1-dependent cytotoxicity of (*R*)-CR8 suggested ubiquitin ligase-mediated degradation of one or more essential proteins, we performed quantitative proteome-wide mass spectrometry to evaluate protein abundance after treating cells with (*R*)-CR8. Of the quantified proteins (more than 8,000) cyclin K was the only protein that consistently showed a decrease in abundance after addition of (*R*)-CR8 (Fig. 1c, Extended Data Fig. 1f, g). As expected, (*R*)-CR8 did not alter the levels of cyclin K mRNA (Extended Data Fig. 1h) and the (*R*)-CR8-induced degradation of cyclin K could be rescued by inhibition of the E1 ubiquitin-activating enzyme (using MLN7243), inhibition of cullin neddylation (MLN4924) and inhibition of the proteasome (MG132) (Fig. 1d). Together, these results suggest that (*R*)-CR8 triggers rapid proteasomal degradation of cyclin K (Fig. 1e) through the activity of a DDB1-containing cullin-RING ubiquitin ligase.

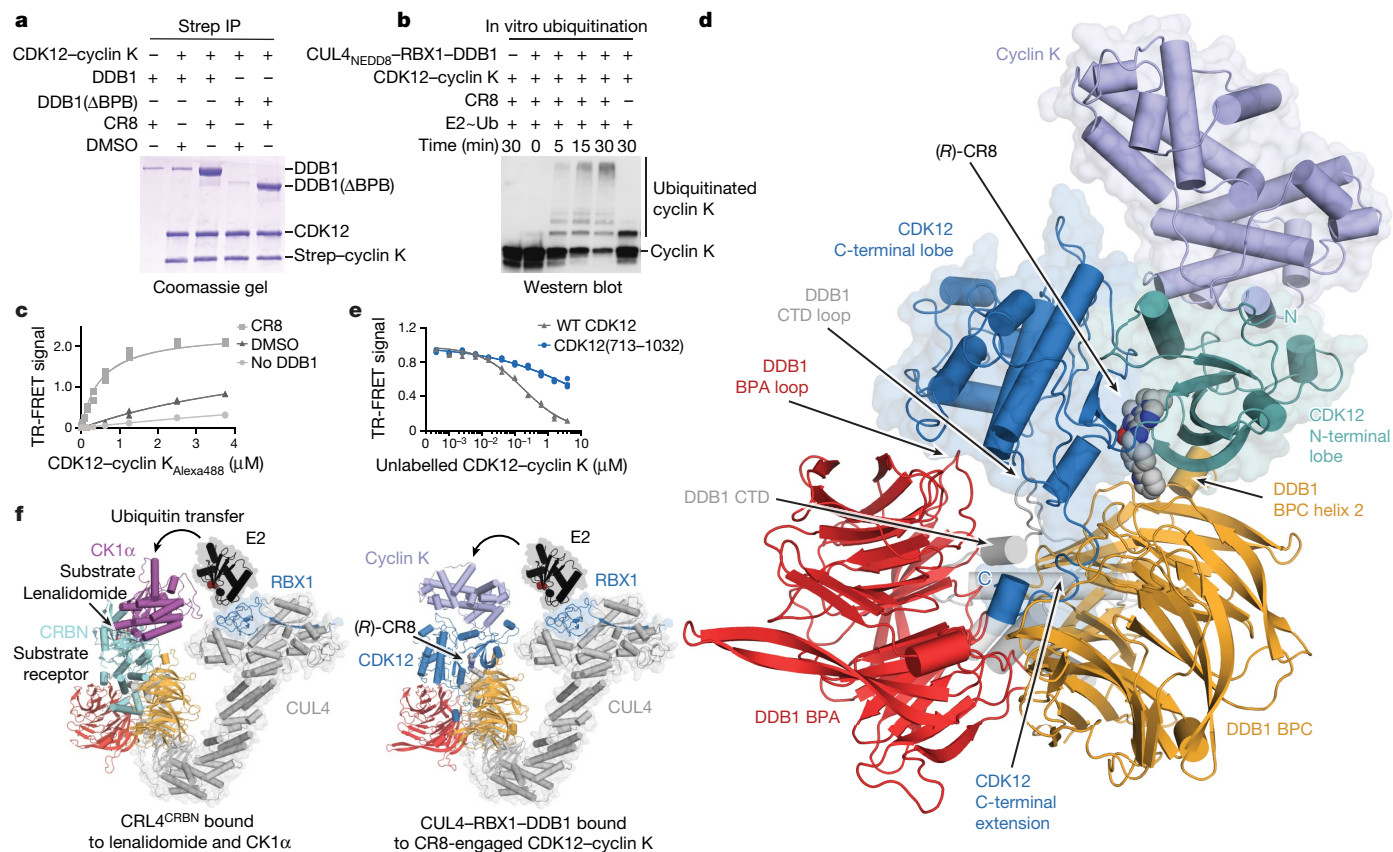
To dissect the molecular machinery that is required for (*R*)-CR8 toxicity, we performed genome-wide and E3 ubiquitin ligase-focused CRISPR-Cas9 resistance screens (Fig. 1f, Extended Data Fig. 2a, b). sgRNAs that target *DDB1*, *CUL4B*, *RBX1*, the cullin-RING activator *NEDD8* and the NEDD8-activating enzyme (*NAE1* and *UBA3*) were substantially enriched in the (*R*)-CR8-resistant cell population. As all of these proteins are required for the activity of cullin-RING ligases, our results provide genetic evidence for the involvement of a functional CUL4-RBX1-DDB1 ubiquitin ligase complex in mediating (*R*)-CR8 cytotoxicity.

Thus far, all known cullin-RING ligases engage their substrates through specific substrate receptors, and DDB1 serves as an adaptor protein that is able to bind over 20 such receptors (also known as DDB1-CUL4-associated-factors, DCAFs)<sup>17,18</sup> to recruit them to the CUL4-RBX1 ligase core. As no DCAFs were identified in our (*R*)-CR8 resistance screens, we constructed a fluorescent reporter of cyclin K stability (Extended Data Fig. 2c), in which the (*R*)-CR8-mediated degradation of endogenous cyclin K (Fig. 1d, e) could be recapitulated with cyclin K fused to enhanced green fluorescent protein (eGFP) (cyclin K<sub>eGFP</sub>) (Extended Data Fig. 2d-f). Using the stability reporter, in which the extent of degradation can be determined by measuring the levels of cyclin K<sub>eGFP</sub> normalized to mCherry expression, we found that both (*S*)-CR8 and (*R*)-CR8 facilitated the degradation of cyclin K<sub>eGFP</sub> to the same extent (Extended Data Fig. 2g; henceforth, CR8 refers to (*R*)-CR8). We then performed a genome-wide CRISPR-Cas9 screen for genes involved in cyclin K reporter stability and validated the involvement of DDB1 in CR8-mediated, but not CR8-independent, degradation of cyclin K (Fig. 1g, Extended Data Fig. 2h-j). In addition, we identified CDK12—a known target of CR8<sup>19</sup> that depends on the interaction with cyclin K for its activity<sup>20</sup>—as a crucial component for CR8-induced destabilization of cyclin K<sub>eGFP</sub> (Fig. 1g, Extended Data Fig. 2h-k).

As neither the cyclin K<sub>eGFP</sub> stability reporter screen nor the CR8 resistance screen identified a substrate receptor, we performed additional CRISPR screens targeting 29 genes that encode known DCAFs or DCAF-like candidate proteins in four different cell lines. Although sgRNAs targeting the previously identified components of the CUL4-RBX1-DDB1 complex consistently caused resistance to CR8, a DCAF substrate receptor could not be identified (Extended Data Fig. 3).

### CR8 directs CDK12 to the ligase core

As none of our genetic screens identified a DCAF that is required for cyclin K degradation, we tested whether the CR8-engaged CDK12-cyclin K complex directly binds one of the CUL4-RBX1-DDB1 ligase components in the absence of a substrate receptor. We therefore performed in vitro co-immunoprecipitation experiments using recombinantly purified proteins. The kinase domain of CDK12



**Fig. 2 | CR8-bound CDK12 binds DDB1 in a DCAF-like manner.**

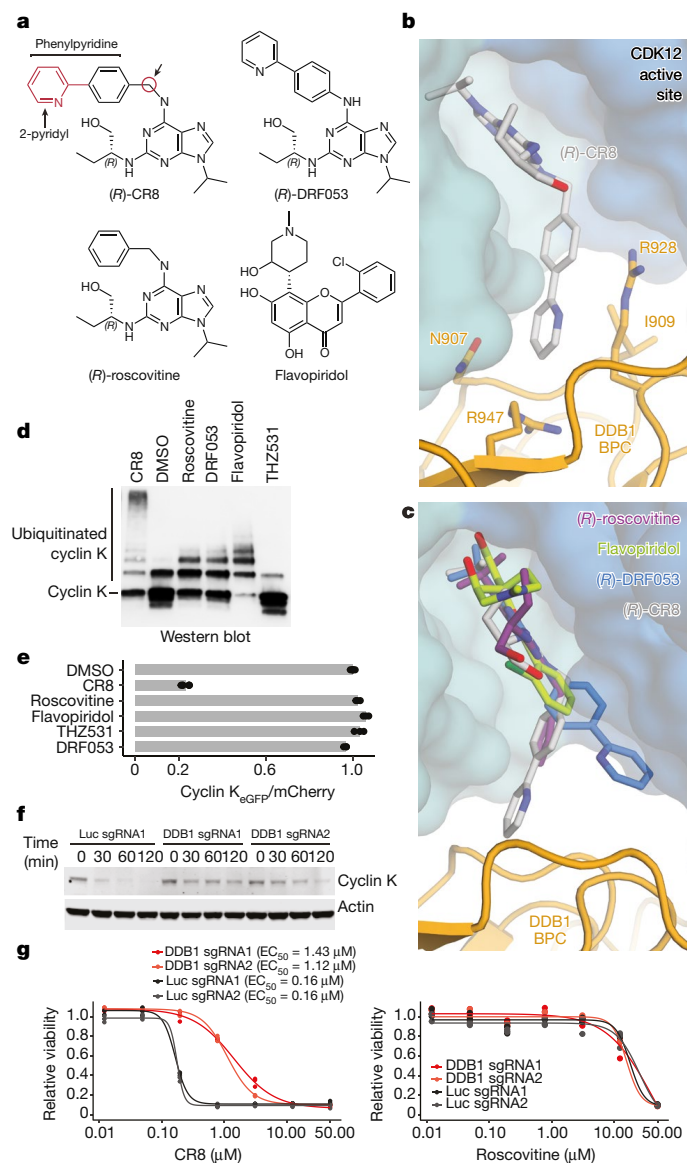
**a**, Co-immunoprecipitation (IP) experiments with recombinant proteins ( $n = 3$ ). Strep, streptavidin. **b**, In vitro ubiquitination of cyclin K by the CUL4<sup>NEDD8</sup>-RBX1-DDB1 ubiquitin ligase core ( $n = 2$ ). In E2-Ub, - represents the thioester bond between the active site of the E2 and the C terminus of ubiquitin. **c**, TR-FRET signal for CDK12-cyclin K<sub>Alexa488</sub> titrated to DDB1 tagged with terbium-coupled streptavidin (DDB1<sub>terbium</sub>) in DMSO or 10  $\mu$ M CR8 ( $n = 3$ ). ‘No DDB1’ only contains terbium-coupled streptavidin and shows concentration-dependent

fluorophore effects. **d**, Cartoon representation of the crystal structure of DDB1( $\Delta$ BPB)-(R)-CR8-CDK12-cyclin K. CTD, C-terminal domain. **e**, TR-FRET counter-titration of unlabelled wild-type (WT) or mutant CDK12-cyclin K (0–10  $\mu$ M) into the preassembled DDB1<sub>terbium</sub>-CR8-CDK12-cyclin K<sub>Alexa488</sub> complex ( $n = 3$ ). **f**, Structural models of CUL4<sup>CRBN</sup> bound to lenalidomide and CK1 $\alpha$  (left) and CUL4-RBX1-DDB1 (CRL4) bound to the (R)-CR8-CDK12-cyclin K complex (right). A cysteine residue at the active site of the E2 enzyme (red spheres) binds ubiquitin through a thioester bond.

(CDK12(713–1052)) bound to cyclin K(1–267) did not markedly enrich DDB1 over the bead-binding control in the absence of CR8, whereas equimolar amounts of CR8 led to stoichiometric complex formation (Fig. 2a). The DDB1  $\beta$ -propeller domain A (BPA) and  $\beta$ -propeller domain C (BPC)<sup>17</sup>, which are otherwise involved in DCAF binding, were sufficient for drug-induced recruitment of CDK12-cyclin K. DDB1  $\beta$ -propeller domain B (BPB), which binds CUL4 and is not involved in DCAF binding, was dispensable for the interaction (Fig. 2a). In vitro ubiquitination assays confirmed that the CUL4A-RBX1-DDB1 ligase core alone is sufficient to drive robust ubiquitination of cyclin K (Fig. 2b). Quantification of the interaction showed that CR8 stimulated binding between CDK12-cyclin K and DDB1 in the range of 100–500 nM, depending on the experimental set-up (Fig. 2c, Extended Data Fig. 4). Although a weak interaction between CDK12-cyclin K and DDB1 was still detectable in vitro in the absence of the drug, CR8 strengthened complex formation by 500- to 1,000-fold as estimated by isothermal titration calorimetry (ITC) (Extended Data Fig. 4f–k). Thus, our data indicate that CR8-engaged CDK12-cyclin K is recruited to the CUL4-RBX1-DDB1 ligase core through DDB1, and CR8 tightens the complex sufficiently to drive CR8-induced degradation of cyclin K in the absence of a canonical DCAF substrate receptor.

We then crystallized CDK12(713–1052)-cyclin K(1–267) bound to CR8 and a truncated version of DDB1 that lacks the BPB domain ( $\Delta$ BPB), and determined the structure of this complex at 3.5  $\text{\AA}$  resolution (Fig. 2d, Extended Data Table 1). In the structure, CDK12 forms extensive

protein-protein interactions (around 2,100  $\text{\AA}^2$ ) with DDB1. CR8 binds the active site of CDK12 and bridges the CDK12-DDB1 interface, whereas cyclin K binds CDK12 on the opposite side and does not contact DDB1. The N-terminal and C-terminal lobes of CDK12 are proximal to DDB1 residues located in a loop of the BPA domain (amino acids 111–114), helix 2 of the BPC domain (amino acids 986–990) and a loop in the C-terminal domain (amino acids 1078–1081), which are otherwise involved in DCAF binding (Extended Data Fig. 5). In addition, the C-terminal extension of CDK12 binds the cleft between the DDB1 domains BPA and BPC—a hallmark binding site for interactions between DDB1 and DCAFs (Extended Data Fig. 5a–d, i). The density for this region could only be tentatively assigned, probably owing to the presence of multiple conformations, but the CDK12 C-terminal tail clearly engages with DDB1 and assumes a conformation that is different from those seen in isolated CDK12-cyclin K structures<sup>19,21</sup> (Extended Data Fig. 6a, b, d). Structure-guided mutational analyses combined with time-resolved fluorescence resonance energy transfer (TR-FRET) assays were used to assess the contribution of these interactions to the CR8-dependent formation of the CDK12-DDB1 complex (Fig. 2e, Extended Data Fig. 5e). Together, our data show that CDK12 assumes the role of a glue-induced substrate receptor and places cyclin K in a position that is typically occupied by CRL4 substrates (Fig. 2f). This renders the binding of CDK12-cyclin K to DDB1 mutually exclusive with that of DCAFs and provides a structural framework that explains why a canonical substrate receptor is dispensable for cyclin K ubiquitination.



**Fig. 3 | A surface-exposed 2-pyridyl moiety of CR8 confers glue degrader activity.** **a**, Chemical structures of CDK inhibitors. Arrows indicate differences between (R)-CR8, (R)-DRF053 and (R)-roscovitine. **b**, Close-up view of the DDB1-CR8-CDK12 interface. The phenylpyridine moiety of CR8 contacts DDB1 residues. **c**, (R)-roscovitine (Protein Data Bank (PDB) entry 2A4L), (R)-DRF053 and flavopiridol (3BLR) in the active site of CDK12 in the DDB1-CR8-CDK12-cyclin K complex through superposition of kinase domains or the purine moiety (for DRF053). **d**, In vitro ubiquitination of the CDK12-cyclin K complex by CUL4<sup>NEDD8</sup>-RBX1-DDB1 in the absence (DMSO) or presence of the indicated compounds (all 2 μM) (*n* = 2). **e**, Flow cytometry analysis of the degradation of cyclin K<sub>eGFP</sub> in HEK293T-Cas9 cells treated with 1 μM of the indicated compounds for 2 h (*n* = 3). **f**, Immunoblots of cyclin K in HEK293T-Cas9 cells transfected with the indicated sgRNAs and treated with 1 μM CR8 (*n* = 2). **g**, Drug sensitivity of sgRNA-transfected HEK293T-Cas9 cells after three days of treatment with CR8 (left) or roscovitine (right) (*n* = 3). The half-maximum effective concentration (EC<sub>50</sub>) values are shown for cells treated with CR8.

### The CDK12-DDB1 interface imparts selectivity

CR8 is a pleiotropic CDK inhibitor that is reported to bind CDK1, CDK2, CDK3, CDK5, CDK7, CDK9 and CDK12<sup>6,19</sup>, yet in cells we observed selective destabilization of cyclin K in the presence of the drug. As cyclin K is reported to associate with CDK9, CDK12 and CDK13<sup>19</sup>, we tested whether the other cyclin K-dependent kinases are also recruited to DDB1. The

closely related CDK13 (90.8% sequence identity)—but not the more divergent CDK9 (45.5% sequence identity) (Extended Data Fig. 7a–c)—was recruited to DDB1 in the presence of CR8, albeit with a lower binding affinity (Extended Data Fig. 7d–f). Analogously, in vitro ubiquitination of cyclin K was less productive for CDK13 than CDK12 (Extended Data Fig. 7g). The key difference in primary sequence between CDK9 and CDK12 or CDK13 is in the C-terminal extension (Extended Data Fig. 7a, b), which in our structure nestles up against the BPA and BPC regions of DDB1 (Fig. 2d, Extended Data Fig. 5i). Mutations in, or truncation of, the C-terminal extension of CDK12 abolished basal binding between CDK12 and DDB1, whereas complex formation could still be facilitated by CR8 to a varying extent (Extended Data Fig. 7h, i). Hence, our data show that the pan-selective CDK inhibitor CR8 induces specific protein-protein interactions between CDK12 or CDK13 and DDB1 and suggest that the C-terminal extension, though contributing to binding, is not essential for drug-dependent kinase recruitment.

### CR8 phenylpyridine confers glue activity

CR8 occupies the ATP-binding pocket of CDK12 and forms discrete contacts with residues in the BPC domain of DDB1 (around 150 Å<sup>2</sup>) through its hydrophobic phenylpyridine ring system (Fig. 3a, b). Mutations of the DDB1 residues Ile909, Arg928 and Arg947 each diminished drug-induced recruitment of the kinase (Extended Data Fig. 5f), highlighting the contribution of the phenylpyridine moiety to complex formation. To evaluate the structure-activity relationship that underlies the gain-of-function activity of CR8, we probed other CDK inhibitors for their ability to drive complex formation between DDB1 and CDK12. DRF053<sup>22</sup>, an inhibitor related to CR8 that carries a differently linked phenylpyridine ring system (Fig. 3a, c), induced binding with a twofold lower affinity than CR8 (Extended Data Fig. 8a). Roscovitine<sup>23</sup>, the parent compound of CR8, which lacks the 2-pyridyl substituent but retains the phenyl ring proximal to Arg928 (Fig. 3a, c), also facilitated complex formation—albeit with an apparent binding affinity threefold lower than CR8 (Extended Data Fig. 8a). The rank order of binding affinity that we observed in our TR-FRET assay correlated with the degree of cyclin K ubiquitination in vitro; DRF053 and roscovitine showed less processive ubiquitination (Fig. 3d). As neither DRF053 nor roscovitine induced degradation of the cyclin K<sub>eGFP</sub> reporter in cells (Fig. 3e), our results show that the presence and correct orientation of the 2-pyridyl moiety on the surface of CDK12 confer the gain-of-function activity of CR8 that leads to cyclin K degradation.

To investigate whether any bound ligand could in principle drive the interaction of CDK12 with DDB1, we tested the endogenous CDK nucleotide cofactor ATP for its ability to promote complex formation. ATP neither facilitated nor abrogated the interaction over basal binding observed in the presence of dimethyl sulfoxide (DMSO) (Extended Data Fig. 6c), suggesting that although the nucleotide-bound conformation of CDK12 seems incompatible with the recruitment of DDB1 (Extended Data Fig. 6b), its C-terminal extension is free to adopt multiple conformations<sup>21</sup>. THZ531<sup>24</sup>, a bulky covalent inhibitor of CDK12 and CDK13 that is predicted to clash with DDB1 (Extended Data Fig. 6d–f), locks the CDK12 C-terminal extension in a conformation that is incompatible with DDB1 recruitment (Extended Data Fig. 6d). Consistently, THZ531 further decreased the TR-FRET signal and diminished cyclin K ubiquitination in vitro below the levels of the DMSO control<sup>24</sup> (Fig. 3d, Extended Data Fig. 6c). Flavopiridol<sup>25</sup>—an inhibitor that is derived from a natural product and is structurally distinct from CR8 (Fig. 3a, c)—also stimulated the binding of CDK12-cyclin K to DDB1 (Extended Data Fig. 8a). Although flavopiridol led to moderate ubiquitination of cyclin K in vitro (Fig. 3d), it did not degrade cyclin K in cells (Fig. 3e). Our results thus show that the interactions between DDB1 and different inhibitor compounds display substantial plasticity and that structurally diverse surface-exposed moieties in CR8, DRF053, roscovitine and flavopiridol can facilitate CDK12-cyclin K recruitment.

Small differences in their ability to stabilize the DDB1–CDK12 complex translate, in an almost binary manner, into the cellular degradation of cyclin K or lack thereof. This behaviour is reminiscent of CRL4<sup>CRBN</sup> and thalidomide analogues<sup>11,26</sup>, in which an apparent affinity threshold must be overcome to drive drug-induced degradation of the target protein.

### Cyclin K degradation adds to CR8 toxicity

Finally, to examine how CRL4-mediated degradation of cyclin K contributes to CR8 cytotoxicity compared to non-degradative CDK inhibition, we analysed CR8 toxicity in wild-type HEK293T-Cas9 cells and cells that were pretreated with MLN4924 (a NEDD8-activating enzyme inhibitor), subject to DCAF overexpression or genetically depleted of DDB1. Global inhibition of the activity of cullin-RING ligases by MLN4924 had only minor effects on cell viability (Extended Data Fig. 9a), but resulted in decreased sensitivity to CR8 (Extended Data Fig. 9b), showing that the neddylation of cullin-RING ligases substantially contributes to CR8 toxicity. Overexpression of the substrate receptor CRBN also affected the sensitivity of cells to CR8 and decreased the degradation of cyclin K (Extended Data Fig. 9c–g), presumably by reducing the free pool of DDB1. As expected, CR8-induced degradation of cyclin K was dependent on DDB1 (Fig. 3f) and, consistently, we found that cytotoxicity of CR8—but not that of the other CDK inhibitors—was tenfold lower in cells depleted of DDB1 (Fig. 3g, Extended Data Fig. 9h). Together, the data demonstrate that the CRL4-dependent gain-of-function glue degrader activity of CR8 strongly contributes to its cellular potency and provides an additional layer of orthologue-specific CDK inactivation through cyclin K degradation.

Kinase inhibitors have long been suspected to have a degradation component to their mode of action<sup>27,28</sup>, and our work provides a characterization and structural dissection of how a kinase inhibitor scaffold acquires degrader properties. Molecular glue degraders have thus far only been shown to engage substrate-recruiting E3 ligase modules. CDK12 is not a constitutive E3 ligase component, but rather serves as a drug-induced substrate receptor, linking DDB1 to the ubiquitination target. CR8 thus bypasses the requirement for a canonical DCAF and instead hijacks the essential adaptor protein DDB1. Although cyclin K is the primary ubiquitination target, CDK12 may become subject to autoubiquitination after prolonged exposure to CR8, in a similar manner to canonical DCAFs<sup>29,30</sup>.

Whereas previously reported molecular glue degraders engage a ligandable pocket on the ligase to recruit target proteins, CR8 instead binds the ATP pocket of CDK12 and does not rely on an independent ligand-binding site on DDB1 (Extended Data Fig. 4h). This suggests that the repertoire of target proteins and ubiquitin ligases accessible to targeted degradation can be expanded through target-binding small molecules that induce de novo contacts with a ligase or strengthen existing weak protein–protein interactions. Kinase inhibitors in particular often show poor selectivity, and small-molecule-induced inactivation of kinases that leverages complementary protein–protein interfaces offers a path towards improved drug selectivity—which might, for example, facilitate the selective inactivation of CDK12, an emerging therapeutic target<sup>31</sup>.

The gain-of-function glue degrader activity of CR8 is attributed to a 2-pyridyl moiety exposed on the kinase surface. Mutations of single residues that are exposed on the protein surface have been shown to promote the formation of higher-order protein complexes; the haemoglobin Glu to Val mutation, for example, induces polymerization in sickle cell anaemia<sup>32</sup>. Accordingly, single-residue mutations that are designed to increase surface hydrophobicity give rise to ordered protein assemblies<sup>33</sup>. Bound compounds—for example, enzyme inhibitors—can in principle mimic such amino acid changes and thereby have strong effects on the protein interaction landscape, suggesting that compound-induced protein–protein interactions may be more common than previously recognized. Together, our results suggest that the modification of surface-exposed regions in target-binding small molecules is a rational strategy that could be used to develop molecular glue degraders for a given protein target.

### Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41586-020-2374-x>.

1. Stanton, B. Z., Chory, E. J. & Crabtree, G. R. Chemically induced proximity in biology and medicine. *Science* **359**, eaao5902 (2018).
2. Chopra, R., Sadok, A. & Collins, I. A critical evaluation of the approaches to targeted protein degradation for drug discovery. *Drug Discov. Today Technol.* **31**, 5–13 (2019).
3. Yu, C. et al. High-throughput identification of genotype-specific cancer vulnerabilities in mixtures of barcoded tumor cell lines. *Nat. Biotechnol.* **34**, 419–423 (2016).
4. Corsello, S. M. et al. Discovering the anticancer potential of non-oncology drugs by systematic viability profiling. *Nat. Cancer* **1**, 235–248 (2020).
5. Ghandi, M. et al. Next-generation characterization of the Cancer Cell Line Encyclopedia. *Nature* **569**, 503–508 (2019).
6. Bettayeb, K. et al. CR8, a potent and selective, roscovitine-derived inhibitor of cyclin-dependent kinases. *Oncogene* **27**, 5797–5807 (2008).
7. Fink, E. C. & Ebert, B. L. The novel mechanism of lenalidomide activity. *Blood* **126**, 2366–2369 (2015).
8. Lu, G. et al. The myeloma drug lenalidomide promotes the cereblon-dependent destruction of Ikaros proteins. *Science* **343**, 305–309 (2014).
9. Krönke, J. et al. Lenalidomide causes selective degradation of IKZF1 and IKZF3 in multiple myeloma cells. *Science* **343**, 301–305 (2014).
10. Matuskiela, M. E. et al. A novel cereblon modulator recruits GSPT1 to the CRL4<sup>CRBN</sup> ubiquitin ligase. *Nature* **535**, 252–257 (2016).
11. Petzold, G., Fischer, E. S. & Thomä, N. H. Structural basis of lenalidomide-induced CK1 $\alpha$  degradation by the CRL4<sup>CRBN</sup> ubiquitin ligase. *Nature* **532**, 127–130 (2016).
12. Fischer, E. S. et al. Structure of the DDB1–CRBN E3 ubiquitin ligase in complex with thalidomide. *Nature* **512**, 49–53 (2014).
13. Han, T. et al. Anticancer sulfonamides target splicing by inducing RBM39 degradation via recruitment to DCAF15. *Science* **356**, eaal3755 (2017).
14. Uehara, T. et al. Selective degradation of splicing factor CAPER $\alpha$  by anticancer sulfonamides. *Nat. Chem. Biol.* **13**, 675–680 (2017).
15. Faust, T. B. et al. Structural complementarity facilitates E7820-mediated degradation of RBM39 by DCAF15. *Nat. Chem. Biol.* **16**, 7–14 (2020).
16. Bondeson, D. P. et al. Catalytic in vivo protein knockdown by small-molecule PROTACs. *Nat. Chem. Biol.* **11**, 611–617 (2015).
17. Angers, S. et al. Molecular architecture and assembly of the DDB1–CUL4A ubiquitin ligase machinery. *Nature* **443**, 590–593 (2006).
18. Lee, J. & Zhou, P. DCAFs, the missing link of the CUL4–DDB1 ubiquitin ligase. *Mol. Cell* **26**, 775–780 (2007).
19. Bösen, C. A. et al. The structure and substrate specificity of human Cdk12/Cyclin K. *Nat. Commun.* **5**, 3505 (2014).
20. Cheng, S.-W. G. et al. Interaction of cyclin-dependent kinase 12/CrkRS with cyclin K1 is required for the phosphorylation of the C-terminal domain of RNA polymerase II. *Mol. Cell. Biol.* **32**, 4691–4704 (2012).
21. Dixon-Clarke, S. E., Elkins, J. M., Cheng, S. W. G., Morin, G. B. & Bullock, A. N. Structures of the CDK12/CycK complex with AMP–PNP reveal a flexible C-terminal kinase extension important for ATP binding. *Sci. Rep.* **5**, 17122 (2015).
22. Oumata, N. et al. Roscovitine-derived, dual-specificity inhibitors of cyclin-dependent kinases and casein kinases 1. *J. Med. Chem.* **51**, 5229–5242 (2008).
23. Meijer, L. et al. Biochemical and cellular effects of roscovitine, a potent and selective inhibitor of the cyclin-dependent kinases cdc2, cdk2 and cdk5. *Eur. J. Biochem.* **243**, 527–536 (1997).
24. Zhang, T. et al. Covalent targeting of remote cysteine residues to develop CDK12 and CDK13 inhibitors. *Nat. Chem. Biol.* **12**, 876–884 (2016).
25. Sedlacek, H. et al. Flavopiridol (L86 8275; NSC 649890), a new kinase inhibitor for tumor therapy. *Int. J. Oncol.* **9**, 1143–1168 (1996).
26. Sievers, Q. L. et al. Defining the human C2H2 zinc finger degrome targeted by thalidomide analogs through CRBN. *Science* **362**, eaat0572 (2018).
27. Jones, L. H. Small-molecule kinase downregulators. *Cell Chem. Biol.* **25**, 30–35 (2018).
28. Schreiber, S. L. A chemical biology view of bioactive small molecules and a binder-based approach to connect biology to precision medicines. *Isr. J. Chem.* **59**, 52–59 (2019).
29. Ito, T. et al. Identification of a primary target of thalidomide teratogenicity. *Science* **327**, 1345–1350 (2010).
30. Fischer, E. S. et al. The molecular basis of CRL4<sup>DDB2/CSA</sup> ubiquitin ligase architecture, targeting, and activation. *Cell* **147**, 1024–1039 (2011).
31. Johnson, S. F. et al. CDK12 inhibition reverses de novo and acquired PARP inhibitor resistance in BRCA wild-type and mutated models of triple-negative breast cancer. *Cell Rep.* **17**, 2367–2381 (2016).
32. Dykes, G. W., Crepeau, R. H. & Edelstein, S. J. Three-dimensional reconstruction of the 14-filament fibers of hemoglobin S. *J. Mol. Biol.* **130**, 451–472 (1979).
33. Garcia-Seisdedos, H., Empereur-Mot, C., Elad, N. & Levy, E. D. Proteins evolve on the edge of supramolecular self-assembly. *Nature* **548**, 244–247 (2017).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature Limited 2020



# Article

## Methods

### Data reporting

No statistical methods were used to predetermine sample size. The experiments were not randomized and the investigators were not blinded to allocation during experiments and outcome assessment.

### Mammalian cell culture

The human HEK293T cell lines were provided by the Genetic Perturbation Platform, Broad Institute and the K562-Cas9, THP1-Cas9 and P31FUJ-Cas9 cell lines were provided by Z. Tothova (Broad Institute). MOLT-4 cells were purchased from ATCC and HEK293T-Cas9 cells<sup>26</sup> and MM1S-Cas9 cells<sup>34</sup> were previously published. Sf9 cells were purchased from Thermo Fisher Scientific and Hi5 cells were purchased from Expression Systems. HEK293T, K562-Cas9, THP1-Cas9, P31FUJ-Cas9, HEK293T-Cas9, MM1S-Cas9 and MOLT-4 cell lines were mycoplasma-negative and authenticated by STR profiling. Sf9 and Hi5 cells were authenticated by the vendor. HEK293T cells were cultured in Dulbecco's modified Eagle's medium (DMEM (Gibco) and all other cell lines in RPMI (Gibco), with 10% fetal bovine serum (FBS) (Invitrogen), glutamine (Invitrogen) and penicillin–streptomycin (Invitrogen) at 37 °C and 5% CO<sub>2</sub>.

### Compounds

(R)-CR8 (3605) was obtained from Tocris, (S)-CR8 (ALX-270-509-M005) and flavopiridol (ALX-430-161-M005) from Enzo Life Sciences, roscovitine (HY-30237), THZ531 (HY-103618) and LDC00067 (HY-15878) from MedChem Express and DRF053 (D6946) from Sigma.

### Antibodies

The following antibodies were used in this study: anti-cyclin K (Bethyl Laboratories, A301-939A for full length cyclin K), anti-cyclin K (Abcam, ab251652, for cyclin K(1–267)), anti-β-actin (Cell Signaling, 3700), anti-CRBN (Sigma Prestige, HPA045910), anti-mouse 800CW (LI-COR Biosciences, 926-32211), anti-rabbit 680LT (LI-COR Biosciences, 925-68021) and anti-rabbit IgG antibodies (Abcam, ab6721).

### Reporter vectors

The following reporters were used in this study: Artichoke (SFFV. BsmBICloneSite-17aaRigidLinker-eGFP.IRES.mCherry.cppt.EF1α. PuroR, Addgene 73320) for genome-wide screen and validation experiments; Cilantro 2 (PGK.BsmBICloneSite-10aaFlexibleLinker-eGFP.IRES. mCherry. cppt.EF1α.PuroR, Addgene 74450) for degradation kinetics; sgBFP (sgRNA.SFFV.tBFP) for validation of drug–E3 ligase pairs; sgRFP657 (sgRNA.EFS.RFP657) for validation of drug–E3 ligase pairs; and sgPuro (pXPRO03, Addgene 52963) for drug-sensitivity assays.

### Oligonucleotides

List of all oligonucleotides used in this study can be found in Supplementary Table 1.

### Bioinformatic screen

We computed Pearson correlations of the toxicity of PRISM repurposing compounds in 8 doses and 578 cell lines<sup>4</sup> with gene expression and copy-number variation of all detectable protein-coding genes of matched cell lines from The Cancer Cell Line Encyclopedia (CCLE)<sup>5</sup>. A z score was computed for each pair of compounds, dose toxicity and genomic feature (gene expression or copy-number variation) across all cell lines. For each compound–genomic feature pair, the most extreme correlations are ranked from negative to positive. To focus on novel relationships between compounds and genes, we restricted genes to a curated list of 499 E3 ligase components and compounds that are not an 'EGFR inhibitor', 'RAF inhibitor' or 'MDM inhibitor' on the basis of PRISM repurposing annotation<sup>5</sup>. Hit compounds were selected if either the z score was less than –6 or the compound was

ranked in the top 15 with a z score less than –4. The resulting list of 158 E3 gene–compound pairs was further curated and shortened manually to 96 E3 gene–compound pairs, which included 95 unique E3 ligases and 85 unique compounds.

### Cloning and lentiviral packaging of sgRNAs targeting 95 E3 ligases

sgRNAs targeting E3 ligases were selected from the human Brunello CRISPR library<sup>35</sup>. A total of 170 pairs of oligonucleotides (IDT) targeting 95 E3 ligases were annealed and cloned into the sgRNA.SFFV.tBFP (guide ID A) or sgRNA.EFS.RFP657 (guide ID B) fluorescent vectors in a 96-well format using previously published protocols<sup>36</sup>. In brief, vectors were linearized with BsmBI (New England Biolabs) and gel-purified with the Spin Miniprep Kit (Qiagen). Annealed oligonucleotides were phosphorylated with T4 polynucleotide kinase (New England Biolabs) and ligated into the linearized and purified vector backbones with T4 DNA ligase (New England Biolabs). Constructs were transformed into XL10-Gold ultracompetent *Escherichia coli* (Stratagene/Agilent Technologies), plasmids were purified using the Miniprep Kit (Qiagen), and the guide sequence was confirmed by Sanger sequencing. For validation of the primary screen, virus was produced in a 96-well format. In brief, 11,000 HEK293T cells were seeded per well in 100 μl DMEM medium supplemented with 10% FBS and penicillin–streptomycin–glutamine. The next day a packaging mix was prepared in a 96-well plate consisting of 500 ng psPAX2, 50 ng pVSV-G and 17 ng sgRNA backbone in 5 μl OptiMem (Invitrogen) and incubated for 5 min at room temperature. This mix was combined with 0.1 μl TransIT-LT1 (Mirus) in 5 μl OptiMem, incubated for 30 min at room temperature and then applied to cells. Two days after transfection, dead cells were removed by centrifugation and lentivirus-containing medium was collected and stored at –80 °C before use.

### Validation of drug–E3 ligase pairs from the bioinformatic screen

K562-Cas9, OVK16-Cas9, A564-Cas9, ES2-Cas9 and MOLM13-Cas9 cell lines were individually transduced with 192 sgRNAs targeting 95 E3 ligases in a 96-well plate format. Exactly 3,000 cells per well were plated in 100 μl RPMI supplemented with 10% fetal calf serum (FCS) and penicillin–streptomycin–glutamine and 30 μl per well of virus supernatant was added. The medium was changed 24 hours after infection. After three days, the percentage of sgRNA-transduced cells was determined by flow cytometry. If more than 60% of cells were transduced, untransduced cells were added to bring the level below 60%. Eight days after infection, the cell density was measured and adjusted to  $1.5 \times 10^5$  cells per ml with RPMI. For treatment, 50 μl of sgRNA-transduced cells were seeded into each well of a 384-well plate with preplated DMSO or cognate drug in three concentrations (0.1 μM, 1 μM or 10 μM) with the Agilent BRAVO Automated Liquid Handling Platform. Plates were sealed with White Rayon adhesive sealing tape (Thermo Fisher Scientific) and grown for three days. Adherent cell lines were trypsinized and resuspended in 50 μl RPMI with Matrix WellMate (Thermo Fisher Scientific). Suspension cells were directly subjected to analysis. Around 10 μl of cell suspension was subjected to flow analysis with a FACSCanto equipped with a high throughput sampler (BD Biosciences). The percentage of sgRNA-transduced cells in the drug-treatment wells was normalized to the DMSO control. Wells with fluorescent drug and samples with fewer than 120 viable cells or less than 6% fluorescent cells were removed from the analysis. All E3–drug pairs were ranked on the basis of the number of experimental conditions (cell line and drug dose) with more than 50% of sgRNA-transduced cells in drug-treatment wells in comparison with the corresponding DMSO control wells.

### Validation of DDB1-resistance phenotype

For validation experiments, virus was produced in a six-well plate format, as described above with the following adjustments:  $2.5 \times 10^5$  HEK293T cells per well in 2 ml DMEM medium, 3 μl per well TransIT-LT1,

15  $\mu$ l per well OPTI-MEM, 500 ng per well of the desired plasmid, 500 ng per well psPAX2 and 50 ng per well pVSV-G in 32.5  $\mu$ l per well OPTI-MEM. After collecting the virus,  $10 \times 10^3$  HEK293T-Cas9 cells in 100  $\mu$ l DMEM medium were transduced with 10  $\mu$ l of virus supernatant. The transduced HEK293T-Cas9 cells were then mixed with untransduced control cells in a 1:9 ratio. Nine days after sgRNA transduction, cells were treated for three days with DMSO or 1  $\mu$ M CR8 and analysed by flow cytometry to determine the percentage of BFP<sup>+</sup> cells. sgRNAs targeting *DDB1* provide partial depletion of DDB1 (50% *DDB1* alleles modified, reducing DDB1 levels by roughly 50%), which suggests selection towards heterozygous or hypomorphic clones.

#### Whole-proteome quantification using tandem mass tag mass spectrometry

Around  $10 \times 10^6$  MOLT-4 cells were treated with DMSO (triplicate) or 1  $\mu$ M CR8 (single replicate) for 1 h or 5 h and later were collected by centrifugation. Samples were processed, measured and analysed as described before<sup>37</sup>. Data are available in the PRIDE repository (PXD016187 and PXD016188).

#### Quantitative PCR

HEK293T-Cas9 cells were treated with DMSO or 1  $\mu$ M CR8 for 2 h, collected by centrifugation, washed with phosphate-buffered saline (PBS) and snap-frozen at  $-80^\circ\text{C}$ . mRNA was isolated using a QIAGEN RNA kit (Qiagen, 74106). For cDNA synthesis, total RNA was reverse-transcribed using a High-Capacity cDNA Reverse Transcription Kit (Thermo Fisher Scientific) before quantitative PCR (qPCR) analysis with TaqMan Fast Advanced Master Mix (Thermo Fisher Scientific, 4444557) for *CCNK* (TaqMan, Hs00171095\_m1, Life Technologies) and *GAPDH* (TaqMan, Hs02758991\_g1). Reactions were run and analysed on a CFX96 Real Time system (Bio-Rad).

#### Immunoblots for whole-protein lysate

Cells were washed with PBS and lysed (150 mM NaCl, 50 mM Tris (pH 7.5), 1% NP-40, 1% glycerol, 1 $\times$  Halt Cocktail protease and phosphatase inhibitors) for 20 min on ice. The insoluble fraction was removed by centrifugation, the protein concentration was quantified using a BCA protein assay kit (Pierce), and an equal amount of lysate was run on SDS-PAGE 4–12% Bis-Tris Protein Gels (NuPAGE, Thermo Fisher Scientific) and then transferred to nitrocellulose membrane with a Trans-Blot Turbo System (Bio-Rad). Membranes were blocked in Odyssey Blocking Buffer/PBS (LI-COR Biosciences) and incubated with primary antibodies overnight at  $4^\circ\text{C}$ . The membranes were then washed in Tris-buffered saline with Tween 20 (TBS-T), incubated for 1 h with secondary IRDye-conjugated antibodies (LI-COR Biosciences) and washed three times in TBS-T for 5 min before near-infrared western blot detection on an Odyssey Imaging System (LI-COR Biosciences).

#### Cyclin K reporter stability analysis

HEK293T-Cas9 cells expressing the cyclin  $K_{\text{eGFP}}$  degradation reporter were transduced with experimental sgRNAs. Nine days after infection, the cells were dosed for 2 h with DMSO or 1  $\mu$ M CR8 and the fluorescent signal was quantified by flow cytometry (CytoFLEX, Beckman or LSR-Fortessa flow cytometer, BD Biosciences). Using FlowJo (flow cytometry analysis software, BD Biosciences), the geometric mean of the eGFP and mCherry fluorescent signal for round and mCherry-positive cells was calculated. The ratio of eGFP to mCherry was normalized to the average of three DMSO-treated controls.

#### Genome-wide CRISPR screen for CR8 resistance

Five per cent (v/v) of the human genome-wide CRISPR-KO Brunello library with 0.4  $\mu$ l polybrene  $\text{ml}^{-1}$  (stock of 8  $\text{mg ml}^{-1}$ ) was added to  $1.5 \times 10^8$  HEK293T-Cas9 cells in 75 ml medium and transduced (2,400 rpm, 2 h,  $37^\circ\text{C}$ ). Twenty-four hours after infection, sgRNA-transduced cells were selected with 2  $\mu\text{g ml}^{-1}$  puromycin for two days. On the ninth day

after infection, cells were treated with either DMSO ( $n=1$ ) or 1  $\mu$ M CR8 ( $n=1$ ) and cultured for an additional three days. Resistant live cells were selected by gently washing away detached dead cells from the medium. Cell pellets were resuspended in multiple direct lysis buffer reactions (1 mM CaCl<sub>2</sub>, 3 mM MgCl<sub>2</sub>, 1 mM EDTA, 1% Triton X-100, Tris pH 7.5, with freshly supplemented 0.2  $\text{mg ml}^{-1}$  proteinase K) with  $1 \times 10^6$  cells per 100- $\mu$ l reaction. The sgRNA sequence was amplified in a first PCR reaction with eight staggered forward primers. Direct lysed cells (20  $\mu$ l) were mixed with 0.04U Titanium Taq (Takara Bio 639210), 0.5 $\times$  Titanium Taq buffer, 800  $\mu$ M dNTP mix, 200 nM SBS3-Stagger-pXPRO03 forward primer and 200 nM SBS12-pXPRO03 reverse primer in a 50- $\mu$ l reaction (cycles: 5 min at  $94^\circ\text{C}$ , 15  $\times$  (30 s at  $94^\circ\text{C}$ , 15 s at  $58^\circ\text{C}$ , 30 s at  $72^\circ\text{C}$ ), 2 min at  $72^\circ\text{C}$ ). Exactly 2  $\mu$ l of the first PCR reaction was used as the template for 15 cycles of the second PCR, in which Illumina adapters and barcodes were added (0.04U Titanium Taq, 1 $\times$  Titanium Taq buffer, 800  $\mu$ M dNTP mix, 200 nM P5-SBS3 forward primer, 200 nM P7-barcode-SBS12 reverse primer). An equal amount of all samples was pooled and subjected to preparative agarose electrophoresis followed by gel purification (Qiagen). Eluted DNA was further purified by NaOAc and isopropanol precipitation. Amplified sgRNAs were quantified using the Illumina NextSeq platform (Genomics Platform, Broad Institute). Read counts for all guides targeting the same gene were used to generate *P* values. Hits enriched in the resistant population with a false discovery rate (FDR)  $< 0.05$  and enriched more than fivefold are labelled on the plot<sup>38</sup> (Fig. 1f).

#### Bison CRISPR screen for CR8 resistance

The Bison CRISPR library targets 713 E1, E2 and E3 ubiquitin ligases, deubiquitinases and control genes and contains 2,852 guide RNAs. It was cloned into pXPRO03 as previously described<sup>35</sup> by the Broad Institute Genetic Perturbation Platform (GPP). The virus for the library was produced in a T-175 flask format, as described above with the following adjustments:  $1.8 \times 10^7$  HEK293T cells in 25 ml complete DMEM medium, 244  $\mu$ l TransIT-LT1, 5 ml OPTI-MEM, 32  $\mu$ g library, 40  $\mu$ g psPAX2 and 4  $\mu$ g pVSV-G in 1 ml OPTI-MEM. Ten per cent (v/v) of the Bison CRISPR library was added to  $6 \times 10^6$  HEK293T-Cas9 cells in triplicates and transduced. Samples ( $n=3$ ) were processed as described above for the genome-wide resistance screen.

#### Genome-wide CRISPR screen for cyclin K reporter stability

A single clone of cyclin  $K_{\text{eGFP}}$  HEK293T-Cas9 was transduced with the genome-wide Brunello library as described above with the following modification:  $4.5 \times 10^8$  cyclin  $K_{\text{eGFP}}$  HEK293T-Cas9 cells in 225 ml medium. Nine days later, cells were treated with CR8 ( $n=3$ ) or DMSO ( $n=3$ ) for at least 2 h and the cyclin K stable population was separated using fluorescence-activated cell sorting (FACS). Four populations were collected (top 5%, top 5–15%, lowest 5–15% and lowest 5%) on the basis of the cyclin  $K_{\text{eGFP}}$  to mCherry mean fluorescent intensity (MFI) ratio on an MA900 Cell Sorter (Sony). Sorted cells were collected by centrifugation and subjected to direct lysis as described above. The screen was analysed as described below by comparing stable populations (top 5% eGFP/mCherry expression) to unstable populations (lowest 15% eGFP/mCherry expression). Hits enriched in the cyclin K stable population with FDR  $< 0.05$  are labelled on the plot (Fig. 1g).

#### Data analysis of the pooled CRISPR screens

The data analysis pipeline comprised the following steps. (1) Each sample was normalized to the total number of reads. (2) For each guide, the ratio of reads in the stable versus the unstable sorted gate was calculated, and sgRNAs were ranked. (3) The ranks for each guide were summed for all replicates. (3) The gene rank was determined as the median rank of the four guides targeting it. (4) *P* values were calculated by simulating a distribution with guide RNAs that had randomly assigned ranks over 100 iterations. R scripts can be found in the Supplementary Information.

### Screen with arrayed DCAF library

An arrayed DCAF library (targeting DCAF substrate receptors, DCAF-like and control genes) was constructed as described above with the appropriate oligonucleotides (Supplementary Table 1). K562-Cas9, P31FUJ-Cas9, THP1-Cas9 and MM1S-Cas9 cells were individually transduced and treated with DMSO or 1  $\mu$ M CR8 (K562-Cas9, P31FUJ-Cas9, THP1-Cas9) or 0.1  $\mu$ M CR8 (MM1S-Cas9). The analysis was performed as described above for the validation of the DDB1-resistance phenotype.

### Protein purification

Human wild-type and mutant versions of DDB1 (Uniprot entry Q16531), CDK12 (Q9NYV4, K965R) and cyclin K (O75909) were subcloned into pAC-derived vectors<sup>39</sup> and recombinant proteins were expressed as *N*-terminal His<sub>6</sub>, His<sub>6</sub>-Spy, StrepII or StrepII-Avi fusions in *Trichoplusia ni* High Five insect cells using the baculovirus expression system (Invitrogen)<sup>40</sup>.

Wild-type or mutant forms of full-length or BPB domain deletion ( $\Delta$ BPB: amino acids 396–705 deleted) constructs of His<sub>6</sub>-DDB1 and StrepII-Avi-DDB1 were purified as previously described for DDB1-DCAF complexes<sup>12</sup>. High Five insect cells co-expressing truncated versions of wild-type or mutant His<sub>6</sub>-CDK12 (amino acids 713–1052 or 713–1032) and His<sub>6</sub>- or His<sub>6</sub>-Spy-tagged cyclin K (amino acids 1–267) were lysed by sonication in 50 mM Tris-HCl (pH 8.0), 500 mM NaCl, 10% (v/v) glycerol, 10 mM MgCl<sub>2</sub>, 10 mM imidazole, 0.25 mM tris(2-carboxyethyl)phosphine (TCEP), 0.1% (v/v) Triton X-100, 1 mM phenylmethylsulfonylfluoride (PMSF) and 1 $\times$  protease inhibitor cocktail (Sigma). After ultracentrifugation, the soluble fraction was passed over HIS-Select Ni<sup>2+</sup> affinity resin (Sigma), washed with 50 mM Tris-HCl (pH 8.0), 1 M NaCl, 10% (v/v) glycerol, 0.25 mM TCEP and 10 mM imidazole and eluted in 50 mM Tris-HCl (pH 8.0), 200 mM NaCl, 10% (v/v) glycerol, 0.25 mM TCEP and 250 mM imidazole. When necessary, affinity tags were removed by overnight tobacco etch virus (TEV) protease treatment. In cases of HIS-Select Ni<sup>2+</sup> affinity-purified CDK12-cyclin K that was not subjected to TEV cleavage, the pH of the eluate was adjusted to 6.8 before ion-exchange chromatography. StrepII-tagged versions of CDK12-cyclin K were affinity-purified using Strep-Tactin Sepharose (IBA), omitting imidazole in lysis, wash and elution buffers, supplementing the elution buffer with 2.5 mM desthiobiotin (IBA GmbH) and using 50 mM Tris-HCl (pH 6.8) throughout.

For ion-exchange chromatography, affinity-purified proteins were diluted in a 1:1 ratio with buffer A (50 mM Tris-HCl (pH 6.8), 10 mM NaCl, 2.5% (v/v) glycerol and 0.25 mM TCEP) and passed over an 8-ml Poros 50HQ column. The flow-through was again diluted in a 1:1 ratio with buffer A and passed over an 8-ml Poros 50HS column. Bound proteins were eluted by a linear salt gradient mixing buffer A and buffer B (50 mM Tris-HCl (pH 6.8), 1 M NaCl, 2.5% (v/v) glycerol and 0.25 mM TCEP) over 15 column volumes to a final ratio of 80% buffer B. Poros 50HS peak fractions containing the CDK12-cyclin K complex were concentrated and subjected to size-exclusion chromatography in 50 mM HEPES (pH 7.4), 200 mM NaCl, 2.5% (v/v) glycerol and 0.25 mM TCEP. The concentrated proteins were flash-frozen in liquid nitrogen and stored at  $-80^{\circ}\text{C}$ .

### Co-immunoprecipitation assay

The purified CDK12<sup>His6</sup>-cyclin K<sup>StrepII</sup> complex was mixed with equimolar concentrations of full-length His<sub>6</sub>-tagged DDB1 or TEV-cleaved DDB1 ( $\Delta$ BPB) (5  $\mu$ M) in the presence of 5  $\mu$ M (*R*)-CR8 or DMSO in immunoprecipitation (IP) buffer (50 mM HEPES (pH 7.4), 200 mM NaCl, 0.25 mM TCEP and 0.05% (v/v) Tween-20) containing 1 mg ml<sup>-1</sup> bovine serum albumin. The solution was added to Strep-Tactin MacroPrep beads (IBA GmbH) pre-equilibrated in IP buffer and incubated for 1 h at 4  $^{\circ}\text{C}$  on an end-over-end shaker. The beads were extensively washed with IP buffer, and the bound protein was eluted with IP buffer containing 2.5 mM desthiobiotin for 1 h at 4  $^{\circ}\text{C}$  on an end-over-end

shaker. Eluted proteins were separated by SDS-PAGE and stained with Coomassie blue.

### Crystallization and data collection

The protein solution for crystallization contained 70  $\mu$ M TEV-cleaved DDB1 ( $\Delta$ BPB), 80  $\mu$ M (*R*)-CR8 and 80  $\mu$ M TEV-cleaved CDK12-cyclin K in 50 mM HEPES (pH 7.4), 200 mM NaCl and 0.25 mM TCEP. Crystals were grown by vapour diffusion in drops containing 200 nl DDB1 ( $\Delta$ BPB)-(*R*)-CR8-CDK12(713–1052)-cyclin K(1–267) complex solution mixed with 200 nl of reservoir solution containing 0.9 M ammonium citrate tribasic (pH 7.0) in two-well-format sitting drop crystallization plates (Swissci). Plates were incubated at 19  $^{\circ}\text{C}$  and crystals appeared 5–13 days after set-up. Crystals were flash-cooled in liquid nitrogen in reservoir solution supplemented with 25% (v/v) glycerol as a cryoprotectant before data collection. Diffraction data were collected at the Swiss Light Source (SLS; beamline PXI) with an Eiger 16M detector (Dectris) at a wavelength of 1  $\text{\AA}$  and a crystal cooled to 100 K. Data were processed with DIALS, scaled with AIMLESS supported by other programs of the CCP4 suite<sup>41</sup> and converted to structure factor amplitudes with STARANISO<sup>42</sup>, applying a locally weighted  $\text{CC}_{1/2} = 0.3$  resolution cut-off.

### Structure determination and model building

The DDB1 ( $\Delta$ BPB)-(*R*)-CR8-CDK12(713–1052)-cyclin K(1–267) complex formed crystals belonging to space group *P*3<sub>1</sub>21, with three complexes in the crystallographic asymmetric unit. Their structure was determined using molecular replacement in Phaser<sup>43</sup> with a search model derived from PDB entry 6H0F for DDB1 ( $\Delta$ BPB) and PDB entry 4NST for CDK12-cyclin K. The initial model was improved by iterative cycles of building with Coot<sup>44</sup>, and refinement using phenix.refine<sup>45</sup> or autoBUSTER<sup>46</sup>, with ligand restraints generated using eLBOW through phenix.ready\_set<sup>47</sup>. The final model was produced by refinement with autoBUSTER. Analysis with MolProbity<sup>48</sup> indicates that 93.9% of the residues in the final model are in favourable regions of the Ramachandran plot, with 0.6% outliers. Data processing and refinement statistics are provided in Extended Data Table 1. Interface analysis was performed using PISA<sup>49</sup>.

### Biotinylation of DDB1

Purified full-length StrepII-Avi-DDB1 was biotinylated in vitro at a concentration of 8  $\mu$ M by incubation with final concentrations of 2.5  $\mu$ M BirA enzyme and 0.2 mM D-biotin in 50 mM HEPES (pH 7.4), 200 mM NaCl, 10 mM MgCl<sub>2</sub>, 0.25 mM TCEP and 20 mM ATP. The reaction was incubated for 1 h at room temperature and stored at 4  $^{\circ}\text{C}$  for 14–16 h. Biotinylated DDB1 (DDB1<sub>biotin</sub>) was purified by gel-filtration chromatography and stored at  $-80^{\circ}\text{C}$  (around 20  $\mu$ M).

### TR-FRET

Increasing concentrations of Alexa488-SpyCatcher-labelled<sup>26</sup> His<sub>6</sub>-Spy-cyclin K in complex with His<sub>6</sub>-CDK12 (CDK12-cyclin K<sub>Alexa488</sub>) were added to a mixture of DDB1<sub>biotin</sub> at 50 nM, terbium-coupled streptavidin at 4 nM (Invitrogen) and compounds at 10  $\mu$ M (final concentrations) in 384-well microplates (Greiner, 784076) in a buffer containing 50 mM Tris (pH 7.5), 150 mM NaCl, 0.1% pluronic acid and 0.5% DMSO (see also figure legends). Before TR-FRET measurements, reactions were incubated for 15 min at room temperature. After excitation of terbium fluorescence at 337 nm, emissions at 490 nm (terbium) and 520 nm (Alexa488) were measured with a 70- $\mu$ s delay to reduce background fluorescence and the reactions were followed by recording 60 data points of each well over 1 h using a PHERAstar FS microplate reader (BMG Labtech). The TR-FRET signal of each data point was extracted by calculating the 520:490 nm ratio. Data were analysed with Prism 7 (GraphPad) assuming equimolar binding of DDB1<sub>biotin</sub> to CDK12-cyclin K<sub>Alexa488</sub> using the equations described previously<sup>8</sup>.

Counter-titrations with unlabelled proteins were carried out by mixing 500  $\mu$ M CDK12-cyclin K<sub>Alexa488</sub> with 50 nM DDB1<sub>biotin</sub> in the presence of 4 nM terbium-coupled streptavidin and 1  $\mu$ M compound for

titrations with unlabelled DDB1 or 12.5  $\mu\text{M}$  compound for titrations with unlabelled CDK12. After incubation for 15 min at room temperature, increasing amounts of unlabelled CDK12–cyclin K or DDB1 (0–10  $\mu\text{M}$ ) were added to the preassembled CDK12–cyclinK<sub>Alexa488</sub>–DDB1<sub>biotin</sub> complexes in a 1:1 volume ratio and incubated for 15 min at room temperature. TR-FRET data were acquired as described above. The 520:490 nm ratios were plotted to calculate the half maximal inhibitory concentrations (IC<sub>50</sub>) assuming a single binding site using Prism 7 (GraphPad). IC<sub>50</sub> values were converted to the respective inhibition constant (K<sub>i</sub>) values as described previously<sup>50</sup>. Three technical replicates were carried out per experiment.

#### CUL4–RBX1–DDB1 reconstitution and in vitro CUL4 neddylation

In vitro CRL4 reconstitution and CUL4 neddylation were performed as described<sup>11</sup>. CUL4<sub>His6</sub>–RBX1<sub>His6</sub> at 3.5  $\mu\text{M}$  was incubated with His<sub>6</sub>-tagged DDB1 at 3  $\mu\text{M}$  in a reaction mixture containing 3.8  $\mu\text{M}$  NEDD8, 50 nM NAE1–UBA3 (E1), 30 nM UBC12 (E2), 1 mM ATP, 50 mM Tris (pH 7.5), 100 mM NaCl, 2.5 mM MgCl<sub>2</sub>, 0.5 mM DTT and 5% (v/v) glycerol for 1.5 h at room temperature. Neddylated and gel-filtration-purified CUL4–RBX1–DDB1 (CUL4<sub>NEDD8</sub>–RBX1–DDB1) was concentrated to 7.6  $\mu\text{M}$ , flash-frozen and stored at –80 °C.

#### In-vitro ubiquitination assays

In vitro ubiquitination was performed by mixing CUL4<sub>NEDD8</sub>–RBX1–DDB1 at 70 nM with a reaction mixture containing compounds at 2  $\mu\text{M}$ , CDK12–cyclin K at 500 nM, E1 (UBA1, Boston Biochem) at 50 nM, E2 (UBCH5a, Boston Biochem) at 1  $\mu\text{M}$  and ubiquitin at 20  $\mu\text{M}$ . Reactions were carried out in 50 mM Tris (pH 7.5), 150 mM NaCl, 5 mM MgCl<sub>2</sub>, 0.2 mM CaCl<sub>2</sub>, 1 mM ATP, 0.1% Triton X-100 and 0.1 mg ml<sup>-1</sup> bovine serum albumin (BSA), incubated for 0–30 min at 30 °C and analysed by western blot using anti-cyclin K and anti-rabbit IgG antibodies. Blots were scanned on an Amersham 600 CCD-based imaging system (GE Life Sciences).

#### ITC

ITC experiments were performed at 25 °C on a VP-ITC isothermal titration calorimeter (Microcal). Purified and TEV-cleaved CDK12–cyclin K and DDB1( $\Delta$ BPB) were exhaustively dialysed in 50 mM HEPES (pH 7.4), 150 mM NaCl, 0.25 mM TCEP and 0.5% DMSO and loaded into the sample cell at a final concentration of 10–50  $\mu\text{M}$ . Kinase inhibitors (CR8 or roscovitine) were diluted from a 100 mM DMSO stock solution to 100–500  $\mu\text{M}$  in buffer containing 50 mM HEPES (pH 7.4), 150 mM NaCl and 0.25 mM TCEP. The final DMSO concentration was 0.5%. Titrations with 100–500  $\mu\text{M}$  compound were performed typically through about 30 injections of 6–10  $\mu\text{l}$  at 210-s intervals from a 300- $\mu\text{l}$  syringe rotating at 300 rpm. An initial injection of the ligand (4  $\mu\text{l}$ ) was made and discarded during data analysis. For probing DDB1–CDK12–cyclin K complex formation, DDB1( $\Delta$ BPB) (100  $\mu\text{M}$ , in the syringe) was titrated into the cell containing CDK12–cyclin K (10  $\mu\text{M}$ ) or CDK12–cyclin K (10  $\mu\text{M}$ ) pre-incubated with CR8 (30  $\mu\text{M}$ ). The heat change accompanying the titration was recorded as differential power by the instrument and determined by integration of the peak obtained. Titrations of ligand to buffer only and buffer into protein were performed to allow baseline corrections. The heat change was fitted using nonlinear least-squares minimization to obtain the dissociation constants (K<sub>d</sub>), the enthalpy of binding ( $\Delta H$ ) and the stoichiometry (N). Between one and three replicates were performed per titration.

#### Bioluminescence resonance energy transfer analyses

Bioluminescence resonance energy transfer (BRET) experiments were performed using a NanoBRET PPI starter kit (Promega N1821) according to the manufacturer's instructions and as previously described<sup>51</sup>.

#### Drug-sensitivity assays

HEK293T-Cas9 cells were resuspended at 0.15  $\times 10^6$  per ml and plated on a 384-well plate with 50  $\mu\text{l}$  per well and MLN4924, MLN7243 or MG132

with or without CR8 serially diluted with D300e Digital Dispenser (Tecan).

HEK293T-Cas9 cells (0.625  $\times 10^6$  cells per well of a 6-well plate) were seeded the day before transfection. The following day, 2.5  $\mu\text{g}$  of pRSF91-GFP or pRSF91-CRBN<sup>9</sup> plasmid DNA was mixed with 250  $\mu\text{l}$  Opti-Mem and 7.5  $\mu\text{l}$  Transit-LT1 (Mirus Bio) according to the manufacturer's protocol. Two days (48 h) after transfection, cells were resuspended at 0.15  $\times 10^6$  cells per ml and plated on a 384-well plate with 50  $\mu\text{l}$  per well.

HEK293T-Cas9 cells were transduced with sgRNAs targeting either *DDB1* or luciferase in a pXPR003 backbone (GPP) (Supplementary Table 1). After nine days of puromycin selection, cells were replated into a 96-well format with 2  $\times 10^4$  cells per well and CR8 and roscovitine were serially diluted with D300e Digital Dispenser (Tecan).

After three days of drug exposure, cell viability was assessed using the CellTiter-Glo luminescent assay (Promega, G7572) on an EnVision Multilabel Plate Reader (Perkin Elmer) or CLARIOstar Plus, MARS 3.4 (BMG LabTech). Cell viabilities were calculated relative to DMSO controls. The half-maximum effective concentration (EC<sub>50</sub>) values were derived from standard four-parameter log-logistic curves fitted with the 'dr4pl' R package.

#### Cyclin K reporter stability analysis with CRBN overexpression

HEK293T-Cas9 cells expressing the cyclin K<sub>eGFP</sub> degradation reporter were transiently transfected with pLX307-Luc or pLX307-CRBN (for flow experiment) as described above and 48 h after infection were treated with CR8 for 2 h and analysed by flow cytometry.

#### Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

#### Data availability

Structural data have been deposited in the PDB under the accession code 6TD3. Proteome quantification data are available in the PRIDE repository (PXD016187 and PXD016188). Additional ITC data are provided in Supplementary Fig. 1. Uncropped gel and western blot source data are shown in Supplementary Fig. 2 and the flow cytometry gating strategy is shown in Supplementary Fig. 3.

#### Code availability

The code necessary to reproduce the statistical analysis is included in the Supplementary Information.

34. Sievers, Q. L., Gasser, J. A., Cowley, G. S., Fischer, E. S. & Ebert, B. L. Genome-wide screen identifies cullin-RING ligase machinery required for lenalidomide-dependent CRL4<sup>CRBN</sup> activity. *Blood* **132**, 1293–1303 (2018).
35. Doench, J. G. et al. Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR–Cas9. *Nat. Biotechnol.* **34**, 184–191 (2016).
36. Sanjana, N. E., Shalem, O. & Zhang, F. Improved vectors and genome-wide libraries for CRISPR screening. *Nat. Methods* **11**, 783–784 (2014).
37. Donovan, K. A. et al. Thalidomide promotes degradation of SALL4, a transcription factor implicated in Duane Radial Ray syndrome. *eLife* **7**, e38430 (2018).
38. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. B* **57**, 289–300 (1995).
39. Abdulrahman, W. et al. A set of baculovirus transfer vectors for screening of affinity tags and parallel expression strategies. *Anal. Biochem.* **385**, 383–385 (2009).
40. Li, T., Pavletich, N. P., Schulman, B. A. & Zheng, N. High-level expression and purification of recombinant SCF ubiquitin ligases. *Methods Enzymol.* **398**, 125–142 (2005).
41. Winn, M. D. et al. Overview of the CCP4 suite and current developments. *Acta Crystallogr. D* **67**, 235–242 (2011).
42. Tickle, I. J. et al. STARANISO (Global Phasing, 2018).
43. McCoy, A. J. et al. Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674 (2007).
44. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D* **66**, 486–501 (2010).
45. Afonine, P. V. et al. Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr. D* **68**, 352–367 (2012).
46. Bricogne, G. et al. BUSTER (Global Phasing, 2017).

47. Moriarty, N. W., Grosse-Kunstleve, R. W. & Adams, P. D. electronic Ligand Builder and Optimization Workbench (eLBOW): a tool for ligand coordinate and restraint generation. *Acta Crystallogr. D* **65**, 1074–1080 (2009).
48. Chen, V. B. et al. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D* **66**, 12–21 (2010).
49. Krissinel, E. & Henrick, K. Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.* **372**, 774–797 (2007).
50. Cer, R. Z., Mudunuri, U., Stephens, R. & Lebeda, F. J.  $IC_{50}$ -to- $K_i$ : a web-based tool for converting  $IC_{50}$  to  $K_i$  values for inhibitors of enzyme activity and ligand binding. *Nucleic Acids Res.* **37**, W441–W445 (2009).
51. Sperling, A. S. et al. Patterns of substrate affinity, competition, and degradation kinetics underlie biological activity of thalidomide analogs. *Blood* **134**, 160–170 (2019).

**Acknowledgements** We thank S. Cavadini and A. Schenk for support during electron microscopy data collection and processing; M. Kolesnikov for help with ITC experiments; the Broad Institute PRISM (particularly M. Kocak); the Compound Management, Cancer Data Science, Walk-up Sequencing, Genetic Perturbation Platform and Flow Facility teams (particularly P. Rogers), K. DeRuff for the introduction to the Agilent BRAVO Automated Liquid Handling Platform, and J. Kennedy for providing sgRNA.SFFV:tBFP and sgRNA.EFS.RFP657 backbones. We acknowledge the Paul Scherrer Institute for provision of synchrotron radiation beam time at beamline PXI of the SLS and thank T. Tomizaki for assistance. We are grateful to all members of the Ebert group, particularly R. Belizaire, S. Koochaki, P. Miller and C. Zou; as well as R. Nowak and P. Tsvetkov, for discussions. This work was supported by the NIH (grants R01HL082945, P01CA108631 and P50CA206963), the Howard Hughes Medical Institute, the Edward P. Evans Foundation and the Leukaemia and Lymphoma Society (to B.L.E.); funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program grant agreement no. 666068, the Gebert R uf Stiftung (GRS-057/14) and the Novartis Research Foundation (to N.H.T.); and NIH grants NCI R01CA214608 and R01CA218278 and a Mark Foundation Emerging Leader Award (to E.S.F.). M.S. received funding from the European Union's Horizon 2020 Research and Innovation Program under the Marie Skłodowska-Curie grant agreement no. 702642; Z.K. was supported by the European Union's Horizon 2020 Research and Innovation Program under the Marie Skłodowska-Curie grant agreement no. 765445; G.P. was supported by the Human Frontier Science Program (HFSP Long-Term Fellowship LT000210/2014) and the European Molecular Biology Organization (EMBO Advanced Fellowship aALTF 761-2016); A.S.S. was supported by a DF/HCC K12 grant, a

Conquer Cancer Foundation Young Investigator Award and an award from the Wong Family Foundation; R.S.S. was supported by an Intermediate Fellowship from the Kay Kendall Leukaemia Fund and an Advanced Clinician Scientist Fellowship from Cancer Research UK; and S.M.C. received funding from grants KL2 TRO02542 and K08 CA230220.

**Author contributions** M.S. designed and performed functional genomics studies with the help of J.K., R.S.S. and E.C.F.; Z.K. and G.P. designed and carried out structural, biochemical and biophysical studies with the help of D.S.; M.S., Y.-D.L., M.M. and Q.L.S. designed and performed cell validation experiments with the help of A.S.S., J.A.G. and M.J.; K.A.D. performed the mass spectrometry experiments; M.S., L.R.W. and S.M.C. performed bioinformatic PRISM analysis; R.B. and G.P. performed structure refinement with the help of Z.K.; D.G., C.S., S.F., T.R.G., E.S.F., N.H.T. and B.L.E. supervised the project; and Z.K., G.P., M.S., B.L.E. and N.H.T. wrote the manuscript with input from all authors.

**Competing interests** B.L.E. has received research funding from Celgene and Deerfield and consulting fees from GRAIL. He serves on the scientific advisory boards for and holds equity in Skyhawk Therapeutics and Exo Therapeutics. E.S.F. is a founder and/or member of the scientific advisory board and equity holder of C4 Therapeutics and Civetta Therapeutics, and is a consultant to Novartis, AbbVie and Pfizer. N.H.T. receives funding from the Novartis Research Foundation and is a member of the scientific advisory board of Monte Rosa Therapeutics. The Fischer laboratory receives or has received research funding from Novartis, Deerfield and Astellas. S.F. has had a consulting or advisory role, received honoraria, research funding and/or travel/accommodation expenses funding from the following for-profit companies: Bayer, Roche, Amgen, Eli Lilly, PharmaMar, AstraZeneca and Pfizer. R.B. is now an employee of Monte Rosa Therapeutics. S.M.C. and T.R.G. receive research funding from Bayer HealthCare. T.R.G. was formerly a consultant and equity holder in Foundation Medicine (acquired by Roche). T.R.G. also is a consultant to GlaxoSmithKline and is a founder of Sherlock Biosciences.

#### Additional information

**Supplementary information** is available for this paper at <https://doi.org/10.1038/s41586-020-2374-x>.

**Correspondence and requests for materials** should be addressed to N.H.T. or B.L.E.

**Peer review information** *Nature* thanks Ivan Dikic, Frank Sicheri and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

**Reprints and permissions information** is available at <http://www.nature.com/reprints>.



# Rational discovery of molecular glue degraders via scalable chemical profiling

Cristina Mayor-Ruiz<sup>1</sup>, Sophie Bauer<sup>1,7</sup>, Matthias Brand<sup>1,7</sup>, Zuzanna Kozicka<sup>2,3</sup>, Marton Siklos<sup>1</sup>, Hana Imrichova<sup>1</sup>, Ines H. Kaltheuner<sup>4</sup>, Elisa Hahn<sup>1</sup>, Kristina Seiler<sup>1</sup>, Anna Koren<sup>1</sup>, Georg Petzold<sup>2</sup>, Michaela Fellner<sup>5</sup>, Christoph Bock<sup>1,6</sup>, André C. Müller<sup>1</sup>, Johannes Zuber<sup>5</sup>, Matthias Geyer<sup>4</sup>, Nicolas H. Thomä<sup>2</sup>, Stefan Kubicek<sup>1</sup> and Georg E. Winter<sup>1</sup>✉

**Targeted protein degradation is a new therapeutic modality based on drugs that destabilize proteins by inducing their proximity to E3 ubiquitin ligases. Of particular interest are molecular glues that can degrade otherwise unligandable proteins by orchestrating direct interactions between target and ligase. However, their discovery has so far been serendipitous, thus hampering broad translational efforts. Here, we describe a scalable strategy toward glue degrader discovery that is based on chemical screening in hyponeedlylated cells coupled to a multi-omics target deconvolution campaign. This approach led us to identify compounds that induce ubiquitination and degradation of cyclin K by prompting an interaction of CDK12–cyclin K with a CRL4B ligase complex. Notably, this interaction is independent of a dedicated substrate receptor, thus functionally segregating this mechanism from all described degraders. Collectively, our data outline a versatile and broadly applicable strategy to identify degraders with nonobvious mechanisms and thus empower future drug discovery efforts.**

A dependency on traditional antagonistic/agonistic pharmacologic concepts has limited the reach of small-molecule strategies. This in turn has contributed to a stagnation in therapeutic innovation caused by an inability to pursue some of the best-validated targets in life-threatening diseases such as cancer. A promising alternative to pharmacologic modulation of protein activity is to regulate the abundance of disease-causing proteins, for instance by chemically redirecting them to E3 ubiquitin ligases for degradation. A strong emphasis is currently placed on heterobifunctional degraders, often referred to as PROTACs (proteolysis targeting chimeras), which simultaneously bind target and ligase with dedicated warheads that are connected by a flexible linker<sup>1–6</sup>. While this modular design principle allows for a straightforward alteration of the degraded neosubstrate by switching the respective warhead, the overall proteomic space that is degradable via PROTACs is intrinsically limited to target proteins that can effectively be liganded via small organic molecules.

Conversely, molecular glue degraders were shown to degrade unligandable proteins. Mechanistic characterization of the clinically approved thalidomide analogs (IMiDs) as molecular glue degraders exemplify how this particular molecular pharmacology can expand the therapeutically actionable space in the human proteome<sup>7</sup>. Binding of IMiDs to the CRL4<sup>CRBN</sup> E3 ligase causes recruitment of selected zinc finger transcription factors, leading to their ubiquitination and subsequent proteasomal degradation<sup>8–11</sup>. It is worth noting that IMiDs have per se no measurable binding affinity to the degraded transcription factors. However, they orchestrate molecular recognition between ligase and transcription factors by inducing several protein–protein interactions proximal to the binding interface. Mechanistically similar to IMiDs, certain aryl sulfonamides around the clinically tested compound indisulam have recently

been shown to act as molecular glues between the CRL4<sup>DCAF15</sup> ligase and the splicing factor RBM39, again causing the targeted degradation of the latter<sup>12–17</sup>.

In both cases, the molecular glue mechanism of action therefore enables the destabilization of target proteins otherwise considered unligandable and thus outside the reach of both traditional small-molecule inhibitors and likely also of heterobifunctional degraders. However, the discovery of molecular glue degraders has thus far unfortunately been serendipitous, and a lack of rational identification strategies has stymied its general applicability.

To address this limitation, we here describe a scalable strategy toward glue degrader discovery by comparative chemical screening in hyponeedlylated cellular models with broadly abrogated ligase activity. In a proof of concept, this led to the identification of a new glue degrader that reprograms the CRL4<sup>DCAF15</sup> ligase. Of note, this strategy also enabled the discovery of three small molecules that induce the destabilization of cyclin K. Integrating orthogonal functional genomics approaches with proteomics, drug-affinity chromatography and biochemical reconstitution experiments revealed an unprecedented mechanism of action that converges on a direct, drug-induced proximity between CDK12–cyclin K and DDB1 that causes ubiquitination and proteasomal degradation of cyclin K. Of note, drug-induced recruitment of CDK12–cyclin K is functionally independent of a dedicated substrate receptor (SR), which functionally differentiates this mechanism from previously characterized glue degraders.

## Results

**Comparative drug profiling in hyponeedlylated cell lines.** Currently known molecular glue degraders function by chemically redirecting cullin-RING ligases (CRLs). CRLs make up the largest

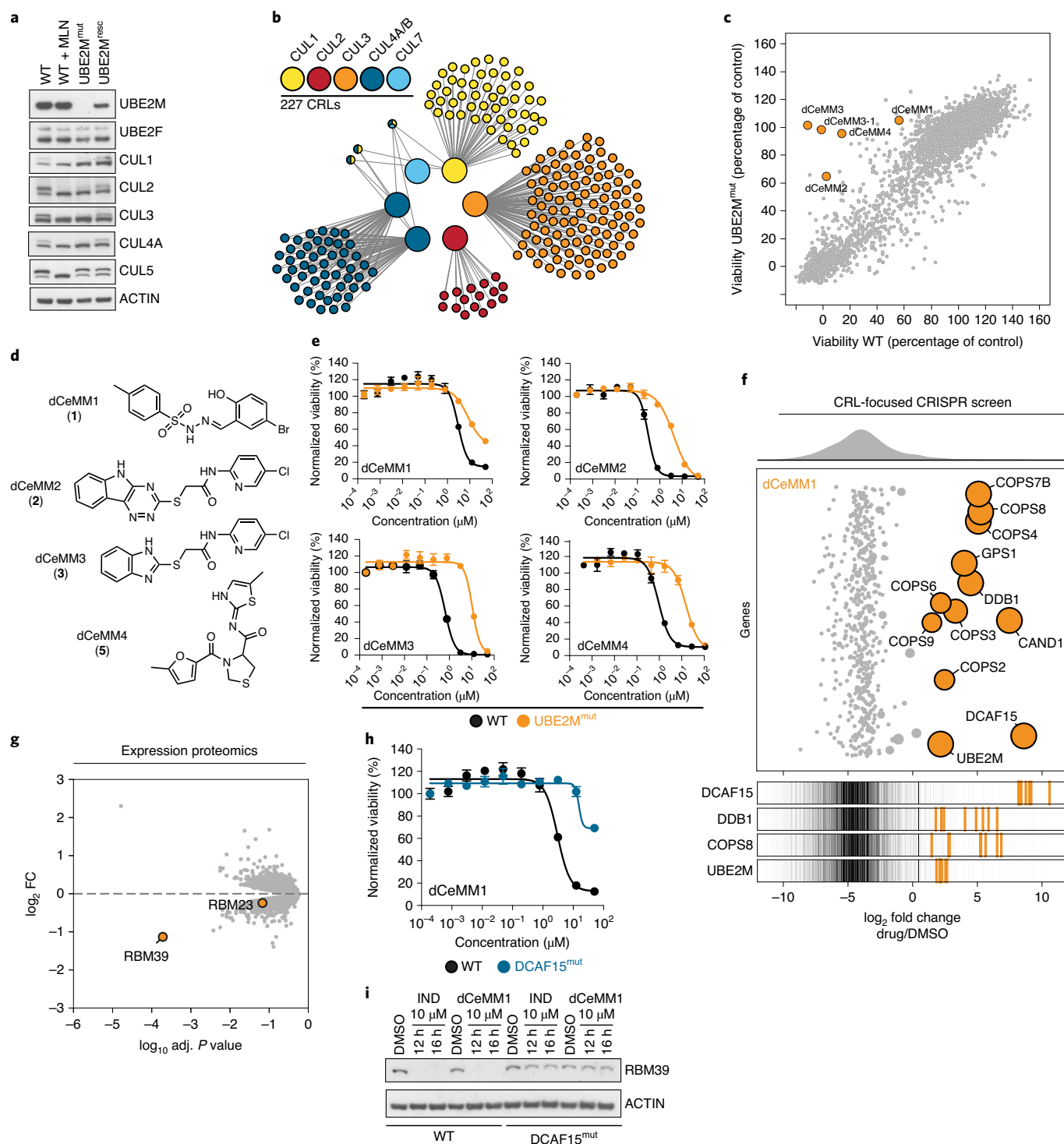
<sup>1</sup>CeMM Research Center for Molecular Medicine of the Austrian Academy of Sciences, Vienna, Austria. <sup>2</sup>FMI Friedrich Miescher Institute for Biomedical Research, Basel, Switzerland. <sup>3</sup>Faculty of Science, University of Basel, Basel, Switzerland. <sup>4</sup>Institute of Structural Biology, University of Bonn, Bonn, Germany. <sup>5</sup>Research Institute of Molecular Pathology, Vienna BioCenter, Vienna, Austria. <sup>6</sup>Department of Laboratory Medicine, Medical University of Vienna, Vienna, Austria. <sup>7</sup>These authors contributed equally: Sophie Bauer, Matthias Brand. ✉e-mail: [gwinter@cemm.oew.ac.at](mailto:gwinter@cemm.oew.ac.at)

family of E3 ubiquitin ligases and are diversified by modular assembly of SRs and adaptor proteins around different cullin backbones, leading to an overall complexity consisting of more than 250 CRLs. Small-molecule degraders typically function by rewiring CRL SRs such as CRBN or DCAF15. CRL hijacking is a strategy of evolutionary dimensions, given that it is also used by several viruses. In fact, certain viral proteins evolved to reprogram CRL activity independently of SRs by directly engaging other CRL components<sup>18</sup>. For instance, direct association of different viral proteins with the CRL4 adaptor protein DDB1 was shown to prompt the degradation of host proteins such as STAT transcription factors<sup>19–22</sup>.

Under physiological conditions, the activity of most CRLs is dependent on the reversible attachment of the ubiquitin-like protein NEDD8 to the respective cullin scaffold. NEDD8 conjugation is orchestrated by an E1 enzyme (NAE), two E2s (UBE2M and UBE2F) and several E3 enzymes. The neddylation E1 enzyme was identified in plants through a genetic screen of resistance to the molecular glue hormone auxin<sup>23</sup>. Similarly, we recently identified that CRISPR–Cas9-induced mutation of UBE2M causes resistance to known small-molecule degraders by inactivation of several CRLs including CRL4<sup>CRBN</sup> and CRL2<sup>VHL</sup> (ref. <sup>24</sup>). Here, we set out to characterize the hypomorphic phenotype of the mutated UBE2M allele in the near-haploid KBM7 cells. Neddylation levels of most cullins were abrogated comparably to pharmacologic NAE inhibition with the exception of cullin 5, which is known to be predominantly neddylation by UBE2F<sup>25,26</sup>. No compensatory upregulation of UBE2F was detected, and levels of cullin neddylation could be restored on genetic rescue via stable UBE2M expression (Fig. 1a). While pharmacologic NAE inhibition is toxic in KBM7 cells, UBE2M<sup>mut</sup> cells did not show a notable proliferative defect despite the promiscuous impairment of CRL activity affecting most of the expressed ligases (Fig. 1b and Extended Data Fig. 1a–c). We hypothesized that comparative chemical profiling in hyponeddylated versus neddylation-proficient cells would outline a strategy to screen for small molecules that are functionally linked to any of the >200 expressed CRLs. To this end, we screened a library of approximately 2,000 cytostatic/cytotoxic small molecules both in wildtype (WT) and UBE2M<sup>mut</sup> KBM7 cells, which led us to identify several chemical scaffolds that appeared to be functionally dependent on uninterrupted neddylation levels (Fig. 1c and Supplementary Table 1). Dose-ranging validation that also extended to isogenic UBE2M<sup>wt</sup> and UBE2M<sup>mut</sup> HCT116 cells led us to prioritize four chemical scaffolds, dCeMM1 (1), dCeMM2 (2), dCeMM3 (3) and the closely related dCeMM3-1 (4), as well as dCeMM4 (5), for further mechanistic validation (Fig. 1d,e and Extended Data Fig. 1d–f). Of note, all four compounds were devoid of any previous functional annotation. To investigate the mechanism of action of dCeMM1, we conducted a focused CRISPR–Cas9 screen using a custom-designed single-guide RNA library covering all known CRLs and associated regulators (Supplementary Table 2 and Extended Data Fig. 1g). This confirmed the requirement on UBE2M, but most notably revealed a critical dependence of dCeMM1 efficacy on the CRL4<sup>DCAF15</sup> ligase complex (Fig. 1f). This is in line with previous reports that certain serendipitously identified and structurally different sulfonamides can act as DCAF15 molecular glues capable of degrading the RRM containing proteins RBM39 and RBM23 (refs. <sup>12,13,15–17</sup>). Coupling cellular dCeMM1 treatment with quantitative proteomics via isobaric tagging revealed that dCeMM1 exclusively destabilizes RBM39 (Fig. 1g and Supplementary Table 3). Further corroborating the mechanism as DCAF15 glue degrader, CRISPR-induced frameshift mutations in DCAF15 rendered cells insensitive to dCeMM1 treatment and rescued RBM39 destabilization (Fig. 1h,i and Extended Data Fig. 1h). Collectively, this outlines that comparative drug profiling in hyponeddylated cellular models enables the identification of new molecular glue degraders that follow an established mechanism of action.

**Identification of structurally distinct cyclin K degraders.** To characterize dCeMM2/3/4, we again coupled cellular treatment with quantitative expression proteomics. We found that all compounds lead to a pronounced destabilization of cyclin K alongside a milder destabilization of both associated kinases CDK12 and CDK13 (Fig. 2a and Supplementary Table 3). No other CDKs or cyclins were notably destabilized (Extended Data Fig. 2a,b). We next validated cyclin K destabilization of all three scaffolds via time-resolved immunoblots. Near-complete cyclin K degradation was evident already after 2 h (Fig. 2b) whereas levels of CDK12/13 appeared not acutely affected, but only destabilized on prolonged drug incubation. Retained UBE2M dependency of CDK12/13 degradation, however, indicated that this delayed destabilization still occurs via a CRL-dependent mechanism that could be direct or indirect in nature (Extended Data Fig. 2c). Moreover, coupling cellular treatment with washout experiments revealed a reversible binding mode (Extended Data Fig. 2d). To develop a structure–activity relationship for the assayed scaffolds, we tested a total of 53 structurally related analogs in dose-ranging viability assays in KBM7 WT and UBE2M<sup>mut</sup> cells. This informed not only on sites amenable for derivatization of functionalized analogs, but also led us to identify structurally similar small molecules with no measurable activity that are thus ideally suited negative controls. Inactive analogs dCeMM2X (6), dCeMM3X (7) and dCeMM4X (8) fail to induce pronounced cyclin K destabilization and do not affect KBM7 cells in dose-ranging viability assays (Extended Data Fig. 2e–g). In line with a mechanism of targeted degradation, induced cyclin K destabilization was dependent on an active proteasome, could be rescued by pharmacologic NAE inhibition and by blocking the ubiquitin-activating enzyme UBA1 (Fig. 2c).

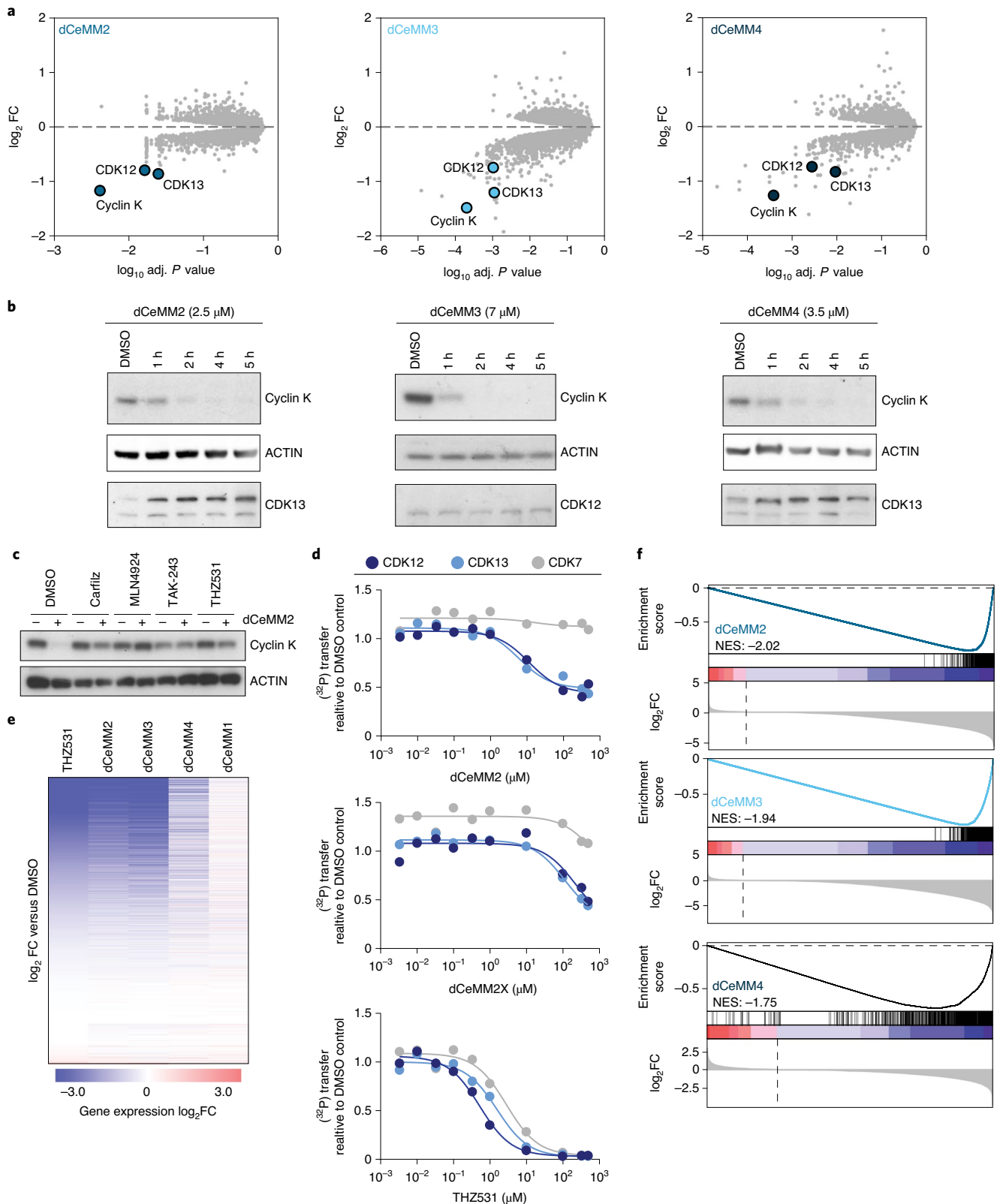
Of note, saturating the ATP binding pocket of CDK12/13 by treatment with the selective covalent CDK12/13 inhibitor THZ531 similarly rescued cyclin K destabilization<sup>27</sup>. This suggests that CDK12/13 active site engagement is required for the cyclin K destabilizing effect of the identified hit compounds (Fig. 2c). In line with this observation, recombinant kinase assays uncovered a measurable inhibitory activity of dCeMM2/3/4 on the enzymatic activity of CDK12/13 with a remarkable selectivity over CDK7. However, the tenfold reduced potency compared to THZ531 likely does not account for the strong cellular efficacy (Fig. 2d and Extended Data Fig. 3a). We next aimed to further corroborate CDK12/13–cyclin K as phenotypically relevant target of dCeMM2/3/4. To that end, we conducted RNA sequencing after 5 h of cellular treatment with all identified small molecules and compared the associated, cell-count normalized transcriptional footprint with transcriptional profiles obtained after selective CDK12/13 inhibition with THZ531. On association with cyclin K, CDK12 and CDK13 are known to phosphorylate the carboxyterminal domain of RNA Pol II, licensing it for processive transcription elongation on a genome-wide scale<sup>28</sup>. In line with a functional impairment of CDK12/13–cyclin K complexes, we observed a global transcriptional downregulation following treatment with dCeMM2/3/4, as well as profound induction of apoptosis, without a phase-specific cell-cycle arrest (Fig. 2e, Extended Data Fig. 3b,c and Supplementary Table 4). Gene Set Enrichment Analysis (GSEA) further validated a profound phenotypic similarity between cellular treatment with dCeMM2/3/4 and selective pharmacologic inhibition of CDK12/13 (Fig. 2f and Extended Data Fig. 4a–c). Proteomics profiling at a later timepoint (12 h) enabled functional enrichment analysis of the differentially expressed protein–protein interaction networks prompted by cyclin K degradation. In line with the proposed mechanism, ‘regulation of cell cycle’ and ‘RNA polymerase II CTD heptapeptide repeat kinase activity’ were the main gene ontology biological processes/molecular functions affected (Extended Data Fig. 4d,e and Supplementary Table 3). Collectively, intersection of quantitative proteomics and transcriptomics functionally validated cyclin K as a proteolytically



**Fig. 1 | Screening in hyponeddylated cells identifies new molecular glue degrader rewiring the CRL4<sup>DCAF15</sup> ligase. a**, Protein levels of UBE2M, UBE2F and neddylation status of different cullin backbones in WT, MLN4924-treated (1  $\mu\text{M}$ , 1 h), UBE2M-deficient and UBE2M-reconstituted KBM7 cells.

**b**, SRs expressed in KBM7 cells that associate with hyponeddylated cullin scaffolds. **c**, Primary screening data comparing DMSO-normalized viability of WT and UBE2M<sup>mut</sup> KBM7 cells treated with an (approximately) 2,000 cytotoxic/cytostatic small-molecule library for 3 d. Doses tested were 10  $\mu\text{M}$  and 500 nM. **d**, Chemical structure of the four prioritized chemical scaffolds (dCeMM1/2/3/4). dCeMM3-1 was structurally identical to dCeMM3 with a Br instead of the Cl. **e**, Dose-resolved, DMSO-normalized viability after 3-d dCeMM1/2/3/4 treatment in WT and UBE2M<sup>mut</sup> KBM7 cells. Mean  $\pm$  s.e.m.;  $n=3$  independent treatments. EC50s dCeMM1/2/3/4 ( $\mu\text{M}$ ) WT = 3, 0.3, 0.6 and 0.4 and UBE2M<sup>mut</sup> = 8, 4.2, 10.7 and 7. **f**, CRL-focused CRISPR resistance screen for dCeMM1. Top, bubble plot displaying median sgRNA enrichment over DMSO, bubble size indicates significance. Bottom, sgRNAs enrichment targeting indicated genes, background indicates distribution of all sgRNAs. Results shown are the median of two independent screens. **g**, Expression proteomics after dCeMM1 treatment (25  $\mu\text{M}$ , 12 h). **h**, DMSO-normalized viability after 3-d dCeMM1 treatment in WT and DCAF15<sup>mut</sup> KBM7 cells. Mean  $\pm$  s.e.m.;  $n=3$  independent treatments. **i**, RBM39 levels in WT and DCAF15<sup>mut</sup> KBM7 cells after indisulam (IND) or dCeMM1 treatment.





**Fig. 2** | dCeMM2/3/4 are new and structurally different cyclin K degraders. **a**, DMSO-normalized expression proteomics after 5 h dCeMM2/3/4 treatment (2.5, 7 and 3.5  $\mu$ M) in KBM7 cells. Limma statistical analysis was used. **b**, Cyclin K levels after dCeMM2/3/4 treatment in KBM7 cells. **c**, dCeMM2 (2.5  $\mu$ M, 5 h) destabilizes cyclin K. 30 min pretreatment with 1  $\mu$ M carfilzomib, 1  $\mu$ M MLN4924, 10  $\mu$ M TAK243 or 1  $\mu$ M THZ531 rescues cyclin K destabilization. **d**, Recombinant kinase assays of dCeMM2/2X and THZ531 inhibition on enzymatic activity of CDK12/13/7. Mean  $\pm$  s.d.,  $n=2$ . **e**, dCeMM2/3/4 induce global transcriptional downregulation with phenotypic similarity to CDK12/13 inhibition by THZ531. Heatmap displays DMSO-normalized  $\log_2$ FC in gene expression for 27,051 transcripts, ranked by THZ531  $\log_2$ FC. **f**, GSEA of 984 genes ( $\log_2$ FC  $< -4$  and adj.  $P < 0.05$ ) significantly downregulated after THZ531 treatment in comparison to dCeMM2/3/4 (FDR  $< 0.001$ ). GSEA preranked function was used (1,000 permutations).

destabilized target of dCeMM2/3/4. The convergence of all three scaffolds on the degradation of the same targets was unexpected given the structural differences between dCeMM2/3 and dCeMM4.

### Cyclin K degradation is mediated via CRL4B ligase complex.

Next, we wanted to understand the molecular mechanism of drug-induced cyclin K destabilization and in particular identify the involved ubiquitin ligase complex. To that end, KBM7 cells were again subjected to CRISPR–Cas9-induced mutagenesis using the aforementioned CRL-focused sgRNA library (Extended Data Figs. 5 and 6 and Supplementary Table 2). Mutagenized cell pools were selected via continuous exposure to dCeMM2/3/4 to identify loss of function (LOF) mutations that allow clonal outgrowth and thus to identify genes that are functionally required for the anti-proliferative effects of the assayed compounds. In line with the initial chemical profiling in hyponeddylated cells, we again identified UBE2M as hit in all tested conditions. Moreover, we identified mutations in three additional genes that constitute a partial CRL complex: the cullin scaffold CUL4B, the adaptor protein DDB1 and the E2 ubiquitin-conjugating enzyme UBE2G1. In addition to UBE2G1, known to extend ubiquitin chains, we also identified UBE2Z, an E2 conjugating enzyme known to prime ubiquitin chains, alongside the E1 ubiquitin-activating enzyme UBA6 (Extended Data Fig. 5a–c). It is worth noting that, compared to previous profiling of the new CRL4<sup>DCAF15</sup> modulator dCeMM1 with the identical library, each of these screens failed to identify a consensus CRL SR. To exclude involvement of an orphan SR, we turned to genome-scale CRISPR–Cas9 screens using a previously reported library<sup>29</sup>. We again found a pronounced enrichment of the top four genes of the CRL-focused sgRNA library (DDB1, CUL4B, UBE2M and UBE2G1), which was particularly evident in the most stringently, dCeMM3-selected screen (Fig. 3a and Extended Data Figs. 5b and 6e). No similarly enriched genes were identified that would indicate the involvement of a hitherto orphan SR. Hence, these results point to a new type of molecular glue degrader mechanism that is SR independent. Validating the pooled screening approach, targeted CRISPR-induced inactivation of CUL4B and UBE2G1 strongly abrogated drug-induced cyclin K degradation (Fig. 3b,c and Extended Data Fig. 7a). The essentiality of DDB1 prevented the isolation of individual LOF clones, but pooled DDB1 inactivation in KBM7 cells with doxycycline inducible Cas9 expression similarly abrogated the degradation of cyclin K (Fig. 3d). Inactivation of DDB1, UBE2G1 and UBE2M also strongly rescued cytotoxicity induced by dCeMM2/3/4, but not THZ531. This further corroborates a functional link between cyclin K degradation and the antiproliferative consequences of the tested compounds and provides another layer of evidence that the cellular efficacy of dCeMM2/3/4 is independent of their weak inhibitory activity in CDK12 kinase assays (Figs. 1e and 3e,f and Extended Data Fig. 7b–e). Finally, despite sharing >80% sequence identity, dCeMM2/3/4 appear to be selectively dependent on CUL4B over CUL4A, suggesting a unique situation of nonredundancy of these cullin scaffolds (Extended Data Fig. 7f–h). Of note, dCeMM1/2/3/4 showed no direct effect on CUL4B abundance, or on levels of the consensus CUL4B substrate H2A-K119ub1 (Extended Data Fig. 7i,j).

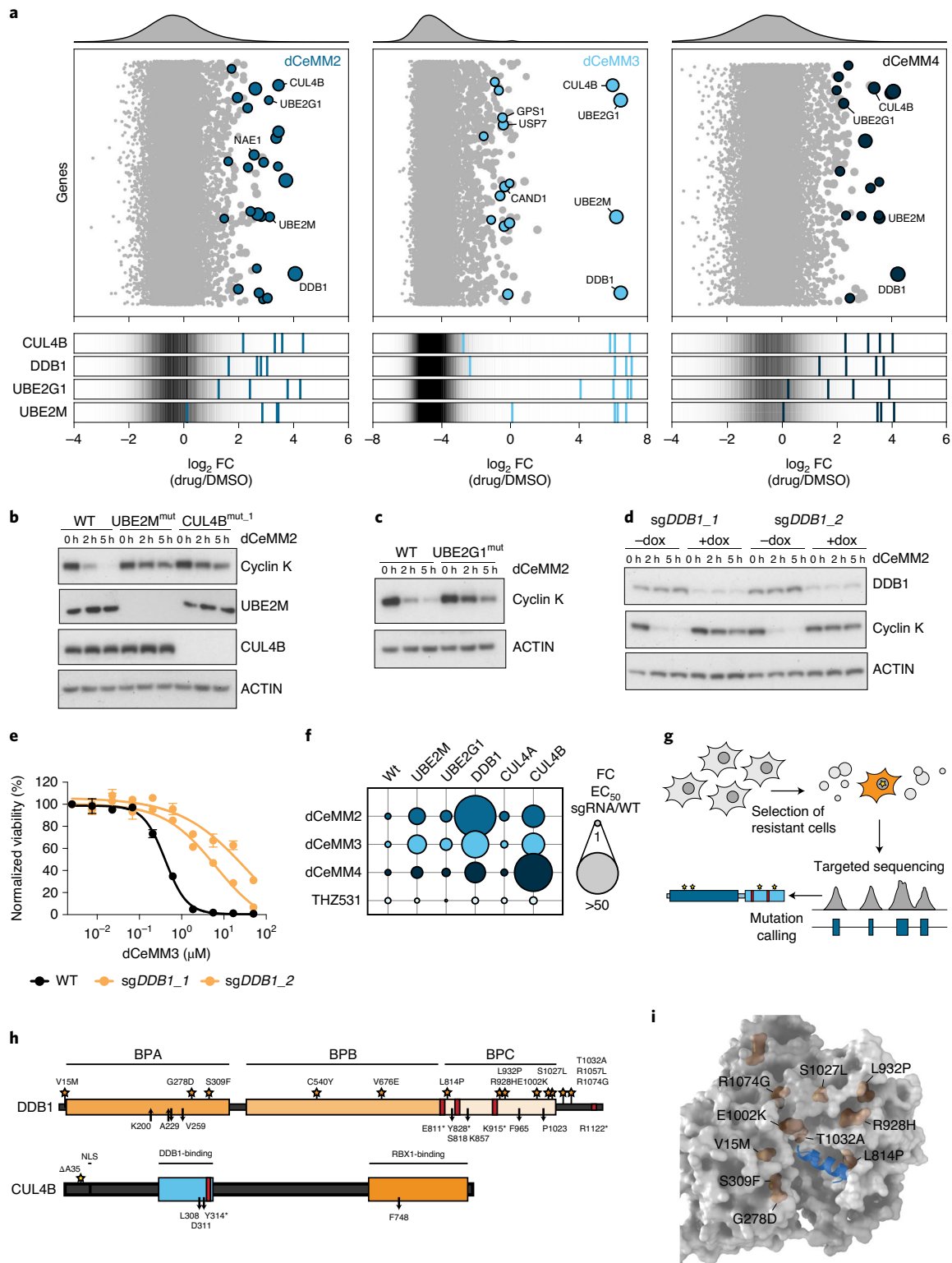
To gain further confidence in the identified genes, we exposed various cell types to excess concentrations of the identified hit compounds to isolate drug-resistant clones. To increase the mutational depth, KBM7 and MV4-11 cells were additionally engineered to carry full LOF mutations in the DNA mismatch repair protein MLH1 (Extended Data Fig. 8a,b). Drug-resistant subclones were isolated in batches and subjected to targeted next-generation sequencing of DDB1, CUL4A, CUL4B, UBE2M, UBE2G1, CDK12/13 and cyclin K via a hybrid-capture-based approach coupled to next-generation sequencing (Fig. 3g). Supporting the CRISPR screening data, we identified a profound enrichment of missense mutations or focal

deletions particularly in the BPA and BPC domains of DDB1 (Fig. 3h and Supplementary Table 5). Most identified mutations cluster inside a pocket formed between BPA and BPC that has previously been shown to engulf N-terminal helices of viral proteins such as the SV5V protein (Fig. 3i), thus highlighting it as a putative actionable site for induced protein–protein interactions<sup>21,22</sup>. Isolation of an HCT116 clone harboring the DDB1 L932P mutation allowed us to test effects on dCeMM2/3/4-induced cyclin K degradation. Indeed, cyclin K degradation was strongly reduced in the mutant background that was also characterized by a subtle DDB1 destabilization (Extended Data Fig. 8c). Additional mutations in dCeMM2/3/4-selected populations were also identified in the DDB1-binding domain of CUL4B (Fig. 3h and Supplementary Table 5). Finally, we identified point mutations in CDK12 and CDK13, albeit outside the kinase domain. Again, we managed to isolate a clone mutant for CDK13 (P1043H) with impaired cyclin K degradation following dCeMM2/3/4 treatment (Extended Data Fig. 8d–f and Supplementary Table 5). Collectively, orthogonal functional genomics profiling and targeted validation uncovered a critical involvement of a partial CRL4B complex in drug-induced cyclin K degradation. Again, the highly congruent chemogenomic profile of the three tested molecules confirms a unifying mechanism of action, despite their different chemical structures. A hotspot of spontaneously arising resistance mutations in a DDB1 pocket previously described as a critical interaction site with viral proteins establishes a rationale for an analogous dCeMM2/3/4-mediated association between CDK12/13–cyclin K and DDB1 that is independent of a dedicated SR.

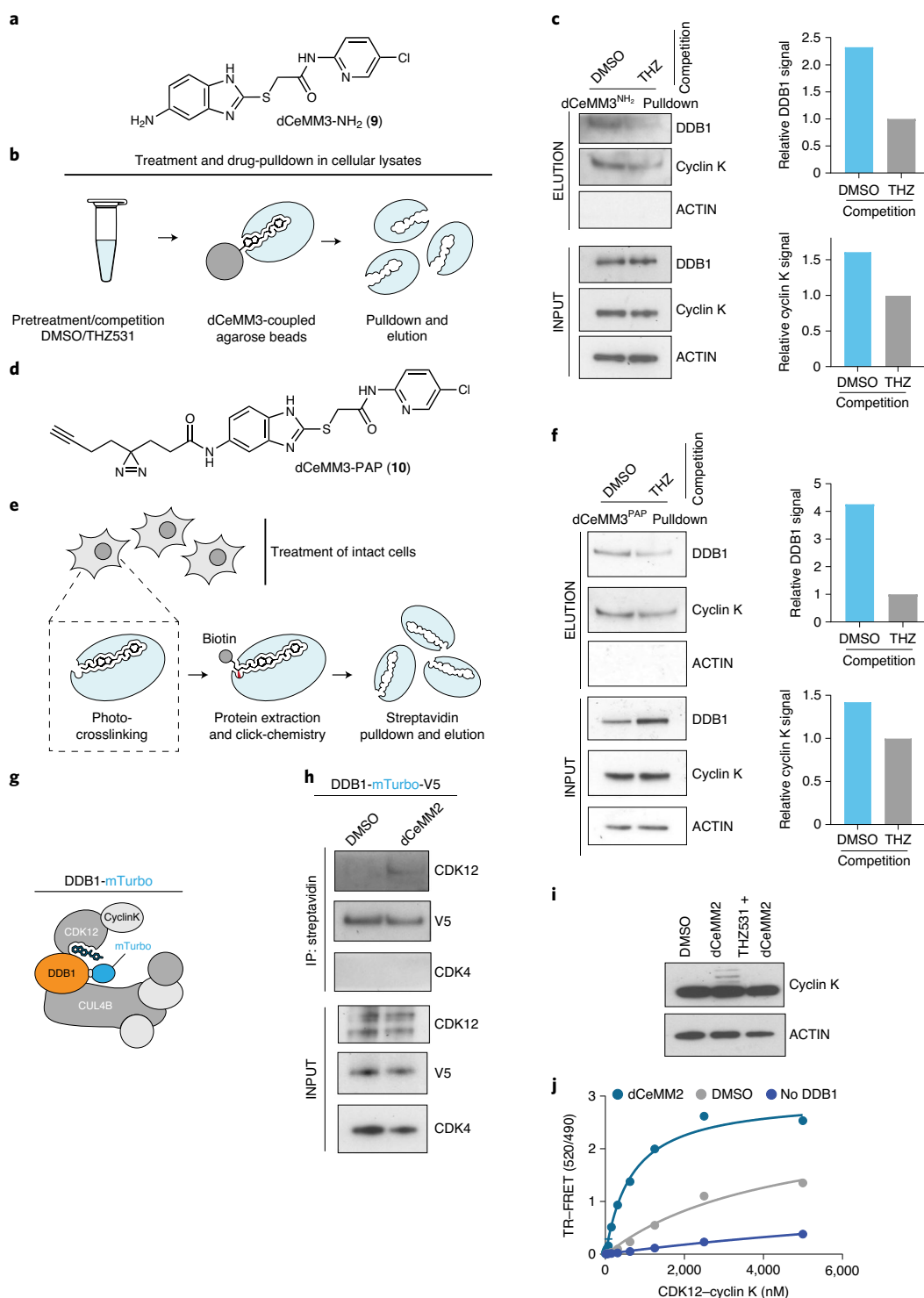
### Drug-induced dimerization between DDB1 and CDK12–cyclin K.

To validate if the drug-induced degradation of cyclin K via the CRL4B–DDB1 complex is mediated via direct drug engagement, we initially set out to perform drug-affinity chromatography using tethered analogs of dCeMM3. To that end, we devised two complementary strategies. First, we synthesized dCeMM3-NH2 (**9**) by installing a free amine on dCeMM3. This allowed for immobilizing dCeMM3-NH2 on sepharose beads, and purification of interacting proteins out of whole cell lysates (Fig. 4a,b). Coupling dCeMM3-NH2 pull-downs with immunoblotting revealed both cyclin K and DDB1 as interacting proteins (Fig. 4c). Treating lysates with THZ531 and thus covalently blocking the active sites on CDK12/13 prevented DDB1 and cyclin K enrichment via the dCeMM3-NH2 resin. This is in line with the fact that (1) cellular cotreatment with THZ531 was able to abrogate drug-induced cyclin K degradation (Fig. 2c), as well as (2) with a weak inhibitory activity of dCeMM2/3/4 in biochemical CDK12 activity assays (Fig. 2d). Next, we devised an alternative method for drug-target enrichment that is based on cellular treatment with dCeMM3-PAP (**10**), a functionalized analog of dCeMM3 containing a photoactive diazirine moiety and an alkyne handle (Fig. 4d). Following cellular treatment with dCeMM3-PAP, cells were ultravioletly crosslinked, lysed and the alkyne handle in dCeMM3-PAP was biotinylated for subsequent immobilization on streptavidin beads and identification of drug-interacting proteins via immunoblots. Again, DDB1 and cyclin K were enriched in the eluates, indicating that target engagement also occurs in intact cells and that this interaction is independent of postlysis artifacts (Fig. 4e,f). Also in this setup, covalently saturating the CDK12/13 binding site with THZ531 competed with dCeMM3-PAP enrichment (Fig. 4e,f). While a certain bias of compound derivatization cannot be excluded for tethering approaches, both strategies suggest simultaneous drug engagement with the destabilized cyclin K and the CRL4 adaptor protein DDB1 in a CDK12/13-dependent manner.

Next, we aimed to provide additional evidence of a direct, drug-induced association of CDK12–cyclin K and DDB1–CUL4B. Of note, under steady-state cellular conditions, cyclin K does not appear as a CUL4B substrate (Extended Data Fig. 9b). CRL



**Fig. 3 | Induced cyclin K degradation is mediated via a CRL4B ligase complex in a SR-independent manner. a**, Genome-wide CRISPR dCeMM2/3/4 resistance screens. Top, bubble plot displaying median sgRNA enrichment over DMSO, bubble size indicates significance. Bottom, sgRNA enrichment targeting indicated genes, background indicates distribution of all sgRNAs. **b–d**, dCeMM2-induced cyclin K degradation (2.5  $\mu$ M) is rescued in UBE2M-, CUL4B- (**b**), UBE2G1- (**c**) and DDB1-deficient cells (**d**). **e**, DMSO-normalized viability in WT and 3-d doxycycline (dox) pretreated *sgDDB1\_1* and *sgDDB1\_2* dox-inducible Cas9 KBM7 cells after 3-d dCeMM3 treatment. Mean  $\pm$  s.e.m.;  $n = 3$  independent treatments. **f**, WT-normalized FC in  $EC_{50}$  based after dose-resolved 3-d viability experiments on exposure to the indicated drugs in the indicated genetic backgrounds. **g**, Targeted hybrid-capture approach coupled to next-generation sequencing to identify mutations in spontaneously dCeMM2/3/4-resistant cells. **h**, Depiction of DDB1 and CUL4B mutations identified by hybrid-capture sequencing in drug-resistant cell pools. Stars indicate point mutations. Red bars indicate premature stop codons. Arrows indicate frameshift mutations. **i**, Structure of DDB1 (gray) in complex with SV5V peptide (amino acids 21–39, blue), PDB 2HYE. DDB1 point mutations located within 20 Å of the peptide are highlighted in orange.



**Fig. 4 | dCeMM2/3/4 induce proximity between CUL4B–DDB1 and CDK12/13–cyclin K.** **a**, dCeMM3-NH<sub>2</sub> chemical structure (dCeMM3 tethered analog). **b**, Drug-affinity chromatography strategy based on probe-coupled agarose beads pull-downs after DMSO or THZ531 (competition) pretreatment in lysates. **c**, Cyclin K and DDB1 enrichment in dCeMM3-NH<sub>2</sub>-based pull-downs. For quantification, eluted protein was normalized to protein available (input panels). THZ531-competed (100  $\mu$ M, 1 h) ratios were set to 1. **d**, Chemical structure of dCeMM3-PAP (PAP, photoaffinity probe). **e**, Drug-target enrichment strategy based on cellular dCeMM3-PAP cotreatment with DMSO or THZ531 (100  $\mu$ M, competition) after Carfilzomib pretreatment (10  $\mu$ M, 30 min). **f**, Cyclin K and DDB1 enrichment in dCeMM3-PAP-based pull-downs. For quantification, eluted protein was normalized to protein available (input panels). THZ531-competed (100  $\mu$ M, 1 h) ratios were set to 1. **g**, Proximity labeling strategy to assess drug-induced dimerization in intact cells based on the biotin ligase miniTurbo (mTurbo). **h**, Biotin-labeled CDK12 enrichment following 1 h DMSO or dCeMM2 treatment in the presence of carfilzomib (10  $\mu$ M) in human embryonic kidney cells transfected with DDB1-mTurbo fusion. **i**, Direct cyclin K immunoblotting of carfilzomib-pretreated KBM7 cells after 2 h DMSO, dCeMM2 or THZ531 + dCeMM2 (10  $\mu$ M) treatments. **j**, TR-FRET signal for CDK12-Alexa488-cyclin K (0–5  $\mu$ M) titrated to <sup>terbium</sup>-DDB1 in DMSO or 10  $\mu$ M dCeMM2. ‘No DDB1’ only contains streptavidin-terbium. Data are means  $\pm$  s.d. ( $n = 3$ ).  $K_{\text{apparent}}$  (nM): DMSO: not determined, ‘no DDB1’: not determined, dCeMM2 = 628.

decomposition after cellular lysis is known to challenge biochemical analysis but few approaches have been shown to overcome these problems<sup>30,31</sup>. We opted to devise a cellular strategy based on enzyme-catalyzed proximity labeling via the efficient biotin ligase miniTurbo (mTurbo)<sup>32</sup>. This assay enables recording dynamic changes in molecular proximity of an mTurbo-tagged bait protein by streptavidin purification of covalently biotinylated, proximal proteins. We transiently expressed C-terminally tagged DDB1 mTurbo-fusions in 293T cells. Cells were treated for 1 h with dCeMM2 or vehicle (DMSO) control, including a 30-min biotin labeling pulse (Fig. 4g,h). In line with a drug-induced molecular proximity, CDK12 interactions with DDB1 were only identified in the presence of dCeMM2, but not after DMSO treatment. (Fig. 4h). Consistently, treatment with dCeMM2 induced polyubiquitination of cyclin K in a CDK12/13-dependent manner (Fig. 4i and Supplementary Fig. 9b). To provide further evidence for a direct, drug-induced binding of CDK12–cyclin K to DDB1, we reconstituted this interaction using recombinant proteins. Drug-induced proximity between CDK12–cyclin K and DDB1 was detected using time-resolved fluorescence energy transfer (TR-FRET). While weak binding between DDB1 and CDK12–cyclin K was observed in the presence of vehicle control (DMSO), treatment with dCeMM2 facilitated an interaction of DDB1 and CDK12–cyclin K with a  $K_{\text{apparent}}$  of 628 nM (Fig. 4j and Extended Data Fig. 9c). A similar affinity was measured for the dCeMM4-induced interaction between CDK12–cyclin K and DDB1, while inactive analogs of both hit compounds failed to strengthen the interaction in a comparable manner (Extended Data Fig. 9d).

Collectively, orthogonal experimental strategies allowed us to validate drug-dependent induction of proximity between CDK12–cyclin K and DDB1, which functionally leads to cyclin K ubiquitination and degradation. Phenotypically, cyclin K degradation by dCeMM2/3/4 validated a previously established hypersensitivity of T-cell malignancies for genetic cyclin K ablation (Extended Data Fig. 9e–g)<sup>33</sup>. Moreover, we observed a correlation between cyclin K degradation and sensitivity, as well as a synergistic potential when combining cyclin K degradation with DNA damage inducing agents (Extended Data Figs. 9e–g and 10).

## Discussion

Molecular glue degraders represent an intriguing strategy to deliver on the promise of chemically degrading disease-relevant proteins that have not been amenable to traditional, inhibitor-centric pharmacologic approaches. A key limitation to the systematic exploration and evaluation of molecular glue degraders as therapeutic strategies has thus far been their serendipitous identification and a lack of rational discovery principles. To address this bottleneck, we here introduce differential chemical profiling in hyponeedylated cell lines that are characterized by a promiscuous defect in CRL activity. Lending proof of concept to this approach, we first identified dCeMM1, a glue degrader of RBM39 that functions by redirecting the activity of the CRL4<sup>DCAF15</sup> ligase. Our strategy also led to the identification of three small molecules (dCeMM2/3/4) that appeared to be functionally independent of a dedicated SR, and that we mechanistically characterized via a subsequent multi-omics target deconvolution campaign.

To that end, we performed a comprehensive intersection of quantitative proteomics and functional genomic strategies, drug-affinity chromatography and cellular proximity assays to reveal an unprecedented and therapeutically yet unexplored mechanism for molecular glue-induced target degradation. Binding of dCeMM2/3/4 to CDK12–cyclin K prompts a dimerization with a DDB1–CUL4B E3 complex. Biochemical reconstitution experiments suggest a baseline affinity between DDB1 and CDK12–cyclin K that is strongly increased on dCeMM2/3/4 addition. While future studies will address the underlying details from a structural perspective, this

observation mechanistically differentiates the described cyclin K glue degraders from the well-studied thalidomide analogs where no baseline affinity between ligase and neosubstrate is measurable in the absence of the respective small molecule<sup>34</sup>. dCeMM2/3/4-induced dimerization between CDK12–cyclin K and DDB1–CUL4B leads to the positioning of cyclin K in the ubiquitination zone of the associated E2 ubiquitin-conjugating enzymes. UBE2Z and UBE2G1 function as the ubiquitin priming and extending E2s, respectively, with UBA6 being the main E1 ubiquitin-activating enzyme. Consequently, cyclin K gets ubiquitinated and degraded by the proteasome via a mechanism that is independent of a dedicated CRL SR. Further research will explore the translational potential of targeted cyclin K degradation as single-agent therapy and in combination with DNA damage inducing agents.

In summary, the presented data provides a rational and scalable framework for the discovery of cellularly active molecular glue degraders with established and nonobvious mechanisms of action. Collectively, we expect this approach and derivatives that employ a different phenotypic readout to substantially increase the arsenal of molecular glue degraders. In contrast to the more commonly used heterobifunctional PROTACs, such molecules have the potential to induce the elimination of unligandable, disease-relevant proteins that are otherwise considered ‘undruggable’.

## Online content

Any Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41589-020-0594-x>.

Received: 5 May 2020; Accepted: 15 June 2020;

Published online: 3 August 2020

## References

- Winter, G. E. et al. DRUG DEVELOPMENT. Phthalimide conjugation as a strategy for in vivo target protein degradation. *Science* **348**, 1376–1381 (2015).
- Bondeson, D. P. et al. Catalytic in vivo protein knockdown by small-molecule PROTACs. *Nat. Chem. Biol.* **11**, 611–617 (2015).
- Zengerle, M., Chan, K. H. & Ciulli, A. Selective small molecule induced degradation of the BET bromodomain protein BRD4. *ACS Chem. Biol.* **10**, 1770–1777 (2015).
- Sakamoto, K. M. et al. Protacs: chimeric molecules that target proteins to the Skp1-Cullin-F box complex for ubiquitination and degradation. *Proc. Natl Acad. Sci. USA* **98**, 8554–8559 (2001).
- Schneekloth, J. S. Jr. et al. Chemical genetic control of protein levels: selective in vivo targeted degradation. *J. Am. Chem. Soc.* **126**, 3748–3754 (2004).
- Toure, M. & Crews, C. M. Small-Molecule PROTACs: new approaches to protein degradation. *Angew. Chem. Int. Ed. Engl.* **55**, 1966–1973 (2016).
- Ito, T. et al. Identification of a primary target of thalidomide teratogenicity. *Science* **327**, 1345–1350 (2010).
- Lu, G. et al. The myeloma drug lenalidomide promotes the cereblon-dependent destruction of Ikaros proteins. *Science* **343**, 305–309 (2014).
- Kronke, J. et al. Lenalidomide causes selective degradation of IKZF1 and IKZF3 in multiple myeloma cells. *Science* **343**, 301–305 (2014).
- Sievers, Q. L. et al. Defining the human C2H2 zinc finger degrome targeted by thalidomide analogs through CRBN. *Science* **362**, eaat0572 (2018).
- Gandhi, A. K. et al. Immunomodulatory agents lenalidomide and pomalidomide co-stimulate T cells by inducing degradation of T cell repressors Ikaros and Aiolos via modulation of the E3 ubiquitin ligase complex CRL4(CRBN.). *Br. J. Haematol.* **164**, 811–821 (2014).
- Han, T. et al. Anticancer sulfonamides target splicing by inducing RBM39 degradation via recruitment to DCAF15. *Science* **356**, eaal3755 (2017).
- Uehara, T. et al. Selective degradation of splicing factor CAPERalpha by anticancer sulfonamides. *Nat. Chem. Biol.* **13**, 675–680 (2017).
- Bussiere, D. E. et al. Structural basis of indisulam-mediated RBM39 recruitment to DCAF15 E3 ligase complex. *Nat. Chem. Biol.* **16**, 15–23 (2020).
- Ting, T. C. et al. Aryl sulfonamides degrade RBM39 and RBM23 by recruitment to CRL4-DCAF15. *Cell Rep.* **29**, 1499–1510.e6 (2019).
- Faust, T. B. et al. Structural complementarity facilitates E7820-mediated degradation of RBM39 by DCAF15. *Nat. Chem. Biol.* **16**, 7–14 (2020).

17. Du, X. et al. Structural basis and kinetic pathway of RBM39 recruitment to DCAF15 by a sulfonamide molecular glue E7820. *Structure* **27**, 1625–1633.e3 (2019).
18. Petroski, M. D. & Deshaies, R. J. Function and regulation of cullin-RING ubiquitin ligases. *Nat. Rev. Mol. Cell Biol.* **6**, 9–20 (2005).
19. Andrejeva, J., Poole, E., Young, D. F., Goodbourn, S. & Randall, R. E. The p127 subunit (DDB1) of the UV-DNA damage repair binding protein is essential for the targeted degradation of STAT1 by the V protein of the paramyxovirus simian virus 5. *J. Virol.* **76**, 11379–11386 (2002).
20. Lin, G. Y., Paterson, R. G., Richardson, C. D. & Lamb, R. A. The V protein of the paramyxovirus SV5 interacts with damage-specific DNA binding protein. *Virology* **249**, 189–200 (1998).
21. Li, T., Chen, X., Garbutt, K. C., Zhou, P. & Zheng, N. Structure of DDB1 in complex with a paramyxovirus V protein: viral hijack of a propeller cluster in ubiquitin ligase. *Cell* **124**, 105–117 (2006).
22. Li, T., Robert, E. I., van Breugel, P. C., Strubin, M. & Zheng, N. A promiscuous alpha-helical motif anchors viral hijackers and substrate receptors to the CUL4-DDB1 ubiquitin ligase machinery. *Nat. Struct. Mol. Biol.* **17**, 105–111 (2010).
23. Leyser, H. M. et al. Arabidopsis auxin-resistance gene AXR1 encodes a protein related to ubiquitin-activating enzyme E1. *Nature* **364**, 161–164 (1993).
24. Mayor-Ruiz, C. et al. Plasticity of the cullin-RING ligase repertoire shapes sensitivity to ligand-induced protein degradation. *Mol. Cell* **75**, 849–858.e8 (2019).
25. Wei, D. & Sun, Y. Small RING finger proteins RBX1 and RBX2 of SCF E3 ubiquitin ligases: the role in cancer and as cancer targets. *Genes Cancer* **1**, 700–707 (2010).
26. Huang, D. T. et al. E2-RING expansion of the NEDD8 cascade confers specificity to cullin modification. *Mol. Cell* **33**, 483–495 (2009).
27. Zhang, T. et al. Covalent targeting of remote cysteine residues to develop CDK12 and CDK13 inhibitors. *Nat. Chem. Biol.* **12**, 876–884 (2016).
28. Greenleaf, A. L. Human CDK12 and CDK13, multi-tasking CTD kinases for the new millennium. *Transcription* **10**, 91–110 (2019).
29. Doench, J. G. et al. Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. *Nat. Biotechnol.* **34**, 184–191 (2016).
30. Reitsma, J. M. et al. Composition and regulation of the cellular repertoire of SCF ubiquitin ligases. *Cell* **171**, 1326–1339.e14 (2017).
31. Reichermeier, K. M. et al. PIKES analysis reveals response to degraders and key regulatory mechanisms of the CRL4 network. *Mol. Cell* **77**, 1092–1106.e9 (2020).
32. Branon, T. C. et al. Efficient proximity labeling in living cells and organisms with TurboID. *Nat. Biotechnol.* **36**, 880–887 (2018).
33. Meyers, R. M. et al. Computational correction of copy number effect improves specificity of CRISPR-Cas9 essentiality screens in cancer cells. *Nat. Genet.* **49**, 1779–1784 (2017).
34. Chamberlain, P. P. & Cathers, B. E. Cereblon modulators: low molecular weight inducers of protein degradation. *Drug Discov. Today Technol.* **31**, 29–34 (2019).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature America, Inc. 2020

## Methods

**Cell lines and generation of mutants.** KBM7 cells with the specified genetic backgrounds were grown in IMDM supplemented with 10% FBS and 1% penicillin/streptomycin (pen/strep). HCT116 and 293T cells were grown in DMEM 10% FBS and 1% pen/strep. MV4;11, MOLT4, Loucy, JURKAT, CCRFCEM, KOPTK1, DND41, CUTTL1, HSB2, HBPALL, MOLT16, SKW3 and SUPT11 were grown in RPMI 10% FBS and 1% pen/strep. KBM7 and HCT116 cells expressing Cas9 were generated using Lenti\_Cas9\_Blasti (Addgene 52962). LentiGuide-Puro (Addgene 52963) was used to express sgRNAs against *UBE2M*, *DCAF15* and *UBE2G1*. LentiGuide-Puro-IRES-mCherry (modified from Addgene 52963) was used to express sgRNAs against *CUL4A* and *DDB1*, lentiCRISPRv2 (Addgene 52961) against *CUL4B* and lentiCRISPR2v2GFP (Addgene no. 82416) against *MLH1*. Lenti-PGK-Hygro-DEST-UBE2M was generated by Gateway cloning (*UBE2M* cDNA from pDEST17-Ube2M Addgene 15798 was cloned into pDONR221 by BP reaction, and then transferred to the destination vector Addgene 19066 by LR reaction), and used in *UBE2M*<sup>mut</sup> cells to generate *UBE2M*<sup>resc</sup> KBM7 cells. sgRNA sequences, plasmids and primers used to confirm single mutants are listed in Supplementary Table 6.

**Reagents.** Key reagents used were dCeMM1 (Enamine Z2065717981, purity 97%), dCeMM2 (ChemDiv 8006-2691, purity 90%), dCeMM3 (ChemSpace PB54687009, purity 99%), dCeMM4 (Enamine Z1126858802, purity 91%) and THZ531 (MedChemExpress HY-103618, purity 99.49%).

Other reagents were Carfilzomib (Selleckchem, S2853), MLN4924 (Selleckchem, S7109), TAK243 (MedChemExpress, HY-100487), indisulam (Sigma-Aldrich, SML1225) and doxycycline (VWR, A2951.0025).

**Phenotypic screen based on differential viability.** WT and *UBE2M*<sup>mut</sup> KBM7 cells were treated with 2,000 cytostatic/cytotoxic small molecules, mostly of unknown function or target, from the PLACEBO in-house collection. Briefly, the screening was divided in three parts: (1) primary screening, (2) follow up and (3) validation. During the primary screening, cells were treated with 10  $\mu$ M and 500 nM of every compound. DMSO was used as a negative control, YM155 1  $\mu$ M was used as a cytotoxic positive control, and the degrader ARV-771 (MedChemExpress, HY-100972) was used as a differential viability positive control at 50 and 250 nM (WT cells are sensitive and *UBE2M*<sup>mut</sup> cells are resistant). After 3 d of treatment, cell viability was measured (CellTiter Glo, Promega). The screen was performed in 384-well plates (Corning 3712) with the compound dispensing system Echo 550 and for cell seeding the Liquid Handler Multidrop Combi. Control compounds were included in several wells, in all the plates. We considered as hits those small molecules that inhibited the proliferation of WT cells, while sparing *UBE2M*<sup>mut</sup> isogenic counterparts. More than 20 compounds fulfilled this criterion and were followed up by testing the dose-ranging differential effect. dCeMM1, dCeMM2, dCeMM3 and dCeMM4 were selected for further validation due to fold change (FC) significance.

**Cell viability assays.** KBM7<sup>WT</sup>, mutant KBM7 clones (*UBE2M*<sup>mut</sup>, *UBE2M*<sup>resc</sup>, *UBE2G1*<sup>mut</sup> *CUL4A*<sup>mut</sup>, *CUL4B*<sup>mut</sup>) and 3-d doxycycline pretreated KBM7<sup>Cas9</sup> sg*DDDB1* cells were seeded at a cell density of 50,000 cells per ml in 96-well plates with DMSO or drug, in triplicates. Same conditions were used for viability assays with HCT116 and leukemia cell lines. Drugs used: dCeMM1, dCeMM2 2 $\times$ , dCeMM3 3 $\times$ , dCeMM4 4 $\times$  or THZ531. Cells were treated for 3 d, after which cell viability was assessed according to the manufacturer's protocol (CellTiter Glo, Promega G7570). Survival curves and half-maximum effective concentration (EC<sub>50</sub>) values were calculated by best-fit analysis of the log<sub>10</sub> drug concentration to FC of drug-treated cells over DMSO-treated cells. All survival assays included technical triplicates per sample, per experiment.

**Western blot analysis.** PBS-washed cell pellets were lysed in 50 mM Tris pH 7.9, 8 M Urea and 1% CHAPS and incubated with shaking at 4 °C for at least 30 min. Then 20  $\mu$ g of supernatants were run and transferred for detection. Antibodies used were CUL1 (1:1,000, Santa Cruz Biotechnology, sc-1276), CUL2 (1:500, Sigma-Aldrich, SAB2501565-100), CUL3 (1:1,000, Cell Signaling Technology, 2759), CUL4A (1:1,000, Cell Signaling Technology, 2699S), CUL4B (1:1,000, Proteintech, 12916-1-AP), CUL5 (1:1,000, Santa Cruz Biotechnology, sc-373822), UBE2M (1:1,000, Santa Cruz Biotechnology, sc-390064), DDB1 (1:1,000, Cell Signaling Technology, 5428S), CCNK (1:5,000, Bethyl, A301-939A), CDK12 (1:1,000, Cell Signaling Technology, 11973S), CDK13 (1:1,000, Bethyl, A301-458A), RBM39 (1:500, Santa Cruz Biotechnology sc-376531), V5 (1:1,000 Cell Signaling Technology, 13202), Ubiquityl-Histone H2A (K119) (1:1,000, Cell Signaling, 8240-20), MLH1 (1:1,000, Cell Signaling Technology, 3515T), ACTIN (1:10,000, Sigma-Aldrich, A5441) and VINCULIN (1:1,000, Santa Cruz Biotechnology, sc-25336). Secondary antibodies used were antimouse/-rabbit/-goat (1:10,000, Jackson ImmunoResearch 115-035-003, 111-035-003 and 705-035-003) and antirabbit Alexa Fluor 488 (1:1,000, Invitrogen, A21441).

**Apoptosis induction: annexinV/propidium iodide.** KBM7 cells were treated with DMSO, dCeMM2 (2.5  $\mu$ M), dCeMM3 (7  $\mu$ M) or dCeMM4 (3.5  $\mu$ M) for 4, 8 and 12 h. To assess apoptosis induction AnnexinV/PI staining was used (BD Bioscience

556547). Then, 5  $\times$  10<sup>5</sup> cells were collected, pelleted by centrifugation, washed with PBS and resuspended in 1 $\times$  binding buffer at a concentration of 10<sup>6</sup> cells per ml, preparing a sufficient volume to have 100  $\mu$ l per sample. Next, 5  $\mu$ l of staining solution was added per sample and incubated for 20 min at room temperature in the dark. Then, 400  $\mu$ l of 1 $\times$  binding buffer was added and cells were analyzed (within 1 h) by flow cytometry. Samples were analyzed on an LSR Fortessa (BD Biosciences) and data analysis was performed using FlowJo v\_10.6.1.

**Cell-cycle analysis by propidium iodide staining.** Two million KBM7 cells were treated with DMSO, dCeMM2 (2.5  $\mu$ M), dCeMM3 (7  $\mu$ M) or dCeMM4 (3.5  $\mu$ M) for the indicated times and fixed in cold 70% ethanol for 30 min at 4 °C. After addition of RNase (100  $\mu$ g ml<sup>-1</sup>) and propidium iodide (50  $\mu$ g ml<sup>-1</sup>), cells were incubated for 1 h at 37 °C before checking cell-cycle profiles by flow cytometry. Samples were analyzed on an LSR Fortessa (BD Biosciences) and data analysis was performed using FlowJo v\_10.6.1.

**Expression proteomics.** We compared overall proteome-wide changes in KBM7WT cells treated with dCeMM1 (25  $\mu$ M), dCeMM2 (2.5  $\mu$ M), dCeMM3 (7  $\mu$ M) or dCeMM4 (3.5  $\mu$ M) for 5 and 12 h, using quantitative proteomics based on isobaric tagging.

**Sample preparation.** Fifty million KBM7WT cells in duplicates per condition were collected, washed four times with ice-cold PBS, the supernatant aspirated and pellets snap frozen in liquid nitrogen. Each washed cell pellet was lysed separately in 40  $\mu$ l lysis buffer as previously described<sup>24</sup>. Offline Fractionation via reversed phase–high-performance liquid chromatography at high pH and two-dimensional–tandem reverse-phase liquid chromatography mass spectrometry were performed as previously described<sup>24</sup>.

**Data analysis.** Acquired raw data files were processed using the Proteome Discoverer v.2.2.0.388 platform, using the Sequest HT database search engine and Percolator validation software node (v.3.04) to remove false positives with a false discovery rate (FDR) of 1% on peptide and protein level under strict conditions. Searches were performed with full tryptic digestion against the human SwissProt database v.2017.06.06 (20,456 sequences and appended known contaminants) with a maximum of two allowed miscleavage sites. Oxidation (+15.994 Da) of methionine and deamidation (+0.984 Da) of glutamine and asparagine were set as variable modification, while carbamidomethylation (+57.021 Da) of cysteine residues and tandem mass tag (TMT) 10-plex labeling of peptide N termini and lysine residues were set as fixed modifications. Data was searched with mass tolerances of  $\pm$ 10 ppm and  $\pm$ 0.025 Da on the precursor and fragment ions, respectively. Results were filtered to include peptide spectrum matches with Sequest HT cross-correlation factor (Xcorr) scores of  $\geq$ 1 and high peptide confidence assigned by Percolator. MS2 signal-to-noise values (S/N) values of TMT-reporter ions were used to calculate peptide/protein abundance values. Peptide spectrum matches with precursor isolation interference values of  $\geq$ 50% and average TMT-reporter ion S/N  $\leq$  10 were excluded from quantitation. Only unique peptides were used for TMT quantitation. Isotopic impurity correction was applied and results exported for analysis for differential expression using limma R package<sup>35</sup>. Before performing limma analysis all proteins lacking reporter ion intensity values were removed and all protein TMT-reporter channels normalized to equal median abundance. TMT ratios with P/Q values (after FDR correction) lower than 0.01 were considered to be significant. For the calling of destabilized proteins, a log<sub>2</sub>FC threshold (drug/DMSO) of -0.3 was applied.

**CRISPR–Cas9 resistance screens (CRL-focused sgRNA library and genome-scale Brunello sgRNA library).** Design of the CRL-focused sgRNA library. CRL-focused library sgRNAs are listed in Supplementary Table 2.

**Lentivirus production.** Here, 293T cells seeded on 15 cm culture plates 16 h before were transfected with 5  $\mu$ g Brunello pooled library (Addgene 73178, two-vector system), 2.5  $\mu$ g pMD2.G (Addgene 12259) and 3.75  $\mu$ g psPAX2 (Brunello sgRNA library, Addgene 12260) or CRL sgRNA library, using PEI (Polysciences, 24765-1). Viral supernatant was harvested 72 h after transfection and concentrated using Lenti-X-concentrator (Takara). Concentrated viral supernatant was stored in aliquots at -80 °C and titrated following a standard protocol<sup>29</sup> to achieve a multiplicity of infection (MOI) of 0.2–0.3.

**CRL sgRNA library screens.** Twelve million KBM7<sup>Cas9</sup> or KBM7<sup>Cas9</sup> cells were transduced at MOI 0.3, yielding a calculated library representation of 635 cells per sgRNA (library representation, 2.5 million cells). For transduction, 100  $\mu$ l of concentrated viral supernatant was added to three million cells in 3 ml IMDM and 8  $\mu$ g ml<sup>-1</sup> polybrene in a 12-well plate. Plate was centrifuged at 2,000 r.p.m. for 1 h at 30 °C in a benchtop centrifuge and then incubated at 37 °C overnight. Transduction efficiency was titrated following a standard protocol<sup>29</sup>. Pools were selected with 1  $\mu$ g ml<sup>-1</sup> puromycin for 5 d (KBM7Cas9) or 500  $\mu$ g ml<sup>-1</sup> neomycin for 8 d (iCas9 cells, followed by 5 d of cas9 expression by doxycycline 0.5  $\mu$ g ml<sup>-1</sup>). Independent resistance screens were performed with both mutant libraries in duplicates, using drugs at starting concentrations of 4 $\times$  EC50: dCeMM1 9  $\mu$ M, dCeMM2 0.9  $\mu$ M,

dCeMM3 2.8  $\mu\text{M}$  and dCeMM4 1.25  $\mu\text{M}$  and a respective DMSO control. Every 4 d, cells were counted and reseeded to 2.5 million cells in 5 ml, applying fresh drug. Drug concentrations were dynamically adjusted to the growth curves to yield a consistent impact on cell proliferation. Drug-resistant pools were harvested after 14 d of treatment, snap frozen in liquid nitrogen and stored at  $-80^{\circ}\text{C}$ .

**Brunello pooled library screens.** Here, 250 million KBM7<sup>Cas9</sup> cells or were transduced at an MOI 0.23, yielding a calculated library representation of 668 cells per sgRNA (library representation, 50 million cells). For transduction, 20  $\mu\text{l}$  of concentrated viral supernatant was added to 5 million cells in 1.5 ml IMDM and 8  $\mu\text{g ml}^{-1}$  polybrene in six-well plates. Plates were centrifuged at 2,000 r.p.m. for 1 h at  $30^{\circ}\text{C}$  in a benchtop centrifuge, 0.5 ml IMDM were added and then incubated at  $37^{\circ}\text{C}$  overnight. The next day, transduced cells were pooled and diluted. Pools were selected with 1  $\mu\text{g ml}^{-1}$  puromycin for 5 d. Independent resistance screens were performed with the library using drugs at starting concentrations of 4 $\times$  EC50: dCeMM2 0.9  $\mu\text{M}$ , dCeMM3 2.8  $\mu\text{M}$  and dCeMM4 1.25  $\mu\text{M}$ . Selective drug treatment was performed on 50 million cells per drug at a seeding density of 500,000 cells per ml. Every 5 d, cells were pooled, counted and reseeded to 50 million cells in 100 ml, applying fresh drug. Drug concentrations were dynamically adjusted to the growth curves to yield a consistent impact on cell proliferation. Drug-resistant pools were harvested after 15 d of treatment, snap frozen in liquid nitrogen and stored at  $-80^{\circ}\text{C}$ .

**Library preparation for next-generation sequencing.** Genomic DNA (gDNA) was extracted from cell-frozen pellets with the QIAamp DNA Mini Kit (Qiagen, 51304). PCR on the gDNA templates was done to amplify sgRNA sequences. Primer sequences used available in Supplementary Table 7.

For the screens performed with the CRL library: the isolated gDNA was processed in parallel to yield 5  $\mu\text{g}$  gDNA per 50  $\mu\text{l}$  reaction for the first PCR. One PCR reaction contained 0.75  $\mu\text{l}$  of ExTaq polymerase (Clontech), 2.5  $\mu\text{l}$  10 $\times$  buffer, 4  $\mu\text{l}$  of dNTP mix, 2.5  $\mu\text{l}$  of 10  $\mu\text{M}$  P7 forward primer mix, 2.5  $\mu\text{l}$  of 10  $\mu\text{M}$  condition-specific P5 barcoded primer and water to reach 50  $\mu\text{l}$ . Target amplification was achieved by using: 1 min at  $95^{\circ}\text{C}$  initial denaturation; 30 s at  $95^{\circ}\text{C}$ , 30 s at  $53^{\circ}\text{C}$ , 30 s at  $72^{\circ}\text{C}$ , for 22–24 cycles; 10 min at  $72^{\circ}\text{C}$  final elongation. Specific amplification of the 300 base-pair target was confirmed by agarose gel electrophoresis. PCRs were purified using AMPure XP beads in a 1:1 ratio, following standard protocols. Amounts of 10 ng per condition were used for the second PCR, using the same PCR conditions as for the first PCR but using the primers Rev2\_p5 and Fwd2\_p7 primers and four cycles.

For the Brunello screens, a published protocol was followed<sup>29</sup>. The isolated gDNA was processed in parallel to yield 10  $\mu\text{g}$  gDNA per 100  $\mu\text{l}$  reaction. One PCR reaction contained 1.5  $\mu\text{l}$  of ExTaq polymerase (Clontech), 10  $\mu\text{l}$  10 $\times$  buffer, 8  $\mu\text{l}$  of dNTP mix, 0.5  $\mu\text{l}$  of 100  $\mu\text{M}$  P5 forward primer mix, 10  $\mu\text{l}$  of 5  $\mu\text{M}$  condition-specific P7 barcoded primer and water to reach 100  $\mu\text{l}$ . DNA oligo primers were ordered PAGE purified from Sigma-Aldrich. Target amplification was achieved by using: 1 min at  $95^{\circ}\text{C}$  initial denaturation; 30 s at  $95^{\circ}\text{C}$ , 30 s at  $53^{\circ}\text{C}$ , 30 s at  $72^{\circ}\text{C}$  for 26–27 cycles; 10 min at  $72^{\circ}\text{C}$  final elongation. Specific amplification of the 360-bp target was confirmed by agarose gel electrophoresis. All PCR reactions of a respective condition were pooled and 100  $\mu\text{l}$  were purified using AMPure XP beads in a 1:1 ratio, following standard protocols. Purified amplicon was eluted using 50  $\mu\text{l}$  TE buffer. Final sequencing libraries were pooled in equimolar amounts and sequenced on a HiSeq 3000/4000.

**Next-generation sequencing data analysis.** Raw read files were converted to fastq format using the convert function from bamtools (v.2.5.0)<sup>36</sup>. Sequencing adapters were trimmed using cutadapt (v.2.7.12) with -g CGAAACACCG and --minimum-length = 10 (genome-wide screens only). The 20 bp of spacer sequence was then extracted using fastx toolkit (v.0.0.14) ([http://hannonlab.cshl.edu/fastx\\_toolkit/](http://hannonlab.cshl.edu/fastx_toolkit/)) and aligned to the respective sgRNA index using bowtie2 (v.2.2.4)<sup>37</sup> allowing for one mismatch in the seed sequence. Spacers were counted using the bash command 'cut -f3 (0) | sort | uniq -c' on the sorted SAM files. A count table with all conditions was then assembled, and the counts +1 were converted to counts per-million to normalize for sequencing depth.  $\log_2$ -normalized FCs compared to DMSO were calculated for each spacer. Statistical analysis was performed using the STARS algorithm v.1.3 (ref. <sup>29</sup>). For this, spacers were rank-ordered based on  $\log_2\text{FC}$  and tested with the parameters -thr 10 -dir P against a null hypothesis of 10,000 random permutations. Genes with a  $q$  value of  $<0.05$  were called hits.

**RNA extraction, synthetic RNA spike-in and RNA sequencing.** Ten million KBM7 cells were treated in triplicate for 5 h with DMSO, dCeMM1 (25  $\mu\text{M}$ ), dCeMM2 (2.5  $\mu\text{M}$ ), dCeMM3 (7  $\mu\text{M}$ ), dCeMM4 (3.5  $\mu\text{M}$ ) or THZ531 (600 nM). Total RNA was extracted with the RNeasy mini kit (Qiagen 74106), and to account for global differences in transcriptional output, we employed an exogenous spike-in approach, according to the manufacturer's protocol (SIRV-Set3 Iso Mix E0/ERCC from Lexogen, 051). RNA amount was determined using Qubit RNA HS Kit (Thermo, Q32852). Then, 1.5  $\mu\text{g}$  RNA per condition was subjected to Poly(A) enrichment (Lexogen no. 039). RNA-seq library prep was performed according to the manufacturer's protocol (Corall Total RNA-Seq Library Prep Kit, Lexogen

095). Next, 2 ng RNA were used as starting material. Before the endpoint PCR, a quantitative PCR was done (PCR Add-on Kit for Illumina, Lexogen 020) to determine the optimal number of PCR cycles (16 cycles). Final sequencing libraries were analyzed on an Agilent 2100 Bioanalyzer, pooled in equimolar amounts and sequenced on a HiSeq 3000/4000.

**RNA sequencing data analysis.** Next-generation sequencing libraries were sequenced by the Biomedical Sequencing Facility at CeMM using the Illumina HiSeq 3000 platform and 50 bp single-end configuration. Raw reads were trimmed using Trimmomatic (v.0.32)<sup>38</sup>. The following parameters were used: HEADCROP:13 ILLUMINACLIP:epignome\_adapters\_2\_add.fa:2:10:4:1:true SLIDINGWINDOW:4:1 MAXINFO:16:0.40 MINLEN:18. Trimmed reads were mapped to the hg38/GRCh38 assembly of the human reference genome and to SIRV-Set 3 sequences using STAR aligner v.2.5.2b (ref. <sup>39</sup>) using the following parameters: --outFilterType BySJout --outFilterMultimapNmax 20 --alignSJoverhangMin 8 --alignSJDBoverhangMin 1 --outFilterMismatchNmax 999 --outFilterMismatchNoverLmax 0.6 --alignIntronMin 20 --alignIntronMax 1000000 --alignMatesGapMax 1000000 --outSAMattributes NH HI NM MD --outSAMtype BAM SortedByCoordinate.

Genes were defined based on genome-build GRCh38.p7, genome-build-accession NCBI:GCA\_000001405.22. Sequences and annotations in FASTA and GTF format for SIRV-Set 3 (SIRV Isoform Mix E0 and ERCCs, catalog no. 051) included in SIRV\_Set3\_Sequences\_170612a (ZIP) were downloaded from <https://www.lexogen.com/sirvs/download/>.

Read counts per gene were obtained from the aligned reads using the htseq-count command (v.0.11.2) from the HTSeq framework<sup>40</sup>. The Bioconductor/R packages edgeR, affy and limma were used for normalization and differential gene expression analysis.

GSEA of genes significantly downregulated on THZ531 treatment ( $\log_2\text{FC} < -4$  and adj.  $P < 0.05$ ) in comparison to dCeMM2/3/4 treatments was performed using GSEA v.3.0. To build the gene rankings, genes were ranked based on their expression change (dCeMM2/3/4 treatment versus DMSO), using the  $\log_2\text{FC}$  value from differential gene expression analysis. The following parameters were used: xtools.gsea.GseaPreranked -nperm 1000 -scoring\_scheme weighted -norm meandiv.

The data analysis code is available at [https://github.com/himrichova/MGs\\_RNAseq\\_analysis](https://github.com/himrichova/MGs_RNAseq_analysis).

The RNA-seq data generated for this study have been deposited in the National Center for Biotechnology Information's (NCBI) Gene Expression Omnibus (GEO) and are accessible through the GEO Series accession number GSE142405.

**Recombinant proteins and in vitro kinase assays.** Human glutathione S-transferase (GST)-CDK12 (714-1063)/GST-Cyclin K (1-267), GST-CDK13 (694-1093)/GST-Cyclin K (1-267) and GST-Cdk7 (2-346)/ CyclinH (1-323)/ Mat1 (1-309) proteins were recombinantly expressed in Sf9 insect cells using the MultiBacTurbo system<sup>41</sup>. For full activation GST-CDK12 (714-1063)/GST-Cyclin K (1-267) and GST-CDK13 (694-1093)/GST-Cyclin K (1-267) were coexpressed with CDK-activating kinase from *S. cerevisiae*. After 72 h, cells were collected by centrifugation and resuspended in lysis buffer (50 mM HEPES pH 7.6, 300 mM NaCl, 5% glycerol and 5 mM bMe), followed by sonication. The lysate was cleared by centrifugation in a Beckman Avanti J-26S XP centrifuge with a JA-25.50 rotor (20,000 r.p.m. for 45 min at  $4^{\circ}\text{C}$ ) and applied to a GST Trap FF column (GE Healthcare) equilibrated with lysis buffer using an Äkta Prime chromatography system (GE Healthcare). After application of 10 column volumes (CV) of lysis buffer, protein was eluted in lysis buffer with 10 mM glutathione. After tobacco etch virus protease cleavage overnight, kinase complexes were further purified by size exclusion chromatography (SEC) using a Superdex 200 PG column (GE Healthcare) equilibrated in SEC buffer (20 mM HEPES pH 7.4, 150 mM NaCl, 5% glycerol and 1 mM TCEP). Fractions containing a stoichiometric kinase complex were determined by SDS-PAGE, pooled and concentrated using Amicon filters (Millipore). Proteins were aliquoted, snap frozen in liquid nitrogen and stored at  $-80^{\circ}\text{C}$ .

Radioactive kinase assays were performed using 0.2  $\mu\text{M}$  CDK-Cyclin K complex incubated with 0.2 mM ATP containing 0.45 mCi [<sup>32</sup>P]-g-ATP per ml (Perkin Elmer) and varying concentrations of inhibitory compound for 10 min before starting the kinase reaction by adding the substrate. Reactions were incubated for 20 min at  $30^{\circ}\text{C}$  and stopped by addition of EDTA to a final concentration of 50 mM. Samples were spotted onto Amersham Protran nitrocellulose membrane (GE Healthcare), followed by three washing steps for 5 min with 0.75% (v/v) phosphoric acid. Counts per minute were estimated in a Beckman Liquid Scintillation Counter (Beckman-Coulter) for 1 min.

**Acquired resistance and hybrid capture.** *MLH1* mutant lines. LentiCRISPR2v2GFP (Addgene no. 82416) was used to express sgRNAs against *MLH1* in KBM7 and MV4;11 cells. Transduced pools of cells were sorted for green fluorescent protein expression twice. Individual clones were expanded in 96-well format and treated with 500  $\mu\text{M}$  temozolomide (Sigma-Aldrich, T2577) in mirror plates to nominate *MLH1* mutants. Knock-out was validated in resistant clones by western blot (anti-*MLH1*, Cell Signaling Technology 3515T). Target sequences were amplified,



A-tailing and TOPO-cloning was performed according to the manufacturer's instructions (Thermo Scientific, K457502) and individual alleles were sequenced. Primers listed in Supplementary Table 6.

#### Generation of cells with acquired drug resistance and hybrid-capture protocol.

One hundred million KBM7 and MV4;11 cells (WT and MLH1<sup>mut</sup>) were treated with DMSO, 6  $\mu$ M dCeMM2, 11  $\mu$ M dCeMM3 or 12  $\mu$ M dCeMM4 in 50 ml medium. After 25 d, Ficoll-gradient centrifugation with Lymphocyte Separation Media (Corning, COR25-072-CV) was performed. The following day cells were counted, PBS-washed and pellets frozen. HCT116 cells (MLH1<sup>-/-</sup>) were seeded in 384-well plates at 30,000 cells per ml in the presence of 50  $\mu$ M dCeMM3 or 30  $\mu$ M dCeMM3.1. Nine individual resistant clones were isolated and acquired resistance was validated against dCeMM2, dCeMM3 and dCeMM4. They were sensitive to Paclitaxel. DMSO-treated HCT116 cells (control) and a pool of 500,000 cells from each of the nine HCT116 clones were PBS-washed and pellets frozen. gDNA was extracted using QIAamp DNA Mini Kit (Qiagen 51304). DNA content was determined with Qubit HS Kit (Thermo Fisher Q32854). The pools of resistant cells were subjected to targeted next-generation sequencing of DDB1, CUL4B, UBE2M, UBE2G1, CDK12/13 and CCNK via a hybrid-capture-based approach. For the library preparation, NEBnext Ultra II FS DNA Library Prep Kit for Illumina (NEB, E7805S) was used (protocol for inputs >100 ng). Then, 500 ng of DNA per condition were used as starting material. Fragmentation and bead size selection was performed for insert size of 150–350 bp. PCR enrichment of adaptor-ligated DNA was done with NEBnext Multiplex Oligos for Illumina (Set1 E7335 and Set2 E75000). Five PCR cycles were performed. For the hybrid capture, 500 ng of DNA were used. Hybridization and prep was performed according to the protocol xGen hybridization capture of DNA libraries from IDT. As xGen blocking oligos the xGen Universal Blocker-TS Mix (IDT 1075475) was used. Custom probes against UBE2M, UBE2G1, CUL4A, CUL4B, CDK12, CDK13 and CCNK were used as xGen Lockdown Panel. Hybridization was performed for 16 h. NEBNext High-Fidelity 2 $\times$  PCR Master Mix (NEB, M0541S) was used for the postcapture PCR, using 16–25 PCR cycles. Final sequencing libraries were analyzed on Agilent 2100 Bioanalyzer, pooled according to resistance rate and sequenced on a MiSeq (paired-end sequencing).

**Bioinformatics analysis of hybrid capture.** Raw read files were converted to fastq format using the SamToFastq function from the picard tools package (v.2.9.0) (<http://broadinstitute.github.io/picard/>) with standard settings. Sequencing adapters and low-quality reads were removed using trimmomatic (v.0.36)<sup>38</sup> in PE mode with standard settings. Data preprocessing was performed following the latest version of Best Practices of the Genome Analysis Toolkit (GATK) consortium<sup>42</sup>. Reads were first aligned to hg38 assembly using bwa mem (v.0.7.8) (<http://bio-bwa.sourceforge.net/bwa.shtml>). Duplicate reads were then filtered using the picard tools MarkDuplicates function. The BaseRecalibrator function from the GenomeAnalysisTK package (v.3.7)<sup>43</sup> was applied supplying the hg38\_dbsnp138 known sites of variation from the GATK resource bundle under the --known-sites argument. Quality score recalibration was then applied using the PrintReads function with the optional -BQSR argument. Variant calling was performed using the Mutect2 function (--output\_mode EMIT\_VARIANTS\_ONLY). Variants were annotated using the Ensembl Variant Effect Predictor webtool<sup>44</sup> in its latest release (Ensembl v.98). Variants that affected the coding sequence (inframe\_deletion, inframe\_insertion, frameshift\_variant, missense\_variant, protein\_altering\_variant, stop\_gained) and that had a greater than fivefold enrichment in allele frequency (as determined by Mutect2) in the drug-resistant population compared to unselected cells were considered hits.

**Flow cytometry analysis of CDK13 levels.** WT and CDK13<sup>+/P1043H</sup> HCT116 cells were fixed with cold 70% ethanol for 30 min on ice. After 15 min 5% Triton X-100 permeabilization on ice, cells were incubated with CDK13 antibody (1:200, bethyl, A301-458A) in 0.5% BSA PBS for 1 h at room temperature. After washing twice with 0.5% BSA PBS, cells were incubated with antirabbit Alexa Fluor 488 secondary antibody (Invitrogen, A21441) for 30 min and washed twice with PBS before analysis by flow cytometry.

**Synthesis of dCeMM3-NH2 and pulldown.** Detailed synthesis in Supplementary Information.

**Coupling of dCeMM3-NH2 to NHS-sepharose beads.** Here, 100  $\mu$ l NHS-Activated Sepharose 4 Fast Flow (GE Life Sciences, 17090601) per condition were washed with 500  $\mu$ l DMSO three times (3 min 800 r.p.m. centrifugation at room temperature). Beads were resuspended in 50  $\mu$ l DMSO and 2.5  $\mu$ l dCeMM3<sup>NH2</sup> 10 mM and 0.75  $\mu$ l TEA (Sigma-Aldrich, T0886) were added. After 16–24 h of incubation on roto-shaker at room temperature, the remaining free amino groups on beads were blocked by adding 2.5  $\mu$ l ethanolamine (Sigma-Aldrich, 11016-7) and incubating on a roto-shaker for at least 8 h at room temperature. Beads were centrifuged and 500  $\mu$ l DMSO washed twice before proceeding with drug pulldown.

**Preparation of cell lysates.** Here, 200 million KBM7 cells were resuspended in 2 ml lysis buffer (50 mM Tris pH 7.5, 0.2% NP-40, 5% glycerol, 1.5 mM MgCl<sub>2</sub>, 100 mM NaCl,

1 mM EDTA) supplemented with protease inhibitors (Thermo Scientific, 78437) and benzonase (Merck, US170746-370746) and incubated on ice for 30 min. After centrifugation (full speed, 4 °C, 30 min) supernatant was transferred to a new tube and extracted protein measured with BCA (Fisher Scientific, 23225). Then 3 mg of protein lysate were pretreated with DMSO or THZ531 100  $\mu$ M for 1 h on a roto-shaker at 4 °C.

**Protein affinity purification and elution.** Drug-coupled beads were washed with 1 $\times$  lysis buffer by centrifugation three times. Washed drug-coupled beads were gently resuspended in the pretreated cell lysates for 2 h on a roto-shaker at 4 °C. After incubation, beads were centrifuged, supernatant removed and washed with lysis buffer at 4 °C. Beads were transferred to unplugged mini Bio-Spin Chromatography Columns (BioRad, 732-6207) in washing buffer (50 mM pH 8 HEPES, 150 mM NaCl and 5 mM EDTA) and washed three times with 1 ml buffer II (all steps performed at 4 °C). Columns were transferred to room temperature, plugged and beads incubated with 4% SDS buffer II for 15 min. Unplugged columns were placed in 1.5 ml Eppendorf tubes and centrifuged for 1 min at room temperature to let eluates enter into the tubes. Eluates were analyzed by western blot. Quantification was performed with ImageJ.

#### Synthesis of dCeMM3-photoaffinity probe (dCeMM3-PAP) and pulldown.

dCeMM3-NH<sub>2</sub> was conjugated to a constant side chain consisting of a photosensitive diazirine group and an alkyne handle. Detailed synthesis in Supplementary Information.

Twenty million KBM7<sup>WT</sup> cells per condition were pretreated with either DMSO or 100  $\mu$ M THZ531 (competition condition) for 1 h in the presence of 10  $\mu$ M carfilzomib (Selleckchem, S2853) in serum-free IMDM medium (3 ml per 20 million cells per 10 cm dish). Then, cells were treated with 10  $\mu$ M dCeMM2<sup>PAP</sup> for 1 h. Target proteins were covalently linked to the probe via photocrosslinking using an ultraviolet crosslinker at 4 °C (365 nm wavelength for 10 min). Cell pellets were collected, PBS-washed and snap frozen in liquid nitrogen. For protein extraction, thawed pellets were resuspended in 500  $\mu$ l lysis buffer (NP-40 0.8%, HEPES pH 7.5 50 mM, Glycerol 5%, NaCl 150 mM, MgCl<sub>2</sub> 1.5 mM, SDS 1%, protease inhibitors and benzonase) and incubated on ice for 30 min. Click reaction to conjugate azide-PEG3-biotin to the photoprobe was performed using 1,000  $\mu$ g of protein per sample (1 ml total volume): 20  $\mu$ l of 5 mM Azide-PEG3-biotin (Sigma-Aldrich, 762024-25MG) was added to each sample followed by a mix of 60  $\mu$ l 1.7 mM TBTA, 20  $\mu$ l 50 mM CuSO<sub>4</sub> and 20  $\mu$ l 50 mM TCEP (100  $\mu$ l per sample), and left at room temperature for 2 h. SpinOUT G-600 columns (G-Biosciences, 786-1621) were used to purify protein samples after the click reaction, according to the manufacturer's protocol. Amounts of 700  $\mu$ g per sample were used for the pulldowns. Enrichment of target proteins was done using Pierce High Capacity NeutrAvidin Agarose beads (Thermo Scientific, 29202). After the last wash, step beads were resuspended in 100  $\mu$ l elution Buffer (HEPES pH 8 50 mM, NaCl 150 mM, EDTA 5 mM, SDS 4%), incubated at 75 °C for 30 min and eluted by centrifuging at full speed for 3 min. Eluates were analyzed by western blot. Quantification was performed with ImageJ.

**Biotin proximity labeling with miniTurbo in live cells.** Cloning of DDB1-miniTurbo-V5. attB1-MCS-mCherry-MCS-miniTurbo-MCS-V5-attB2 was synthesized as a gBlock (IDTDNA) and cloned into a Gateway-compatible donor vector (pDONR221) using BP clonease (Invitrogen, 11789-020) to generate pENTR221\_MCS\_mCherry\_miniTurbo\_V5 (miniTurbo sequence obtained from Addgene, 107170). DDB1 was cloned by Gibson assembly (NEB, E5510S) into NdeI+XbaI digested pENTR221\_MCS\_mCherry\_miniTurbo\_V5. Primers to amplify DDB1 with Gibson compatible overhangs are listed in Supplementary Table 6. pcDNA3-FLAG-DDB1 (Addgene, 19918) was used as a template. Resulting plasmid pENTR221\_DDB1\_miniTurbo\_V5 was subsequently cloned into a Gateway-compatible destination vector (Addgene, 19066) using LR clonease (Thermo Fisher Scientific, 11-791-020) to generate pLenti\_DDB1\_miniTurbo\_V5.

**Biotin labeling in live cells and pulldown.** One 10-cm culture plate of 293T cells (70–80% confluent) was transfected with 2  $\mu$ g of pLenti\_DDB1\_miniTurbo\_V5 using Lipofectamine 2000 (Invitrogen, 11668019). On the next day, the plate was expanded to 2 $\times$ 10-cm plates. 48 h after transfection, cells were pretreated with 10  $\mu$ M Carfilzomib for 30 min, then treated with either DMSO or 20  $\mu$ M dCeMM2 for 1.5 h, adding 500  $\mu$ M biotin during the last 30 min. Labeling was stopped by transferring the cells to ice and washing five times with ice-cold PBS<sup>32</sup>. Cells were collected and lysed in 600  $\mu$ l of lysis buffer (50 mM Tris-HCl pH 7.5, 125 mM NaCl, 5% glycerol, 0.2% NP-40, 1.5 mM MgCl<sub>2</sub> and protease inhibitors). After 15 min incubation at room temperature, lysates were clarified by centrifuging at 15,000g for 10 min. For streptavidin pulldown of the biotinylated proteins, 500  $\mu$ g of protein per condition were incubated with 50  $\mu$ l of lysis buffer-washed streptavidin magnetic beads (Thermo Fisher Scientific, 11205D) for 1 h at room temperature on a rotator. Beads were pelleted using a magnetic rack and each bead sample was washed three times with lysis buffer. To elute biotinylated proteins, beads were resuspended in 65  $\mu$ l elution Buffer (HEPES pH 8 50 mM, NaCl 150 mM, EDTA 5 mM, SDS 4%) and incubated at 75 °C for 30 min. Beads were pelleted on a magnetic rack and eluate (60  $\mu$ l) was collected. Eluates were analyzed by western blot.

**Cyclin K ubiquitination in live cells.** Five million KBM7 cells per condition were pretreated for 30 min with Carfilzomib 1  $\mu$ M, then treated with DMSO, dCeMM2 10  $\mu$ M or dCeMM2 10  $\mu$ M + THZ531 10  $\mu$ M for 2 h, collected and processed for western blot analysis. Cells were lysed in 200  $\mu$ l lysis buffer: 50 mM Tris pH 7.5, 150 mM NaCl, 0.1% NP-40, 1 mM EDTA, 5 mM MgCl<sub>2</sub>, 5% glycerol with protease inhibitors, phosphatase inhibitors and NEM 25 mM (Sigma-Aldrich E3876). NEM was included in all washing steps, including PBS washing after collection before lysis. For K-48 pulldown assays, 20 million cells per conditions were used and the protocol provided by the manufacturer was followed (Nanotab Biotechnologies, N1810).

**TR-FRET experiments. Protein purification.** Human WT DDB1 (Uniprot entry Q16531), CDK12 (Q9NYV4) and cyclin K (O75909) were subcloned into pAC-derived vectors<sup>45</sup> and recombinant proteins were expressed as N-terminal His<sub>6</sub>, His<sub>6</sub>-Spy or StrepII-Avi fusions in *Trichoplusia ni* High Five insect cells using the baculovirus expression system (Invitrogen)<sup>46</sup>.

WT full-length StrepII-Avi-DDB1 was purified as previously described for DDB1-DCAF complexes<sup>47</sup>. High Five insect cells coexpressing truncated versions of WT His<sub>6</sub>-CDK12 (713-1032) and His<sub>6</sub>-Spy-tagged cyclin K (amino acids 1-267) were lysed by sonication in 50 mM Tris-HCl (pH 8.0), 500 mM NaCl, 10% (v/v) glycerol, 10 mM MgCl<sub>2</sub>, 10 mM imidazole, 0.25 mM tris(2-carboxyethyl) phosphine (TCEP), 0.1% (v/v) Triton X-100, 1 mM phenylmethylsulfonyl fluoride and 1 $\times$  protease inhibitor cocktail (Sigma). Following ultracentrifugation, the soluble fraction was passed over HIS-Select Ni<sup>2+</sup> affinity resin (Sigma), washed with 50 mM Tris-HCl (pH 8.0), 1 M NaCl, 10% (v/v) glycerol, 0.25 mM TCEP, 10 mM imidazole and eluted in 50 mM Tris-HCl (pH 8.0), 200 mM NaCl, 10% (v/v) glycerol, 0.25 mM TCEP and 250 mM imidazole. The pH of the eluate was adjusted to 6.8 ready for ion exchange chromatography and the affinity purified proteins were diluted in a 1:1 ratio with buffer A (50 mM Tris-HCl (pH 6.8), 10 mM NaCl, 2.5% (v/v) glycerol, 0.25 mM TCEP) and passed over an 8 ml Poros 50HQ column. The flow through was again diluted in a 1:1 ratio with buffer A and passed over an 8 ml Poros 50HS column. Bound proteins were eluted by a linear salt gradient mixing buffer A and buffer B (50 mM Tris-HCl (pH 6.8), 1 M NaCl, 2.5% (v/v) glycerol, 0.25 mM TCEP) over 15 CV to a final ratio of 80% buffer B. Poros 50HS peak fractions containing the CDK12-cyclin K complex were concentrated and subjected to SEC in 50 mM HEPES (pH 7.4), 200 mM NaCl, 2.5% (v/v) glycerol and 0.25 mM TCEP. The concentrated proteins were flash frozen in liquid nitrogen and stored at -80  $^{\circ}$ C.

**Biotinylation of DDB1.** Purified full-length StrepII-Avi-DDB1 was biotinylated in vitro at a concentration of 8  $\mu$ M by incubation with final concentrations of 2.5  $\mu$ M BirA enzyme and 0.2 mM D-biotin in 50 mM HEPES (pH 7.4), 200 mM NaCl, 10 mM MgCl<sub>2</sub>, 0.25 mM TCEP and 20 mM ATP. The reaction was incubated for 1 h at room temperature and stored at 4  $^{\circ}$ C for 14-16 h. Biotinylated DDB1 (biotin-DDB1) was purified by gel filtration chromatography and stored at -80  $^{\circ}$ C (-20  $\mu$ M).

**TR-FRET.** Increasing concentrations of Alexa488-SpyCatcher-labeled<sup>10</sup> His<sub>6</sub>-Spy-cycK/His<sub>6</sub>-CDK12 (Alexa488-cyclin K-CDK12) were added to a mixture of biotinylated DDB1 (biotin-DDB1) at 50 nM, terbium-coupled streptavidin at 4 nM (Invitrogen) and kinase inhibitors at 10  $\mu$ M (final concentrations) in 384-well microplates (Greiner, 784076) in a buffer containing 50 mM Tris (pH 7.5), 150 mM NaCl, 0.1% pluronic acid and 0.5% DMSO (see also figure legends). Forward titrations were carried out by adding increasing concentrations of Alexa488-cycK-CDK12 (0.5-5  $\mu$ M) into premixed 10  $\mu$ M compound, 50 nM biotin-DDB1 and 4 nM terbium-coupled streptavidin. Before TR-FRET measurements, reactions were incubated for 15 min at room temperature. After excitation of terbium (Tb) fluorescence at 337 nm, emissions at 490 nm (Tb) and 520 nm (Alexa488) were measured with a 70- $\mu$ s delay to reduce background fluorescence and the reactions were followed by recording 60 data points of each well over 1 h using a PHERAstar FS microplate reader (BMG Labtech). The TR-FRET signal of each data point was extracted by calculating the 520 to 490 nm ratio. Data were analyzed with Prism 7 (GraphPad) assuming equimolar binding of biotin-DDB1 to Alexa488-cyclin K-CDK12 using the equations described previously<sup>8</sup>. Three technical replicates were carried out per experiment.

**Cyclin K degradation-sensitivity correlations and synergy experiments.** Correlation of degradation and sensitivity (Supplementary Figs. 9f and 10c) in a panel of cell lines was calculated using cyclin K levels by immunoblotting and EC50s by 3-d viability assays. Cyclin K levels were quantified with ImageJ, normalized to KBM7 and plotted against the EC50 of a particular cell line.

For the two-way compound combinations of synergy experiments, compounds were added to 384-well plates in a 6  $\times$  10 matrixed format (six-point titrations, including DMSO control of dCeMM2 and ten-point titrations of DNA damage drugs including DMSO control) were used. Then 1,000 SKW3 cells per well in 50  $\mu$ l were used in 384-well plates. Cell viability was measured 72 h later (CellTiter Glo, Promega G7570) and Bliss scores were calculated as previously described<sup>48</sup>.

**Statistics and reproducibility.** Information regarding error bars, numbers of replicates or samples, and statistical analyses are described in the corresponding

figure legends. Representative results of at least two independent experiments are shown unless otherwise indicated. No statistical methods were used to predetermine sample size. The experiments were not randomized and the investigators were not blinded to allocation during experiments and outcome assessment.

**Reporting Summary.** Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

## Data availability

The authors declare that the data supporting the findings of this study are available within the paper and its supplementary information files. Sequencing of sgRNA cassettes from all the CRISPR-Cas9 screens described in this study (Figs. 1h and 3a and Extended Data Figs. 5a and 6b-e) and hybrid-capture experiments (Fig. 3h and Extended Data Figs. 7h and 8d) have been deposited in the NCBI Sequence Read Archive under the accession code PRJNA599346. The analyzed data are provided in Supplementary Tables 2 and 5, respectively. Expression proteomics data (Figs. 1g and 2a and Extended Data Figs. 2a and 4d,e) are provided in Supplementary Table 3. Searches were performed with full tryptic digestion against the human SwissProt database v.2017.06.06 ([https://www.uniprot.org/statistics/Swiss-Prot%202017\\_06](https://www.uniprot.org/statistics/Swiss-Prot%202017_06)). Structure of DDB1 in complex with SV5V peptide (Fig. 3i) corresponds to PDB 2HYE. RNA-seq data (Fig. 2e,f and Extended Data Fig. 4a-c) have been deposited in the GEO under the accession code GSE142405. The analyzed data are provided in Supplementary Table 4. Source data are provided with this paper.

## Code availability

A detailed description of the bioinformatics analysis used for CRISPR screens, hybrid capture and RNA-seq is available in the Methods. RNA-seq analysis code available at [https://github.com/himrichova/MGS\\_RNAseq\\_analysis](https://github.com/himrichova/MGS_RNAseq_analysis). Source data are provided with this paper.

## References

- Ritchie, M. E. et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **43**, e47 (2015).
- Barnett, D. W., Garrison, E. K., Quinlan, A. R., Stromberg, M. P. & Marth, G. T. BamTools: a C++ API and toolkit for analyzing and managing BAM files. *Bioinformatics* **27**, 1691-1692 (2011).
- Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357-359 (2012).
- Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114-2120 (2014).
- Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15-21 (2013).
- Anders, S., Pyl, P. T. & Huber, W. HTSeq-a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166-169 (2015).
- Bieniossek, C., Imasaki, T., Takagi, Y. & Berger, I. MultiBac: expanding the research toolbox for multiprotein complexes. *Trends Biochem. Sci.* **37**, 49-57 (2012).
- Van der Auwera, G. A. et al. From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. *Curr. Protoc. Bioinformatics* **43**, 11 10 1-11 10 33 (2013).
- McKenna, A. et al. The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297-1303 (2010).
- McLaren, W. et al. The ensembl variant effect predictor. *Genome Biol.* **17**, 122 (2016).
- Abdulrahman, W. et al. A set of baculovirus transfer vectors for screening of affinity tags and parallel expression strategies. *Anal. Biochem.* **385**, 383-385 (2009).
- Li, T., Pavletich, N. P., Schulman, B. A. & Zheng, N. High-level expression and purification of recombinant SCF ubiquitin ligases. *Methods Enzymol.* **398**, 125-142 (2005).
- Fischer, E. S. et al. Structure of the DDB1-CRBN E3 ubiquitin ligase in complex with thalidomide. *Nature* **512**, 49-53 (2014).
- Winter, G. E. et al. Systems-pharmacology dissection of a drug synergy in imatinib-resistant CML. *Nat. Chem. Biol.* **8**, 905-912 (2012).

## Acknowledgements

CeMM and the Winter laboratory are supported by the Austrian Academy of Sciences. This work has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (grant agreement no. 851478). This work was further enabled by funding from the Austrian Science Fund (FWF, project nos. P32125-B and P30271-B28). C.M.-R. is supported by an individual Marie Skłodowska-Curie postdoctoral fellowship (grant agreement no. 796010). Z.K. was supported by a European Union's Horizon 2020 Research and Innovation Program under the Marie Skłodowska-Curie grant agreement number 765445.

G.P. was supported by the Human Frontier Science Program (HFSP Long-Term Fellowship LT000210/2014) and the European Molecular Biology Organization (EMBO Advanced Fellowship ALTF 761-849 2016). We further acknowledge funding awarded from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program grant agreement no. 666068 and from the Novartis Research Foundation to N.H.T. Sequencing was performed at the Biomedical Sequencing facility and proteomics at the Proteomics facility at CeMM. We thank B. Ebert and M. Slabicki for sharing and discussing results ahead of publication.

### Author contributions

C.M.-R. planned and performed most of the experiments, analyzed data and made figures. S.B. provided technical help in many experiments and prepared RNA-seq and hybrid-capture sequencing libraries. M.B. analyzed CRISPR screens and hybrid-capture experiments and made figures. Z.K. developed and conducted TR-FRET assays and performed purification and labeling of recombinant proteins. M.S. synthesized tethered analogs of dCeMM3. H.I. analyzed RNA-seq data. I.H.K. performed recombinant kinase assays. E.H. and K.R. provided technical help in acquired resistance and experiments with dCeMM3-PAP. A.K. provided technical help in the chemical screen. G.P. performed purification and labeling of recombinant proteins. M.F. cloned the CRL sgRNA library. C.B. provided input on bioinformatic analysis. A.C.M. supervised proteomics

experiments. J.Z. supported sgRNA design for the CRL library and supervised the cloning strategy. M.G. supervised recombinant kinase assays. N.H.T. gave critical input to the manuscript and supervised TR-FRET experiments. S.K. supervised and helped to design chemical profiling screen. G.E.W. wrote the manuscript, planned and supervised the presented research and has overall project responsibility.

### Competing interests

C.M.-R. and G.E.W. are listed as inventors of a patent application for glue discovery in neddylation-deficient cellular systems. C.M.-R., S.K. and G.E.W. are listed as inventors of patent applications covering the chemical space of dCeMM2/3/4. M.B., S.K., G.E.W. and CeMM are founders and equity holders of Proxygen.

### Additional information

**Extended data** is available for this paper at <https://doi.org/10.1038/s41589-020-0594-x>.

**Supplementary information** is available for this paper at <https://doi.org/10.1038/s41589-020-0594-x>.

**Correspondence and requests for materials** should be addressed to G.E.W.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

Li, Z., Pinch, B.J., Olson, C.M., Donovan, K.A., Nowak, R.P., Mills, C.E., Scott, D.A., Doctor, Z.M., Eleuteri, N.A., Chung, M., et al. (2020). Development and Characterization of a Wee1 Kinase Degradator. *Cell Chem. Biol.* 27, this issue, 57–65.

Mayor-Ruiz, C., and Winter, G.E. (2019). Identification and characterization of cancer vulnerabilities via targeted protein degradation. *Drug Discov. Today Technol.* 37, 81–90.

Mayor-Ruiz, C., Jaeger, M.G., Bauer, S., Brand, M., Sin, C., Hanzl, A., Mueller, A.C., Menche, J., and Winter, G.E. (2019). Plasticity of the Cullin-RING Ligase Repertoire Shapes Sensitivity to Ligand-Induced Protein Degradation. *Mol. Cell* 75, 849–858.e8.

Nowak, R.P., DeAngelo, S.L., Buckley, D., He, Z., Donovan, K.A., An, J., Safaee, N., Jedrychowski,

M.P., Ponthier, C.M., Ishoey, M., et al. (2018). Plasticity in binding confers selectivity in ligand-induced protein degradation. *Nat. Chem. Biol.* 14, 706–714.

Olson, C.M., Jiang, B., Erb, M.A., Liang, Y., Doctor, Z.M., Zhang, Z., Zhang, T., Kwiatkowski, N., Boukhali, M., Green, J.L., et al. (2018). Pharmacological perturbation of CDK9 using selective CDK9 inhibition or degradation. *Nat. Chem. Biol.* 14, 163–170.

Paiva, S.L., and Crews, C.M. (2019). Targeted protein degradation: elements of PROTAC design. *Curr. Opin. Chem. Biol.* 50, 111–119.

Potjeywd, F., Turner, A.W., Beri, J., Rectenwald, J.M., Norris-Drouin, J.L., Cholensky, S.H., Margolis, D.M., Pearce, K.H., Herring, L.E., and James, L.I. (2020). Degradation of Polycomb Repressive Complex 2 with an EED-Targeted

Bivalent Chemical Degradator. *Cell Chem. Biol.* 27, this issue, 47–56.

Zhang, L., Riley-Gillis, B., Vijay, P., and Shen, Y. (2019). Acquired Resistance to BET-PROTACs (Proteolysis-Targeting Chimeras) Caused by Genomic Alterations in Core Components of E3 Ligase Complexes. *Mol. Cancer Ther.* 18, 1302–1311.

Zoppi, V., Hughes, S.J., Maniaci, C., Testa, A., Gmaschitz, T., Wieshofer, C., Koegl, M., Ricking, K.M., Daniels, D.L., Spallarossa, A., and Ciulli, A. (2019). Iterative Design and Optimization of Initially Inactive Proteolysis Targeting Chimeras (PROTACs) Identify VZ185 as a Potent, Fast, and Selective von Hippel-Lindau (VHL) Based Dual Degradator Probe of BRD9 and BRD7. *J. Med. Chem.* 62, 699–726.

## Taking the Brakes Off Targeted Protein Degradation

Zuzanna Kozicka,<sup>1,2</sup> Georg Petzold,<sup>1</sup> and Nicolas H. Thomä<sup>1,\*</sup>

<sup>1</sup>Friedrich Miescher Institute for Biomedical Research, Basel, Switzerland

<sup>2</sup>University of Basel, Basel, Switzerland

\*Correspondence: [nicolas.thoma@fmi.ch](mailto:nicolas.thoma@fmi.ch)

<https://doi.org/10.1016/j.chembiol.2019.12.010>

**PCSK9 heightens LDL cholesterol levels by chaperoning the liver LDL receptor to lysosomes for degradation. In this issue of *Cell Chemical Biology*, Petrilli et al. (2020) identify novel PCSK9 ligands and convert them into a proof-of-concept degrader, offering a unique way to modulate this key protein-protein interaction.**

Targeted protein degradation (TPD) is an emerging paradigm in pharmacology with the promise to efficiently inactivate previously intractable drug targets (reviewed in [Chopra et al., 2019](#)). Whereas classical inhibitors bind and block the active site of enzymatic proteins or receptors, degrader compounds recruit the target protein to a ubiquitin ligase and trigger the ubiquitination and subsequent degradation of the target in a catalytic manner. This can be achieved with bifunctional degrader molecules (PROTACs) that consist of a target-recruiting ligand connected through a covalent linker to a ubiquitin ligase-binding compound ([Chopra et al., 2019](#)). Molecular glue degraders, which are small, drug-like compounds that induce target-ligase interactions by binding at the protein-protein interface without the need for two ligandable pockets, represent another type of degraders, whose rational design is less intuitive than that of PROTACs

([Chopra et al., 2019](#)). An additional class of degraders, albeit much further from clinical translation than PROTACs or molecular glues, achieves degradation by target recruitment to the proteasome through coupling target binders to hydrophobic moieties ([Neklesa et al., 2011](#)).

To date, an overwhelming majority of small-molecule degraders are repurposed inhibitors, with only a few examples of degraders published where no prior inhibitor was available ([Gechijian et al., 2018](#)). Given that for degradation only a binder, and not a functional inhibitor, is required, the true potential of TPD for difficult targets has not yet been realized. Fulfilling the promise of expanding the repertoire of druggable proteins would require shifting the focus toward high-throughput screening approaches that allow the identification of a broad range of binders for a given target. The work of [Petrilli et al. \(2020\)](#) presented in this issue of *Cell Chemical Biology* illustrates a

comprehensive yet untraditional drug discovery workflow, whereby a binder-oriented screen yields a novel ligand of PCSK9 that is subsequently converted into a proof-of-concept PCSK9 degrader.

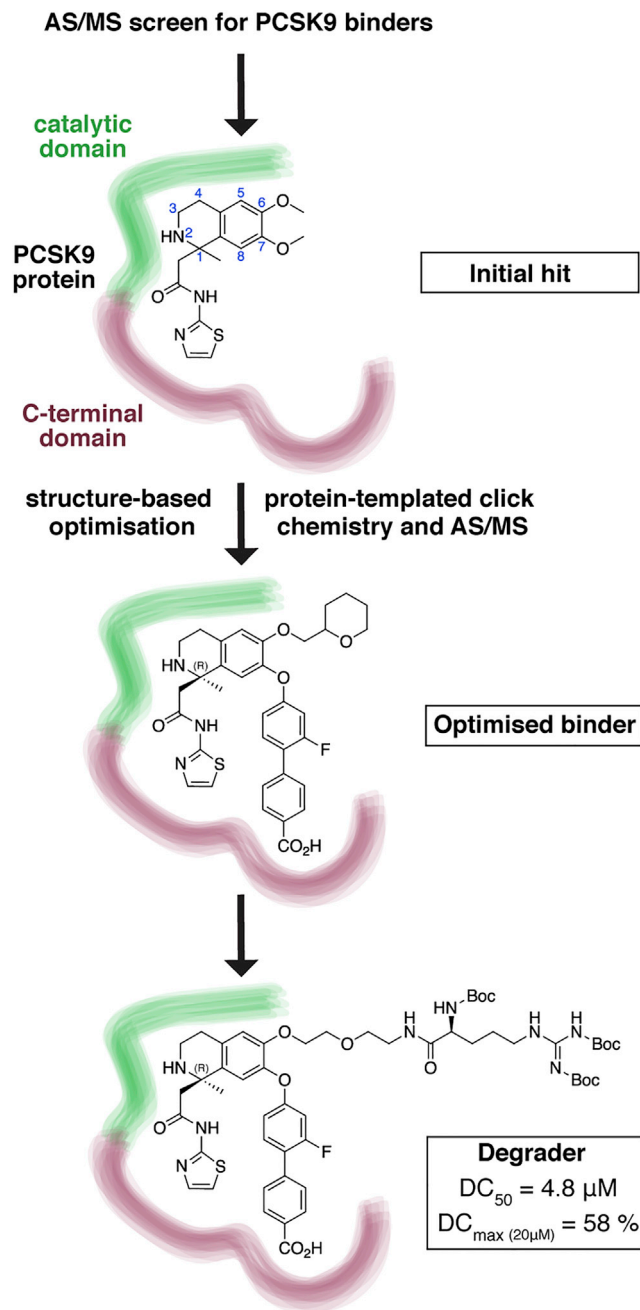
Low-density lipoprotein (LDL) cholesterol is an important risk factor in coronary heart disease. Statins are the most widely used cholesterol-lowering drugs and are known to block an early step in sterol biosynthesis by inhibiting HMG-CoA reductase. Although commonly prescribed across the world for cardiovascular disease prevention, many patients fail to achieve satisfactory LDL cholesterol levels following statin treatment. The discovery that the serine protease proprotein convertase subtilisin kexin 9 (PCSK9) affects LDL cholesterol levels by regulating the abundance of the LDL receptor protein has paved the way for the development of novel strategies for the treatment of hypercholesterolemia ([Cohen et al., 2006](#)). PCSK9 binds the LDL receptor



and targets it to lysosomes for degradation, which results in reduced uptake of LDL particles from the blood and higher levels of circulating LDL cholesterol. Accordingly, patients with loss-of-function PCSK9 mutations have extraordinarily low LDL cholesterol levels and up to 88% lower risk of coronary heart disease (Cohen et al., 2006). Preventing the PCSK9-LDL receptor interaction can therefore lower the risk of cardiovascular events. Many approaches are currently being pursued to modulate PCSK9 levels, including siRNA, mRNA translation inhibitors, gene editing, and PCSK9 vaccines (reviewed in Seidah et al., 2019). The only clinically approved treatment, however, involves monoclonal antibodies (evolocumab and alirocumab) that do not fully recapitulate the PCSK9 knockout phenotype and, although effective, remain a rather costly approach (Sabatine et al., 2017).

Small-molecule therapeutics could yield more cost-effective treatments. However, such approaches have been precluded by the fact that the PCSK9 active site is nonessential for LDL receptor downregulation and by challenges inherent to developing small-molecule inhibitors of protein-protein interactions (PPIs). The catalytic domain of PCSK9 binds the epidermal growth factor-like domain A (EGF-A) of the LDL receptor and the relatively flat and featureless surface has thus far rendered the interface intractable by small molecules (Kwon et al., 2008). Even though peptides capable of blocking the interaction have shown considerable promise (Zhang et al., 2017), they are yet to be optimized and developed into orally bioavailable drugs.

Petrilli et al. (2020) initially sought to find small molecules that disrupt the



**Figure 1. The Workflow of the Study Featuring Chemical Structures of the Developed Compounds and Degradation Data for the Proof-of-Concept PCSK9 Degrader**

PCSK9-LDL receptor interaction but failed to identify such compounds despite considerable efforts involving biochemical and cell-based assays and several diverse libraries. Likely with TPD in mind, the authors refocused their hit-finding campaign and instead pursued PCSK9 binders, irrespective of their binding pocket or the predicted pharmacological

effect. An affinity selection/mass spectrometry (AS/MS)-based screening approach yielded one validated hit compound (Figure 1). The identified compound bound to a previously undiscovered pocket located between the catalytic and C-terminal domain of PCSK9 close to where the reported loss-of-function mutations occur in the protein. Guided by their structural data, the authors (Petrilli et al., 2020) systematically optimized the ligand focusing on the C6 and C7 positions of the isoquinoline ring and successfully engineered several additional non-covalent interactions with PCSK9 residues, giving rise to high-affinity ligands ( $K_i = 29$  nM for the optimized binder; Figure 1). In a parallel optimization effort, the authors used the PCSK9 protein to template a 1,3-dipolar Huisgen cycloaddition that only gives the respective 1,2,3-triazole products if the azide and alkyne fragments screened are bound to pockets in close proximity and with the correct orientation. While this target-guided ligand optimization approach did not yield the most potent binders in the study, it provided distinct structure-activity relationship data and illustrated a streamlined fragment-based ligand optimization workflow combining protein templated click chemistry with AS/MS. The authors further employed a cellular thermal shift assay (CETSA) to corroborate efficient target engagement in cell lysates.

As these novel PCSK9 binders failed to modulate the interaction with the LDL receptor, the authors (Petrilli et al., 2020) attempted to turn their inactive molecules into active bifunctional degraders to enable drug-induced PCSK9 degradation. Coupling the PCSK9 binders to ligands recruiting VHL, a cullin-RING ligase substrate receptor commonly

employed in TPD, or cIAP, a non-cullin E3 ligase, did not yield active PROTAC molecules for reasons unclear. Nonetheless, as a proof of concept, the authors coupled their ligand to *tert*-butyl carbamate-protected arginine (Arg(Boc)<sub>3</sub>), a small chemical degron that reportedly elicits degradation in a ubiquitin-independent manner by recruiting the target protein directly to the proteasome (Shi et al., 2016). Tethering the Arg(Boc)<sub>3</sub> moiety to the solvent-exposed C6 position on the isoquinoline ring resulted in an active PCSK9 degrader compound (Figure 1). Structural characterization of the protein-degrader complex revealed that protein-ligand interactions were maintained and the proteasome-recruiting moiety was solvent exposed. Experiments in HEK293 cells overexpressing PCSK9 demonstrated degradation of the target protein (DC<sub>50</sub> = 4.8 μM, DC<sub>max</sub> = 58%) and confirmed that the compound has no adverse effects on cell viability.

This proof-of-concept degrader clearly shows that TPD is a viable strategy to modulate PCSK9 protein levels and paves the way for the further development of degraders for this clinically relevant target. As the Arg(Boc)<sub>3</sub> moiety-based recruitment has certain translational limitations, further optimization of PROTACs through exploring different ligase and linker combinations, or the development of molecular glue degraders of PCSK9, could yield compounds with immediate clinical applicability.

More generally, the example of the PCSK9-LDL receptor interaction highlights the difficulties inherent in developing PPI inhibitors. Despite recent progress in the field, many highly clinically relevant PPIs remain challenging and inhibition of one protein partner is often not enough to have a functional consequence on the interaction. The work of Petrilli et al. (2020) illustrates a promising strategy for targeting an intractable PPI through leveraging drug-induced degradation of one of the interacting proteins.

This study by Petrilli et al. (2020) also exemplifies an interesting workflow that

does not simply aim to identify an inhibitor of the target, or a disruptor of a particular PPI, but is considerably more open-minded in the search for small-molecule binders regardless of their immediate functional consequences. Such an extended focus can reveal the true potential of TPD, whereby inactive binders can be repurposed to drive degradation of otherwise intractable proteins. Furthermore, adopting binder-oriented screening approaches more widely in drug discovery campaigns could also allow the identification of compounds that affect protein activity in underappreciated but therapeutically relevant ways, including altering the cellular stabilities or interactomes of their targets (Schreiber, 2019).

The work presented in this issue of *Cell Chemical Biology* by Petrilli et al. (2020) showcases the development of a proof-of-concept PCSK9 degrader from a hit compound that binds to a previously undescribed pocket on the protein. Although active compounds that could be directly translated into the clinic are yet to be reported, the present findings provide a much-needed framework that will likely expedite the development of degraders for cardiovascular disease prevention. As degraders combine the advantages of small-molecular inhibitors with the ability to control protein levels typically limited to antibody-based and genetic approaches, it would be important to perform *in vivo* studies (once optimized compounds are available) to answer whether degradation can indeed recapitulate the PCSK9 knockout phenotype. In the not-too-distant future, PCSK9 degraders may thus hopefully become a novel class of small-molecule drugs for the treatment of hypercholesterolemia.

#### DECLARATION OF INTERESTS

N.H.T. receives funding from the Novartis Research Foundation and is a Scientific Advisory Board member of Monte Rosa Therapeutics.

#### REFERENCES

- Chopra, R., Sadok, A., and Collins, I. (2019). A critical evaluation of the approaches to targeted protein degradation for drug discovery. *Drug Discov. Today. Technol.* *31*, 5–13.
- Cohen, J.C., Boerwinkle, E., Mosley, T.H., Jr., and Hobbs, H.H. (2006). Sequence variations in PCSK9, low LDL, and protection against coronary heart disease. *N. Engl. J. Med.* *354*, 1264–1272.
- Gechijian, L.N., Buckley, D.L., Lawlor, M.A., Reyes, J.M., Paulk, J., Ott, C.J., Winter, G.E., Erb, M.A., Scott, T.G., Xu, M., et al. (2018). Functional TRIM24 degrader via conjugation of ineffectual bromodomain and VHL ligands. *Nat. Chem. Biol.* *14*, 405–412.
- Kwon, H.J., Lagace, T.A., McNutt, M.C., Horton, J.D., and Deisenhofer, J. (2008). Molecular basis for LDL receptor recognition by PCSK9. *Proc. Natl. Acad. Sci. USA* *105*, 1820–1825.
- Neklesa, T.K., Tae, H.S., Schneekloth, A.R., Stulberg, M.J., Corson, T.W., Sundberg, T.B., Raina, K., Holley, S.A., and Crews, C.M. (2011). Small-molecule hydrophobic tagging-induced degradation of HaloTag fusion proteins. *Nat. Chem. Biol.* *7*, 538–543.
- Petrilli, W.L., Adam, G.C., Erdmann, R.S., Abeywickrema, P., Agnani, V., Ai, X., Baysarowich, J., Byrne, N., Caldwell, J.P., Chang, W., et al. (2020). From Screening to Targeted Degradation: Strategies for the Discovery and Optimization of Small Molecule Ligands for PCSK9. *Cell Chem. Biol.* *27*, this issue, 32–40.
- Sabatine, M.S., Giugliano, R.P., Keech, A.C., Honarpour, N., Wiviott, S.D., Murphy, S.A., Kuder, J.F., Wang, H., Liu, T., Wasserman, S.M., et al.; FOURIER Steering Committee and Investigators (2017). Evolocumab and Clinical Outcomes in Patients with Cardiovascular Disease. *N. Engl. J. Med.* *376*, 1713–1722.
- Schreiber, S.L. (2019). A chemical biology view of bioactive small molecules and a binder-based approach to connect biology to precision medicines. *Isr. J. Chem.* *59*, 52–59.
- Seidah, N.G., Prat, A., Piriolo, A., Catapano, A.L., and Norata, G.D. (2019). Novel strategies to target proprotein convertase subtilisin kexin 9: beyond monoclonal antibodies. *Cardiovasc. Res.* *115*, 510–518.
- Shi, Y., Long, M.J.C., Rosenberg, M.M., Li, S., Kobjack, A., Lessans, P., Coffey, R.T., and Hedstrom, L. (2016). Boc<sub>3</sub>Arg-Linked Ligands Induce Degradation by Localizing Target Proteins to the 20S Proteasome. *ACS Chem. Biol.* *11*, 3328–3337.
- Zhang, Y., Ultsch, M., Skelton, N.J., Burdick, D.J., Beresini, M.H., Li, W., Kong-Beltran, M., Peterson, A., Quinn, J., Chiu, C., et al. (2017). Discovery of a cryptic peptide-binding site on PCSK9 and design of antagonists. *Nat. Struct. Mol. Biol.* *24*, 848–856.

## STRUCTURAL BIOLOGY

## Mechanisms of OCT4-SOX2 motif readout on nucleosomes

Alicia K. Michael<sup>1\*</sup>, Ralph S. Grand<sup>1\*</sup>, Luke Isbel<sup>1\*</sup>, Simone Cavadini<sup>1</sup>, Zuzanna Kozicka<sup>1,2</sup>, Georg Kempf<sup>1</sup>, Richard D. Bunker<sup>1†</sup>, Andreas D. Schenk<sup>1</sup>, Alexandra Graff-Meyer<sup>1</sup>, Ganesh R. Pathare<sup>1</sup>, Joscha Weiss<sup>1</sup>, Syota Matsumoto<sup>1</sup>, Lukas Burger<sup>1,3</sup>, Dirk Schübeler<sup>1,2‡</sup>, Nicolas H. Thomä<sup>1‡</sup>

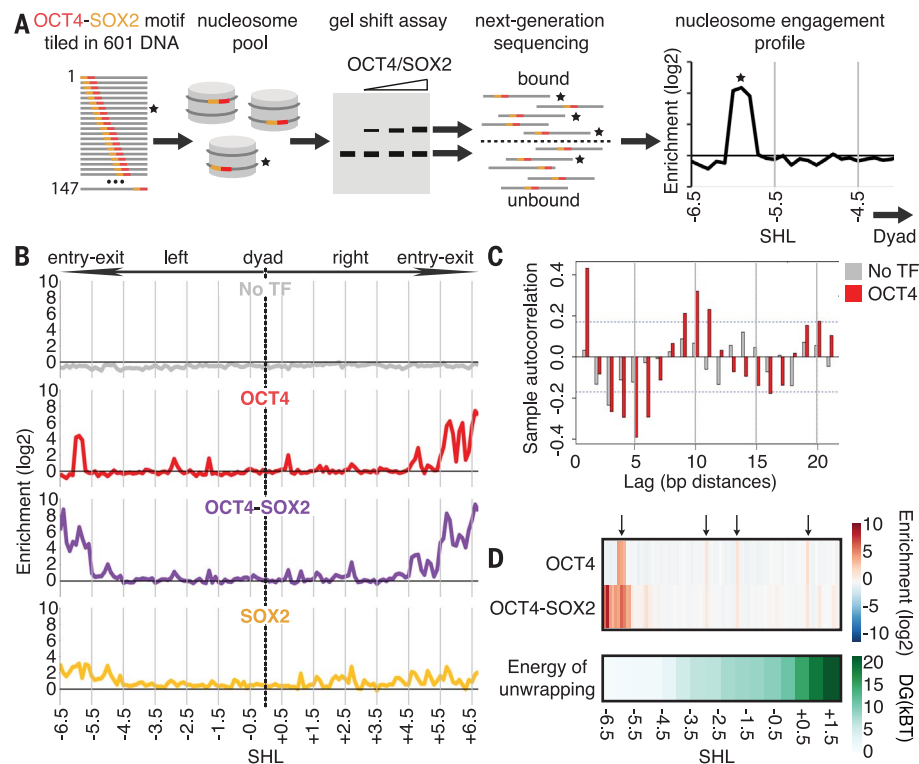
Transcription factors (TFs) regulate gene expression through chromatin where nucleosomes restrict DNA access. To study how TFs bind nucleosome-occupied motifs, we focused on the reprogramming factors OCT4 and SOX2 in mouse embryonic stem cells. We determined TF engagement throughout a nucleosome at base-pair resolution in vitro, enabling structure determination by cryo-electron microscopy at two preferred positions. Depending on motif location, OCT4 and SOX2 differentially distort nucleosomal DNA. At one position, OCT4-SOX2 removes DNA from histone H2A and histone H3; however, at an inverted motif, the TFs only induce local DNA distortions. OCT4 uses one of its two DNA-binding domains to engage DNA in both structures, reading out a partial motif. These findings explain site-specific nucleosome engagement by the pluripotency factors OCT4 and SOX2, and they reveal how TFs distort nucleosomes to access chromatinized motifs.

**T**ranscription factors (TFs) regulate gene expression and govern cell identity by interacting with specific sequence motifs. OCT4 and SOX2 serve as reprogramming TFs (1) that cooperate as critical mediators of pluripotency (2–4). Although they are not obligate heterodimers in solution, the combined OCT4-SOX2 motif drives OCT4-SOX2 complex formation (5, 6).

Chromatin restricts DNA access (7, 8), but a specialized subset of TFs, termed pioneer factors, can engage chromatinized motifs to trigger cell-fate changes (9). Several TFs have been shown to bind motifs embedded in nucleosomes in vitro (1, 10, 11); however, the nucleosome architecture, with histones H2A, H2B, H3, and H4 and its two DNA gyres (12), restricts TF access to >95% of nucleosomal DNA (13). Two extreme scenarios for nucleosomal TF-engagement have been put forward: TF binding without changing the nucleosomal architecture (10, 14) or TF-mediated changes to the nucleosome by distorting the histone core, looping the DNA, or taking advantage of nucleosome unwrapping dynamics at the entry-exit sites (15, 16).

OCT4 binding is predicted to be incompatible with the nucleosome architecture on the basis of its engagement with free DNA (17–19), although partial motifs have been identified where OCT4 engages only a portion of its binding site to maintain nucleosome integrity (1). Despite being critical for genome

regulation, the structural and mechanistic principles governing nucleosome engagement by single or multiple TFs have yet to be determined.



**Fig. 1. SeEN-seq identifies preferred binding sites of OCT4 and OCT4-SOX2 across the nucleosome.**

(A) Principle of SeEN-seq. A library of TF motif-containing nucleosome positioning sequences is assembled into nucleosomes and incubated with TF(s). TF-bound and unbound nucleosome complexes are separated by EMSA and sequenced, revealing position-specific enrichments. Stars indicate a specific example sequence at each step of the assay. (B) SeEN-seq enrichment ( $n = 3$  replicates) for nucleosome pool with or without TFs. Further details are provided in fig. S1C. Motif position is indicated by superhelix location (SHL) that describes where the minor groove faces away (SHLs  $\pm 1$ ,  $\pm 2$ , etc.) or toward (SHLs  $\pm 1.5$ ,  $\pm 2.5$ , etc.) the histone octamer. (C) Autocorrelation analysis of OCT4 enrichment; dashed lines indicate 95% confidence interval. (D) DNA nucleosome unwrapping energy (15) versus OCT4 and OCT4-SOX2 SeEN-seq enrichment profile. Arrows indicate enriched positions in OCT4. DG, delta Gibbs free energy; kBT, 0.6 kcal/mol, where  $T = 300$  K.

<sup>1</sup>Friedrich Miescher Institute for Biomedical Research, Maulbeerstrasse 66, 4058 Basel, Switzerland. <sup>2</sup>Faculty of Science, University of Basel, Petersplatz 1, 4003 Basel, Switzerland. <sup>3</sup>Swiss Institute of Bioinformatics, 4058 Basel, Switzerland.

\*These authors contributed equally to this work.

†Present address: Monte Rosa Therapeutics, Aeschenvorstadt 36, 4051 Basel, Switzerland.

‡Corresponding author. Email: nicolas.thoma@fmi.ch (N.H.T.); dirk@fmi.ch (D.S.)

(NGS), which generated single molecule counts that approximate motif affinity as a function of position.

A range of protein concentrations of either OCT4, SOX2 (residues 37 to 118), or OCT4-SOX2 together were assessed, with a total of 4752 conditions measured with high reproducibility (Fig. 1B and fig. S1, A to C). The trends in OCT4 or SOX2 binding at selected positions were validated by fluorescence polarization measurements (fig. S2, A to E). Previously, two OCT4 motif locations were tested and found to provide similar OCT4 access to the nucleosome-embedded motif (27). Although this was recapitulated in SeEN-seq, our comprehensive examination of all motif locations reveals clear OCT4 preference for nucleosomal DNA entry-exit sites as well as discrete preferential binding sites ~1 to 3 bp in width throughout the nucleosome (Fig. 1B and fig. S1C). SOX2 shows less differential enrichment in SeEN-seq, in line with published data (10), with some degree of preferred binding toward the entry-exit sites and near the dyad (Fig. 1B and fig. S1C). Given the small enrichment amplitudes for SOX2 alone, we focused on the robust and differential binding activity seen for OCT4 and how it is affected by SOX2. OCT4 and SOX2 cooperate strongly, which results in up to 650-fold increased binding compared with that caused

by OCT4 alone (fig. S1C). This effect is most evident at entry-exit sites with weaker cooperativity observed at internal sites (fig. S1C).

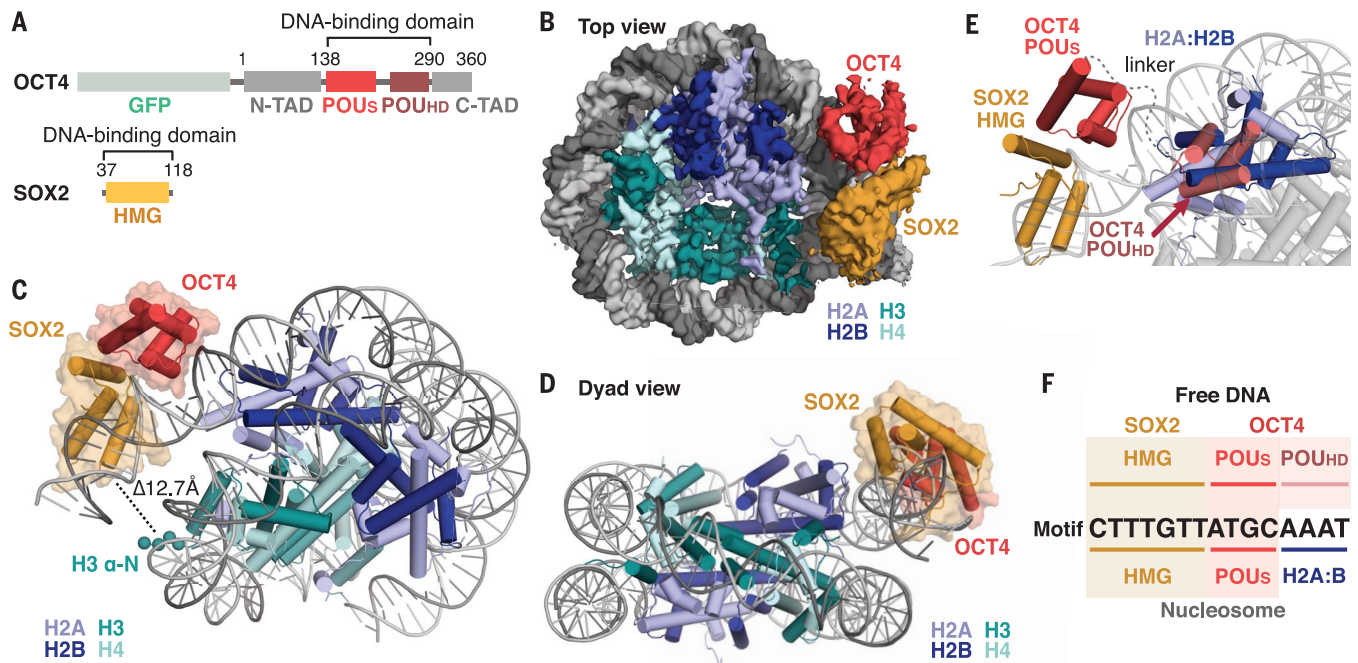
Although OCT4-SOX2 binding appears roughly symmetrical across the dyad, for OCT4 alone, enrichment in the right half of the nucleosome [superhelix locations (SHLs) +4 to +6.5] is more pronounced compared with that in the left (SHLs -4 to -6.5) (Fig. 1B and fig. S1C). A notable difference is the motif orientation on either side of the dyad; on the left, the OCT4 portion of the motif is closest to the dyad, whereas on the right it is oriented toward the entry-exit site of the nucleosome. Positions enriched for OCT4 alone and OCT4-SOX2 exhibited stronger binding at discrete sites with 10-bp periodicity across the nucleosome (Fig. 1C and fig. S2, F and G). Both OCT4 and OCT4-SOX2 show a trend of stronger binding at the entry-exit site than at the dyad (Fig. 1D). This would be expected if OCT4 and OCT4-SOX2 binding was facilitated by nucleosomal breathing, which is more pronounced toward the nucleosomal entry-exit sites (15). However, binding is not simply a function of nucleosomal breathing, because in the presence of OCT4 alone, motifs on the left half of the nucleosome (SHLs -6.5 to -6.0) are not bound across 5 bp, whereas adjacent motifs more proximal to the dyad are tightly bound (Fig. 1, B

and D, and fig. S2H). These entry-exit site loci are unbound by OCT4 alone, but they are bound cooperatively with SOX2 (Fig. 1D and fig. S2H). Thus, spatial orientation of the motif, cooperativity, and nucleosomal breathing dynamics all govern OCT4-SOX2 binding.

### Structure of OCT4-SOX2 bound at SHL -6 shows DNA release from the histone core

To dissect mechanisms of nucleosome engagement, we performed structural studies with OCT4-SOX2, which co-bind key target genes *in vivo* and have a cooperative effect *in vitro* (fig. S2, I to L) (22). For structural studies, we focused on sites that show discrete OCT4 SeEN-seq enrichment and cooperative binding in the presence of SOX2 (fig. S3A). A site 57 bp from the dyad (SHL -6) enabled structure determination from ~90,000 particles at an overall resolution of 3.1 Å (Fig. 2, A and B, and fig. S3). The nucleosome core and discrete domains of OCT4 [POU-specific (POUs)] and SOX2 [high-mobility group (HMG)] were well resolved (see fig. S3F for local resolution), which allowed conservative refinement with reference restraints to existing high-resolution OCT4 and SOX2 DNA-bound crystal structures (Fig. 2, B and C; fig. S4; and fig. S5, A to D) (18, 19).

In the OCT4-SOX2-NCP<sup>SHL-6</sup> structure, OCT4 and SOX2 together remove the DNA from core

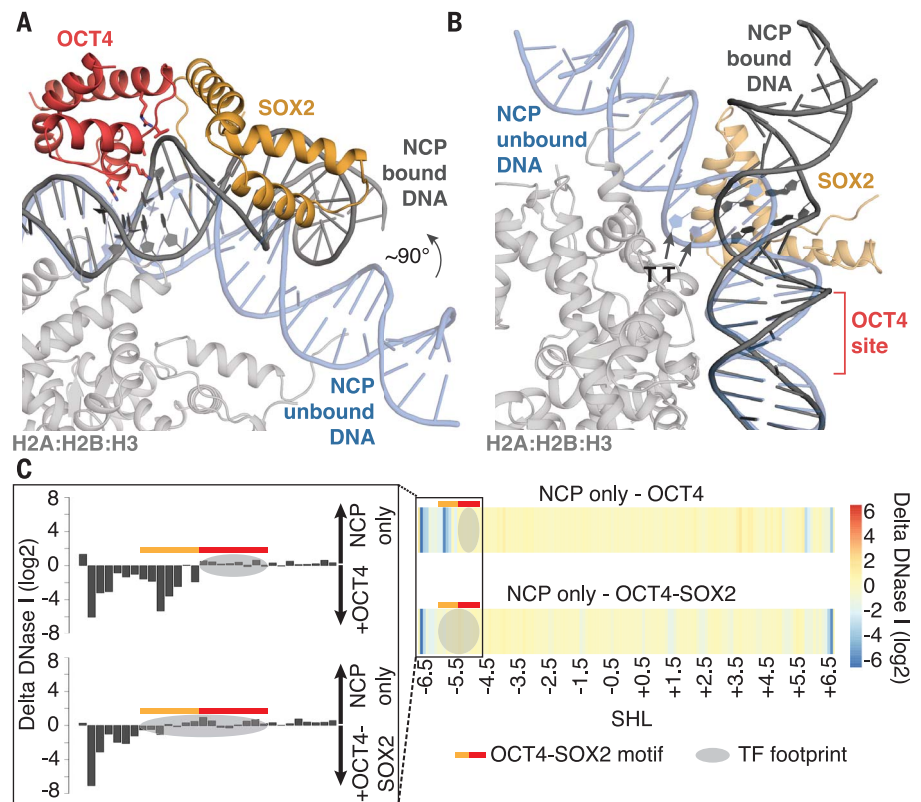


**Fig. 2. Cryo-EM structure of OCT4-SOX2-NCP<sup>SHL-6</sup> complex.** (A) Domain schematic of OCT4 and SOX2 constructs. GFP, green fluorescent protein; TAD, transactivation domain. (B) Cryo-EM map of OCT4-SOX2-NCP<sup>SHL-6</sup> at 3.1-Å resolution. (C) Model of the OCT4-SOX2-NCP<sup>SHL-6</sup> complex. The H3 N-terminal helix and tail (shown as spheres) stabilize DNA at the entry-exit in a canonical nucleosome (31). The presence of OCT4-SOX2 increases the distance between H3 (Arg<sup>41</sup>) and the nearest

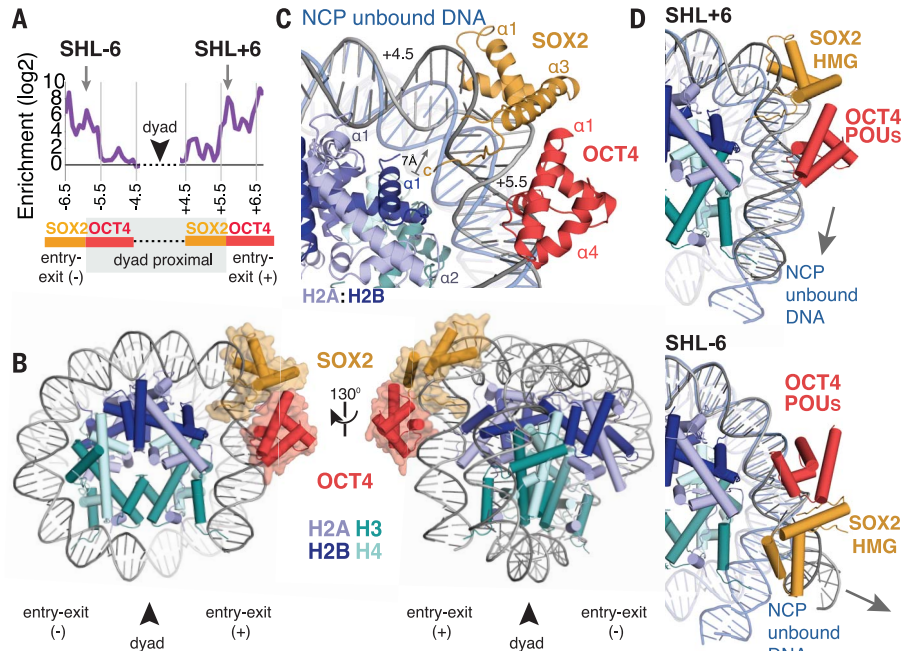
DNA base (T143) by 12.7 Å as compared with the unbound nucleosome. (D) Dyad view of the OCT4-SOX2-NCP<sup>SHL-6</sup> structure. (E) In the depicted OCT4-SOX2 arrangement, a model of POUHD engagement with its motif shows significant clash with the H2A:H2B dimer. Overlay of free DNA-bound structure (PDB: 104X) with the nucleosome-bound structure, aligned on the DNA (see also fig. S5E). (F) Schematic of free DNA versus the observed nucleosome-binding mode of OCT4-SOX2.



**Fig. 3. OCT4-SOX2 lifts the entry-exit DNA away from the histone core.** (A) Comparison of the unbound nucleosome DNA (blue) with the OCT4-SOX2 bound nucleosome structure (gray). The DNA is kinked  $\sim 90^\circ$  away from the histones. Residues at the OCT4-DNA interface are shown as sticks. POU5 motif nucleotides are shown as ribose and base rings. (B) SOX2 kinks the nucleosomal DNA away from the histones. SOX2 uses conserved (Phe<sup>48</sup> and Met<sup>49</sup>) residues to intercalate a TT base step, indicated by arrows (see also fig. S8A). OCT4 is removed for clarity. (C) Difference in DNase I digestion across the nucleosome in the presence of OCT4 or OCT4-SOX2 (right). Zoom-in view around the OCT4-SOX2 motif (left).



**Fig. 4. OCT4-SOX2 bound at SHL +6 induces localized DNA distortion.** (A) Depiction of OCT4-SOX2 SeEN-seq profile illustrating symmetric positions used for structure determination (arrows). As the OCT4-SOX2 motif is tiled across the nucleosome, the orientation of the proteins in relation to the dyad is inverted. (B) OCT4-SOX2-NCP<sup>SHL+6</sup> model. (C) Details of SOX2-induced DNA kink. SOX2 binding locally distorts DNA near SHL +5. The C terminus of SOX2 approaches the histone core when bound in this orientation, distorting DNA away from the stabilizing H2B residues [N terminus (Arg<sup>34</sup>) and  $\alpha 1$ ]. The 7-Å movement indicated is calculated between the phosphate backbone at Cys<sup>123</sup> in the bound and unbound DNA. SHLs are shown (SHLs +4.5 and +5.5). (D) Comparison of the DNA trajectory in the OCT4-SOX2-NCP<sup>SHL+6</sup> (SHL +6) (top) versus OCT4-SOX2-NCP<sup>SHL-6</sup> (SHL -6) (bottom) structures. The DNA trajectory at the entry-exit site of the TF-bound structures is indicated with an arrow.



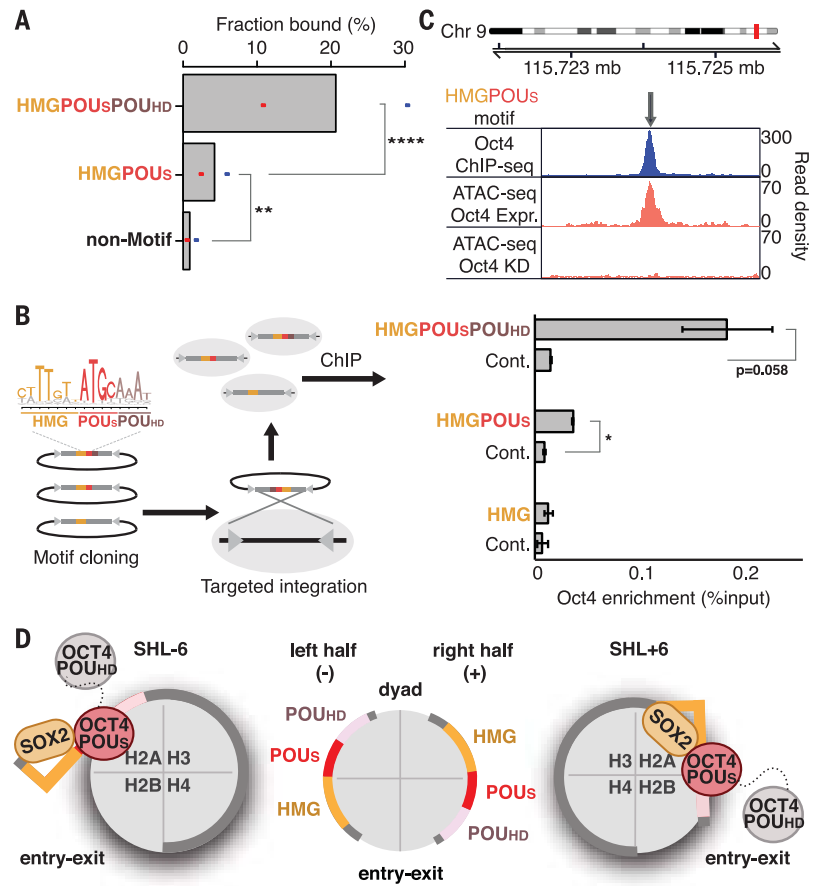
histones (Fig. 2, C and D). OCT4 has a bipartite DNA-binding domain composed of a POU5 and POU-homeodomain (POUHD) separated by 17 residues, whereas SOX2 utilizes an HMG domain (18). When nucleosome-bound, the OCT4-POU5 and SOX2-HMG DNA-binding

domains engage major and minor grooves, respectively, with protein-DNA interactions consistent with those previously seen for the individual OCT4-POU5 domain and SOX2 on free DNA (fig. S5E). The DNA remains attached and straightened around the OCT4 site but is

detached around the SOX2 motif (Fig. 2 and fig. S5F). SOX2 kinks the DNA and, together with OCT4, synergistically releases DNA from the core histones (movie S1). OCT4-SOX2 forms no discernable histone contacts, with the DNA separating them from the nucleosome core

**Fig. 5. The HMG-POUS partial motif is bound in vivo and opens chromatin.**

(A) Fraction of the top thousand motifs, ranked by match to the position weight motif, that are bound by Oct4 (2-fold ChIP-seq enrichment). Mean and individual values from two separate datasets are shown (33, 34). Significant enrichment is detected for full and partial motifs compared with loci without a motif (lowest 1000 genomic matches to the Oct4-Sox2 motif).  $**P < 0.01$ ;  $****P < 0.0001$ . (Pearson's chi-square test, indicated for dataset with least significance). (B) Schematic representation of targeted motif insertion and Oct4 ChIP-qPCR enrichment at the ectopic insertion locus and endogenous control (Cont.) locus.  $*P < 0.05$ ; error bars indicate SEM of at least two biological replicates. (C) Representative genomic region (chr9: 115723244 to 115725243) that shows binding of Oct4 to a partial HMG-POUS motif (top) and Oct4-dependent chromatin accessibility (bottom). Expr., Oct4-expressing cells; KD, Oct4-knockdown cells. (D) Graphical representation of the nucleosome distortion induced by OCT4-SOX2 when bound at dyad-symmetric sites.



(Fig. 2, C and D). Density for the OCT4 transactivation domains was not observed, which is consistent with similar nucleosome-affinity measurements for full-length OCT4 and OCT4 DNA-binding domain only (fig. S5G).

OCT4 recognizes a partial motif, engaging DNA with its POU domain, whereas the POUHD is not engaged. On free DNA, both POU domains engage the major groove spanning 8 bp on opposite sides of the DNA and would clash with either the histones or DNA gyres at all locations on the nucleosome (19, 28) (Fig. 2E and fig. S6A). To stabilize the OCT4-SOX2-NCP<sup>SHL-6</sup> complex for imaging, GraFix cross-linking was necessary and a cross-link was evident between H2A and H2B (fig. S6B) (29). To test if POUHD access was blocked by the cross-linking of histones, we solved an ~4-Å map of the non-cross-linked OCT4-SOX2-NCP<sup>SHL-6</sup> nucleosome (fig. S7, A to E), which resulted in a largely indistinguishable model (root mean square deviation, 1.3 Å; fig. S7, F and G). As the OCT4 POUHD motif is occluded by H2A-H2B, the binding mechanism involving only the OCT4 POU domain is consistent with partial motif engagement (Fig. 2F) (1). Partial motif recognition, however, does not necessarily render TF binding compatible with the nucleosomal DNA structure (fig. S5F).

In the context of the nucleosome, SOX2 competes with histones for DNA binding and kinks DNA by ~90° at SHL -6.5 away from the histones, similar to HMG domains on free DNA (Fig. 3, A and B, and fig. S8A) (18). This is accomplished by intercalation of the SOX2 Phe<sup>48</sup> and Met<sup>49</sup> wedge at the TT base step (18). Variant SOX2 motifs that lessen distortion of the DNA induced by SOX2 have been described and may facilitate nucleosome-compatible dyad binding (1, 10, 30). However, with the canonical SOX2 motif used here, SOX2 facing the entry-exit site is not compatible with the canonical nucleosome architecture and triggers DNA release (31).

Despite disruption of histone-DNA contacts at SHL -6.5, no histone rearrangements were observed after OCT4-SOX2 binding and DNA release across 14 bp (fig. S8, B and C). To verify OCT4-SOX2 binding and DNA release using an orthogonal approach, we performed deoxyribonuclease I (DNaseI) footprinting in the absence and presence of OCT4 or OCT4-SOX2. This revealed increased digestion at the nucleosomal entry-exit site (SHLs -7 to -5.5) in the presence of OCT4-SOX2, in line with DNA detachment and partial motif binding (Fig. 3C and fig. S9, A and B). Notably, OCT4 alone triggers DNA release, and OCT4 and OCT4-SOX2 also destabilize the opposite nucleo-

somal entry-exit site (SHLs +5.5 to +7), which is also evident in the cryo-electron microscopy (cryo-EM) map (Fig. 2 and fig. S3). The structures, footprinting profiles (Fig. 3), and thermal stability assays (fig. S9C) together support the idea that OCT4 and OCT4-SOX2 release DNA from the histones and have a global effect on the DNA structure of the nucleosome.

#### OCT4-SOX2 bound at SHL +6 induces minimal distortion to nucleosomal DNA

The SeEN-seq profile suggests that OCT4-SOX2 engagement depends on motif orientation (Fig. 1B). To examine this structurally, we utilized the same position but with the motif inverted across the dyad axis, i.e., SHL +6 (Fig. 4A). Doing so places the ~90° kink-inducing SOX2 motif in a dyad-proximal orientation and OCT4 closer to the entry-exit site. The SHL +6 site was enriched for OCT4-SOX2 binding in SeEN-seq, and the use of this position enabled cryo-EM structure determination of an OCT4-SOX2-NCP<sup>SHL+6</sup> complex at an overall resolution of 3.5 Å (Fig. 4B and fig. S10; see fig. S10E for local resolution). The map allowed unambiguous rigid body docking of the nucleosome and of SOX2 (figs. S11 and S12). OCT4 (POU) density was less continuous but sufficient to dock a C $\alpha$  model (fig. S11D). The resulting

OCT4-SOX2 interface was consistent with previously determined structures (Fig. 2, fig. S11C, and materials and methods).

In the structure, OCT4 engages its motif with the POU domain only (Fig. 4B), and the POUHD is unable to access its motif in the observed DNA configuration (fig. S13). OCT4-SOX2 together give rise to an extended DNA-binding surface across 11 bp that further bends the nucleosomal DNA by  $\sim 90^\circ$  at the SOX2 site (SHL +5) and straightens the DNA near the OCT4 site (SHL +6), producing an L-shaped DNA arrangement (Fig. 4B). This conformation locally lifts the duplex away from the histones and DNA gyre, but, in contrast to the reversed orientation, does not fully release the DNA from histones at the entry-exit site. At the SOX2 motif with the DNA locally detached, SOX2 helices 1 and 2 widen the minor groove, and the C terminus (residues 110 to 114) wedges between the DNA and histones (SHL +5) (Fig. 4C). Despite partial engagement of an internal-nucleosomal motif, the SOX2 and OCT4 (POUS) DNA interactions and induced DNA distortions are again similar to those previously seen on free DNA (fig. S11C). Within one helical turn, on either side of OCT4-SOX2, these DNA distortions are largely absorbed into the canonical DNA trajectory of the nucleosome (Fig. 4D). The histone core architecture again shows no substantial distortion. The DNA at the opposite nucleosomal entry-exit site (SHLs -7 to -5.5) appears to be disordered in the cryo-EM map (fig. S10E and fig. S14). In both structures (Figs. 2 and 4), OCT4 binds a partial DNA motif through its POU domain and, along with SOX2, affects the entire nucleosomal DNA structure to varying extents.

### The HMG-POUS partial motif is sufficient for TF engagement and the opening of chromatin in vivo

Previous work has identified that either *Oct4* or *Sox2* alone engage reduced-complexity motifs on nucleosomes during reprogramming (1). A recent study has also identified weaker-scoring motifs for SOX proteins on nucleosome-sized fragments (32). The structures now reveal that OCT4-SOX2 partial motif engagement is utilized in both orientations on the nucleosome. This led us to test whether OCT4-POUS and SOX2-HMG domains are sufficient to engage chromatin in vivo (Fig. 5A and fig. S15A). Through in-depth analysis of existing chromatin immunoprecipitation sequencing (ChIP-seq) datasets (33, 34), we found that the partial HMG-POUS motif is sufficient to drive genomic binding, although the full motif was bound more frequently (Fig. 5A). To test this experimentally, we introduced full and partial motifs at a defined genomic position in mouse embryonic stem cells (mESCs) by recombination-mediated cassette exchange

(35) (Fig. 5B). Motifs were introduced in the SHL -6 position of the 601 sequence (see materials and methods), and *Oct4* binding was determined. This revealed significant *Oct4* enrichment at both the full and partial motifs but not in the control (Fig. 5B and fig. S15B). Thus, single motifs recapitulate genome-wide *Oct4* binding to partial motifs.

Next, we asked if binding creates open chromatin. Comparing accessibility in mESCs (36) revealed that full and partial motifs can both generate accessible chromatin in an OCT4-dependent manner (Fig. 5C and fig. S15, C and D). Consistently, the same effect is evident upon knockdown of *Oct4* or *Sox2*, which shows that both TFs are required for full accessibility at these loci (37) (fig. S15, E to H). This confirms the cooperativity observed in SeEN-seq; however, the local effect is expected to be highly context dependent, as additional proteins contribute to accessibility in vivo (36). We interpret our structures to depict an initial encounter complex between OCT4-SOX2 and the nucleosome en route to open chromatin. Upon nucleosome removal, OCT4 is expected to engage a full motif with its two POU domains, thereby accounting for the stronger enrichment of the full versus the partial motif. Together, genome-wide binding, single-locus insertion, and genome-wide accessibility data demonstrate that the OCT4-POUS domain is sufficient to engage and open chromatin in conjunction with Sox2. This reveals the potential for such nucleosome-compatible motifs to function as bona fide binding sites beyond the ability to initially engage closed chromatin.

### Discussion

The structures illustrate binding mechanisms of OCT4-SOX2 at two positions on the nucleosome. At SHL  $\pm 6$ , the structures depict OCT4-SOX2 near the entry-exit sites, where both TFs cooperate to access DNA. At the SHL -6 site, where SOX2 faces the entry-exit site, OCT4-SOX2 releases the DNA duplex from the histones. In the OCT4-SOX2-NCP<sup>SHL+6</sup> structure, where SOX2 faces the dyad, the nucleosomal DNA assumes an L-shaped trajectory and is not fully released from the histones (Fig. 5D). The SHL -6 structure demonstrates that partial motif recognition and DNA release are not mutually exclusive (Figs. 2 and 3), whereas the SHL +6 structure depicts how more-internally bound sites can be accommodated without full removal of nucleosomal DNA ends (Fig. 4). We consistently find only the OCT4 POUS engaged, with the POUHD motif occluded by the nucleosome architecture. The OCT4 POUS and POUHD domains could, in principal, engage the full OCT4 motif if the DNA was further unwrapped from the histones, which we do not observe in our structures (Figs. 2 and 4) or DNaseI experiments (Fig. 3C). Thus, partial motif recognition allows TFs to minimize DNA

unwrapping when engaging nucleosomal sites, although partial motifs do not fully preempt nucleosome distortions. We show that partial OCT4 motifs, in conjunction with SOX2, are recognized in vivo and create open chromatin (Fig. 5).

SeEN-seq binding profiles, combined with the structural data, allow us to rationalize OCT4-SOX2 engagement throughout the nucleosome. OCT4 and OCT4-SOX2 bind best at the nucleosomal entry-exit sites, where DNA breathing is expected to facilitate access. We observe distinct exceptions from such end-binding behavior for OCT4 in particular, where narrow regions of pronounced binding are juxtaposed to nonengaged regions. To correlate these accessibility profiles to the structure, we computationally translated isolated OCT4 POU or POUHD domains along the DNA of the unbound nucleosome and calculated predicted atomic clash scores at each position (fig. S16). A comparison of OCT4 SeEN-seq data to the POU domain nucleosome-clash scores revealed that the solvent accessibility of the POU—but not of the POUHD—correlates with OCT4 binding (fig. S16, A and B). The solvent accessibility for OCT4 POU is also a good predictor for OCT4-SOX2 engagement (fig. S16, C and D). The presence of SOX2 enables tight OCT4-SOX2 binding at the nucleosome ends but has a limited effect on more-internal sites (Fig. 1). At nucleosome ends, SOX2 can also drive binding at motifs where the OCT4 POU is inward-facing and OCT4 alone does not bind (fig. S2H). Our structural and functional findings are consistent with a model where cooperative binding between OCT4 and SOX2 not only strengthens DNA binding affinity but also triggers additional DNA distortions that must be accommodated. These distortions are better tolerated at the entry-exit sites, where nucleosomal DNA is more loosely bound (38). Whereas partial motifs delimit the TF footprint and DNA unwrapping, the available structures (13) show that protein domains bound to nucleosomal DNA retain their free DNA-binding mode (Figs. 2 and 4). TF-induced DNA distortions intrinsically destabilize the nucleosome (10), which likely facilitates the binding of additional factors and disrupts the internucleosomal interactions of higher-order chromatin (15, 39) (fig. S17).

The OCT4-SOX2 structures and the accompanying in vitro and in vivo evidence provide a framework by which TFs use nucleosomal DNA distortion and not histone rearrangement to access parts of their motif. The degree of DNA distortion imposed on the nucleosome architecture depends on the position of the motif. Our structures suggest principal recognition mechanisms for nucleosome-incompatible TFs as well as for those TFs accommodated on the nucleosome without DNA release, illustrating how TFs can read out chromatinized binding sites.

## REFERENCES AND NOTES

1. A. Soufi *et al.*, *Cell* **161**, 555–568 (2015).
2. D. J. Rodda *et al.*, *J. Biol. Chem.* **280**, 24731–24737 (2005).
3. K. Takahashi, S. Yamanaka, *Cell* **126**, 663–676 (2006).
4. V. Malik *et al.*, *Nat. Commun.* **10**, 3477 (2019).
5. D. C. Ambrosetti, C. Basilico, L. Dailey, *Mol. Cell. Biol.* **17**, 6321–6329 (1997).
6. T. Kumar Mistri *et al.*, *Biophys. J.* **100**, 74a (2011).
7. C. C. Adams, J. L. Workman, *Mol. Cell. Biol.* **15**, 1405–1421 (1995).
8. L. A. Mirny, *Proc. Natl. Acad. Sci. U.S.A.* **107**, 22534–22539 (2010).
9. K. S. Zaret, J. S. Carroll, *Genes Dev.* **25**, 2227–2241 (2011).
10. F. Zhu *et al.*, *Nature* **562**, 76–81 (2018).
11. M. Fernandez Garcia *et al.*, *Mol. Cell* **75**, 921–932.e6 (2019).
12. R. K. McGinty, S. Tan, *Chem. Rev.* **115**, 2255–2273 (2015).
13. S. Matsumoto *et al.*, *Nature* **571**, 79–84 (2019).
14. L. A. Cirillo *et al.*, *EMBO J.* **17**, 244–254 (1998).
15. B. Fierz, M. G. Poirier, *Annu. Rev. Biophys.* **48**, 321–345 (2019).
16. G. Li, M. Levitus, C. Bustamante, J. Widom, *Nat. Struct. Mol. Biol.* **12**, 46–53 (2005).
17. J. Huertas, C. M. MacCarthy, H. R. Schöler, V. Cojocaru, *Biophys. J.* S0006-3495(20)30032-1 (2020).
18. A. Reményi *et al.*, *Genes Dev.* **17**, 2048–2059 (2003).
19. D. Esch *et al.*, *Nat. Cell Biol.* **15**, 295–301 (2013).
20. X. Yu, M. J. Buck, *Genome Res.* **29**, 107–115 (2019).
21. G. D. Stormo, Z. Zuo, Y. K. Chang, *Brief. Funct. Genomics* **14**, 30–38 (2015).
22. L. A. Boyer *et al.*, *Cell* **122**, 947–956 (2005).
23. X. Chen *et al.*, *Cell* **133**, 1106–1117 (2008).
24. N. Tapia *et al.*, *Sci. Rep.* **5**, 13533 (2015).
25. A. Khan *et al.*, *Nucleic Acids Res.* **46**, D260–D266 (2018).
26. P. T. Lowary, J. Widom, *J. Mol. Biol.* **276**, 19–42 (1998).
27. S. Li, E. B. Zheng, L. Zhao, S. Liu, *Cell Rep.* **28**, 2689–2703.e4 (2019).
28. D. C. Williams Jr., M. Cai, G. M. Clore, *J. Biol. Chem.* **279**, 1449–1457 (2004).
29. H. Stark, *Methods Enzymol.* **481**, 109–126 (2010).
30. P. Scaffidi, M. E. Bianchi, *J. Biol. Chem.* **276**, 47296–47302 (2001).
31. K. Luger, A. W. Mäder, R. K. Richmond, D. F. Sargent, T. J. Richmond, *Nature* **389**, 251–260 (1997).
32. M. P. Meers, D. H. Janssens, S. Henikoff, *Mol. Cell* **75**, 562–575.e5 (2019).
33. C. Chronis *et al.*, *Cell* **168**, 442–459.e20 (2017).
34. Z. Liu, W. L. Kraus, *Mol. Cell* **65**, 589–603.e9 (2017).
35. F. Lienert *et al.*, *Nat. Genet.* **43**, 1091–1097 (2011).
36. H. W. King, R. J. Klose, *eLife* **6**, e22631 (2017).
37. E. T. Friman *et al.*, *eLife* **8**, e50087 (2019).
38. M. A. Hall *et al.*, *Nat. Struct. Mol. Biol.* **16**, 124–129 (2009).
39. T. Schalch, S. Duda, D. F. Sargent, T. J. Richmond, *Nature* **436**, 138–141 (2005).

## ACKNOWLEDGMENTS

**Funding:** N.H.T. and D.S. acknowledge support from the Novartis Research Foundation, the European Research Council under the European Union's (EU) Horizon 2020 research and innovation program grant agreement (N.H.T., 666068; D.S., 667951), and the Swiss National Science Foundation (N.H.T., Sinergia-CRSII3\_160734/1 and SNF 31003A\_179541; D.S., 310030B\_176394). A.K.M. acknowledges the Human Frontier Science Program. A.K.M. and R.S.G. acknowledge EMBO Long-Term Fellowships. R.S.G. and L.I. acknowledge the EU Horizon 2020 Research and Innovation Program under the Marie Skłodowska-Curie grant (R.S.G., 705354; L.I., 748760). L.I. acknowledges National Health and Medical Research Council CJ Martin Fellowship APP1148380. Z.K. acknowledges EU Horizon 2020 Marie Skłodowska-Curie ESR grant 765445. **Author contributions:** A.K.M., R.S.G., and L.I. developed SeEN-seq. L.I. and R.S.G. performed NGS library preparation, data analysis, and recombinase-mediated cassette exchange (RMCE) insertions. A.K.M. purified

proteins, assembled nucleosomes, and performed EMSA and biochemical assays. L.I. and L.B. performed analysis of periodicity and genomic datasets. A.K.M. prepared cryo-EM samples. A.K.M. and S.C. performed cryo-EM and analysis. Z.K. performed fluorescence polarization assays. G.K. and R.D.B. prepared atomic models with A.K.M. and S.C. R.S.G. performed DNaseI footprinting assay and analysis. A.D.S. and A.G.-M. provided technical support for cryo-EM. G.R.P., J.W., and S.M. contributed nucleosome preparations and reagents. A.K.M., R.S.G., L.I., D.S., and N.H.T. wrote the manuscript. The research was directed by D.S. and N.H.T. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** Plasmids and cell lines generated by this study are available from the Friedrich Miescher Institute for Biomedical Research under a material transfer agreement. The electron density reconstructions and final models were deposited with the EM Database (accession codes: EMD-10406, EMD-10408, and EMD-10864) and with the Protein Data Bank (PDB) (accession codes: 6T90, 6T93, and 6YOV). All other data are available in the main text or the supplementary materials.

## SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/368/6/498/1460/suppl/DC1  
Materials and Methods  
Figs. S1 to S17  
Table S1  
References (40–83)  
MDAR Reproducibility Checklist  
Movie S1

[View/request a protocol for this paper from Bio-protocol.](#)

22 January 2020; accepted 16 April 2020  
Published online 23 April 2020  
10.1126/science.abb0074

# Structural mechanism of cGAS inhibition by the nucleosome

<https://doi.org/10.1038/s41586-020-2750-6>

Received: 25 February 2020

Accepted: 22 July 2020

Published online: 10 September 2020

 Check for updates

Ganesh R. Pathare<sup>1,2,5</sup>, Alexiane Decout<sup>3,5</sup>, Selene Glück<sup>3</sup>, Simone Cavadini<sup>1,2</sup>, Kristina Makasheva<sup>4</sup>, Ruud Hovius<sup>4</sup>, Georg Kempf<sup>1,2</sup>, Joscha Weiss<sup>1,2</sup>, Zuzanna Kozicka<sup>1,2</sup>, Baptiste Guey<sup>3</sup>, Pauline Melenec<sup>3</sup>, Beat Fierz<sup>4</sup>, Nicolas H. Thomä<sup>1,2,✉</sup> & Andrea Ablasser<sup>3,✉</sup>

The DNA sensor cyclic GMP–AMP synthase (cGAS) initiates innate immune responses following microbial infection, cellular stress and cancer<sup>1</sup>. Upon activation by double-stranded DNA, cytosolic cGAS produces 2'3' cGMP–AMP, which triggers the induction of inflammatory cytokines and type I interferons<sup>2–7</sup>. cGAS is also present inside the cell nucleus, which is replete with genomic DNA<sup>8</sup>, where chromatin has been implicated in restricting its enzymatic activity<sup>9</sup>. However, the structural basis for inhibition of cGAS by chromatin remains unknown. Here we present the cryo-electron microscopy structure of human cGAS bound to nucleosomes. cGAS makes extensive contacts with both the acidic patch of the histone H2A–H2B heterodimer and nucleosomal DNA. The structural and complementary biochemical analysis also find cGAS engaged to a second nucleosome in *trans*. Mechanistically, binding of the nucleosome locks cGAS into a monomeric state, in which steric hindrance suppresses spurious activation by genomic DNA. We find that mutations to the cGAS–acidic patch interface are sufficient to abolish the inhibitory effect of nucleosomes in vitro and to unleash the activity of cGAS on genomic DNA in living cells. Our work uncovers the structural basis of the interaction between cGAS and chromatin and details a mechanism that permits self–non-self discrimination of genomic DNA by cGAS.

In the cytoplasm of mammalian cells, the enzyme cGAS is crucial for the detection of double-stranded DNA (dsDNA) during infection<sup>2</sup>. On binding dsDNA, cGAS synthesizes the second messenger 2'3' cyclic GMP–AMP (cGAMP), which in turn stimulates antiviral and pro-inflammatory responses through the adaptor protein stimulator of interferon genes (STING)<sup>2–7,10</sup>. In addition, the nucleus contains a pool of cGAS that associates strongly with chromatin (Extended Data Fig. 1a)<sup>8,9,11</sup>. The chromatinized state of intact genomic DNA has been reported to limit cGAS activity<sup>12,13</sup>, and cGAS has been found to bind more tightly to nucleosomes than to the corresponding naked DNA duplexes<sup>13</sup>. Here we sought a mechanism that explains how cGAS can be juxtaposed to nucleosomal DNA without undergoing activation.

## A minimal inhibitory histone unit for cGAS

We tested whether histones, the building blocks of nucleosomes, may regulate cGAS inside the nucleus. Treatment of cells with aclarubicin robustly evicts core histones from chromatin<sup>14</sup>; in particular, histones H2A and H2B (Extended Data Fig. 1b). Notably, disruption of nucleosomes by aclarubicin also led to mobilization of nuclear cGAS (Extended Data Fig. 1b). Proximity-ligation assays (PLAs) further indicated prominent association of cGAS with histones in situ, which was

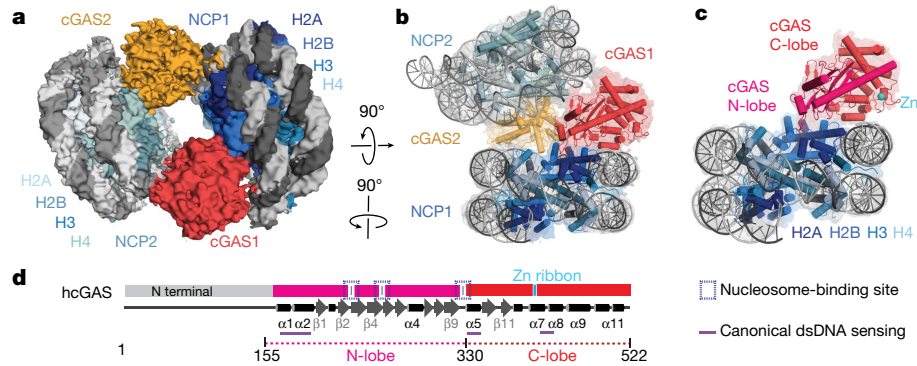
partially lost upon aclarubicin treatment (Extended Data Fig. 1c, d). Thus, histones appear to dynamically engage cGAS in the nucleus.

Consistent with previous work<sup>13</sup>, functional analysis of the in vitro enzymatic activity of cGAS revealed that mononucleosomes (hereafter nucleosomes) inhibited DNA-induced cGAMP synthesis (Extended Data Fig. 1e). Likewise, compact chromatin fibres (12-mer nucleosome arrays) suppressed cGAS activity (Extended Data Fig. 1e). H2A–H2B dimers also had an inhibitory effect, but neither H2A or H2B monomers nor H3 or H4 monomers did (Extended Data Fig. 1f, g). Thus, H2A–H2B dimers on their own can suppress cGAS, albeit with weaker overall potency than fully assembled nucleosomes (Extended Data Fig. 1h). Additional features of chromatin are therefore necessary to exert maximal inhibition.

## Overall structure of the cGAS–NCP complex

To determine how cGAS interacts with nucleosomes, we pursued structural studies. A 1.5:1 molar mixture of human cGAS (residues 155–522) with a 147-bp 601 DNA nucleosome core particle (NCP) resulted in heterogeneous particle distribution (Extended Data Fig. 2a–d). To select for and stabilize more homogeneous cGAS–NCP complexes, we combined gradient centrifugation with chemical crosslinking (GraFix)<sup>15</sup>. Both wild-type (WT) cGAS and cGAS K394E, a mutant

<sup>1</sup>Friedrich Miescher Institute for Biomedical Research, Basel, Switzerland. <sup>2</sup>University of Basel, Basel, Switzerland. <sup>3</sup>Global Health Institute, Swiss Federal Institute of Technology Lausanne (EPFL), Lausanne, Switzerland. <sup>4</sup>Institute of Chemical Sciences and Engineering, Swiss Federal Institute of Technology Lausanne (EPFL), Lausanne, Switzerland. <sup>5</sup>These authors contributed equally: Ganesh R. Pathare, Alexiane Decout. ✉e-mail: nicolas.thoma@fmi.ch; andrea.ablasser@epfl.ch

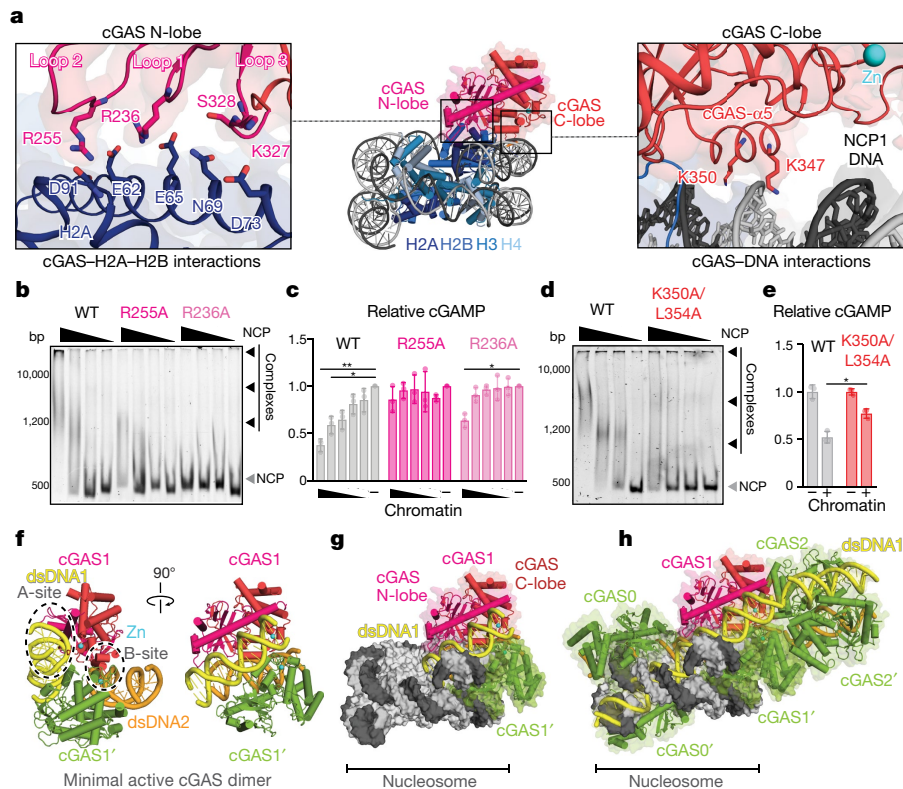


**Fig. 1 | Cryo-electron microscopy structure of cGAS bound to nucleosomes.** **a**, 3D reconstruction of the complex containing two cGAS protomers, cGAS1 (red) and cGAS2 (orange), and two nucleosomal core particles, NCP1 and NCP2, respectively. **b**, **c**, Ribbon diagrams of the NCP1-cGAS1-cGAS2-NCP2 complex (**b**) and the cGAS1-NCP1 complex (**c**) fit into corresponding electron-density maps. The two lobes of cGAS, N-lobe and C-lobe, are

shown in pink and red, respectively. **d**, Schematic domain architecture for human cGAS (hcGAS) as previously defined<sup>23</sup>. Residue numbers are shown and the dotted line indicates the construct used for structural analysis. The dsDNA-sensing regions that are involved in cGAS activation are underlined in purple, and the nucleosome-binding regions that are involved in cGAS inhibition are marked by blue dashed boxes.

impaired in dsDNA-mediated cGAS dimerization<sup>16</sup>, were used for structure determination. For the cGAS K394E mutant, we obtained a 4.1 Å reconstruction that revealed two NCPs organized in an NCP1-cGAS1-cGAS2-NCP2 sandwich arrangement with an expected molecular

weight of around 560 kDa, consistent with the most prominent peak fraction in multi-angle light scattering under non-crosslinked conditions (Fig. 1a, b, Extended Data Fig. 3, Extended Data Table 1a, Supplementary Videos 1, 2). The two individual nucleosomes are held



**Fig. 2 | The cGAS1-NCP1 complex and structural mechanism of inhibition.**

**a**, Magnified view of the cGAS1(K394E)-NCP1 complex bipartite interactions, cGAS-histone interactions (left) and cGAS-nucleosomal DNA interactions (right). **b**, EMSA gel showing the interaction of nucleosomes (40 ng  $\mu\text{l}^{-1}$ ) with a concentration gradient of WT, R255A cGAS and R236A cGAS (from 50 to 6 ng  $\mu\text{l}^{-1}$ ; 1:2 step dilutions); the black arrowheads indicate higher-order cGAS-NCP complexes. **c**, In vitro cGAMP synthesis of WT cGAS, R255A cGAS and R236A cGAS with or without a concentration gradient of chromatin (from 5 to 0.3125 nM; 1:2 step dilutions) normalized by cGAMP levels in the absence of chromatin for each individual mutant. **d**, EMSA gel showing the interaction of nucleosomes (40 ng  $\mu\text{l}^{-1}$ ) with increasing concentrations of WT or K350A/L354A cGAS (from 100 to 12 ng  $\mu\text{l}^{-1}$ ; 1:2 step dilution). **e**, In vitro cGAMP synthesis of WT and K350A/L354A hcGAS with or without chromatin (5 nM) normalized by cGAMP levels in the absence of

chromatin for each individual mutant. Data are representative for three independent experiments showing similar results (**b**, **d**) or mean  $\pm$  s.d. of  $n = 3$  independent experiments (**c**, **e**). One-way analysis of variance (ANOVA) with post-hoc Dunnett multiple comparison test: \*\* $P = 0.0092$ , \* $P = 0.092$  (WT) and \* $P = 0.0311$  (R236A) (**c**) and two-tailed Student's  $t$ -test; \* $P = 0.0192$  (**e**). The data points are from independent experiments. **f**, Overview of active hcGAS-DNA 2:2 complex with two distinct dsDNA-binding surfaces (A-site and B-site)<sup>16</sup> (Protein Data Bank (PDB) ID: 4LEY). **g**, Superposition of the hcGAS-dsDNA (**f**) and cGAS1-NCP1 complexes illustrating the incompatibility of DNA ligand binding (dsDNA1 in yellow) to cGAS in the nucleosome-bound configuration. **h**, Model based on superpositioning of the cGAS1-NCP1 complex onto DNA-bound cGAS oligomers as previously defined<sup>22</sup> (PDB: 5N6I) (**h**). For gel source data, see Supplementary Fig. 1.

# Article

together by two cGAS protomers. While the first cGAS protomer and its corresponding NCP (designated cGAS1 and NCP1) are well resolved, the second nucleosome–cGAS pair (NCP2 and cGAS2) is less ordered (Extended Data Fig. 3e). In the dimeric NCP1–cGAS1–cGAS2–NCP2 arrangement, each cGAS protomer interacts with the histone octamer of one NCP through histones H2A and H2B and the nucleosomal DNA (for example, cGAS1 and NCP1), while contacting the second nucleosome (for example, cGAS1 and NCP2) primarily through interactions with the nucleosomal DNA (Fig. 1a, b). In the WT cGAS structure, we observed a similar overall structural arrangement, with the NCP1–cGAS1–cGAS2–NCP2 complex at 5.1 Å and a NCP1–cGAS1 structure at 4.7 Å resolution following focused 3D classification (Extended Data Figs. 4, 5, Extended Data Table 1). Given the structural similarity, the higher-resolution cGAS K394E mutant was used for subsequent analysis (Extended Data Figs. 3, 6).

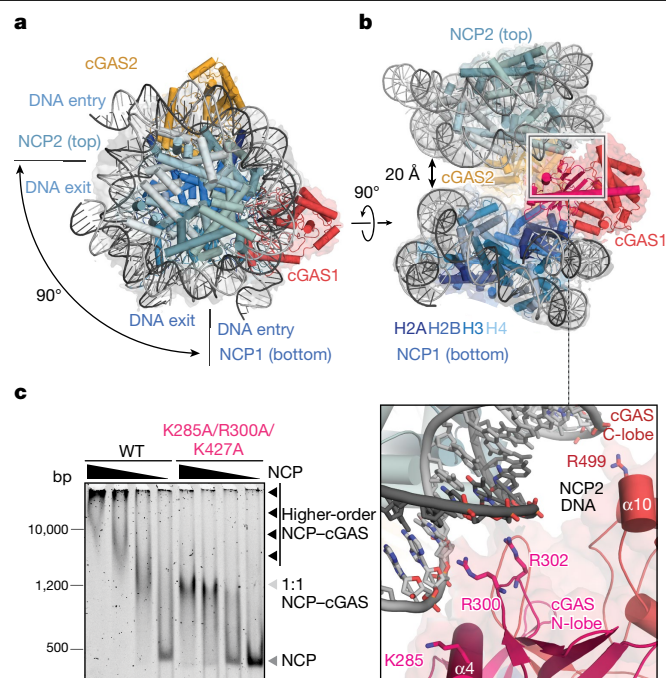
## Structural insights into the cGAS1–NCP1 complex

For the cGAS K394E mutant, focused 3D classification of cGAS1–NCP1 yielded a structure at a resolution of 3.1 Å (Fig. 1c, Extended Data Fig. 3c, f, Extended Data Table 1a). This revealed the binding interface between cGAS and the nucleosome, which is in large parts contributed by three contact surfaces on cGAS that interact with the acidic patch of H2A–H2B—a common site involved in protein–nucleosome assemblies<sup>17</sup> (Fig. 2a, Extended Data Fig. 6a–d, Supplementary Video 3): (1) loop 1 (residues 234–237), a canonical acidic patch contact involving R236 of cGAS forming a salt bridge to E62 and E65 of H2A (Fig. 2a, Extended Data Fig. 6b); (2) loop 2 (residues 255–258), a proximal cGAS loop that connects  $\beta$ 4 and  $\alpha$ 4, with R255 forming a salt bridge to H2A residues D91, E93 and E62 (Fig. 2a, Extended Data Fig. 6c); and (3) loop 3 (residues 328–330), which together with the C-terminal end of cGAS helix  $\alpha$ 5 (residues 354–356) forms multiple interactions with H2A helix  $\alpha$ 3 (residues 64–71) (Fig. 2a, Extended Data Fig. 6d). In addition to these protein–protein interactions, the cGAS1 C-lobe also forms localized DNA backbone contacts with NCP1, engaging the nucleosomal DNA around super-helical location 6 (SHL6) through residues K347 and K350<sup>18</sup> (Fig. 2a (right panel), Extended Data Fig. 6e, f).

To investigate the functional importance of the observed interfaces between cGAS and NCP1, we performed site-directed mutagenesis and carried out electromobility shift assays (EMSA). Mutations of cGAS residues that engage the nucleosomal acidic patch (R255A and R236A) completely abrogated nucleosome binding (Fig. 2b). In *in vitro* enzymatic activity, we found that both cGAS R255A and cGAS R236A were no longer inhibited by chromatin (Fig. 2c). To corroborate the relevance of the acidic patch interaction for cGAS inhibition, we made use of a peptide derived from latency-associated nuclear antigen (LANA), a well-known acidic patch binder<sup>19</sup>. In the presence of the LANA peptide, but not a corresponding mutant peptide, cGAS was competed off from the nucleosome and regained *in vitro* DNA-induced activity in the presence of chromatin, as judged by robust cGAMP synthesis (Extended Data Fig. 7a, b). Of note, cGAS residues K350 and L354, which contact the nucleosomal DNA of NCP1 (Extended Data Fig. 6f, i), also had significant effects on nucleosome binding and inhibition when mutated (Fig. 2d, e, Extended Data Fig. 6f, i). Thus, cGAS is anchored to chromatin through a bipartite interface on nucleosomes composed of the acidic patch and nucleosomal DNA contacts, respectively.

## Mechanism of cGAS inhibition by NCP1

In canonical binding of dsDNA, two separate surfaces on cGAS, designated A-site and B-site, interact with two individual strands of DNA to promote the assembly of a 2:2 cGAS–DNA complex—the minimal active enzymatic unit<sup>20–23</sup> (Fig. 2f). Moreover, a third DNA-binding site, designated C-site, has been proposed to facilitate cGAS oligomerization

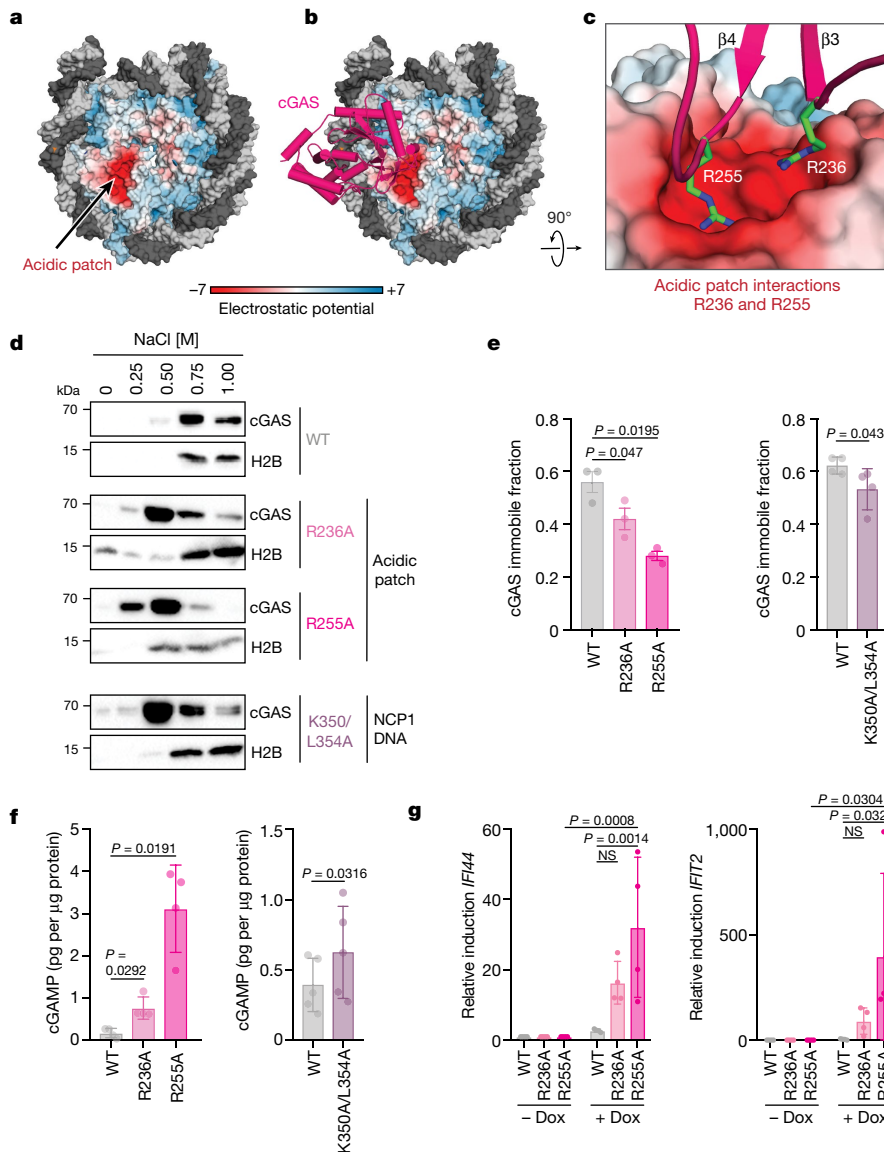


**Fig. 3 | cGAS interactions with the second nucleosome in *trans*.** **a, b**, The NCP1–cGAS1–cGAS2–NCP2 di-nucleosomal arrangement is shown. **a** Magnified view detailing the interactions between the N-lobe of cGAS1 (pink), the C-lobe of cGAS1 (red) and the nucleosomal DNA of NCP2 (grey) (**b**, bottom) is also displayed. **c**, EMSA gel showing the interaction of nucleosomes with increasing concentrations of WT or K285A/R300A/K427A cGAS (100 to 12 ng  $\mu$ l<sup>-1</sup>; 1:2 step dilution). The arrowheads highlight free nucleosomes (dark grey), complexed nucleosomes (black) and a putative 1:1 cGAS:NCP assembly (light grey). The experiment shown in **c** was independently repeated three times with similar results. For gel source data, see Supplementary Fig. 1.

in liquid-phase condensation<sup>24</sup>. In the NCP-bound configuration, the cGAS A-site, including the zinc thumb, faces the histone octamer disc. Nucleosomal DNA interactions are further enforced by residues that are essential for cGAS dimerization (for example, K394) (Extended Data Fig. 4c, d), although the K394-containing loop and the zinc-finger motif play only a minor role in nucleosome binding (Extended Data Figs. 4c, d, 7f, g). The key cGAS–NCP interaction originates from the B-site (for example, R236, K254, R255 and S328), which also contributes to nucleosomal DNA binding (for example, K347 and L354). The cGAS active site in our structure points away from NCP1, towards the solvent and NCP2, and is principally accessible (Fig. 2).

Nucleosome binding hence confers cGAS inactivation in three essential ways (Fig. 2g, h, Supplementary Video 4): first, owing to steric clashes with both the nucleosomal DNA and histones H2A and H2B, cGAS cannot engage dsDNA at the interface between lobe 1 and lobe 2; second, key residues on cGAS that are required for DNA binding and dimerization are tied up in interactions with the nucleosome and, thus, are not available for canonical dsDNA binding and activation (Extended Data Fig. 4c, d); and third, both histones and nucleosomal DNA sterically prevent dimerization of cGAS, an essential prerequisite for enzymatic activity<sup>16,21,22</sup>. Importantly, the steric restrictions imparted by H2A–H2B are sufficiently pronounced to explain the inability of cGAS to undergo dsDNA-dependent activation in the presence of this histone dimer. The structure thereby provides the mechanism of cGAS inhibition by H2A–H2B, while identifying additional contacts and inhibitory principles that are specific to cGAS inhibition by the nucleosome (Extended Data Fig. 6g–i).

We next dissected the contributions of nucleosomal DNA as opposed to linker DNA to cGAS binding and activation. Fluorescence polarization assays revealed that cGAS binds more tightly to nucleosomes with long



**Fig. 4 | Effect of structure-based mutations on cellular cGAS activity.** **a, b**, Electrostatic surface representation of the NCP disc surface alone (**a**) or with cGAS (pink ribbon) (**b**)<sup>31</sup>. The electrostatic potential is shown from red (-7) to blue (+7) k T/e. **c**, A magnified view of contacts between cGAS and the acidic patch of the nucleosome. **d**, Differential nuclear salt fractionation probed for cGAS and H2B by immunoblot from HeLa cGAS KO cells reconstituted with doxycycline-inducible WT cGAS or cGAS mutants after 2 days of doxycycline treatment. The experiments in **d** were independently repeated at least three times with similar results. **e**, HeLa cGAS KO cells were transfected with WT cGAS–GFP or mutant cGAS–GFP and the immobile fraction of nuclear-localized cGAS was assessed by FRAP. Data are mean  $\pm$  s.d. of  $n = 3$  (left) and  $n = 4$  (right) independent experiments. One-way ANOVA with post-hoc Dunnett multiple comparison test (left) or two-tailed Student’s *t*-test (right). **f**, cGAMP production from HeLa cGAS KO cells reconstituted with doxycycline-inducible WT cGAS or cGAS mutants after 2 days of doxycycline treatment. Data are mean  $\pm$  s.d. of  $n = 4$  (left) and  $n = 5$  (right) independent experiments. One-way ANOVA with post-hoc Dunnett multiple comparison test (left) or two-tailed Student’s *t*-test (right). **g**, HeLa cGAS KO cells reconstituted with doxycycline (Dox)-inducible WT cGAS or cGAS mutants and treated with doxycycline for 24 h were co-cultured with BJ fibroblasts for 24 h. Cells were lysed and mRNA levels of *IFI44* (left) and *IFIT2* (right) were assessed as indicated. Data are presented as fold induction relative to cells without doxycycline and shown are the mean  $\pm$  s.d. of  $n = 4$  independent experiments. Two-way ANOVA with post-hoc Tukey multiple comparison test; NS, not significant. Individual data points are from biological replicates. For gel source data, see Supplementary Fig. 1.

overhangs than to those without or with only short overhangs, probably owing to the presence of additional DNA-binding sites (Extended Data Fig. 7d, e). We then assessed the catalytic activity of cGAS (WT) and the cGAS acidic patch mutants, R236A and R255A, on nucleosomes with and without a 80-bp dsDNA overhang. Whereas WT cGAS and cGAS mutants robustly synthesized cGAMP on naked dsDNA, they remained inactive in the presence of nucleosomes that lacked a DNA overhang (Extended Data Fig. 7c). Thus, nucleosomal DNA is not a good substrate for cGAS activation. Nucleosomes carrying 80-bp long linker DNA still failed to activate WT cGAS, but elicited activation of both cGAS R236A and cGAS R255A (Extended Data Fig. 7c). Thus, *in vitro* WT cGAS preferentially binds to the NCP over linker DNA, limiting its enzymatic activity.

### In trans interaction between cGAS and NCP2

The cGAS1 protomer also binds to a second nucleosome (for example, cGAS1 and NCP2) predominantly through protein–DNA contacts with the nucleosomal DNA at SHL3. The two NCPs are held about 20 Å apart, with the DNA entry or exit sites of the two nucleosomes pointing roughly 90° away (Fig. 3a). The interaction with the second nucleosome is mediated by phosphate backbone contacts involving conserved cGAS1 residues K285 and K299–R302 (Fig. 3b). Compared to

the protein–DNA contacts, fewer and less well-ordered cGAS1–NCP2 protein–protein interactions were observed between a  $\beta$ -hairpin loop (cGAS residues 365–369) extending towards the C-terminal tail of H2B, as well as inter-nucleosomal contacts between two N-terminal tails of histone H4 (Extended Data Fig. 8a, b). The *in trans* nucleosome interaction interface on cGAS is largely provided by the C-site<sup>24</sup> (Extended Data Fig. 8c–f), which supports higher-order cGAS–NCP assemblies in this structure.

To validate the observed *in trans* cGAS1–NCP2 interface, we performed EMSAs. We found that the combined mutation of the NCP2-interacting motifs on cGAS (K285, R300 and K427) still allowed cGAS to interact with nucleosomes, as indicated by a prominent EMSA gel shift that probably reflects a 1:1 cGAS1–NCP1 complex (Fig. 3c). However, all higher-order cGAS–NCP assemblies readily detected with WT cGAS were lost when the secondary nucleosome-binding site was mutated. Consistent with the preserved ability to bind to nucleosomes *in vitro*, mutations to cGAS in the *trans* interface (K285A/R300A/K427A) had no detectable effect on cGAS intranuclear tethering in reconstituted HeLa cGAS knockout (KO) cells (Extended Data Fig. 8g). Hence, while the bipartite cGAS1–NCP1 interface forms the primary anchoring motif between cGAS and the nucleosome, the secondary cGAS1–NCP2 interface contributes to the formation of higher-order complexes.



## Effect of structure-based mutations on cellular activity

To determine the functional relationship between nucleosome binding and cGAS inhibition in cells, we focused on motifs on cGAS that interact with the acidic patch (R255A and R236A) (Fig. 4a–c) and with the nucleosomal DNA in *cis* (that is, NCPI–cGAS1; K350A and L354A), the two key *in vitro* interfaces for cGAS–nucleosome binding (Fig. 2a). Consistent with our *in vitro* assays and extending recent work<sup>9</sup>, the cGAS mutants R236A and R255A as well as K350A/L354A were strongly defective in nuclear tethering when expressed in HeLa cGAS KO cells (Fig. 4d). Using fluorescence recovery after photobleaching (FRAP), we detected differences between the mutants in their degree of intranuclear mobility, with cGAS R255A showing the highest, R236A intermediate and K350A/L354A the lowest mobility relative to WT cGAS (Fig. 4e, Extended Data Fig. 9a). Notably, the degree of dissociation correlated well with cellular cGAS responses, with R255A expression triggering the highest cGAMP levels, followed by R236A, and cGAS K350/L354A showing negligible activity (Fig. 4f).

We next examined whether the expression of the two most striking cGAS mutants, R236A and R255A, stimulates a type I interferon response. Activation of cGAS not only promotes conventional, cell-autonomous signalling but also elicits cellular activation *in trans* through the transfer of cGAMP<sup>25</sup>. We found that cGAS mutants triggered only modest upregulation of interferon-stimulated genes when induced in a synchronized manner in mono-cultures of HeLa cells (Extended Data Fig. 9b). This effect may be due to negative-feedback regulation at the level of STING<sup>26,27</sup>, resulting in non-responsiveness towards intracellular cGAMP accumulation over time (Extended Data Fig. 9c, d). By contrast, in co-culture with human BJ fibroblasts, which serve as naive acceptor cells, the expression of cGAS mutants in HeLa cells induced strong upregulation of interferon-stimulated genes and WT cGAS had no such effect (Fig. 4g). Collectively these findings suggest that disrupting the interaction of cGAS with the acidic patch of nucleosomes is in itself sufficient to trigger innate immune activation.

## Discussion

We provide the structural basis for cGAS inhibition by nucleosomes: a bipartite interface involving contacts to both the acidic patch of H2A–H2B dimers and the nucleosomal DNA that ‘traps’ cGAS in an inactive state, in which cGAS can neither engage dsDNA in a manner required for canonical dsDNA sensing nor undergo the dimerization/oligomerization reaction that is required for its catalytic activity.

On the basis of our work, we propose that cGAS uses a ‘missing-self’ recognition strategy to reliably discriminate between self and non-self DNA: instead of focusing on pathogen-specific features that promote activation as is the case for many pattern recognition receptors<sup>28</sup>, cGAS exploits the suppressive activity of nucleosomes, leveraging essentially ‘inbuilt identifiers’ of eukaryotic genomes, to avert aberrant activity. The motifs responsible for cGAS interactions with the nucleosome are well conserved within cGAS homologues that utilize DNA for the regulation of their catalytic activity (Extended Data Fig. 9e). We propose that the inhibitory interaction of cGAS with nucleosomes is a key element of a multi-layered regulation strategy that allows cGAS to reside in the nucleus without undergoing persistent activation<sup>29,30</sup>.

## Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information,

acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41586-020-2750-6>.

1. Ablasser, A. & Chen, Z. J. cGAS in action: expanding roles in immunity and inflammation. *Science* **363**, eaat8657 (2019).
2. Sun, L., Wu, J., Du, F., Chen, X. & Chen, Z. J. Cyclic GMP–AMP synthase is a cytosolic DNA sensor that activates the type I interferon pathway. *Science* **339**, 786–791 (2013).
3. Wu, J. et al. Cyclic GMP–AMP is an endogenous second messenger in innate immune signaling by cytosolic DNA. *Science* **339**, 826–830 (2013).
4. Ablasser, A. et al. cGAS produces a 2'-5'-linked cyclic dinucleotide second messenger that activates STING. *Nature* **498**, 380–384 (2013).
5. Gao, P. et al. Cyclic [G(2'-5')pA(3',5')p] is the metazoan second messenger produced by DNA-activated cyclic GMP–AMP synthase. *Cell* **153**, 1094–1107 (2013).
6. Diner, E. J. et al. The innate immune DNA sensor cGAS produces a noncanonical cyclic dinucleotide that activates human STING. *Cell Rep.* **3**, 1355–1361 (2013).
7. Zhang, X. et al. Cyclic GMP–AMP containing mixed phosphodiester linkages is an endogenous high-affinity ligand for STING. *Mol. Cell* **51**, 226–235 (2013).
8. Gentili, M. et al. The N-terminal domain of cGAS determines preferential association with centromeric DNA and innate immune activation in the nucleus. *Cell Rep.* **26**, 2377–2393. e13 (2019).
9. Volkman, H. E., Cambier, S., Gray, E. E. & Stetson, D. B. Tight nuclear tethering of cGAS is essential for preventing autoreactivity. *eLife* **8**, e47491 (2019).
10. Barber, G. N. STING: infection, inflammation and cancer. *Nat. Rev. Immunol.* **15**, 760–770 (2015).
11. Orzalli, M. H. et al. cGAS-mediated stabilization of IFI16 promotes innate signaling during herpes simplex virus infection. *Proc. Natl Acad. Sci. USA* **112**, E1773–E1781 (2015).
12. Lahaye, X. et al. NONO detects the nuclear HIV capsid to promote cGAS-mediated innate immune activation. *Cell* **175**, 488–501.e22 (2018).
13. Zierhut, C. et al. The cytoplasmic DNA sensor cGAS promotes mitotic cell death. *Cell* **178**, 302–315.e23 (2019).
14. Pang, B. et al. Drug-induced histone eviction from open chromatin contributes to the chemotherapeutic effects of doxorubicin. *Nat. Commun.* **4**, 1908 (2013).
15. Stark, H. GraFix: stabilization of fragile macromolecular complexes for single particle cryo-EM. *Methods Enzymol.* **481**, 109–126 (2010).
16. Li, X. et al. Cyclic GMP–AMP synthase is activated by double-stranded DNA-induced oligomerization. *Immunity* **39**, 1019–1031 (2013).
17. McGinty, R. K. & Tan, S. Nucleosome structure and function. *Chem. Rev.* **115**, 2255–2273 (2015).
18. Civril, F. et al. Structural mechanism of cytosolic DNA sensing by cGAS. *Nature* **498**, 332–337 (2013).
19. Barbera, A. J. et al. The nucleosomal surface as a docking station for Kaposi's sarcoma herpesvirus LANA. *Science* **311**, 856–861 (2006).
20. Abe, T. & Barber, G. N. Cytosolic-DNA-mediated, STING-dependent proinflammatory gene induction necessitates canonical NF- $\kappa$ B activation through TBK1. *J. Virol.* **88**, 5328–5341 (2014).
21. Zhang, X. et al. The cytosolic DNA sensor cGAS forms an oligomeric complex with DNA and undergoes switch-like conformational changes in the activation loop. *Cell Rep.* **6**, 421–430 (2014).
22. Andreeva, L. et al. cGAS senses long and HMGB/TFAM-bound U-turn DNA by forming protein–DNA ladders. *Nature* **549**, 394–398 (2017).
23. Zhou, W. et al. Structure of the human cGAS–DNA complex reveals enhanced control of immune surveillance. *Cell* **174**, 300–311.e11 (2018).
24. Xie, W. et al. Human cGAS catalytic domain has an additional DNA-binding interface that enhances enzymatic activity and liquid-phase condensation. *Proc. Natl Acad. Sci. USA* **116**, 11946–11955 (2019).
25. Ablasser, A. et al. Cell intrinsic immunity spreads to bystander cells via the intercellular transfer of cGAMP. *Nature* **503**, 530–534 (2013).
26. Dobbs, N. et al. STING activation by translocation from the ER is associated with infection and autoinflammatory disease. *Cell Host Microbe* **18**, 157–168 (2015).
27. Konno, H., Konno, K. & Barber, G. N. Cyclic dinucleotides trigger ULK1 (ATG1) phosphorylation of STING to prevent sustained innate immune signaling. *Cell* **155**, 688–698 (2013).
28. Janeway, C. A. Jr. Approaching the asymptote? Evolution and revolution in immunology. *Cold Spring Harb. Symp. Quant. Biol.* **54**, 1–13 (1989).
29. Ablasser, A. & Hur, S. Regulation of cGAS- and RLR-mediated immunity to nucleic acids. *Nat. Immunol.* **21**, 17–29 (2020).
30. Guey, B. et al. BAF restricts cGAS on nuclear DNA to prevent innate immune activation. *Science* **369**, 823–828 (2020).
31. Dolinsky, T. J. et al. PDB2PQR: an automated pipeline for the setup of Poisson–Boltzmann electrostatics calculations. *Nucleic Acids Res.* **32**, W665–W667 (2004).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature Limited 2020

## Methods

### Cell culture and generation of modified cell lines

Cells were maintained in DMEM (Life Technologies) containing 10% (v/v) FCS, 1% (v/v) penicillin (10,000 IU)/streptomycin (10 mg) (BioConcept), 4.5 g/l D-glucose and 2 mM L-glutamine. HeLa cells are from Sigma (93021013-1VL) and grown under 5% CO<sub>2</sub> and 20% O<sub>2</sub>. Foreskin fibroblasts (BJ-5ta) were purchased from the American Type Culture Collection (ATCC; CRL-4001) and cultured at 5% O<sub>2</sub>. cGAS KO HeLa cells were generated according to the CRISPR–Cas9 technology as described previously<sup>32</sup>. The single guide RNA sequences ((5′-3′) forward: CAC CGA GAC TCG GTG GGA TCC ATC G; (5′-3′) reverse: AAA CCG ATG GAT CCC ACC GAG TCT C) were cloned into the plasmid pSpCas9(BB)-2A-GFP (PX458) (52961, AddGene). Plasmid (1 μg) was transfected into HeLa cells with Lipofectamine 2000 (Life Technologies). Single cells were plated into wells of a 96-MW plate. Single cells were selected for GFP expression and expanded to get clones, which were tested for the KO phenotype by sequencing and immunoblotting. Clones without cGAS were functionally validated. Lentiviral vectors were produced as described previously and HEK 293T cells used for this purpose were a kind gift from Dr. D. Trono, EPFL<sup>33</sup>. In brief, HEK 293T cells were transfected with pCMVDR8.74, pMD2.G plasmids and the puromycin-selectable lentiviral vector pTRIPZ containing the open reading frame of the protein of interest by the calcium phosphate precipitation method. The supernatant containing lentiviral particles was harvested at 48 h and 72 h, pooled and concentrated by ultracentrifugation. All cell lines used were checked for mycoplasma contamination by PCR on a regular base and no contamination was found.

### Mutagenesis PCR

Point mutants of human cGAS (amino acids 155–522) were generated by site-directed PCR mutagenesis based on the QuickChange Primer Design method (Agilent) using PrimeSTAR Max DNA Polymerase (Takara) and suitable primers. Each mutated gene was cloned into a pTRIPZ vector for lentivirus production, a pET28 vector for expression of a C-terminal 6 × His-Halo fusion protein in *Escherichia coli* and a pIRESneo3 vector for fluorescence recovery after photobleaching (FRAP) experiments.

### Fractionation of cellular nuclei

BJ-5ta fibroblasts ( $n = 800,000$ ) were seeded into 10-cm culture dishes. After 2 days, cells were treated with doxorubicin (D1515, Sigma) or aclarubicin (FBM-10-1099, Focus Biomolecules) as indicated, and differential salt fractionations were obtained as follows: cells were lysed in a lysis buffer containing 10 mM HEPES (pH 7.4), 10 mM KCl, 0.5% NP-40 with protease inhibitors (cOmplete, Mini, EDTA-free Protease Inhibitor Cocktail; I1836170001, Sigma). Lysates were cleared by centrifugation and supernatant A was recovered. Pellet was resuspended in a lysis buffer containing 20 mM HEPES (pH 7.4) and 0.25 M NaCl. Supernatant B was recovered after centrifugation. The same procedure was repeated as described above with lysis buffers containing increasing NaCl concentrations (0.5 M, 0.75 M and 1 M NaCl).

### PLA

The Duolink In Situ Detection Reagents Red Kit (DUO94001, DUO92002 and DUO 92004, Sigma) was used for the proximity ligation using the following antibodies: histone H2B (1:150; ab52484, Abcam), histone H4 (1:150; ab31830, Abcam) and cGAS (1:150; D1D3G; I5102, Cell Signaling). Cells were seeded onto coverslips (12 mm, Roth) at 40,000 cells per coverslip, fixed with 100% (v/v) methanol for 3 min and blocked with 5% BSA in PBS for 1 h at room temperature. Cells were incubated with the primary antibody in PBS containing 5% (w/v) BSA for 16 h at 4 °C in a humid chamber. After washing with 1 × Buffer A (Sigma), cells were processed for the PLA. Briefly, cells were incubated in anti-mouse IgG Duolink In Situ PLA Probe MINUS (1:5; Sigma) and anti-rabbit IgG

Duolink In Situ PLA Probe PLUS (1:5; Sigma) for 1 h at 37 °C. Thereafter, cells were washed with Buffer A and incubated in 1 × Duolink ligation buffer containing DNA ligase (1:40; Sigma) for 30 min at 37 °C. After incubation, cells were washed in Buffer A and incubated in 1 × amplification buffer (Sigma) with DNA polymerase (1:80; Sigma) for 100 min at 37 °C. Cells were washed in Buffer A and incubated for 30 min at 37 °C in 1 × Detection Solution Red (Sigma). Cells were then washed in Buffer B (Sigma). Cells were mounted with the medium Duolink In Situ Mounting Medium with DAPI (Sigma). Images were acquired using a ×63/1.4 oil objective on a confocal laser scanning microscope (LSM700, Zeiss). Confocal imaging was performed with Z-sections for at least 10 randomly chosen fields. Maximum intensity projection was applied on the images. The number of PLA-positive signal per cell within the DAPI-positive area was counted using the Cell Counter plugin in Fiji.

### cGAMP measurement

HeLa cells reconstituted with WT cGAS or cGAS mutants were plated ( $0.075 \times 10^6$  cells/ml) in the presence of doxycycline (0.1–1 μg/ml) for 2 days. Cells were harvested by trypsinization (trypsin-EDTA (0.05%), Life Technologies) for 5 min. Cell pellets were lysed in RIPA lysis buffer containing 50 mM Tris, 150 mM NaCl, 1% (w/v) sodium deoxycholate, 0.03% (v/v) SDS, 0.005% (v/v) Triton X-100, 5 mM EDTA, 2 mM sodium orthovanadate and cOmplete Protease Inhibitor Cocktail (Roche) (pellet from one well of a six-well plate in 130 μl of RIPA) for 30 min on ice. Lysed cells were centrifuged for 5 min at 18,200 g and 4 °C. Diluted supernatants were used for cGAMP ELISA assay (Cayman 2′-3′-cGAMP ELISA kit, 501700) according to the manufacturer's instructions. Protein concentration in the supernatant was measured using BCA Pierce Protein assay kit and was used to normalize cGAMP concentration.

### Immunoblotting

Protein extracts were loaded into 10% or 15% SDS-polyacrylamide gels. cGAS was blotted onto nitrocellulose membrane (0.45 μm, Bio-Rad) and histones were transferred on polyvinylidene difluoride membrane (0.2 μm, Bio-Rad). The primary antibody was incubated in 5% BSA diluted in PBS 1 × overnight at 4 °C. The secondary antibodies anti-mouse or anti-rabbit horseradish peroxidase (HRP)-conjugated antibodies were incubated for 1 h at room temperature. Proteins were visualized with the enhanced chemiluminescence substrate ECL (Pierce, Thermo Scientific) and imaged using the ChemiDox XRS Bio-Rad Imager. The following antibodies were used in this study: histone H2B (1:1,000; ab52484, Abcam), histone H2A (1:1,000; ab18255, Abcam), histone H4 (1:1,000; ab31830, Abcam); cGAS (1:1,000, D1D3G; I5102, Cell Signaling), histone H3 (1:1,000; 9715, Cell Signaling), STING (1:1,000; D2P2F, Cell Signaling), lamin A/C (1:1,000; SAB4200236, Sigma), GAPDH (1:1,000; AM4300, Life Technologies), donkey anti-rabbit HRP (1:5,000; 711-036-152, Jackson ImmunoResearch) and donkey anti-mouse HRP (1:5,000; 715-036-151, Jackson ImmunoResearch).

### Confocal imaging of endogenous cGAS and H2B

Cells ( $n = 40,000$ ) were seeded on coverslips (12 mm, Roth). At 48 h after seeding, cells were fixed with 100% (v/v) methanol for 3 min and blocked with PBS containing 5% (w/v) BSA for 1 h at room temperature. Cells were incubated with the primary antibodies (1:150; rabbit anti-cGAS (I5102S, CST) and mouse anti-H2B (1:150; ab52484, Abcam)) in PBS containing 5% (w/v) BSA for 16 h at 4 °C in a humid chamber. Afterwards, coverslips were washed with PBS three times and then incubated in PBS containing 5 μg/ml of DAPI and the secondary antibodies (1:1,000; goat anti-rabbit IgG (H+L) Alexa Fluor 488 conjugate (A11008, Thermo) and donkey anti-mouse IgG (H+L) Alexa Fluor 568 conjugate (A10037, Thermo)). At 1 h post-incubation, coverslips were washed three times with PBS 1 × and mounted on microscope slides (15545650, Thermo) using Fluoromount-G (O100-01, SouthernBiotech). Images were acquired using a ×63/1.40 HC Plan-Apochromat oil immersion objective on a SP8-STED 3 × confocal microscope (Leica).

# Article

cGAS labelled with Alexa Fluor 488 was imaged with an excitation laser of 488 nm and an emission window of 492–532 nm, detected with a hybrid detector. Nuclei counterstained with DAPI were imaged with an excitation laser of 405 nm and an emission window of 410–480 nm, detected by a photomultiplier detector. H2B was labelled with Alexa Fluor 568 and imaged with an excitation laser of 561 nm and an emission window of 560–620 nm, detected by a hybrid detector. Images were acquired with a voxel size of  $0.0655 \times 0.0655 \times 1 \mu\text{m}^3$ .

## FRAP

HeLa cGAS KO cells were plated on 35-mm glass bottom culture dishes (part number P35G-1.5-14-C, MatTek Corporation) at 5,000 cells per dish. On the next day, cells were transfected with plasmids encoding hcGAS-GFP WT or cGAS mutants (pIRESneo3 hs cGAS GFP, pIRESneo3 hs cGAS K350A L354A GFP, pIRESneo3 hs cGAS R236A GFP and pIRESneo3 hs cGAS R255A GFP) using Lipofectamin 2000 following the manufacturer's instructions. After 24 h, cells were used for FRAP experiments, which were performed on a Zeiss LSM 710 confocal microscope at 37 °C with a W-Plan Apochromat  $\times 63/1.0$  objective. A circle with a diameter of 1.33  $\mu\text{m}$  (10 pixels) within the hcGAS-GFP signal located inside the nucleus was partially photobleached with a 488-nm laser (100% power) with 20 iterations within 0.200 s. Time-lapse images were acquired over a 20-s time course after photobleaching with 0.200-s intervals with a laser power between 0.4% and 0.6%. Images were processed by Fiji and normalized to FRAP Analyser (software developed at the University of Luxembourg) using the single normalization + full-scale method. FRAP data were fitted to a binding + diffusion circular model using the FRAP Analyser. Immobile fraction extracted data were plotted in GraphPad Prism 8.

## Fluorescence polarization

Fic-labelled 21-bp dsDNA oligonucleotide (5'-Fic-GACCTTTGTTATGCAACCTAA-3') was used as a fluorescent tracer. Increasing amounts of WT or K394E cGAS (0.3–2,500 nM) were mixed with tracer (10 nM final concentration) in a 384-well microplate (784076, Greiner) at room temperature. The interaction was measured in a buffer containing 20 mM HEPES pH 7.4, 500  $\mu\text{M}$  TCEP, 40 mM NaCl, 10 mM KCl and 0.1% (v/v) pluronic acid. PHERAstar FS microplate reader (BMG Labtech) equipped with a fluorescence polarization filter unit was used to determine the changes in fluorescence polarization. The polarization units were converted to fraction bound as described previously<sup>34</sup>. The fraction bound was plotted versus cGAS concentration and fitted assuming a 1:1 binding model to determine the dissociation constant ( $K_d$ ) using Prism 7 (GraphPad). Since the oligonucleotide that was used contained a fluorescent label, we refer to these as apparent  $K_d$  ( $K_{app}$ ). All measurements were performed in triplicates. For the competitive titration assays, the cGAS bound to the fluorescent oligo tracer was back-titrated with unlabelled dsDNA (21, 147, 167 and 227 bp) or nucleosomes (146, 167 and 227 bp). These counter-titration experiments were carried out by mixing tracer (10 nM) and cGAS (300 nM), and titrating increasing concentration of the unlabelled competitor (0–2.5  $\mu\text{M}$ ). The fraction bound was plotted versus competitor concentration and the data were fitted with a non-linear regression curve to obtain the  $\text{IC}_{50}$  values in Prism 7 (GraphPad). At least two technical replicates were performed per experiment.

## Cellular activation assays

HeLa cells reconstituted with WT cGAS or cGAS mutants were seeded ( $0.075 \times 10^6$  cells/ml) in six-well plates in the presence of doxycycline (1  $\mu\text{g}/\text{ml}$ ) for 16 h or 40 h. Stimulation with dsDNA (90 bp) was carried out as previously described<sup>33</sup>. Briefly, dsDNA (1.6  $\mu\text{g}/\text{ml}$ ) was transfected using Lipofectamine 2000 (Life Technologies) and cells were harvested 4 h later. For co-culture studies, HeLa cells were treated with doxycycline as described above. After 24 h, cells were collected, washed, and 19,000 cells were mixed with human BJ fibroblasts ( $0.095 \times 10^6$  cells/ml) and incubated overnight.

## Quantitative real-time PCR

Cells were lysed using RLT buffer (79216, Qiagen). RNA was extracted according to the manufacturer's protocol (Qiagen RNeasy Mini kit) and treated with RNase-free DNase (ENO521, Thermo Scientific). RNA (500 ng) was reverse transcribed (RevertAid, EP0442, Thermo Fisher Scientific) and analysed by RT-quantitative PCR in duplicates or triplicates using the Maxima SYBR Green/ROX qPCR Master Mix (K0223, Thermo Fisher Scientific). The quantitative PCR reactions were run on a QuantStudio 5 Real-Time PCR system. GAPDH was used for normalization. Primer sequences (5'-3'): *GAPDH*: forward GAGTCAACGGATTTGGTCGT, reverse GACAAGCTTCCCGTTCTCAG; *IFI44*: forward GAT GTG AGC CTG TGA GGT CC, reverse CTT TAC AGG GTC CAG CTC CC; *IFI2*: forward GCGTGAAGAAGGTGAAGAGG, reverse GCAGGTAGG CATTGTTGGT; and *CGAS*: forward GCACGTGAAGATTCTGCACC, reverse TGACTCAGAGGATTTTCTCGG. The sequence of the sense strand of the 90-mer DNA is as follows (5'-3'): TACAGATCTACTAGTGATCTATGACTGATCTGTACATGATCTACATACAGATCTACTAGTGATCTATGACTGATCTGTACATGATCTACA.

## Expression and purification of recombinant cGAS

Truncated human cGAS (155–522) WT or mutants were expressed and purified from *E. coli* strain BL21 (DE3). Plasmids expressing His<sup>6</sup>-Halo-tagged truncated human cGAS were induced with 2 mM IPTG at 18 °C for 20 h. Bacteria were collected by centrifugation and lysed by sonication in lysis buffer (20 mM HEPES pH 8.0, 300 mM NaCl, 20 mM imidazole, 1 mM DTT and protease inhibitor). After centrifugation, clear lysate was incubated with Ni-NTA beads (Qiagen), washed with lysis buffer and 20 mM HEPES pH 8.0, 1 M NaCl, 20 mM imidazole, 1 mM DTT and eluted with 20 mM HEPES pH 7.5, 500 mM NaCl and 250 mM imidazole. Eluted cGAS was subjected to size-exclusion chromatography using a Superdex 200 16/60 column in 20 mM HEPES pH 7.5, 300 mM KCl and 1 mM DTT. The protein was flash frozen in liquid nitrogen and stored at –80 °C.

## EMSA

For the cGAS mutants, biotinylated nucleosomes (31583, Active motifs) were incubated with serial dilutions of recombinant cGAS at room temperature for 30 min in PBS in a sample volume of 10  $\mu\text{l}$ . The binding reactions contained 40  $\mu\text{g ml}^{-1}$  of nucleosomes and cGAS proteins ranged from 100 to 12  $\mu\text{g ml}^{-1}$  with twofold increase. After the reaction, 5  $\mu\text{l}$  glycerol was added. Reactions were detected by electrophoresis on a 5% PAGE gel in 0.5 $\times$  TBE buffer at 10 mA for 1 h 15 min. The gels were incubated for 15 min in SYBR Safe containing PBS and were scanned using the Typhoon FLA-9500 imager (GE Healthcare) and imaged using the ChemiDoc XRS Bio-Rad Imager and Image Lab 6.0.0 software. For the EMSA using the LANA peptide, biotinylated nucleosomes (40  $\mu\text{g ml}^{-1}$ ; 31583, Active motifs) were incubated with the LANA from 0.6  $\text{mg ml}^{-1}$  to 78  $\mu\text{g ml}^{-1}$ , twofold dilutions at room temperature for 5 min in PBS for a sample volume of 4  $\mu\text{l}$ . Recombinant cGAS WT (40 pmol) in 4  $\mu\text{l}$  PBS was added and incubated for 30 min at room temperature. After the reaction, 5  $\mu\text{l}$  glycerol was added. Reactions were detected by electrophoresis on a 5% PAGE gel in 0.5 $\times$  TBE buffer at 10 mA for 1 h 15 min. The gels were incubated for 15 min in SYBR Safe containing PBS and were scanned using the ChemiDoc XRS Bio-Rad Imager and Image Lab 6.0.0 software (Bio-Rad).

## cGAS in vitro competition assay

Human cGAS (50 nM) was mixed with recombinant histones and H2A–H2B dimer (5–0.3  $\mu\text{M}$ ) or nucleosomes (75–2 nM) or nucleosome fibres (6.5–0.1 nM) in 10 mM HEPES pH 8.0, 10 mM KCl and 1 mM  $\text{MgCl}_2$ . 0.1  $\text{mg ml}^{-1}$  HT-DNA, 10  $\mu\text{Ci}$  [ $\alpha$ -<sup>32</sup>P]ATP and 1 mM GTP were added and left to react for 12 h at 37 °C. Reaction solution (1  $\mu\text{l}$ ) was spotted onto TLC plates (HPTLC silica gel, 60 Å pores, F<sub>254</sub>; 1055480001, Merck Millipore), and the nucleotides were separated with 5 mM  $\text{NH}_4\text{Cl}$  17%

EtOH as the mobile phase at 25 °C for 30 min. The plates were visualized by autoradiography and scanned using the Typhoon FLA-9500 Imager (ImageQuantTool, GE Healthcare). Images were processed using Image Lab 6.0.0 software (Bio-Rad) to quantify the intensity of the spots corresponding to cGAMP. After normalizing by cGAMP levels in the absence of chromatin for each individual mutant, the  $IC_{50}$  was calculated using GraphPad Prism.

### Large-scale production of 601 DNA

Production of 601 DNA was performed as previously described<sup>35</sup>. Briefly, a plasmid carrying 32 copies of the Widom '601' sequence, each flanked by EcoRV restriction sites, was purified from an 8L 2YT culture of transformed *E. coli* DH5 $\alpha$  cells using alkaline lysis, followed by isopropanol precipitation, RNase A treatment of the suspended pellet and subsequent chromatography on Sepharose 6. After isopropanol precipitation, the 601 sequences were released from the plasmid by digestion with EcoRV (12 ml total volume containing 132  $\mu$ l EcoRV for 40 h). The 601 DNA fragment was isolated by PEG precipitation with 14.5% PEG, and further purified by ethanol-acetate precipitation and subsequent chloroform-phenol extraction to yield 9.2 mg of 601 DNA.

### Nucleosome preparation

Histones were prepared and octamers made as previously described<sup>36</sup>. Nucleosomes were reconstituted through overnight gradual dialysis from TEK2000 (10 mM Tris pH 7.5, 1 mM EDTA and 2 M KCl) into TEK10 (10 mM Tris pH 7.5, 1 mM EDTA and 10 mM KCl) with dialysis buttons using octamer:DNA ratios of 1.3, 1.4 and 1.5 each in 0.9 ml total volume containing 6.7  $\mu$ M DNA. After recovering the material from the dialysis buttons, the nucleosomes were concentrated using Amicon centrifugal concentrators with a 30 kDa MWCO to yield a nucleosome concentration of 0.86 mg ml<sup>-1</sup>.

### Cryo-electron microscopy sample preparation

601Widom sequence NCPs (147 bp) and purified human cGAS were mixed in a 1:1.5 molar ratio in gel filtration buffer (20 mM HEPES pH 7.4, 300 mM KCl and 250  $\mu$ M TCEP) and dialysed for 24 h against low-salt dialysis buffer (20 mM HEPES pH 7.4, 50 mM KCl and 250  $\mu$ M TCEP). Thereafter, the dialysed complex was concentrated using an Amicon Ultra 0.5-ml centrifugal filter (Merck Millipore) and applied to a GraFix<sup>45</sup> gradient of 10–30% sucrose containing top solution (20 mM HEPES pH 7.4, 50 mM KCl and 250  $\mu$ M TCEP, 10% w/v sucrose) and bottom solution (20 mM HEPES pH 7.4, 50 mM KCl and 250  $\mu$ M TCEP, 30% w/v sucrose and 1.5% glutaraldehyde). The gradient ultracentrifugation was carried out at 30,000 r.p.m. for 18 h at 0 °C using an AH-650 swinging-bucket rotor. Fractions (100  $\mu$ l) were collected and analysed by both native PAGE and SDS–PAGE. Thereafter, the peak fractions were combined and dialysed overnight (20 mM HEPES pH 7.4, 50 mM KCl and 250  $\mu$ M TCEP) to remove sucrose. The resulting complex sample was concentrated with an Amicon-Ultra 0.5-ml centrifugal filter to 1 mg ml<sup>-1</sup> as determined by measuring protein concentration at Abs280. Quantifoil holey carbon grids (R 1.2/1.3 200-mesh, Quantifoil Micro Tools) were glow discharged with Solarus plasma cleaner (Gatan) for 30 s in a H<sub>2</sub>/O<sub>2</sub> environment. Sample (3  $\mu$ l) was applied to grids and blotted for 3 s at 4 °C and 100% humidity in a VitroBot Mark IV (FEI), and immediately plunged into liquid ethane.

### Cryo-electron microscopy data acquisition

First, for the cGAS (K394E)–NCP complex, two data sets were collected for GraFix-crosslinked samples using the Titan Krios electron microscope (Thermo Fisher Scientific) at 300 keV, with zero energy loss (slit 20 eV). Automatic data collection was done using EPU (Thermo Fisher Scientific) on a Cs-corrected (CEOS GmbH), with micrographs recorded using a Gatan K2 summit direct electron detector (Gatan). The acquisition was performed at a nominal magnification of  $\times$ 130,000 in EFTEM nanoprobe mode, yielding a pixel size of 0.86 Å at the specimen level. All

data sets were recorded with the 100- $\mu$ m objective aperture and with a total dose of 45 e<sup>-</sup>/Å<sup>2</sup>, recording 40 frames. The targeted defocus values ranged from –0.5 to –2  $\mu$ m. Similarly, few micrographs were recorded for the two non-crosslinked samples, the WT cGAS–NCP complex and the dimerization mutant (K394E) cGAS–NCP complex.

Second, for the cGAS(WT)–NCP complex, a dataset was collected for GraFix-crosslinked samples using the Glacios (Thermo Fisher Scientific) electron microscope at 200 keV. Automatic data collection was done using EPU (Thermo Fisher Scientific) on a Cs-corrected (CEOS GmbH), with micrographs recorded using a Falcon 3EC Direct Electron Detector. The acquisition was performed at a nominal magnification of  $\times$ 150,000 in EFTEM nanoprobe mode, yielding a pixel size of 0.68 Å at the specimen level. All data sets were recorded with the 100  $\mu$ m objective aperture and with a total dose of 35 e<sup>-</sup>/Å<sup>2</sup>, recording 40 frames. The targeted defocus values ranged from –0.5 to –2  $\mu$ m.

### Cryo-electron microscopy image processing

First, for the cGAS (K394E)–NCP complex, on-the-fly evaluation of the data was performed with CryoFLARE (in house development; www.cryoflare.org)<sup>37</sup>. Micrographs below an EPA limit of 5 Å were used for further processing. A total of 2,890 micrographs were acquired in two sessions. Drift correction was performed with the RELION3 motioncor in which a motion-corrected sum of all 40 frames was generated with and without applying a dose-weighting scheme and CTF was fitted using gCTF<sup>38</sup> on the non-dose-weighted sums. A small set of particles (54,000) was picked using cryOLO<sup>39</sup> and imported to cryoSPARC<sup>40</sup>. After 2D classification, an ab initio model was generated. This model was used as an initial 3D map for further 3D classification in RELION3 for the two data sets independently. In data set 1, the particles (13,943) included in the class that contained two cGAS and two NCPs were imported into cryoSPARC<sup>40</sup> and subjected to non-uniform refinement and later refined to 4.1 Å (Extended Data Table 1(a)). In data set 2, the particles (87,323) included in the class that contained one cGAS and one NCP were subjected to local refinement in cryoSPARC and refined to 3.8 Å. The particles that were used for the 4.1 Å and 3.8 Å maps were merged and refined to 3.3 Å in RELION3<sup>41</sup>. CTF refinement and signal subtraction was done for the density, accounting for 1cGAS–1NCP complex. 3D classification followed by non-uniform refinement in cryoSPARC led to a 3.1 Å map.

Second, for the cGAS (WT)–NCP complex, a total of 5,007 movies were acquired for the cGAS (WT)–NCP complex. Full frame motion correction followed by patch CTF was performed using cryoSPARC. Particle picking was done using template picker for 700 images and a 2D template was generated. Selected 2Ds were later used to generate an ab initio model and template picking of particles for the rest of the images in cryoSPARC. A total of 142,743 particles out of 404,087 were selected from 2D classification to do a homo-refinement. 3D hetero-refinement with two classes was carried out with 56,747 (40%) particles, giving a map that was later locally refined to 5.1 Å. The same class with 56,747 particles was further used for particle subtraction for the cGAS1–NCP1 complex and further local refinement was carried out to obtain a 4.7 Å map.

The resolution values reported for all reconstructions are based on the gold-standard Fourier shell correlation (FSC) curve<sup>42</sup>. High-resolution noise substitution has been used for correcting the effects of soft masking for the related FSC curves. All of the maps have been filtered based on local resolution estimated with MonoRes (XMIPP)<sup>43</sup> and were later sharpened using the localdeblur\_sharpen protocol (XMIPP).

### Model building and refinement

A nucleosome model from PDB entry 6R8Y<sup>44</sup> and a human cGAS model from PDB entry 4LEV<sup>16</sup> were used as initial references for the cryo-electron microscopy map interpretation. The models were rigid-body docked using Chimera<sup>45</sup> and COOT<sup>46</sup>. Sequence

reassignment to 6R8Y with the 147-bp 601 Widom sequence<sup>47</sup> along with the human histone sequence was done. The starting model for cGAS (4LEV) was refined against the corresponding crystallographic structure factors with Phenix<sup>48</sup> and Rosetta<sup>49</sup> to resolve some of the geometry outliers. Restraints for the covalently attached crosslinker were generated with JLIgand<sup>50</sup>, Phenix and Rosetta. Model building and refinement of the cryo-electron microscopy structures were carried out iteratively with COOT<sup>46</sup>, Phenix<sup>48</sup> and Rosetta<sup>49</sup> using reference model restraints for cGAS (torsional angles) derived from the template model (see above). The reference model restraints were generated with Phenix<sup>48</sup> and converted to Rosetta constraints. Residues at the interface with the NCP were not restrained. In case of the dimeric nucleosome–cGAS, the refinement was with reference model restraints derived from the higher-resolution monomer complex (torsional angles). Model validation was done with Phenix<sup>48</sup> and MOLPROBITY<sup>51</sup>. Side chains without sufficient density were marked with zero occupancy.

## Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

## Data availability

The electron density reconstructions and corresponding final models for NCP1–cGAS1–cGAS2–NCP2 and NCP1–cGAS1 were deposited in the Electron Microscopy Data Bank (accession codes: EMDB-10694 and EMDB-10695) and in the PDB (accession codes: 6Y5D and 6Y5E). The electron density reconstructions for NCP1–WT cGAS1–WT cGAS2–NCP2 and NCP1–WT cGAS1 were deposited in the Electron Microscopy Data Bank (accession codes: EMDB-11006 and EMDB-11005).

32. Ran, F. A. et al. Genome engineering using the CRISPR–Cas9 system. *Nat. Protocols* **8**, 2281–2308 (2013).
33. Haag, S. M. et al. Targeting STING with covalent small-molecule inhibitors. *Nature* **559**, 269–273 (2018).
34. Marks, B. D. et al. Multiparameter analysis of a screen for progesterone receptor ligands: comparing fluorescence lifetime and fluorescence polarization measurements. *Assay Drug Dev. Technol.* **3**, 613–622 (2005).
35. Hanson, B. L., Alexander, C., Harp, J. M. & Bunick, G. J. Preparation and crystallization of nucleosome core particle. *Methods Enzymol.* **375**, 44–62 (2004).
36. Luger, K., Rechsteiner, T. J. & Richmond, T. J. Preparation of nucleosome core particle from recombinant histones. *Methods Enzymol.* **304**, 3–19 (1999).
37. Schenk, A. D., Cavadini, S., Thomä, N. H. & Genoud, C. Live analysis and reconstruction of single-particle cryo-electron microscopy data with CryoFLARE. *J. Chem. Inf. Model.* **60**, 2561–2569 (2020).
38. Zhang, K. Gctf: real-time CTF determination and correction. *J. Struct. Biol.* **193**, 1–12 (2016).
39. Wagner, T. et al. SPHIRE-crYOLO is a fast and accurate fully automated particle picker for cryo-EM. *Commun. Biol.* **2**, 218 (2019).
40. Punjani, A., Rubinstein, J. L., Fleet, D. J. & Brubaker, M. A. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nat. Methods* **14**, 290–296 (2017).

41. Zivanov, J. et al. New tools for automated high-resolution cryo-EM structure determination in RELION-3. *eLife* **7**, e42166 (2018).
42. Scheres, S. H. RELION: implementation of a Bayesian approach to cryo-EM structure determination. *J. Struct. Biol.* **180**, 519–530 (2012).
43. de la Rosa-Trevin, J. M. et al. Xmipp 3.0: an improved software suite for image processing in electron microscopy. *J. Struct. Biol.* **184**, 321–328 (2013).
44. Matsumoto, S. et al. DNA damage detection in nucleosomes involves DNA register shifting. *Nature* **571**, 79–84 (2019).
45. Pettersen, E. F. et al. UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).
46. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 2126–2132 (2004).
47. Lowary, P. T. & Widom, J. New DNA sequence rules for high affinity binding to histone octamer and sequence-directed nucleosome positioning. *J. Mol. Biol.* **276**, 19–42 (1998).
48. Adams, P. D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 213–221 (2010).
49. Terwilliger, T. C. et al. phenix.mr\_rosetta: molecular replacement and model rebuilding with Phenix and Rosetta. *J. Struct. Funct. Genomics* **13**, 81–90 (2012).
50. Lebedev, A. A. et al. JLIgand: a graphical tool for the CCP4 template-restraint library. *Acta Crystallogr. D Biol. Crystallogr.* **68**, 431–440 (2012).
51. Davis, I. W. et al. MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res.* **35**, W375–W383 (2007).
52. Waterhouse, A. M., Procter, J. B., Martin, D. M., Clamp, M. & Barton, G. J. Jalview version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics* **25**, 1189–1191 (2009).

**Acknowledgements** We thank N. Jordan and J. Ricci for technical assistance. A.A. received grants from the SNF (BSSGIO-155984 and 31003A\_159836), NCCR in Chemical Biology, and was supported by the European Union's Horizon 2020 Research and Innovation program grant agreement (grant no: 804933, ImAging). N.H.T. was supported by the European Research Council (ERC) under the European Union's Horizon 2020 Research and Innovation program grant agreement (grant no. 666068), the Novartis Research Foundation and by the SNF (31003A\_179541). B.F. is supported by the European Union's Horizon 2020 Research and Innovation program grant agreement (grant no: 724022, chromo-SUMMIT) and the NCCR in Chemical Biology. Z.K. was supported by the European Union's Horizon 2020 Research and Innovation program under the Marie Skłodowska-Curie grant agreement no. 765445.

**Author contributions** G.R.P. conducted the cryo-electron microscopy experiments and data processing with help from S.C. and G.K. for the model building. A.D. performed the in vitro assays and purified recombinant cGAS. S.G., P.M. and A.A. performed the cellular experiments. K.M. and R.H. purified recombinant histone proteins and assembled the recombinant nucleosomes and chromatin fibres. Electron microscopy samples were prepared by G.R.P. with help from J.W. Z.K. and J.W. performed the fluorescence polarization assays. B.G. performed the FRAP experiments. B.F. supervised experiments related to the in vitro reconstitution of nucleosomes/chromatin and provided valuable discussion. N.H.T. supervised the structural work and the biophysical assays. A.A. supervised the in vitro assays and cellular studies. N.H.T. and A.A. wrote the manuscript with input from all authors.

**Competing interests** A.A. is a member of the scientific advisory board of IFM Therapeutics and scientific co-founder of IFM-Due.

## Additional information

**Supplementary information** is available for this paper at <https://doi.org/10.1038/s41586-020-2750-6>.

**Correspondence and requests for materials** should be addressed to N.H.T. or A.A.

**Peer review information** Nature thanks Osamu Nureki and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

**Reprints and permissions information** is available at <http://www.nature.com/reprints>.