

DISSECTING BMI1 PROTEIN-PROTEIN INTERACTIONS THROUGH CHEMICAL BIOLOGY

by

Felicia Gray

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Chemical Biology)
in the University of Michigan
2015

Doctoral Committee:

Assistant Professor Tomasz Cierpicki, Co-Chair
Assistant Professor Jolanta E. Grembecka, Co-Chair
Associate Professor Elizabeth Lawlor
Professor Anna K. Mapp
Associate Professor Raymond C. Trievel

© Felicia Gray

All rights reserved, 2015

Dedication

For Dorothy Kithinji

Acknowledgements

I first need to acknowledge Dr. Tomasz Cierpicki and Dr. Jolanta Grembecka for their mentorship, patience and support over these past four years. I am grateful for their unending enthusiasm for science and for providing a lab environment in which I could be successful.

I thank my committee members, Dr. Anna Mapp, Dr. Ray Trievel and Dr. Beth Lawlor for providing support and guidance throughout this process. I have always looked forward to my committee meetings and appreciate their time and insight.

I would like to thank the members of the Cierpicki and Grembecka labs, past and present, for creating a collegial and engaging environment in which to learn and work. I have enjoyed working with Shihan, Bhavna, Hyoje, Shirish, Weijiang and Qingjie in tackling interdisciplinary problems. I am especially grateful for Dmitry, Dave, George and Jon for providing both scientific support and opportunities to relax outside of lab.

I thank many friends in the Chemical Biology program: Alison, Hong, Matt W., Matt P., JP, Carol Ann, Elin, Beth, Chinmay and Carrie. I especially thank Ningkun for guiding me as a naïve rotation student and still being my friend afterwards! My graduate experience would not have been the same without the friendship of Paul and Victoria, essentially from the beginning, and I appreciate learning from their personal and scientific insights many, many times over the course of this processes.

I am grateful to have worked with the wonderful women in FEMMES for four years as we have worked towards our shared ambition to increase diversity in STEM fields and to just plain excite young girls about science. This was a very meaningful group for me to have been a part of and I am continuously inspired and enlightened by the spirit of the group. I am excited to see where the program goes with the next generation.

I am thankful to Noah who has tremendous patience and who challenges me to be my best person. I thank him especially for reading this thesis and providing his critical insight. And finally, I am forever grateful to my family for providing a solid foundation to follow my passion and supporting my every move in my less-than-linear path to this point. I would not be where I am without Bev's unending faith in me and her insistence that I live up to it! Molly inspires me every day and I am so honored to be her big sister. I thank my Dad and Bebe for their love, enthusiastic interest in my work and for providing me a second home in the Midwest. My mother never gives herself enough credit and I'm sure I cannot do it here but it is through her eternal support and wise council that she has set me up for success and her regular notes of encouragement provided confidence and inspiration.

Table of Contents

Dedication	ii
Acknowledgements	iii
List of Figures	ix
List of Tables	xi
List of Equations	xii
List of Abbreviations	xiii
Abstract	xv
Chapter 1: Introduction	1
A. Motivation	1
B. Transcriptional regulation through chromatin modification	2
C. Polycomb proteins negatively regulate gene transcription	3
<i>C.1. Canonical polycomb gene silencing mechanisms</i>	3
<i>C.2. Non-canonical polycomb gene silencing mechanisms</i>	5
<i>C.3. Architecture of the canonical PRC1 complex</i>	5
D. Polycomb function in biology	7
<i>D.1. Phenotypes of polycomb deletion suggest complex functional roles</i>	7
<i>D.2. Delineating polycomb function through identification of gene targets</i>	8
E. BMI1 is an oncogene implicated in many cancers	9
<i>E.1. Clinical implications of BMI1 overexpression</i>	9
<i>E.2. Mechanistic insight into BMI1 as an oncogene</i>	10

<i>E.3. Genetic inhibition of BMI1 in cancer models supports small molecule inhibitor development for this target</i>	14
F. Targeting protein-protein interactions with small molecule inhibitors	16
<i>F.1. Challenges for small molecule inhibition of protein-protein interactions</i>	16
<i>F.2. Methods of identifying protein-protein interaction inhibitors</i>	17
<i>F.3. Examples of PPI inhibitors for chromatin modifying proteins</i>	18
G. Thesis summary	20
H. References	20
Chapter 2: Structural Characterization of BMI1 Protein-Protein Interaction Domain	37
A. Abstract	37
B. Background	37
C. Results	38
<i>C.1. The second domain of BMI1 interacts directly with PHC2</i>	38
<i>C.2. Optimization of BMI1 second domain construct for structural studies</i>	39
<i>C.3. Characterization of BMI1-binding domain in PHC2</i>	42
<i>C.4. Structure determination of BMI1-PHC2 complex</i>	44
<i>C.5. Design of mutations in BMI1 to disrupt BMI1-PHC2 interaction</i>	49
<i>C.6. BMI1 ULD can form higher order oligomers</i>	50
<i>C.7. Multiple BMI1 protein-protein interactions regulate cellular proliferation</i>	54
D. Discussion and Conclusion	56
<i>D.1. Need for using multiple methodologies to determine structure of BMI1-PHC2 complex</i>	56
<i>D.2. Structural studies of BMI1-PHC2 reveal basis for polycomb ULD specificity</i>	57
<i>D.3. Implications of BMI1 protein-protein interactions on PRC1 architecture</i>	60
Acknowledgements	61
E. Experimental methods	62
F. References	68
Chapter 3: High-throughput Screening to Identify Inhibitors of the BMI1-PHC2 Protein-Protein Interaction	72
A. Abstract	72
B. Background	73
<i>B.1. Rationale for developing small molecule inhibitors of BMI1-PHC2 protein-protein interaction</i>	73
<i>B.2. Is the BMI1-PHC2 interaction a good target for disruption with small molecules?</i> 73	
<i>B.3. Biochemical assays and high-throughput screening strategy</i>	75

C. Results	77
<i>C.1. BMI1 -PHC2 Biochemical Assay Development</i>	77
<i>C.2. Primary Screening</i>	83
<i>C.3. Confirmation Screening</i>	83
<i>C.4. Dose Response Measurements</i>	84
<i>C.5. Characterization of active compounds</i>	85
D. Discussion and Conclusions	94
Acknowledgements	95
E. Experimental methods	95
F. References	101
Chapter 4: Development of Small Molecule Inhibitors of Ring1B/BMI1 E3 Ubiquitin Ligase	104
A. Abstract	104
B. Background	105
<i>B.2. Mechanism of protein ubiquitination</i>	105
<i>B.3. Ring1B/BMI1 E3 ligase structure and substrate specificity</i>	107
<i>B.4. E3 ligases as targets for small molecule inhibitors</i>	108
C. Results	110
<i>C.1. Identification of small molecule fragment ligands of Ring1B/BMI1 fusion protein</i>	110
<i>C.2. NMR-guided SAR of Ring1B/BMI1 ligands</i>	111
<i>C.3. Ring1B/BMI1 ligands are potent and specific inhibitors of in vitro ubiquitin ligase assay activity</i>	124
<i>C.4. Ligand binding site and orientation probed by mutagenesis and NMR studies</i>	126
<i>C.5. Inhibitor binding results in conformational change in Ring1B</i>	131
<i>C.6. Ring1B/BMI1 inhibitors disrupt protein-nucleosome interaction</i>	132
<i>C.7. Preliminary cellular studies demonstrate utility of inhibitors as chemical probes</i>	133
D. Discussion and Conclusion	134
Acknowledgements	136
E. Experimental methods	137
F. References	143
Chapter 5. Discussion	148
A. Conclusions	148
<i>A.1. Structural insights into BMI1 function in PRC1 complex</i>	148
<i>A.2. Multiple approaches to inhibit BMI1 with small molecules</i>	151
B. Future Directions	152

<i>B.1. BMI1 function in protein complexes.....</i>	152
<i>B.2. Future efforts to inhibit BMI1's protein-protein interactions with small molecules</i>	154
<i>B.3. Use of BMI1 inhibitors as chemical tools</i>	154
<i>B.4. Investigation into RING E3 ligase mechanisms</i>	156
<i>B.5. Inhibitors of RING E3 ligases</i>	156
C. References	157
Appendix A	161
A. Abstract	161
B. Introduction	161
<i>B.1. Function of disordered regions in proteins</i>	161
<i>B.2. ¹³C-detected NMR.....</i>	162
<i>B.3. Model system: the menin protein.....</i>	163
C. Results	163
D. Discussion.....	165
E. Materials and Methods:.....	166
F. References	167

List of Figures

Figure 1.1. Canonical hierarchical mechanism of polycomb-mediated gene silencing.	4
Figure 1.2. Domain organization of proteins in the canonical PRC1 complex.	6
Figure 1.3. Contributions of BMI1 to the hallmarks of cancer.....	10
Figure 1.4. Strategies for inhibiting protein-protein interactions in chromatin modification.....	18
Figure 2.1. BMI1 C-terminus interacts directly with PHC2_B.	39
Figure 2.2. Optimization of BMI1 second domain construct using NMR.....	41
Figure 2.3. ITC determination of BMI1-PHC2 binding stoichiometry and affinity.....	42
Figure 2.4. Mapping of BMI1-binding motif of PHC2 using ¹³ C-detected NMR.....	43
Figure 2.5. PHC2 interacts with BMI1 in cells through a conserved fragment.....	44
Figure 2.6. Crystal structure of BMI1 ULD.	45
Figure 2.7. Hybrid solution NMR- x-ray crystal structure of BMI1-PHC2 complex.....	46
Figure 2.8. Structural details of the BMI1-PHC2 interaction.	47
Figure 2.9. PHC2 electron density is at crystallographic interface in crystal structure.....	48
Figure 2.10. BMI1 mutations disrupt PHC2 interaction <i>in vitro</i> and in cells.....	49
Figure 2.11. BMI1 self-associates in solution.	51
Figure 2.12. Crystal structure of BMI1 suggests two homodimer interfaces.	52
Figure 2.13. BMI1-BMI1 interface mutants disrupt self-association but not interaction with PHC2.....	53
Figure 2.14. Multiple BMI1 protein-protein interactions regulate cellular proliferation.	55
Figure 2.15. Basis for binding partner selectivity in ULD-domain polycomb proteins.	59
Figure 2.16. Models for PRC1 organization and oligomerization.....	61
Figure 3.1. BMI1-PHC2 interface is an attractive target for small molecule inhibitors.....	74
Figure 3.2. Features of BMI1-PHC2 interaction with relevance for drug discovery.	75
Figure 3.3. Scheme of fluorescence polarization assay for BMI1-PHC2 interaction.....	78
Figure 3.4. Development of fluorescence polarization assay for HTS.	79

Figure 3.5. FP Assay Performance.	80
Figure 3.6. Scheme of AlphaLISA assay for BMI1-PHC2 interaction.	81
Figure 3.7. Competition experiments of AlphaLISA assay.	82
Figure 3.8. Quality assessment of BMI1-PHC2 AlphaLISA Assay.	82
Figure 3.9. Comparing compound activity in confirmation screening.	84
Figure 3.10. Summary of HTS.	85
Figure 3.11. Validation of HTS hits from fresh powder.	86
Figure 3.12. BI-1-1 binds directly to BMI1 ULD.	89
Figure 3.13. BI-1-7 binding site mapping by NMR spectroscopy.	90
Figure 3.14. BI-2 binding site mapping by NMR spectroscopy.	91
Figure 3.15. BI-3 binding site mapping by NMR spectroscopy.	93
Figure 4.1. Overview of ubiquitination cascade.	106
Figure 4.2. Characterization of Compound 1.	110
Figure 4.3. NMR estimation of binding affinity to guide SAR.	112
Figure 4.4. Characterization of most potent compound by ITC.	124
Figure 4.5. Optimized Ring1B/BMI1 ligand is inhibitor of <i>in vitro</i> ubiquitin ligase activity.	125
Figure 4.6. Inhibitors are specific for Ring1B/BMI1 E3 ubiquitin ligase.	126
Figure 4.7. Mapping of NMR chemical shift perturbations onto Ring1B/BMI1 structure identifies ligand binding site.	127
Figure 4.8. Probing compound binding site by mutagenesis.	128
Figure 4.9. Paramagnetic relaxation enhancement (PRE) studies orient Ring1B/BMI1 ligands in binding site.	130
Figure 4.10. Ring1B-nucleosome interface.	132
Figure 4.11. Ring1B/BMI1 ligand binding prevents interaction with nucleosome.	133
Figure 4.12. Ring1B/BMI1 inhibitors are cell permeable with on-target activity.	134
Figure 4.13. Docking model of Ring1B in complex with compound 69.	135
Figure 5.1. Model for role of BMI1-PHC2 oligomerization in PRC1 spreading.	149
Figure 5.2. Model for role of BMI1-PHC2 oligomerization in PRC1 subunit heterogeneity.	150
Figure 5.3. Two approaches to inhibit BMI1 with small molecules.	151
Figure A.1. Assignment of disordered residues in <i>Nematostella</i> menin.	164
Figure A.2. Applications of ¹³ C-detected NMR approach for rapid characterization of disordered regions of proteins.	166

List of Tables

Table 2.1. Statistics of crystal structure of BMI1 ULD.....	65
Table 2.2. Statistics for NMR Structure Determination.	66
Table 3.1. Confirmed hits from HTS campaign that bind directly to BMI1 ULD.	86
Table 3.2. Structure-activity-relationship for BI-1 series of inhibitors.	88
Table 3.3. Recipe for M9 minimal media for expression of ¹⁵ N-labeled BMI1 ULD.....	96
Table 3.4. Summary of high-throughput screening campaign.....	97
Table 4.1. SAR for 5-phenylpyrrole scaffold.	115
Table 4.2. SAR of 5-indol-4-yl-pyrrole scaffold.	116
Table 4.3. SAR of 5-indol-4-yl-pyrrole-2-carboxylic acid scaffold.	118
Table 4.4. SAR of monosubstituted 5-indol-4-yl-pyrrole-2-carboxylic acid.	120
Table 4.5. SAR of disubstituted 3-phenyl5-indol-4-yl-pyrrole-2-carboxylic acid.	122
Table 4.6. SAR of heterocycle substituents of 5-indol-4-yl-pyrrole scaffold.....	123
Table 4.7. Reaction conditions for ubiquitin-ligase assay.	142

List of Equations

Equation 3.1. Z-factor calculation	77
Equation 4.1. Calculation of chemical shift perturbations.....	139
Equation 4.2. Calculation of K_D and B_{max} for ligand titration.....	140

List of Abbreviations

AlphaLISA	Amplified Luminescence Proximity Homogenous Assay
AML	Acute myeloid leukemia
AUC	Analytical Ultracentrifugation
BME	β -mercaptoethanol
BMI1	B cell-specific Moloney murine leukemia virus integration site 1
BSA	Bovine Serum Albumin
ChIP	Chromatin Immunoprecipitation
CIC	Cancer initiating cells
CML	Chronic myeloid leukemia
D₂O	Deuterium oxide
DMSO	Dimethyl sulfoxide
DTT	Dithiothreitol
EMT	Epidermal-mesenchymal transition
FBDD	Fragment based drug design
FITC	Fluorescein
FP	Fluorescence polarization
FRET	Förster resonance energy transfer
HD1	Homology Domain 1
HECT	Homologous to the E6-AP Carboxyl Terminus
HTS	High-throughput screen
IC₅₀	Half-maximal inhibitory concentration
IPTG	Isopropyl β -D-1-thiogalactopyranoside
ITC	Isothermal titration calorimetry
LIC	Leukemia initiating cells
MEF	Mouse embryonic fibroblasts
NCP	Nucleosome core particle
Ni-NTA	Nickle nitrilotriacetic acid
NMR	Nuclear Magnetic Resonance

PcG	Polycomb group
PDB	Protein Data Bank
PPI	Protein-protein interaction
PRC1	Polycomb Repressive Complex 1
PTMs	Post-translational modifications
RING	Really Interesting New Gene
SAR	Structure Activity Relationship
SD	Standard deviation
Se-Met	Seleno-methionine
TCEP	tris(2-carboxyethyl)phosphine
Trx	Thioredoxin
ULD	Ubiquitin-Like Domain
YAP	Yes-associated protein

Abstract

BMI1 has emerged as a key oncogenic factor in many cancers. BMI1 is best characterized as a component of the vertebrate canonical polycomb repression complex 1 (PRC1). This complex facilitates the mono-ubiquitination of histone H2A and chromatin compaction ultimately leading to transcriptional repression of target genes. BMI1 and its associated PRC1 proteins negatively regulate transcription of hundreds of genes including key tumor suppressor genes. Through these mechanisms BMI1 activity has been linked to unregulated cellular proliferation, tumor metastasis and cancer-initiating cell self-renewal. Accumulated data suggest that BMI1 may be a potential target for pharmacological intervention. Previous work suggested that BMI1 has multiple protein binding partners both within the PRC1 complex and with non-polycomb proteins, although the molecular details and functional significance of these interactions remain poorly characterized. Motivated by the prospects to target these protein-protein interactions with small molecule inhibitors, we pursued a multi-pronged campaign to: 1) characterize BMI1 protein-protein interactions at the molecular level and 2) develop novel chemical tools to explore BMI1 function in both normal and cancer biology.

Using a hybrid solution NMR and X-ray crystallography approach, we solved the 3D structure of BMI1 in complex with its PRC1 binding partner protein PHC2. In conjunction, we performed a detailed biochemical and biophysical characterization of the BMI1 protein-protein interaction domain and demonstrated novel mode of self-association of this domain. Mutagenic disruption of both BMI1-PHC2 and BMI1-BMI1 interactions blocks cellular proliferation demonstrating that multiple protein-protein interactions are critical for BMI1 function. On the basis of these findings, we designed two biochemical assays to quantify the BMI1-PHC2

interaction. We employed these assays as a platform for high-throughput screening to discover small molecule inhibitors of BMI1. Through this screen, we identified three classes of small molecule inhibitors that bind directly to BMI1 to disrupt the BMI1-PHC2 protein-protein interaction. Through mechanistic studies by NMR spectroscopy and mass spectrometry we determined that these molecules represent three different strategies for BMI1 inhibitor development.

As a complementary approach to inhibit BMI1 we have pursued development of small molecule inhibitors of the Ring1B/BMI1 E3 ubiquitin ligase complex. Through extensive structural-activity relationship studies a small fragment-like ligand with millimolar affinity was optimized to yield a potent and specific inhibitor with low micromolar inhibitory activity in an *in vitro* ubiquitin ligase assay. Through mutagenesis and solution NMR studies, we demonstrated that Ring1B/BMI1 inhibitors induce significant protein conformational change and that the inhibitor-bound conformation is incompatible with nucleosome binding by Ring1B. In cellular experiments, Ring1B/BMI1 inhibitors decrease global levels of ubiquitinated H2A in MCF10A cells demonstrating cellular permeability and on-target activity. These molecules represent the first direct-binding inhibitors of Ring1B/BMI1 and have a novel mechanism of action through inserting into the Ring1B RING domain leading to significant conformational change that prevents direct protein-nucleosome interaction.

Overall, this work contributes to the understanding of BMI1 function through characterization of its multiple protein-protein interactions and demonstrates that these interactions can be inhibited by small molecules representing novel strategies to target this protein for development of new chemical tools or potential therapeutics for cancer.

Chapter 1: Introduction

A. Motivation

The work in this dissertation is motivated by the goal of attaining a molecular understanding of the role of the oncogene BMI1 in cancer. BMI1 (B lymphoma Mo-MLV insertion region 1 homolog) has been identified as a critical factor regulating tumor growth and aggressiveness in many pathologies. Despite interest in targeting BMI1 for potential pharmacological intervention, the molecular mechanisms governing BMI1's role in normal and cancer biology remain elusive. Of particular interest is understanding BMI1 within the context of the multi-subunit polycomb repressive complex 1 (PRC1). As a member of this complex, BMI1 negatively regulates transcription of many genes involved in cell fate. Lacking enzymatic activity, BMI1 function is determined by its protein-protein interactions, which are currently incompletely characterized. Comprehensive molecular descriptions of these interactions can therefore provide opportunities for inhibitor development through targeted disruption of protein-protein interactions.

To support these goals, I have focused on developing a structural and biochemical understanding of BMI1's multiple protein-protein interactions to facilitate development of small molecules that can serve as chemical tools to explore BMI1 oncogenic function. By applying these compounds to modulate BMI1 function in various contexts we expect to be able to design new strategies to inhibit this protein in cancer.

In this chapter I will briefly discuss the mechanisms of BMI1 and polycomb proteins in transcriptional silencing and expand on the role of BMI1 in cancer biology. Given the

implications of BMI1 as a potential therapeutic target I will also highlight exciting advances in drug discovery through small molecule inhibitors of protein-protein interactions.

B. Transcriptional regulation through chromatin modification

In eukaryotic cells nuclear DNA is packaged into chromatin fibers through the assembly of nucleosomes. Nucleosomes are the repeating unit of chromatin composed of 145-147 base pairs of DNA wrapped around a protein octamer of four histone proteins, H2A, H2B, H3 and H4. Higher order condensation of nucleosomes into compact particles allows up to six linear feet of DNA to be packaged into a single cell.¹ While such packaging is an efficient storage mechanism for genetic material, access to the underlying DNA is essential for cellular function through gene expression and DNA replication. Thus dynamic processes are required to regulate both global chromatin organization and the local transcriptional activity of individual genes.

A key feature of this process is post-translational modification (PTM) of histone proteins. These mechanisms have been extensively reviewed²⁻⁷ and for brevity I will summarize the overall themes. Histone modifying proteins with enzymatic activity catalyze the addition or removal of methyl, phospo, acetyl or ubiquitin groups, among others, to histone residues. Histone PTMs serve to regulate gene transcription through a number of mechanisms, although a consensus understanding of these processes is not yet defined. Modulation of the biochemical properties of histone residues through modification alters interactions between histones and DNA and between nucleosomes thus changing the local environment of chromatin organization. Additionally, histone PTMs can serve as binding sites for effector proteins with “reader” domains that bind specifically to modified residues. Through association of these proteins with other proteins such as chromatin remodelers or transcription factors the modifications thus serve as signals for activated or repressed gene transcription.⁸ Specific modifications at different residues are associated with active or inactive transcription. For example, histone 3 lysine 4 trimethylation is associated with active transcription whereas histone 3 lysine 27 trimethylation is a mark of repressed transcription.^{1,9} While tempting to interpret these modifications as a “code,” the cellular context such as maturity, phase in cell cycle or other epigenetic memory mechanisms can further dictate the functional outcome of individual modifications.^{6,10,11} Recent work has defined mechanisms of histone PTM cross talk where the modification of one residue predicates or precludes modification of another residue at either a local or distant chromatin

domain. Such crosstalk therefore further serves to coordinate transcriptional profiles and complicates analysis of individual modifications.^{6,11,12}

Of particular interest to this work are the polycomb group (PcG) proteins. PcG proteins are an evolutionary conserved family of proteins responsible for catalyzing histone modifications associated with transcriptional repression.

C. Polycomb proteins negatively regulate gene transcription

C.1. Canonical polycomb gene silencing mechanisms

In the classical understanding of polycomb biology, PcG proteins form two distinct multi-subunit complexes with distinct chromatin modifying functions. In mammals, the polycomb repressive complex 2 (PRC2) is comprised of EZH2, EED and SUZ12¹³⁻¹⁶ and the polycomb repressive complex 1 (PRC1) is made up by Ring1B, BMI1, PHC2 and CBX7.^{17,18} The canonical mammalian polycomb transcriptional repressive mechanism has been reviewed extensively.¹⁹⁻²⁵ Briefly, PRC2 is recruited to chromatin where the SET-domain containing subunit, EZH2, catalyzes the trimethylation of histone 3 lysine 27 (H3K27me3).^{13,26} This methyllysine is subsequently recognized by a chromo-domain containing subunit of PRC1, CBX7,^{26,27} recruiting the remainder of the complex including the heterodimer of Ring1B/BMI1 which catalyzes the monoubiquitination of histone H2A at lysine 119 (H2AK119Ub).^{28,29} The canonical polycomb gene silencing mechanism is illustrated in Figure 1.1.

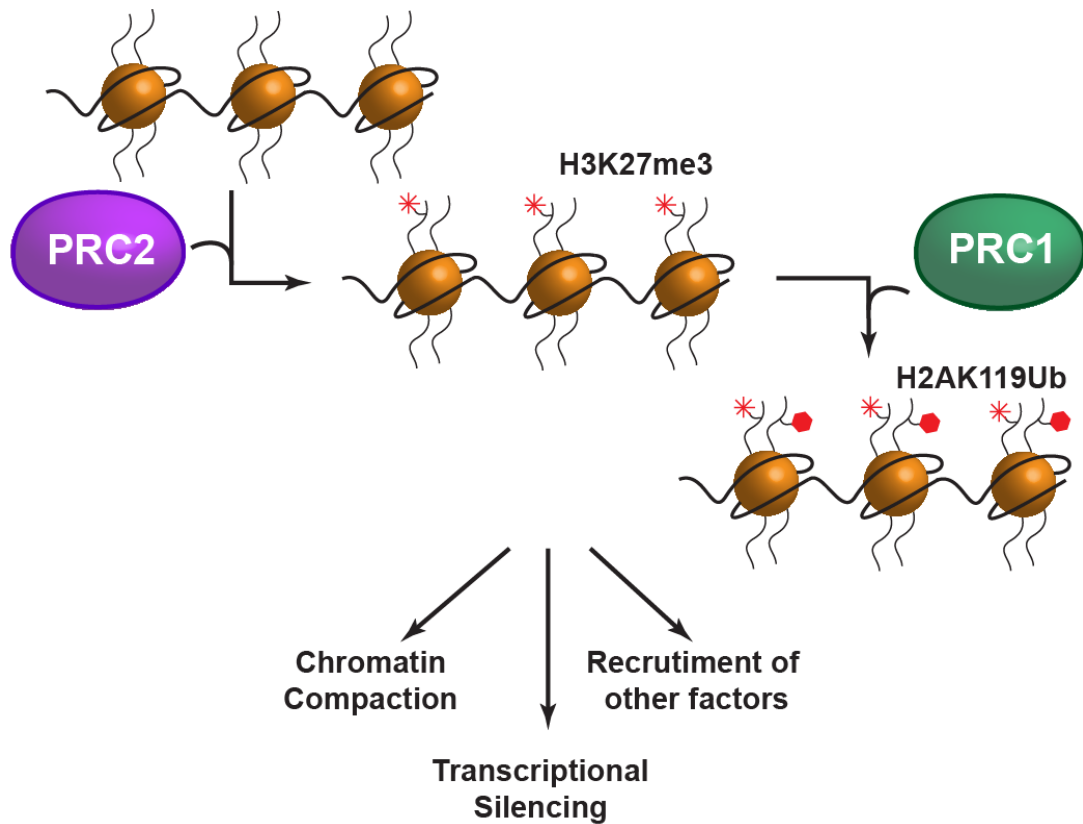


Figure 1.1. Canonical hierarchical mechanism of polycomb-mediated gene silencing.

Histone ubiquitination is one of the most abundant histone modifications in cells and is found on about 10% of endogenous H2A proteins at any given time.³⁰ In mouse embryonic stem cells this mark decorates the promoters of over 500 genes³¹ regulating transcription of genes involved in cell fate and development.³²

While the mechanistic function of histone ubiquitination in polycomb gene silencing is incompletely characterized it has been demonstrated to be critical in gene regulation in some contexts. Cao and co-workers used western blot and gene expression studies to illustrate that knockout of Ring1A or BMI1 decreases global H2A ubiquitination levels and that this is correlated with increased expression of *Hox* genes in mouse embryonic fibroblasts.³³ In similar studies in embryonic stem cells conditional Ring1B knockout abolished global H2A ubiquitination which was correlated with de-repression of polycomb target genes such as *Gata4* and *HoxA7*.³⁴

While H3K27me3, H2AK119Ub and the presence of PcG proteins on chromatin is associated with chromatin compaction and repressed transcription, it is still an open question if PcG activity is correlative with silent chromatin or an active driver of gene silencing.³⁵ Indeed, the Bickmore and Koseki groups identified classes of genes in embryonic stem cells that were repressed by even a catalytically inactive PRC1 complex suggesting that histone ubiquitination is not required for polycomb silencing mechanisms in all contexts.^{31,36}

C.2. Non-canonical polycomb gene silencing mechanisms

Several lines of evidence have emerged challenging the hierarchical mechanism of the canonical polycomb pathway. ChIP and fluorescent microscopy studies demonstrated that PRC1 and PRC2 complexes can occupy distinct genes³⁷⁻⁴⁰ and it was observed that PRC1 can be recruited to chromatin in the absence of H3K27me3.^{39,41-43} Recently, it was demonstrated in multiple contexts that PRC2 can be recruited to chromatin by non-canonical PRC1 complexes through histone ubiquitination.^{44,45} These results flip the polycomb hierarchy and suggests there are intricate regulatory mechanisms involved that remain to be discovered. Finally, in mice and cultured hematopoietic stem cells disruption of PRC1 and PRC2 complexes results in different phenotypes signifying that different transcriptional networks are regulated by the different complexes.⁴⁶ Together these data suggest that polycomb silencing mechanisms are likely highly context-dependent and provide support for the development of chemical tools that can be used to interrogate these systems.

C.3. Architecture of the canonical PRC1 complex

Initial biochemical characterization of the canonical vertebrate PRC1 complex revealed four protein subunits CBX7, Ring1B, PHC2 and BMI1.¹⁸ Subsequent structural and biochemical studies determined some of the direct protein-protein interactions that coordinate overall PRC1 organization. In this context the N-terminal RING domains of Ring1B and BMI1 proteins^{28,29} heterodimerize to form the E3 ubiquitin ligase core of the complex. Ring1B binds directly to CBX7 via a C-terminal ubiquitin-like domain in Ring1B.⁴⁷ The CBX7 chromodomain recognizes methylated histone lysines contributing to gene targeting for the entire complex.⁴⁷⁻⁴⁹ The PHC2 subunit has two conserved domains, a homology domain (HD1) and a C-terminal SAM (sterile alpha motif) domain. SAM domains can form helical polymers⁵⁰⁻⁵² and it is suggest that PHC2

SAM domain polymerization contributes to the clustering of PRC1 complexes associated with chromatin condensation and gene silencing.^{53,54} Prior studies indicated that BMI1 and PHC2 interact directly and that this interaction is mediated by the BMI1 C-terminus.^{55,56} The structure and biochemical characterization of this interaction is detailed in Chapter 2. The overall domain organization of proteins in the canonical PRC1 complex is illustrated in Figure 1.2.

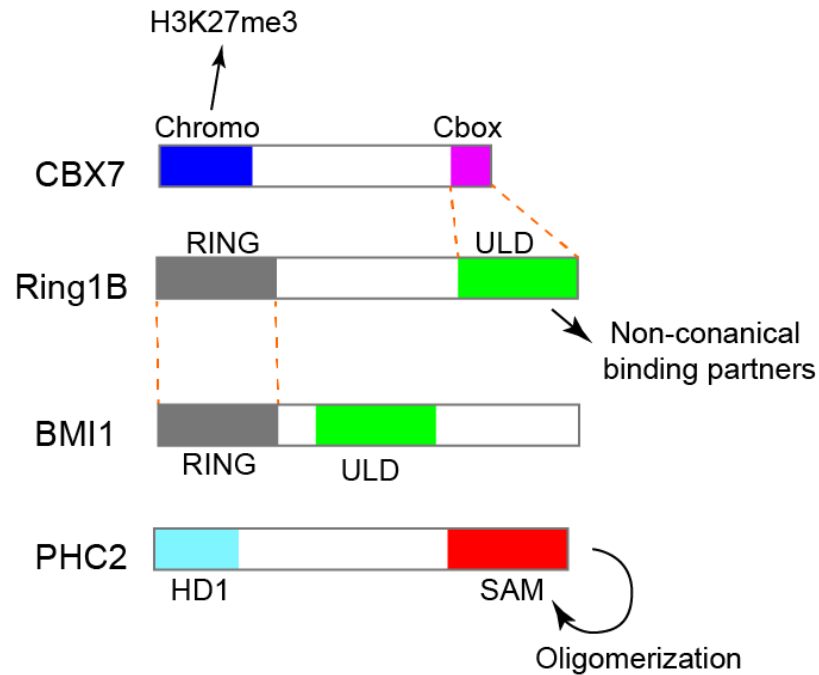


Figure 1.2. Domain organization of proteins in the canonical PRC1 complex.

Characterized domains for each protein are labeled. Orange lines indicate direct protein-protein interactions within the complex that have been structurally characterized prior to this work. Known interactions with other factors are also labeled.

While these subunits represent the best characterized members of the canonical PRC1 complex, recent reports of variant PRC1 complexes with either different orthologs of the canonical PRC1 proteins or with distinct non-polycomb subunits have been identified with non-conserved functions.^{42,57,58} In mammals there are multiple orthologs of PRC1 proteins, including CBX (CBX 2,4,6,7,8), BMI1/PCGF4 (PCGF 1,2,3,4,5,6), PHC2 (PHC 1,2,3) and Ring proteins (Ring1A, Ring1B). Illustrating their distinctive functionalities different CBX orthologs have been found to dominate in certain developmental stages and contribute to regulation of distinct gene targets. For example, CBX7-containing complexes repress pro-differentiation genes in

embryonic stem cells. CBX7 expression decreases with cell differentiation and this is complimented by increasing expression of CBX2 and CBX4 which regulate lineage commitment genes.^{43,59} Further, all PCGF orthologs of BMI1 have conserved RING domains and thus incorporation of these subunits into PRC1 complexes is believed to be mutually exclusive. These PCGF proteins contribute to PRC1 complex heterogeneity through interacting with distinct binding partners, although the cellular context and relevance for all complexes have not been established.^{44,60-62} The mechanisms and functions of this heterogeneity is an area of ongoing research.

D. Polycomb function in biology

D.1. Phenotypes of polycomb deletion suggest complex functional roles

Polycomb genes were originally identified in *Drosophila* as regulating body segmentation of the insect.^{63,64} The family name is derived from the “poly-comb” phenotype observed when these genes were deleted.^{63,65,66} The role of PcG proteins in regulation of organism development is conserved in vertebrates, as explored with PcG knockout mice.^{67,68} All mice with disrupted PRC1 components have skeletal abnormalities and compromised viability. The most extreme phenotype is observed with *Ring1B* knockout which is embryonic lethal. In contrast, *Ring1A* is less ubiquitously expressed and *Ring1A* deletion mice which have developmental aberrations but survive full term suggesting that although these homologs can mechanistically compensate as an E3 ubiquitin ligase, they have different functional roles.^{69,70} *Phc1* knockout mice display postnatal lethality and have impaired hematopoietic activity in conjunction with skeletal and cardiac abnormalities.⁷¹⁻⁷³ *Cbx* knockout mice also have high perinatal mortality, hematopoietic defects and male-to-female sex reversal.⁷⁴ Similarly to Ring protein homologs, Cbx proteins have non-redundant spatial and temporal distributions and do not completely compensate for loss of one protein.⁴³

While loss of *Bmi1* is not embryonic lethal, *Bmi1* deficient mice are also barely viable and surviving animals exhibit severe hematopoietic and immune deficiencies as well as cerebral and skeletal deformities.⁷⁵⁻⁷⁹ Through knockout studies BMI1 was identified as an important stem cell gene^{30,80,81} regulating the self-renewal and pluripotency phenotypes of adult hematopoietic,^{77,78,82-84} neuronal^{79,85} and prostate stem cells.⁸⁶ While not explored in detail to-

date, the extreme phenotypes of *Bmi1* knockout mice suggest that PCGF homologs do not fully compensate for BMI1 function. Overall, these studies demonstrate a critical link between BMI1 molecular mechanisms and many important biological processes.

D.2. Delineating polycomb function through identification of gene targets

Understanding the gene targets of polycomb proteins is critical to understand their function. One approach to target identification is through ChIP analysis of genes modified by H2AK119Ub. However, the incomplete characterization of the role of H2A ubiquitination in polycomb-mediated gene silencing means that identifying polycomb target genes based on analysis of chromatin H2A ubiquitination can lead to both false positives and negatives. First, it has been demonstrated that transcriptional repression of some genes is independent of histone H2A ubiquitination but dependent on PcG chromatin occupancy, whereas regulation of a different group of genes is dependent on histone H2A ubiquitination.^{31,36} Further, other E3 ubiquitin ligase enzymes with activity towards H2A have been identified.^{87,88} Suggesting that not all H2A ubiquitination is the result of polycomb activity, complicating target gene identification by this method.

Another approach to identify target genes is through gene expression profiling following knockdown of polycomb proteins. However, this analysis can be complicated by epigenetic inheritance of chromatin modifications. These mechanisms can contribute to silencing of genes through subsequent generations⁸⁹⁻⁹¹ precluding a complete analysis of the effects of genetic inhibition of polycomb proteins. Given these complications evaluation of PRC1 chromatin occupancy in addition to histone modification and gene expression change following complex disruption may only hint at a description of polycomb target genes. With these caveats in mind, a number of recent studies provide insight into the networks and pathways regulated by vertebrate polycomb proteins.

Bracken and coworkers evaluated chromatin occupancy of PRC1, PRC2 and H3K27me3 through ChIP-on-chip genome wide mapping of from human embryonic fibroblasts.³² This work identified 1000s of co-occupied genes of which almost 300 were de-repressed following siRNA depletion of BMI1. Through analysis of these occupied genes it is understood that PcG proteins regulate many genes involved in cell fate, development and differentiation. For example,

members of the Hedgehog, Notch and Retinoic Acid signaling pathways along with some *Hox* genes were identified as target genes.³² The Jaenisch group similarly performed genome-wide ChIP-based mapping of PRC1 subunits Ring1B and PHC1 in embryonic stem cells, identifying over 1,000 genes with these proteins enriched near the transcription start site.⁹² Gene ontology analysis revealed a bias of occupied genes for factors involved in transcription and developmental processes such as organogenesis, neurogenesis, embryonic development and cell differentiation. Transcription factors suggested to be regulated by PRC1 include members of the *Sox*, *Hox* and *Fox* families which have essential roles in embryonic development and can be misregulated in disease.⁹³⁻⁹⁵ Finally, Endoh et al examined global H2AK119Ub through ChIP-on-chip experiments in embryonic stem cells.³¹ These efforts identified over 500 genes that were enriched for this mark and confirmed that many polycomb target genes are regulators of transcription or development. Together these data provide mechanistic support for PcG activity in regulation of normal organism developmental.

E. BMI1 is an oncogene implicated in many cancers

Complimentarily to BMI1's role in normal biology, many lines of evidence suggest a critical role for BMI1 in cancer biology. BMI1 was originally identified as an oncogene collaborating with *myc* in murine lymphomagenesis. In these models overexpression of *Bmi1* accelerated *myc*-driven lymphoma^{96,97} and could induce de novo lymphomas when expressed by itself.⁹⁸ Subsequent investigations into the oncogenic role of BMI1 have established it as an important factor in many solid and blood cancers through multiple molecular mechanisms.

E.1. Clinical implications of BMI1 overexpression

Many human malignancies have high levels of BMI1 protein. These include solid cancers such as colorectal carcinoma,^{99,100} breast cancer,^{101,102} prostate cancer,¹⁰³ head and neck cancers,^{79,104} hepatocellular carcinoma,¹⁰⁵⁻¹⁰⁸ non-small cell lung carcinoma^{109,110} and hematopoietic diseases such as lymphoma,¹¹¹ and acute myeloid leukemia.¹¹² BMI1 has also been identified as a prognostic indicator in a number of pathologies. For instance, acute and chronic myeloid leukemia patients (AML and CML) with high levels of BMI1 had decreased overall survival and decreased disease free survival compared to patients with lower BMI1 levels.^{112,113} Similar correlations have been identified for patients with head and neck cancers,¹¹⁴

diffuse large B cell lymphoma,¹¹⁵ colorectal cancer,^{100,116} lung cancer,¹¹⁰ liver cancer¹⁰⁶ and tongue cancer.¹¹⁷ Confounding alternative studies have found correlations between low BMI1 levels and poor patient prognosis in breast cancer¹¹⁸ and glioblastoma.¹¹⁹ These data suggest that BMI1 likely is just one oncogenic factor among many that contribute to these various pathologies and the dynamic interplay between the factors can dictate patient outcome.

E.2. Mechanistic insight into BMI1 as an oncogene

In a concise summary of the biological “hallmarks” of cancer Hanahan and Weinberg characterized six cellular capabilities enabling tumor growth and metastatic propagation.¹²⁰ These hallmarks include sustained proliferative signaling, evading growth suppressors, resisting cell death, enabling replicative immortality, inducing angiogenesis and activating invasion and metastasis. Though many mechanistic studies BMI1 has been identified to contribute to a number of these hallmarks. A brief summary is presented below and illustrated in Figure 1.3.

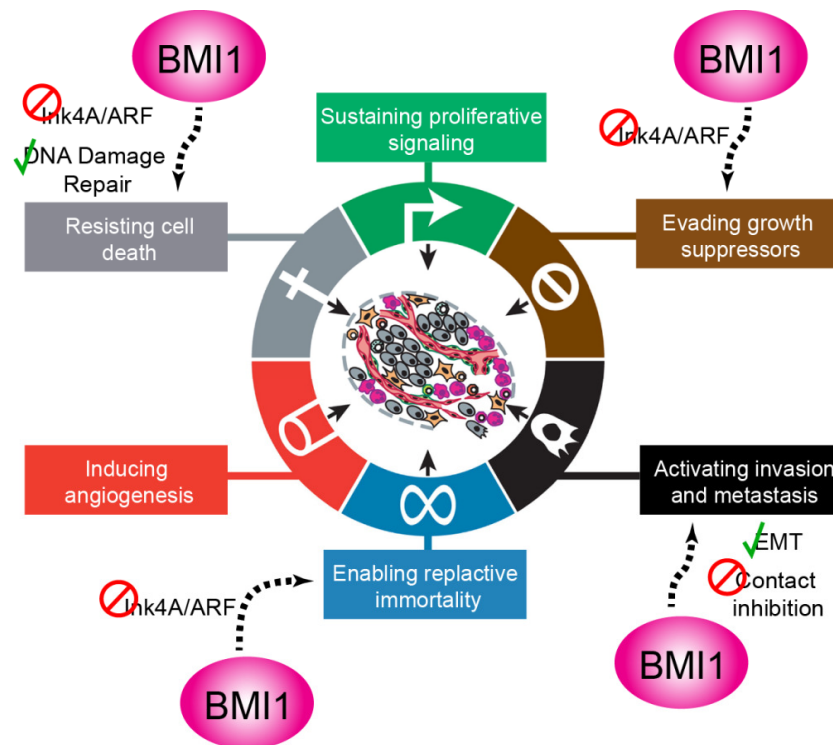


Figure 1.3. Contributions of BMI1 to the hallmarks of cancer.

Figure adapted from Hanahan and Weinberg. {Hanahan, 2000 #695}

E.2.1. BMI1 contributes to cellular avoidance of growth suppression and cell death

Early evidence implicating BMI1 as an oncogene was the demonstration that BMI1 promotes aberrant cellular proliferation through the transcriptional repression of two critical tumor suppressor proteins p16 and p14.¹²¹ In humans the *p16^{Ink4A}* and *p14^{Arf}* genes are encoded in the Ink4A/ARF locus through alternate reading frames.^{122,123} The p16 protein stimulates cellular senescence through the retinoblastoma pathway and the p14 protein promotes cellular apoptosis through the p53 pathway. Therefore repression of transcription of these genes by BMI1 and PcG proteins leads to avoidance of normal cellular senescence and apoptosis pathways. Many studies have demonstrated the knockdown of BMI1 leads to increased expression of *p14* and *p16* in multiple cell lines.^{78,124-133}

While avoidance of cell cycle arrest and apoptosis mechanisms through the Ink4A/ARF pathway is a well-defined mechanism for BMI1 in cancer, there is ample evidence of BMI1 contributing to cancer by Ink4A/ARF-independent channels. For example, BMI1 induced transformation of *Ink4A/Arf*^{-/-} astrocytes, promoting an extended proliferative capacity.¹³⁴ In a mouse model for glioma, mice injected with *Bmi1* wild type *Ink4A/Arf*^{-/-} cells developed brain tumors with rapid morbidity, whereas mice injected with *Bmi1*^{-/-}; *Ink4A/Arf*^{-/-} were protected from this aggressive disease.¹³⁴ Similar results have been found in models for pancreatic cancer,¹³⁵ hepatocellular carcinoma,^{133,136} leukemia,^{137,138} prostate cancer,¹³⁹ Ewing sarcoma¹⁴⁰ and breast cancer.¹⁴¹ Together these data support alternative oncogenic mechanisms for BMI1.

E.2.2. BMI1 contributes to tumor metastasis and aberrant cellular proliferation

A further mechanism for BMI1 contributing to tumor development and cancer progression is through regulation of the epithelial-mesenchymal transition (EMT).^{101,142-144} EMT is a critical pathway promoting cancer metastasis through induction of mesenchymal properties such as motility, invasiveness and resistance to apoptosis in epithelial cells.^{145,146} Mechanistically, EMT is regulated in vertebrates by the Twist and Snail transcription factors which repress expression of *E-cadherin*, leading to reduction of adherens junctions and increased cellular migratory capacity.¹⁴⁶ Recent work by Yang and coworkers established that under hypoxic conditions TWIST1 overexpression leads to *BMI1* upregulation¹⁴³ which enhances the tumor-initiating capacity of head and neck squamous cell carcinoma cells. While not sufficient to

reverse EMT on its own, loss of BMI1 in conjunction with loss of TWIST1 led to increased *E-cadherin* expression inducing reversion of EMT and decreased migration *in vitro* indicating that BMI1 is a direct molecular driver of these processes. Complementarily, BMI1 was shown to suppress transcription of the tumor suppressor *PTEN*¹⁴² leading to activation of the PI3K-AKT signaling pathway and stabilization of the other EMT inducer, SNAIL. Other supporting evidence includes the demonstration that other factors including ER α and ZEB1 regulate *BMI1* and therefore modulate EMT.^{144,147} Together these data demonstrate that BMI1 contributes to EMT through multiple mechanisms and overall promotes cell migration and tumor metastasis.¹⁴⁸ Inhibition of BMI1 may therefore represent a strategy to repress these mechanisms in aggressive pathologies.

Another cellular mechanism contributing to cancer cell outgrowth and invasion is through avoidance of contact inhibition pathways.^{149,150} Contact inhibition is a cellular property which restricts the *in vitro* growth of cells in confluent conditions.^{151,152} Mechanistically, contact inhibition is regulated in part by the Hippo pathway via phosphorylation and inactivation of the downstream effectors YAP and TAZ.¹⁵³ Aberrant activation of these transcriptional co-activators induces expression of proliferation-promoting factors leading to avoidance of contact inhibition. Hsu and Lawlor observed that YAP protein, but not transcript, levels were reduced in Ewing sarcoma family tumor cells with BMI1 knockdown and this effect was correlated with reduced tumorigenic potential.^{140,154} This data suggest that in this tumor context, BMI1 may contribute to irregular cellular proliferation and invasiveness via the Hippo-YAP pathway. The molecular mechanisms of this effect are still under investigation, however this work contributes to a growing body of evidence supporting Ink4A/ARF-independent mechanisms for BMI1 in cancer biology.

E.2.3. BMI1 protects cells from arrest through stimulating DNA damage repair and protection from oxidative stress

A third way in which BMI1 contributes to aberrant cell proliferation is through activation of DNA repair pathways.^{155,156} Multiple lines of evidence suggest that BMI1 and other PRC1 components are recruited to DNA double-strand break (DSB) sites where Ring1B/BMI1 ubiquitinates both histone H2A and phosphorylated H2AX (γ H2AX).¹⁵⁷⁻¹⁶⁰ This activity serves to promote accumulation of other factors in the DSB repair pathway such as BRCA1, RNF168

and 53BP1. While the precise timing of BMI1 recruitment in the repair cascade remains an open question, knockdown of BMI1 leads to high levels of chromosome breaks¹⁵⁸ and cell cycle arrest in G2/M.¹⁵⁹ Further, loss of BMI1 sensitizes cells to ionizing radiation^{160,161} and genotoxic agents.^{158,162} Together these data suggest that BMI1 has an important role in genome protection allowing cells to avoid DNA-damage induced arrest.¹⁶³

BMI1 has also been shown to promote resistance to oxidative stress in a number of cell models.^{156,162} Liu and coworkers demonstrated that hematopoietic stem cell and thymocytes from *Bmi1*^{-/-} mice have increased levels of reactive oxygen species (ROS) which was correlated with increased expression of factors related to redox homeostasis.¹⁵⁶ Further, mitochondria from *Bmi1*^{-/-} thymocytes had reduced oxygen consumption and oxidative capacity indicating that mitochondrial disruption in the absence of BMI1 may be a source of ROS.¹⁵⁶ Similar correlations between reduced *Bmi1* levels and increased ROS were found in embryonic neurons,¹⁶⁴ ovarian cancer cells¹⁶² and hematopoietic cells.¹⁵⁸ In these BMI1-depleted cell culture models increased ROS levels triggered activation of the DNA damage response leading to caspase cleavage and induction of apoptosis.^{156,162,164}

Together this data supports a complex role for BMI1 in both upstream genome protection from oxidative stress and a downstream response to DNA double-strand breaks coordinating their repair and avoidance of apoptosis. Further work is required to sort the precise molecular mechanisms of BMI1 in these pathways.

E.2.4. BMI1 enables self-renewal of cancer initiating cells

The theory of cancer initiating cells (CIC) suggests that there is a subpopulation of tumor cells that possess stem-like properties such as self-renewal and pluripotency.¹⁶⁵⁻¹⁶⁷ These cells are also associated with EMT¹⁶⁸⁻¹⁷⁰ and can sustain tumor growth or establish new tumor populations. Given that CICs are generally resistant to chemo and radiation therapies these cell populations are therefore implicated in tumor relapse and have been proposed to be an important target for novel therapy development.^{165,171}

BMI1 has been identified as a critical factor regulating cancer initiating cells in a number of contexts.^{30,81,172} This was first revealed in studies by the Sauvageau^{83,173}, Morrison⁸⁵ and Clarke⁷⁸ groups in models for normal hematopoiesis and leukemogenesis. In these studies fetal

liver cells from *Bmi1*^{-/-} mice had reduced proliferative capacities *in vitro* and failed to repopulate the hematopoietic compartment of recipient mice.⁷⁸ In contrast, cells derived from *Bmi1* wild type mice engrafted in irradiated recipient mice,⁷⁸ demonstrating that BMI1 is critical for normal hematopoiesis and the maintenance of hematopoietic stem cells (HSCs). In a murine model for acute myeloid leukemia (AML), mice transplanted with fetal liver cells from *Bmi1* wild type or ^{-/-} mice transduced with *Hoxa9-Meis1* all developed leukemia with a similar latency demonstrating that *Bmi1* is not essential for establishing leukemia in primary recipients.⁸³ However, subsequent transplantation of *Bmi1*^{-/-} bone marrow cells from primary recipients into secondary recipients protected mice from AML, suggesting that BMI1 depletion leads to progressive loss of leukemia initiating cells (LIC).⁸³ Supporting *in vitro* evidence demonstrated that loss of BMI1 caused LIC cell cycle arrest in G1 phase, differentiation to macrophages, increased apoptosis and reduced colony formation in methylcellulose assays.⁸³ Subsequently, BMI1 has been implicated in tumorigenesis and the maintenance of CICs in models of prostate cancer,⁸⁶ gliomas,^{174,175} hepatocellular carcinoma¹⁷⁶ and colorectal cancer.¹⁷⁷ Mechanistically, BMI1 regulation of the *Ink4a/Arf* gene products p16 and p14,⁸³ some *Hox* genes and members of the Wnt, Hedgehog and Notch^{32,92} pathways likely contribute to self-renewal and maintenance of CICs in a number of contexts.¹⁷⁸⁻¹⁸¹

The studies outlined above provide detailed mechanistic support for a complicated oncogenic role for BMI1 in many tumor types.

E.3. Genetic inhibition of BMI1 in cancer models supports small molecule inhibitor development for this target

Numerous genetic inhibition studies have demonstrated that many cancer cell types are dependent on BMI1 for continued proliferation and tumorigenesis. In a study comparing the effect of RNAi BMI1 knockdown in a panel of normal and cancer cells the Tollefsbol group established that loss of BMI1 led to acute cell death and growth inhibition of embryonic, breast, ovarian and neuroblastoma cancer cell lines.¹²⁷ In contrast, loss of BMI1 in normal embryonic lung and stem cells, mammary epithelial, lung fibroblasts, skeletal muscle cells and brain cortical neurons led to only moderate growth inhibition and no significant cell death.¹²⁷ This suggests there may be a therapeutic window to inhibit BMI1 in cancer cells without adverse effects on normal cells.

The anti-proliferative effect of BMI1 knockdown has been further confirmed in culture models for many solid and blood cancers. shRNA and siRNA knockdown of BMI1 significantly decreased cellular proliferation of gastric carcinoma,¹⁸² hepatocellular carcinoma,^{108,176,183} breast cancer¹⁴⁰ and colorectal cancer cells.¹⁷⁷ Accordingly, BMI1 knockdown significantly reduced tumor growth in mouse xenograft models for breast cancer,^{184,185} liver cancer,^{108,136,176,183} ovarian carcinoma¹⁸⁶ and Ewing sarcoma.¹⁴⁰ In exciting recent studies, Kreso *et al* demonstrated that reduction of *BMI1* transcript levels in colon cancer initiating cells impairs self-renewal and significantly reduces tumor formation in mouse xenograft models of human colon cancer.¹⁷⁷

In models for hematological malignancies, loss of BMI1 through RNA knockdown reduced colony formation and prevented leukemic transformation by PLZF-RAR α ,¹⁸⁷ and decreased multiple myeloma cell proliferation *in vitro* and tumor formation *in vivo*.¹⁸⁸ Similarly, loss of BMI1 prevented leukemic transformation of cells with the oncogenic AML1-ETO chimeric transcription factor.¹⁸⁹ In a detailed analysis, Rizo and coworkers demonstrated that in primary AML CD34+ cells treated with BMI1 RNAi had impaired self-renewal, colony formation and long-term growth.⁸² BMI1 knockdown accordingly increased expression of the cell cycle regulator *INK4A/ARF* genes products.⁸² The CD34+ cell population has been identified as leukemia initiating cells (LIC) in some AMLs¹⁹⁰⁻¹⁹² and these results support the previous finding that BMI1 knockdown in CD34+ AML cells precluded the formation of AML in secondary mouse recipients.⁸³

Together these studies suggest that inhibiting BMI1 may be a novel therapeutic approach in many cancers including through targeting cancer initiating cells. A complicating factor is the role of BMI1 in normal processes as observed in the hematopoietic, gastrointestinal and neurological abnormalities of BMI1 knockout mice.^{77,78,82,83} We believe these extreme phenotypes may be a result of the total BMI1 knockout approach used in these studies which affects both the E3 ligase capacity of the PRC1 complex as well as complex organization. Novel approaches are needed to target different functions of BMI1 within the PRC1 complex to establish if there is a therapeutic window for therapy development

It is worth noting that while there are reports in the literature of small molecule inhibitors of BMI1 these do not have demonstrated direct binding to BMI1 and likely have other molecular targets. The current BMI1 inhibitors either regulate *BMI1* transcription through an unknown

mechanism¹⁷⁷ or inhibit Ring1B/BMI1-mediated H2A ubiquitination likely through covalent modification of the E2 enzyme.^{193,194} This leaves an exciting opportunity for development of well-validated small molecule inhibitors of this important oncogene. Given that BMI1 serves as a scaffolding protein in the PRC1 complex efforts to develop targeted small molecule inhibitors of this protein should therefore focus on the development of inhibitors of its protein-protein interactions (PPIs).

F. Targeting protein-protein interactions with small molecule inhibitors

Protein-protein interactions are critical for most cellular functions¹⁹⁵ and have been identified as promising therapeutic targets to treat many pathologies. Developing small molecule inhibitors of PPIs has traditionally been regarded as a challenging task, however recent successes both in preclinical and clinical studies has increased enthusiasm for these efforts. These topics have been reviewed extensively¹⁹⁶⁻²⁰⁴ so here I will briefly describe the molecular features of PPIs that present challenges for small molecule inhibition before discussing methods to identify such inhibitors and provide examples of exciting PPI inhibitors for chromatin modifying proteins with potential therapeutic value.

F.1. Challenges for small molecule inhibition of protein-protein interactions

Traditional drug targets are proteins or receptors that have well-defined small molecule binding sites such as G-protein coupled receptors, nuclear receptors, ion channels or enzymes.^{205,206} The majority of FDA approved small molecule therapeutics are modulators of these proteins via orthosteric or allosteric inhibition of natural ligand binding. Therefore there is a bias in experimental methods, small molecule libraries and theoretical understanding to design inhibitors of these targets at the exclusion of inhibitors of protein-protein interactions, of which there are hundreds of thousands of potential targets.^{207,208}

A challenging aspect of inhibiting PPIs is the molecular basis of these interfaces. In contrast to receptor-ligand or enzyme-cofactor interactions, PPI interfaces generally have large contact surfaces, ranging from 1,000-2,000 Å² of buried surface area.²⁰⁹⁻²¹¹ The topography of these interfaces is typically flat, lacking the small molecule binding pockets featured in enzyme or receptor ligand binding sites. The affinity of PPIs can be driven by so-called “hot-spot” interactions, where binding free energy (ΔG) is concentrated at specific residues.²¹²⁻²¹⁴

Alternatively, ΔG is dispersed over a larger area or between multiple direct interfaces necessitating larger, more complex inhibitor molecules.¹⁹⁹ Another feature of some PPI interfaces is conformational flexibility and dynamics which allow proteins to attain different structures for transient contacts or interactions with many different binding partners.^{196,199,215}

From an inhibitor-development perspective these features of PPIs have challenged traditional drug discovery approaches based on enzymatic turnover. Conversely, creative methods have been developed to exploit PPI hot-spots and conformational dynamics yielding novel molecules with utility as both chemical tools and clinical candidates.

F.2. Methods of identifying protein-protein interaction inhibitors

Identifying inhibitors of PPIs has historically relied on high-throughput screening (HTS) using assay platforms systems sensitive to the disruption of the complex.²¹⁶ In brief, such assays include biochemical assays which quantify the PPI through interaction of the minimal interaction components,²¹⁷ “gray-box” assays composed of the reconstituted complete protein complex²¹⁸ and cell-based assays monitoring the PPI in the context of the complex cellular environment. HTS involves screening large libraries of small molecules. However, PPI inhibitor identification by this method can be challenged by commercial libraries which have been developed for traditional pharmacological targets such as GCPRs and enzymes. As such, commercial screening libraries can be ill-suited to identify small molecule inhibitors of PPIs which tend to be larger and more hydrophobic than enzyme inhibitors and such molecules are underrepresented in commercial libraries.^{199,219,220} While there are many examples of successful applications of HTS for PPI drug discovery these platforms can be challenged by protein targets that participate in weak or transient interactions or have complicated biochemical assay systems.

In these cases, alternative strategies have been developed to identify small molecule ligands of interesting protein targets. Such tactics include fragment based screening to identify direct-binding ligands based on biophysical methodologies,²²¹⁻²²³ ligand-tethering to trap low affinity interactions via covalent linkage at a particular protein site²²⁴ and *in silico* screening which requires structural information of the PPI interface.^{225,226} Fragment based screening is described in more detail in Chapter 4. Thinking beyond typical small molecules,

peptidomimetics, stapled peptides and natural products are promising approaches for developing PPI inhibitors.

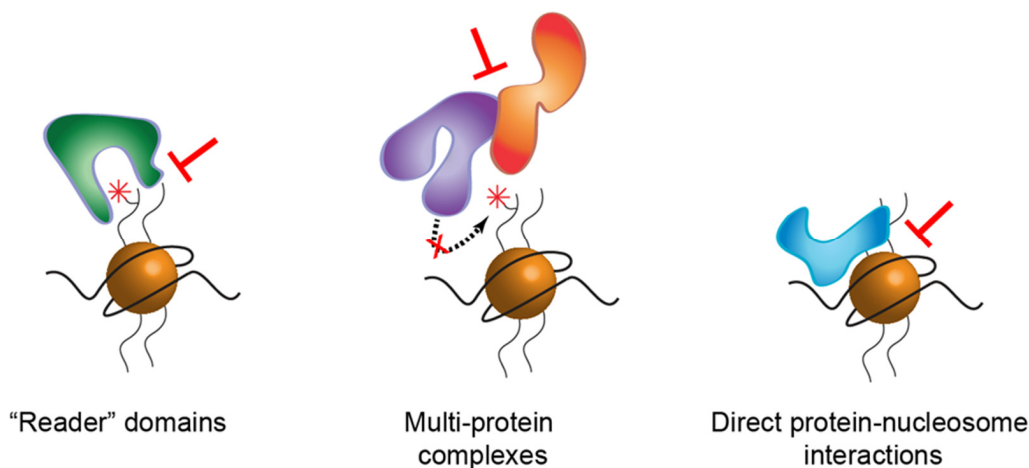


Figure 1.4. Strategies for inhibiting protein-protein interactions in chromatin modification.

Potential targets for inhibition by small molecules are indicated with red bars.

F.3. Examples of PPI inhibitors for chromatin modifying proteins

Chromatin modifying proteins are very attractive targets for inhibitor development in the quest to selectively regulate cellular processes in various pathologies.²²⁷ Many small molecules have been developed to inhibit the enzymes that catalyze histone modification^{228,229} through competition with substrate, cofactors or through allosteric mechanisms. A more difficult challenge is disrupting the protein-protein interactions that drive chromatin recognition and enzyme activity. Figure 1.4 illustrates three strategies for small molecule interference to achieve these goals: 1) disruption of chromatin PTM recognition by “reader” domains; 2) disruption of PPIs in multi-subunit complexes that contribute to catalytic activity of the enzymatic subunit and; 3) disruption of direct protein-nucleosome interactions.

First, recruitment of many chromatin-associated proteins is achieved through recognition of chromatin modifications by reader domains specific for modified residues. Recently exciting molecules targeting these interactions have emerged as both tool compounds and clinical candidates. Some of the most famous examples of these inhibitors are JQ1 and I-BET762 which inhibit acetyl lysine binding by BET-family bromo-domain proteins.^{230,231} JQ1 displaces the BET

family member BRD4 from chromatin leading to reduced *in vitro* cellular proliferation and delayed tumor growth in xenograft models of the BRD4-requiring NUT midline carcinoma. I-BET762 also inhibits bromo-domain proteins and has advanced to phase 1 clinical trials for hematological malignancies.²³² Complementarily, the Frye group has developed UNC1215 which is an inhibitor of methyllysine recognition by the L3MBTL3 protein.²³³ This molecule has been used as a chemical tool to probe the role of lysine methylation in the protein-protein interactions of L3MBTL3.²³⁴ Given the diversity of methyl- and acetyllysine binding domains, the number and scope of chemical tools for this class of proteins will likely expand rapidly.²³⁵

Second, many chromatin modifying enzymes are components of large multi-subunit complexes and protein-protein interactions within the complex can modulate enzymatic activity. As such, an alternative approach to regulate chromatin modification is through disruption of these regulatory protein-protein interactions. A recent example of this is the discovery of inhibitors of the EZH2/EED protein-protein interaction which is required for EZH2-mediated H3K27 methylation.^{236,237} Disruption of this interaction by natural products prevents chromatin modification, relieves PRC2-mediated transcriptional repression and arrests proliferation of EZH2-driven lymphoma cells.²³⁶ Similarly, small molecule inhibition of the menin-MLL interaction prevents chromatin targeting of MLL-fusion proteins with histone methyltransferase activity and reduces transcription of MLL target genes.²³⁸⁻²⁴⁰ In turn, pharmacologic disruption of this interaction suppress *in vitro* proliferation of leukemia cells, reduces tumor growth and prevents leukemia progression in mouse models of MLL leukemia.²³⁸

Lastly, an unmet opportunity in regulating chromatin modification is inhibiting direct interactions between unmodified chromatin and chromatin-binding proteins. This a relatively new area of research due to difficulties determining which proteins within a multi-subunit complex have direct interactions with the nucleosome. Contributions by the Tan group^{241,242} have provided detailed structural and mechanistic insight into these interactions, yet there is a general dearth of such data complicating inhibitor development projects. I expect that as identification and characterization of these interactions are expanded efforts to develop strategies to inhibit them will advance in parallel.

The ever-expanding appreciation for the role of chromatin modification in many pathologies suggest that this areas of research will provide ample opportunity to develop small

molecule inhibitors as tool compounds to illuminate these complicated biological processes and as potential therapeutics.

G. Thesis summary

In this dissertation I describe structural and functional insights into the multiple protein-protein interactions of the PRC1 protein BMI1 and demonstrate different approaches to inhibit these interactions with small molecules. In Chapter 2, I provide a detailed biochemical and biophysical characterization of the BMI1 protein-protein interaction domain and demonstrate through mutagenesis that multiple protein-protein interactions are important for BMI1 regulation of cellular proliferation. In Chapter 3, I report my efforts to identify small molecule inhibitors of the BMI1-PHC2 interaction through a high-throughput screening campaign. In Chapter 4, I detail the development of a novel small molecule inhibitor of the Ring1B/BMI1 E3 ubiquitin ligase through the optimization of a fragment ligand into a potent, specific, and cell-permeable inhibitor of nucleosome ubiquitination through disruption of the interaction between the E3 ligase complex and the nucleosome substrate. Finally, in Appendix A I present an efficient NMR-based method to characterize disordered regions in proteins for structural and functional studies of these proteins.

In summary, this dissertation provides critical insight into BMI1 as a core architectural organizer of the canonical PRC1 complex and demonstrates multiple approaches to modulate BMI1's protein-protein interactions through small molecule inhibitors. These results support efforts to understand BMI1's role outside of the polycomb complex and suggest further promising strategies to target BMI1 with chemical probes or potential therapeutics.

H. References

1. Peterson, C. L. & Laniel, M. A. Histones and histone modifications. *Curr Biol* **14**, R546-551 (2004).
2. Musselman, C. A., Lalonde, M. E., Cote, J. & Kutateladze, T. G. Perceiving the epigenetic landscape through histone readers. *Nat Struct Mol Biol* **19**, 1218-1227 (2012).
3. Tessarz, P. & Kouzarides, T. Histone core modifications regulating nucleosome structure and dynamics. *Nat Rev Mol Cell Biol* **15**, 703-708 (2014).

4. Badeaux, A. I. & Shi, Y. Emerging roles for chromatin as a signal integration and storage platform. *Nat Rev Mol Cell Biol* **14**, 211-224 (2013).
5. Sims, R. J., 3rd & Reinberg, D. Is there a code embedded in proteins that is based on post-translational modifications? *Nat Rev Mol Cell Biol* **9**, 815-820 (2008).
6. Lee, J. S., Smith, E. & Shilatifard, A. The language of histone crosstalk. *Cell* **142**, 682-685 (2010).
7. Bhaumik, S. R., Smith, E. & Shilatifard, A. Covalent modifications of histones during development and disease pathogenesis. *Nat Struct Mol Biol* **14**, 1008-1016 (2007).
8. Strahl, B. D. & Allis, C. D. The language of covalent histone modifications. *Nature* **403**, 41-45 (2000).
9. Ernst, J. & Kellis, M. Discovery and characterization of chromatin states for systematic annotation of the human genome. *Nat Biotechnol* **28**, 817-825 (2010).
10. Berger, S. L. The complex language of chromatin regulation during transcription. *Nature* **447**, 407-412 (2007).
11. Fischle, W., Wang, Y. & Allis, C. D. Histone and chromatin cross-talk. *Curr Opin Cell Biol* **15**, 172-183 (2003).
12. Suganuma, T. & Workman, J. L. Signals and combinatorial functions of histone modifications. *Annu Rev Biochem* **80**, 473-499 (2011).
13. Kuzmichev, A., Nishioka, K., Erdjument-Bromage, H., Tempst, P. & Reinberg, D. Histone methyltransferase activity associated with a human multiprotein complex containing the Enhancer of Zeste protein. *Genes Dev* **16**, 2893-2905 (2002).
14. Kuzmichev, A., Jenuwein, T., Tempst, P. & Reinberg, D. Different EZH2-containing complexes target methylation of histone H1 or nucleosomal histone H3. *Mol Cell* **14**, 183-193 (2004).
15. Pasini, D., Bracken, A. P., Jensen, M. R., Lazzerini Denchi, E. & Helin, K. Suz12 is essential for mouse development and for EZH2 histone methyltransferase activity. *EMBO J* **23**, 4061-4071 (2004).
16. Margueron, R. & Reinberg, D. The Polycomb complex PRC2 and its mark in life. *Nature* **469**, 343-349 (2011).
17. Vandamme, J., Volkel, P., Rosnoblet, C., Le Faou, P. & Angrand, P. O. Interaction proteomics analysis of polycomb proteins defines distinct PRC1 complexes in mammalian cells. *Mol Cell Proteomics* **10**, M110 002642 (2011).
18. Levine, S. S., Weiss, A., Erdjument-Bromage, H., Shao, Z., Tempst, P. & Kingston, R. E. The core of the polycomb repressive complex is compositionally and functionally conserved in flies and humans. *Mol Cell Biol* **22**, 6070-6078 (2002).
19. Simon, J. A. & Kingston, R. E. Mechanisms of polycomb gene silencing: knowns and unknowns. *Nat Rev Mol Cell Biol* **10**, 697-708 (2009).
20. Simon, J. A. & Kingston, R. E. Occupying chromatin: Polycomb mechanisms for getting to genomic targets, stopping transcriptional traffic, and staying put. *Mol Cell* **49**, 808-824 (2013).
21. Muller, J. & Verrijzer, P. Biochemical mechanisms of gene regulation by polycomb group protein complexes. *Curr Opin Genet Dev* **19**, 150-158 (2009).
22. Schwartz, Y. B. & Pirrotta, V. Polycomb silencing mechanisms and the management of genomic programmes. *Nat Rev Genet* **8**, 9-22 (2007).
23. Bantignies, F. & Cavalli, G. Polycomb group proteins: repression in 3D. *Trends Genet* **27**, 454-464 (2011).

24. Di Croce, L. & Helin, K. Transcriptional regulation by Polycomb group proteins. *Nat Struct Mol Biol* **20**, 1147-1155 (2013).
25. Kerppola, T. K. Polycomb group complexes--many combinations, many functions. *Trends Cell Biol* **19**, 692-704 (2009).
26. Cao, R., Wang, L., Wang, H., Xia, L., Erdjument-Bromage, H., Tempst, P., Jones, R. S. & Zhang, Y. Role of histone H3 lysine 27 methylation in Polycomb-group silencing. *Science* **298**, 1039-1043 (2002).
27. Wang, L., Brown, J. L., Cao, R., Zhang, Y., Kassis, J. A. & Jones, R. S. Hierarchical recruitment of polycomb group silencing complexes. *Mol Cell* **14**, 637-646 (2004).
28. Buchwald, G., van der Stoop, P., Weichenrieder, O., Perrakis, A., van Lohuizen, M. & Sixma, T. K. Structure and E3-ligase activity of the Ring-Ring complex of polycomb proteins Bmi1 and Ring1b. *EMBO J* **25**, 2465-2474 (2006).
29. Li, Z., Cao, R., Wang, M., Myers, M. P., Zhang, Y. & Xu, R. M. Structure of a Bmi-1-Ring1B polycomb group ubiquitin ligase complex. *J Biol Chem* **281**, 20643-20649 (2006).
30. Richly, H., Aloia, L. & Di Croce, L. Roles of the Polycomb group proteins in stem cells and cancer. *Cell Death Dis* **2**, e204 (2011).
31. Endoh, M., Endo, T. A., Endoh, T., Isono, K., Sharif, J., Ohara, O., Toyoda, T., Ito, T., Eskeland, R., Bickmore, W. A., Vidal, M., Bernstein, B. E. & Koseki, H. Histone H2A mono-ubiquitination is a crucial step to mediate PRC1-dependent repression of developmental genes to maintain ES cell identity. *PLoS Genet* **8**, e1002774 (2012).
32. Bracken, A. P., Dietrich, N., Pasini, D., Hansen, K. H. & Helin, K. Genome-wide mapping of Polycomb target genes unravels their roles in cell fate transitions. *Genes Dev* **20**, 1123-1136 (2006).
33. Cao, R., Tsukada, Y. & Zhang, Y. Role of Bmi-1 and Ring1A in H2A ubiquitylation and Hox gene silencing. *Mol Cell* **20**, 845-854 (2005).
34. Stock, J. K., Giadrossi, S., Casanova, M., Brookes, E., Vidal, M., Koseki, H., Brockdorff, N., Fisher, A. G. & Pombo, A. Ring1-mediated ubiquitination of H2A restrains poised RNA polymerase II at bivalent genes in mouse ES cells. *Nat Cell Biol* **9**, 1428-1435 (2007).
35. Henikoff, S. & Shilatifard, A. Histone modification: cause or cog? *Trends Genet* **27**, 389-396 (2011).
36. Eskeland, R., Leeb, M., Grimes, G. R., Kress, C., Boyle, S., Sproul, D., Gilbert, N., Fan, Y., Skoultchi, A. I., Wutz, A. & Bickmore, W. A. Ring1B compacts chromatin structure and represses gene expression independent of histone ubiquitination. *Mol Cell* **38**, 452-464 (2010).
37. Yu, M., Mazor, T., Huang, H., Huang, H. T., Kathrein, K. L., Woo, A. J., Chouinard, C. R., Labadorf, A., Akie, T. E., Moran, T. B., Xie, H., Zacharek, S., Taniuchi, I., Roeder, R. G., Kim, C. F., Zon, L. I., Fraenkel, E. & Cantor, A. B. Direct recruitment of polycomb repressive complex 1 to chromatin by core binding transcription factors. *Mol Cell* **45**, 330-343 (2012).
38. Pasini, D., Bracken, A. P., Hansen, J. B., Capillo, M. & Helin, K. The polycomb group protein Suz12 is required for embryonic stem cell differentiation. *Mol Cell Biol* **27**, 3769-3779 (2007).

39. Schoeftner, S., Sengupta, A. K., Kubicek, S., Mechtler, K., Spahn, L., Koseki, H., Jenuwein, T. & Wutz, A. Recruitment of PRC1 function at the initiation of X inactivation independent of PRC2 and silencing. *EMBO J* **25**, 3110-3122 (2006).
40. Vincenz, C. & Kerppola, T. K. Different polycomb group CBX family proteins associate with distinct regions of chromatin using nonhomologous protein sequences. *Proc Natl Acad Sci U S A* **105**, 16572-16577 (2008).
41. Tavares, L., Dimitrova, E., Oxley, D., Webster, J., Poot, R., Demmers, J., Bezstarosti, K., Taylor, S., Ura, H., Koide, H., Wutz, A., Vidal, M., Elderkin, S. & Brockdorff, N. RYBP-PRC1 complexes mediate H2A ubiquitylation at polycomb target sites independently of PRC2 and H3K27me3. *Cell* **148**, 664-678 (2012).
42. Gao, Z., Zhang, J., Bonasio, R., Strino, F., Sawai, A., Parisi, F., Kluger, Y. & Reinberg, D. PCGF homologs, CBX proteins, and RYBP define functionally distinct PRC1 family complexes. *Mol Cell* **45**, 344-356 (2012).
43. Morey, L., Pascual, G., Cozzuto, L., Roma, G., Wutz, A., Benitah, S. A. & Di Croce, L. Nonoverlapping functions of the Polycomb group Cbx family of proteins in embryonic stem cells. *Cell Stem Cell* **10**, 47-62 (2012).
44. Blackledge, N. P., Farcas, A. M., Kondo, T., King, H. W., McGouran, J. F., Hanssen, L. L., Ito, S., Cooper, S., Kondo, K., Koseki, Y., Ishikura, T., Long, H. K., Sheahan, T. W., Brockdorff, N., Kessler, B. M., Koseki, H. & Klose, R. J. Variant PRC1 complex-dependent H2A ubiquitylation drives PRC2 recruitment and polycomb domain formation. *Cell* **157**, 1445-1459 (2014).
45. Kalb, R., Latwiel, S., Baymaz, H. I., Jansen, P. W., Muller, C. W., Vermeulen, M. & Muller, J. Histone H2A monoubiquitination promotes histone H3 methylation in Polycomb repression. *Nat Struct Mol Biol* **21**, 569-571 (2014).
46. Majewski, I. J., Ritchie, M. E., Phipson, B., Corbin, J., Pakusch, M., Ebert, A., Busslinger, M., Koseki, H., Hu, Y., Smyth, G. K., Alexander, W. S., Hilton, D. J. & Blewitt, M. E. Opposing roles of polycomb repressive complexes in hematopoietic stem and progenitor cells. *Blood* **116**, 731-739 (2010).
47. Wang, R., Taylor, A. B., Leal, B. Z., Chadwell, L. V., Ilangovan, U., Robinson, A. K., Schirf, V., Hart, P. J., Lafer, E. M., Demeler, B., Hinck, A. P., McEwen, D. G. & Kim, C. A. Polycomb group targeting through different binding partners of RING1B C-terminal domain. *Structure* **18**, 966-975 (2010).
48. Bernstein, E., Duncan, E. M., Masui, O., Gil, J., Heard, E. & Allis, C. D. Mouse polycomb proteins bind differentially to methylated histone H3 and RNA and are enriched in facultative heterochromatin. *Mol Cell Biol* **26**, 2560-2569 (2006).
49. Kaustov, L., Ouyang, H., Amaya, M., Lemak, A., Nady, N., Duan, S., Wasney, G. A., Li, Z., Vedadi, M., Schapira, M., Min, J. & Arrowsmith, C. H. Recognition and specificity determinants of the human cbx chromodomains. *J Biol Chem* **286**, 521-529 (2011).
50. Kim, C. A., Gingery, M., Pilpa, R. M. & Bowie, J. U. The SAM domain of polyhomeotic forms a helical polymer. *Nat Struct Biol* **9**, 453-457 (2002).
51. Nanyes, D. R., Junco, S. E., Taylor, A. B., Robinson, A. K., Patterson, N. L., Shivarajpur, A., Halloran, J., Hale, S. M., Kaur, Y., Hart, P. J. & Kim, C. A. Multiple polymer architectures of human polyhomeotic homolog 3 sterile alpha motif. *Proteins* **82**, 2823-2830 (2014).
52. Robinson, A. K., Leal, B. Z., Nanyes, D. R., Kaur, Y., Ilangovan, U., Schirf, V., Hinck, A. P., Demeler, B. & Kim, C. A. Human polyhomeotic homolog 3 (PHC3) sterile alpha

- motif (SAM) linker allows open-ended polymerization of PHC3 SAM. *Biochemistry* **51**, 5379-5386 (2012).
53. Robinson, A. K., Leal, B. Z., Chadwell, L. V., Wang, R., Ilangoan, U., Kaur, Y., Junco, S. E., Schirf, V., Osmulski, P. A., Gaczynska, M., Hinck, A. P., Demeler, B., McEwen, D. G. & Kim, C. A. The growth-suppressive function of the polycomb group protein polyhomeotic is mediated by polymerization of its sterile alpha motif (SAM) domain. *J Biol Chem* **287**, 8702-8713 (2012).
 54. Isono, K., Endo, T. A., Ku, M., Yamada, D., Suzuki, R., Sharif, J., Ishikura, T., Toyoda, T., Bernstein, B. E. & Koseki, H. SAM domain polymerization links subnuclear clustering of PRC1 to gene silencing. *Dev Cell* **26**, 565-577 (2013).
 55. Alkema, M. J., Bronk, M., Verhoeven, E., Otte, A., van 't Veer, L. J., Berns, A. & van Lohuizen, M. Identification of Bmi1-interacting proteins as constituents of a multimeric mammalian polycomb complex. *Genes Dev* **11**, 226-240 (1997).
 56. Gunster, M. J., Satijn, D. P., Hamer, K. M., den Blaauwen, J. L., de Bruijn, D., Alkema, M. J., van Lohuizen, M., van Driel, R. & Otte, A. P. Identification and characterization of interactions between the vertebrate polycomb-group protein BMI1 and human homologs of polyhomeotic. *Mol Cell Biol* **17**, 2326-2335 (1997).
 57. Gil, J. & O'Loughlen, A. PRC1 complex diversity: where is it taking us? *Trends Cell Biol* **24**, 632-641 (2014).
 58. Luis, N. M., Morey, L., Di Croce, L. & Benitah, S. A. Polycomb in stem cells: PRC1 branches out. *Cell Stem Cell* **11**, 16-21 (2012).
 59. O'Loughlen, A., Munoz-Cabello, A. M., Gaspar-Maia, A., Wu, H. A., Banito, A., Kunowska, N., Racek, T., Pemberton, H. N., Beolchi, P., Laval, F., Masui, O., Vermeulen, M., Carroll, T., Graumann, J., Heard, E., Dillon, N., Azuara, V., Snijders, A. P., Peters, G., Bernstein, E. & Gil, J. MicroRNA regulation of Cbx7 mediates a switch of Polycomb orthologs during ESC differentiation. *Cell Stem Cell* **10**, 33-46 (2012).
 60. Junco, S. E., Wang, R., Gaipa, J. C., Taylor, A. B., Schirf, V., Gearhart, M. D., Bardwell, V. J., Demeler, B., Hart, P. J. & Kim, C. A. Structure of the polycomb group protein PCGF1 in complex with BCOR reveals basis for binding selectivity of PCGF homologs. *Structure* **21**, 665-671 (2013).
 61. Gearhart, M. D., Corcoran, C. M., Wamstad, J. A. & Bardwell, V. J. Polycomb group and SCF ubiquitin ligases are found in a novel BCOR complex that is recruited to BCL6 targets. *Mol Cell Biol* **26**, 6880-6889 (2006).
 62. Trojer, P., Cao, A. R., Gao, Z., Li, Y., Zhang, J., Xu, X., Li, G., Losson, R., Erdjument-Bromage, H., Tempst, P., Farnham, P. J. & Reinberg, D. L3MBTL2 protein acts in concert with PcG protein-mediated monoubiquitination of H2A to establish a repressive chromatin structure. *Mol Cell* **42**, 438-450 (2011).
 63. Simon, J., Chiang, A. & Bender, W. Ten different Polycomb group genes are required for spatial control of the *abdA* and *AbdB* homeotic products. *Development* **114**, 493-505 (1992).
 64. Orlando, V. & Paro, R. Chromatin multiprotein complexes involved in the maintenance of transcription patterns. *Curr Opin Genet Dev* **5**, 174-179 (1995).
 65. McKeon, J., Slade, E., Sinclair, D. A., Cheng, N., Couling, M. & Brock, H. W. Mutations in some Polycomb group genes of *Drosophila* interfere with regulation of segmentation genes. *Mol Gen Genet* **244**, 474-483 (1994).

66. Lewis, E. B. A gene complex controlling segmentation in *Drosophila*. *Nature* **276**, 565-570 (1978).
67. Sauvageau, M. & Sauvageau, G. Polycomb group genes: keeping stem cell activity in balance. *PLoS Biol* **6**, e113 (2008).
68. van Lohuizen, M. Functional analysis of mouse Polycomb group genes. *Cell Mol Life Sci* **54**, 71-79 (1998).
69. Voncken, J. W., Roelen, B. A., Roefs, M., de Vries, S., Verhoeven, E., Marino, S., Deschamps, J. & van Lohuizen, M. Rnf2 (Ring1b) deficiency causes gastrulation arrest and cell cycle inhibition. *Proc Natl Acad Sci U S A* **100**, 2468-2473 (2003).
70. del Mar Lorente, M., Marcos-Gutierrez, C., Perez, C., Schoorlemmer, J., Ramirez, A., Magin, T. & Vidal, M. Loss- and gain-of-function mutations show a polycomb group function for Ring1A in mice. *Development* **127**, 5093-5100 (2000).
71. Shirai, M., Osugi, T., Koga, H., Kaji, Y., Takimoto, E., Komuro, I., Hara, J., Miwa, T., Yamauchi-Takahara, K. & Takihara, Y. The Polycomb-group gene *Rae28* sustains *Nkx2.5/Csx* expression and is essential for cardiac morphogenesis. *J Clin Invest* **110**, 177-184 (2002).
72. Ohno, Y., Yasunaga, S., Ohtsubo, M., Mori, S., Tsumura, M., Okada, S., Ohta, T., Ohtani, K., Kobayashi, M. & Takihara, Y. *Hoxb4* transduction down-regulates Geminin protein, providing hematopoietic stem and progenitor cells with proliferation potential. *Proc Natl Acad Sci U S A* **107**, 21529-21534 (2010).
73. Ohta, H., Sawada, A., Kim, J. Y., Tokimasa, S., Nishiguchi, S., Humphries, R. K., Hara, J. & Takihara, Y. Polycomb group gene *rae28* is required for sustaining activity of hematopoietic stem cells. *J Exp Med* **195**, 759-770 (2002).
74. Ma, R. G., Zhang, Y., Sun, T. T. & Cheng, B. Epigenetic regulation by polycomb group complexes: focus on roles of CBX proteins. *J Zhejiang Univ Sci B* **15**, 412-428 (2014).
75. Zencak, D., Lingbeek, M., Kostic, C., Tekaya, M., Tanger, E., Hornfeld, D., Jaquet, M., Munier, F. L., Schorderet, D. F., van Lohuizen, M. & Arsenijevic, Y. *Bmi1* loss produces an increase in astroglial cells and a decrease in neural stem cell population and proliferation. *J Neurosci* **25**, 5774-5783 (2005).
76. Lessard, J., Schumacher, A., Thorsteinsdottir, U., van Lohuizen, M., Magnuson, T. & Sauvageau, G. Functional antagonism of the Polycomb-Group genes *eed* and *Bmi1* in hemopoietic cell proliferation. *Genes Dev* **13**, 2691-2703 (1999).
77. van der Lugt, N. M., Domen, J., Linders, K., van Roon, M., Robanus-Maandag, E., te Riele, H., van der Valk, M., Deschamps, J., Sofroniew, M., van Lohuizen, M. & et al. Posterior transformation, neurological abnormalities, and severe hematopoietic defects in mice with a targeted deletion of the *bmi-1* proto-oncogene. *Genes Dev* **8**, 757-769 (1994).
78. Park, I. K., Qian, D., Kiel, M., Becker, M. W., Pihalja, M., Weissman, I. L., Morrison, S. J. & Clarke, M. F. *Bmi-1* is required for maintenance of adult self-renewing haematopoietic stem cells. *Nature* **423**, 302-305 (2003).
79. Leung, C., Lingbeek, M., Shakhova, O., Liu, J., Tanger, E., Saremaslani, P., Van Lohuizen, M. & Marino, S. *Bmi1* is essential for cerebellar development and is overexpressed in human medulloblastomas. *Nature* **428**, 337-341 (2004).
80. Pietersen, A. M. & van Lohuizen, M. Stem cell regulation by polycomb repressors: postponing commitment. *Curr Opin Cell Biol* **20**, 201-207 (2008).
81. Sparmann, A. & van Lohuizen, M. Polycomb silencers control cell fate, development and cancer. *Nat Rev Cancer* **6**, 846-856 (2006).

82. Rizo, A., Olthof, S., Han, L., Vellenga, E., de Haan, G. & Schuringa, J. J. Repression of BMI1 in normal and leukemic human CD34(+) cells impairs self-renewal and induces apoptosis. *Blood* **114**, 1498-1505 (2009).
83. Lessard, J. & Sauvageau, G. Bmi-1 determines the proliferative capacity of normal and leukaemic stem cells. *Nature* **423**, 255-260 (2003).
84. Iwama, A., Oguro, H., Negishi, M., Kato, Y., Morita, Y., Tsukui, H., Ema, H., Kamijo, T., Katoh-Fukui, Y., Koseki, H., van Lohuizen, M. & Nakauchi, H. Enhanced self-renewal of hematopoietic stem cells mediated by the polycomb gene product Bmi-1. *Immunity* **21**, 843-851 (2004).
85. Molofsky, A. V., Pardal, R., Iwashita, T., Park, I. K., Clarke, M. F. & Morrison, S. J. Bmi-1 dependence distinguishes neural stem cell self-renewal from progenitor proliferation. *Nature* **425**, 962-967 (2003).
86. Lukacs, R. U., Memarzadeh, S., Wu, H. & Witte, O. N. Bmi-1 is a crucial regulator of prostate stem cell self-renewal and malignant transformation. *Cell Stem Cell* **7**, 682-693 (2010).
87. Kalb, R., Mallery, D. L., Larkin, C., Huang, J. T. & Hiom, K. BRCA1 is a histone-H2A-specific ubiquitin ligase. *Cell Rep* **8**, 999-1005 (2014).
88. Bhatnagar, S., Gazin, C., Chamberlain, L., Ou, J., Zhu, X., Tushir, J. S., Virbasius, C. M., Lin, L., Zhu, L. J., Wajapeyee, N. & Green, M. R. TRIM37 is a new histone H2A ubiquitin ligase and breast cancer oncoprotein. *Nature* **516**, 116-120 (2014).
89. Bird, A. DNA methylation patterns and epigenetic memory. *Genes Dev* **16**, 6-21 (2002).
90. Jones, P. A. & Baylin, S. B. The fundamental role of epigenetic events in cancer. *Nat Rev Genet* **3**, 415-428 (2002).
91. Weber, M. & Schubeler, D. Genomic patterns of DNA methylation: targets and function of an epigenetic mark. *Curr Opin Cell Biol* **19**, 273-280 (2007).
92. Boyer, L. A., Plath, K., Zeitlinger, J., Brambrink, T., Medeiros, L. A., Lee, T. I., Levine, S. S., Wernig, M., Tajonar, A., Ray, M. K., Bell, G. W., Otte, A. P., Vidal, M., Gifford, D. K., Young, R. A. & Jaenisch, R. Polycomb complexes repress developmental regulators in murine embryonic stem cells. *Nature* **441**, 349-353 (2006).
93. Schepers, G. E., Teasdale, R. D. & Koopman, P. Twenty pairs of sox: extent, homology, and nomenclature of the mouse and human sox transcription factor gene families. *Dev Cell* **3**, 167-170 (2002).
94. Shah, N. & Sukumar, S. The Hox genes and their roles in oncogenesis. *Nat Rev Cancer* **10**, 361-371 (2010).
95. Lehmann, O. J., Sowden, J. C., Carlsson, P., Jordan, T. & Bhattacharya, S. S. Fox's in development and disease. *Trends Genet* **19**, 339-344 (2003).
96. Haupt, Y., Alexander, W. S., Barri, G., Klinken, S. P. & Adams, J. M. Novel zinc finger gene implicated as myc collaborator by retrovirally accelerated lymphomagenesis in E mu-myc transgenic mice. *Cell* **65**, 753-763 (1991).
97. van Lohuizen, M., Verbeek, S., Scheijen, B., Wientjens, E., van der Gulden, H. & Berns, A. Identification of cooperating oncogenes in E mu-myc transgenic mice by provirus tagging. *Cell* **65**, 737-752 (1991).
98. Haupt, Y., Bath, M. L., Harris, A. W. & Adams, J. M. bmi-1 transgene induces lymphomas and collaborates with myc in tumorigenesis. *Oncogene* **8**, 3161-3164 (1993).
99. Kim, J. H., Yoon, S. Y., Kim, C. N., Joo, J. H., Moon, S. K., Choe, I. S., Choe, Y. K. & Kim, J. W. The Bmi-1 oncoprotein is overexpressed in human colorectal cancer and

- correlates with the reduced p16INK4a/p14ARF proteins. *Cancer Lett* **203**, 217-224 (2004).
100. Li, D. W., Tang, H. M., Fan, J. W., Yan, D. W., Zhou, C. Z., Li, S. X., Wang, X. L. & Peng, Z. H. Expression level of Bmi-1 oncoprotein is associated with progression and prognosis in colon cancer. *J Cancer Res Clin Oncol* **136**, 997-1006 (2010).
 101. Guo, B. H., Feng, Y., Zhang, R., Xu, L. H., Li, M. Z., Kung, H. F., Song, L. B. & Zeng, M. S. Bmi-1 promotes invasion and metastasis, and its elevated expression is correlated with an advanced stage of breast cancer. *Mol Cancer* **10**, 10 (2011).
 102. Honig, A., Weidler, C., Hausler, S., Krockenberger, M., Buchholz, S., Koster, F., Segerer, S. E., Dietl, J. & Engel, J. B. Overexpression of polycomb protein BMI-1 in human specimens of breast, ovarian, endometrial and cervical cancer. *Anticancer Res* **30**, 1559-1564 (2010).
 103. van Leenders, G. J., Dukers, D., Hessels, D., van den Kieboom, S. W., Hulsbergen, C. A., Witjes, J. A., Otte, A. P., Meijer, C. J. & Raaphorst, F. M. Polycomb-group oncogenes EZH2, BMI1, and RING1 are overexpressed in prostate cancer with adverse pathologic and clinical features. *Eur Urol* **52**, 455-463 (2007).
 104. Tabor, M. H., Clay, M. R., Owen, J. H., Bradford, C. R., Carey, T. E., Wolf, G. T. & Prince, M. E. Head and neck cancer stem cells: the side population. *Laryngoscope* **121**, 527-533 (2011).
 105. Sasaki, M., Ikeda, H., Itatsu, K., Yamaguchi, J., Sawada, S., Minato, H., Ohta, T. & Nakanuma, Y. The overexpression of polycomb group proteins Bmi1 and EZH2 is associated with the progression and aggressive biological behavior of hepatocellular carcinoma. *Lab Invest* **88**, 873-882 (2008).
 106. Wang, H., Pan, K., Zhang, H. K., Weng, D. S., Zhou, J., Li, J. J., Huang, W., Song, H. F., Chen, M. S. & Xia, J. C. Increased polycomb-group oncogene Bmi-1 expression correlates with poor prognosis in hepatocellular carcinoma. *J Cancer Res Clin Oncol* **134**, 535-541 (2008).
 107. Zhang, R., Xu, L. B., Zeng, H., Yu, X. H., Wang, J. & Liu, C. Elevated expression of Bmi1 in hepatocellular carcinoma with bile duct tumor thrombi. *Hepatogastroenterology* **60**, 2042-2047 (2013).
 108. Ruan, Z. P., Xu, R., Lv, Y., Tian, T., Wang, W. J., Guo, H. & Nan, K. J. Bmi1 knockdown inhibits hepatocarcinogenesis. *Int J Oncol* **42**, 261-268 (2013).
 109. Vonlanthen, S., Heighway, J., Altermatt, H. J., Gugger, M., Kappeler, A., Borner, M. M., van Lohuizen, M. & Betticher, D. C. The bmi-1 oncoprotein is differentially expressed in non-small cell lung cancer and correlates with INK4A-ARF locus expression. *Br J Cancer* **84**, 1372-1376 (2001).
 110. Vrzalikova, K., Skarda, J., Ehrmann, J., Murray, P. G., Fridman, E., Kopolovic, J., Knizetova, P., Hajduch, M., Klein, J., Kolek, V., Radova, L. & Kolar, Z. Prognostic value of Bmi-1 oncoprotein expression in NSCLC patients: a tissue microarray study. *J Cancer Res Clin Oncol* **134**, 1037-1042 (2008).
 111. Bea, S., Tort, F., Pinyol, M., Puig, X., Hernandez, L., Hernandez, S., Fernandez, P. L., van Lohuizen, M., Colomer, D. & Campo, E. BMI-1 gene amplification and overexpression in hematological malignancies occur mainly in mantle cell lymphomas. *Cancer Res* **61**, 2409-2412 (2001).

112. Saady, N. S., Fawzy, I. M., Azmy, E., Goda, E. F., Eneen, A. & Abdul Salam, E. M. BMI1 gene expression in myeloid leukemias and its impact on prognosis. *Blood Cells Mol Dis* **53**, 194-198 (2014).
113. Chowdhury, M., Mihara, K., Yasunaga, S., Ohtaki, M., Takihara, Y. & Kimura, A. Expression of Polycomb-group (PcG) protein BMI-1 predicts prognosis in patients with acute myeloid leukemia. *Leukemia* **21**, 1116-1122 (2007).
114. Allegra, E., Trapasso, S., Pisani, D. & Puzzo, L. The role of BMI1 as a biomarker of cancer stem cells in head and neck cancer: a review. *Oncology* **86**, 199-205 (2014).
115. van Galen, J. C., Muris, J. J., Oudejans, J. J., Vos, W., Giroth, C. P., Ossenkoppele, G. J., Otte, A. P., Raaphorst, F. M. & Meijer, C. J. Expression of the polycomb-group gene BMI1 is related to an unfavourable prognosis in primary nodal DLBCL. *J Clin Pathol* **60**, 167-172 (2007).
116. Du, J., Li, Y., Li, J. & Zheng, J. Polycomb group protein Bmi1 expression in colon cancers predicts the survival. *Med Oncol* **27**, 1273-1276 (2010).
117. Hayry, V., Makinen, L. K., Atula, T., Sariola, H., Makitie, A., Leivo, I., Keski-Santti, H., Lundin, J., Haglund, C. & Hagstrom, J. Bmi-1 expression predicts prognosis in squamous cell carcinoma of the tongue. *Br J Cancer* **102**, 892-897 (2010).
118. Pietersen, A. M., Horlings, H. M., Hauptmann, M., Langerod, A., Ajouaou, A., Cornelissen-Steijger, P., Wessels, L. F., Jonkers, J., van de Vijver, M. J. & van Lohuizen, M. EZH2 and BMI1 inversely correlate with prognosis and TP53 mutation in breast cancer. *Breast Cancer Res* **10**, R109 (2008).
119. Cenci, T., Martini, M., Montano, N., D'Alessandris, Q. G., Falchetti, M. L., Annibali, D., Savino, M., Bianchi, F., Pierconti, F., Nasi, S., Pallini, R. & Larocca, L. M. Prognostic relevance of c-Myc and BMI1 expression in patients with glioblastoma. *Am J Clin Pathol* **138**, 390-396 (2012).
120. Hanahan, D. & Weinberg, R. A. Hallmarks of cancer: the next generation. *Cell* **144**, 646-674 (2011).
121. Jacobs, J. J., Kieboom, K., Marino, S., DePinho, R. A. & van Lohuizen, M. The oncogene and Polycomb-group gene bmi-1 regulates cell proliferation and senescence through the ink4a locus. *Nature* **397**, 164-168 (1999).
122. Chudnovsky, Y., Khavari, P. A. & Adams, A. E. Melanoma genetics and the development of rational therapeutics. *J Clin Invest* **115**, 813-824 (2005).
123. Kim, W. Y. & Sharpless, N. E. The regulation of INK4/ARF in cancer and aging. *Cell* **127**, 265-275 (2006).
124. Smith, K. S., Chanda, S. K., Lingbeek, M., Ross, D. T., Botstein, D., van Lohuizen, M. & Cleary, M. L. Bmi-1 regulation of INK4A-ARF is a downstream requirement for transformation of hematopoietic progenitors by E2a-Pbx1. *Mol Cell* **12**, 393-400 (2003).
125. Itahana, K., Zou, Y., Itahana, Y., Martinez, J. L., Beausejour, C., Jacobs, J. J., Van Lohuizen, M., Band, V., Campisi, J. & Dimri, G. P. Control of the replicative life span of human fibroblasts by p16 and the polycomb protein Bmi-1. *Mol Cell Biol* **23**, 389-401 (2003).
126. Molofsky, A. V., He, S., Bydon, M., Morrison, S. J. & Pardal, R. Bmi-1 promotes neural stem cell self-renewal and neural development but not mouse growth and survival by repressing the p16Ink4a and p19Arf senescence pathways. *Genes Dev* **19**, 1432-1437 (2005).

127. Liu, L., Andrews, L. G. & Tollefsbol, T. O. Loss of the human polycomb group protein BMI1 promotes cancer-specific cell death. *Oncogene* **25**, 4370-4375 (2006).
128. Bruggeman, S. W., Valk-Lingbeek, M. E., van der Stoop, P. P., Jacobs, J. J., Kieboom, K., Tanger, E., Hulsman, D., Leung, C., Arsenijevic, Y., Marino, S. & van Lohuizen, M. Ink4a and Arf differentially affect cell proliferation and neural stem cell self-renewal in Bmi1-deficient mice. *Genes Dev* **19**, 1438-1443 (2005).
129. Bracken, A. P., Kleine-Kohlbrecher, D., Dietrich, N., Pasini, D., Gargiulo, G., Beekman, C., Theilgaard-Monch, K., Minucci, S., Porse, B. T., Marine, J. C., Hansen, K. H. & Helin, K. The Polycomb group proteins bind throughout the INK4A-ARF locus and are disassociated in senescent cells. *Genes Dev* **21**, 525-530 (2007).
130. Dhawan, S., Tschén, S. I. & Bhushan, A. Bmi-1 regulates the Ink4a/Arf locus to control pancreatic beta-cell proliferation. *Genes Dev* **23**, 906-911 (2009).
131. Maertens, G. N., El Messaoudi-Aubert, S., Racek, T., Stock, J. K., Nicholls, J., Rodriguez-Niedenfuhr, M., Gil, J. & Peters, G. Several distinct polycomb complexes regulate and co-localize on the INK4a tumor suppressor locus. *PLoS One* **4**, e6380 (2009).
132. Agherbi, H., Gaussmann-Wenger, A., Verthuy, C., Chasson, L., Serrano, M. & Djabali, M. Polycomb mediated epigenetic silencing and replication timing at the INK4a/ARF locus during senescence. *PLoS One* **4**, e5622 (2009).
133. Chiba, T., Seki, A., Aoki, R., Ichikawa, H., Negishi, M., Miyagi, S., Oguro, H., Saraya, A., Kamiya, A., Nakauchi, H., Yokosuka, O. & Iwama, A. Bmi1 promotes hepatic stem cell expansion and tumorigenicity in both Ink4a/Arf-dependent and -independent manners in mice. *Hepatology* **52**, 1111-1123 (2010).
134. Bruggeman, S. W., Hulsman, D., Tanger, E., Buckle, T., Blom, M., Zevenhoven, J., van Tellingen, O. & van Lohuizen, M. Bmi1 controls tumor development in an Ink4a/Arf-independent manner in a mouse model for glioma. *Cancer Cell* **12**, 328-341 (2007).
135. Bednar, F., Schofield, H. K., Collins, M. A., Yan, W., Zhang, Y., Shyam, N., Eberle, J., Almada, L. L., Olive, K., Bardeesy, N., Fernandez-Zapico, M. E., Nakada, D., Simeone, D. M., Morrison, S. J. & Pasca di Magliano, M. Bmi1 is required for the initiation of pancreatic cancer through an Ink4a-independent mechanism. *Carcinogenesis* (2015).
136. Xu, C. R., Lee, S., Ho, C., Bommi, P., Huang, S. A., Cheung, S. T., Dimri, G. P. & Chen, X. Bmi1 functions as an oncogene independent of Ink4A/Arf repression in hepatic carcinogenesis. *Mol Cancer Res* **7**, 1937-1945 (2009).
137. Yuan, J., Takeuchi, M., Negishi, M., Oguro, H., Ichikawa, H. & Iwama, A. Bmi1 is essential for leukemic reprogramming of myeloid progenitor cells. *Leukemia* **25**, 1335-1343 (2011).
138. Mourgues, L., Imbert, V., Nebout, M., Colosetti, P., Neffati, Z., Lagadec, P., Verhoeven, E., Peng, C., Duprez, E., Legros, L., Rochet, N., Maguer-Satta, V., Nicolini, F. E., Mary, D. & Peyron, J. F. The BMI1 polycomb protein represses cyclin G2-induced autophagy to support proliferation in chronic myeloid leukemia cells. *Leukemia* (2015).
139. Nacerddine, K., Beaudry, J. B., Ginjala, V., Westerman, B., Mattioli, F., Song, J. Y., van der Poel, H., Ponz, O. B., Pritchard, C., Cornelissen-Steijger, P., Zevenhoven, J., Tanger, E., Sixma, T. K., Ganesan, S. & van Lohuizen, M. Akt-mediated phosphorylation of Bmi1 modulates its oncogenic potential, E3 ligase activity, and DNA damage repair activity in mouse prostate cancer. *J Clin Invest* **122**, 1920-1932 (2012).

140. Douglas, D., Hsu, J. H., Hung, L., Cooper, A., Abdueva, D., van Doorninck, J., Peng, G., Shimada, H., Triche, T. J. & Lawlor, E. R. BMI-1 promotes ewing sarcoma tumorigenicity independent of CDKN2A repression. *Cancer Res* **68**, 6507-6515 (2008).
141. Datta, S., Hoenerhoff, M. J., Bommi, P., Sainger, R., Guo, W. J., Dimri, M., Band, H., Band, V., Green, J. E. & Dimri, G. P. Bmi-1 cooperates with H-Ras to transform human mammary epithelial cells via dysregulation of multiple growth-regulatory pathways. *Cancer Res* **67**, 10286-10295 (2007).
142. Song, L. B., Li, J., Liao, W. T., Feng, Y., Yu, C. P., Hu, L. J., Kong, Q. L., Xu, L. H., Zhang, X., Liu, W. L., Li, M. Z., Zhang, L., Kang, T. B., Fu, L. W., Huang, W. L., Xia, Y. F., Tsao, S. W., Li, M., Band, V., Band, H., Shi, Q. H., Zeng, Y. X. & Zeng, M. S. The polycomb group protein Bmi-1 represses the tumor suppressor PTEN and induces epithelial-mesenchymal transition in human nasopharyngeal epithelial cells. *J Clin Invest* **119**, 3626-3636 (2009).
143. Yang, M. H., Hsu, D. S., Wang, H. W., Wang, H. J., Lan, H. Y., Yang, W. H., Huang, C. H., Kao, S. Y., Tzeng, C. H., Tai, S. K., Chang, S. Y., Lee, O. K. & Wu, K. J. Bmi1 is essential in Twist1-induced epithelial-mesenchymal transition. *Nat Cell Biol* **12**, 982-992 (2010).
144. Wei, X. L., Dou, X. W., Bai, J. W., Luo, X. R., Qiu, S. Q., Xi, D. D., Huang, W. H., Du, C. W., Man, K. & Zhang, G. J. ERalpha inhibits epithelial-mesenchymal transition by suppressing Bmi1 in breast cancer. *Oncotarget* (2015).
145. Wu, K. J. & Yang, M. H. Epithelial-mesenchymal transition and cancer stemness: the Twist1-Bmi1 connection. *Biosci Rep* **31**, 449-455 (2011).
146. Martin, A. & Cano, A. Tumorigenesis: Twist1 links EMT to self-renewal. *Nat Cell Biol* **12**, 924-925 (2010).
147. Wellner, U., Schubert, J., Burk, U. C., Schmalhofer, O., Zhu, F., Sonntag, A., Waldvogel, B., Vannier, C., Darling, D., zur Hausen, A., Brunton, V. G., Morton, J., Sansom, O., Schuler, J., Stemmler, M. P., Herzberger, C., Hopt, U., Keck, T., Brabletz, S. & Brabletz, T. The EMT-activator ZEB1 promotes tumorigenicity by repressing stemness-inhibiting microRNAs. *Nat Cell Biol* **11**, 1487-1495 (2009).
148. Tam, W. L. & Weinberg, R. A. The epigenetics of epithelial-mesenchymal plasticity in cancer. *Nat Med* **19**, 1438-1449 (2013).
149. Abercrombie, M. Contact inhibition and malignancy. *Nature* **281**, 259-262 (1979).
150. Silletti, S., Yao, J. P., Pienta, K. J. & Raz, A. Loss of cell-contact regulation and altered responses to autocrine motility factor correlate with increased malignancy in prostate cancer cells. *Int J Cancer* **63**, 100-105 (1995).
151. Abercrombie, M. & Ambrose, E. J. Interference microscope studies of cell contacts in tissue culture. *Exp Cell Res* **15**, 332-345 (1958).
152. Abercrombie, M. & Heaysman, J. E. Observations on the social behaviour of cells in tissue culture. II. Monolayering of fibroblasts. *Exp Cell Res* **6**, 293-306 (1954).
153. Moroishi, T., Hansen, C. G. & Guan, K. L. The emerging roles of YAP and TAZ in cancer. *Nat Rev Cancer* **15**, 73-79 (2015).
154. Hsu, J. H. & Lawlor, E. R. BMI-1 suppresses contact inhibition and stabilizes YAP in Ewing sarcoma. *Oncogene* **30**, 2077-2085 (2011).
155. Gieni, R. S., Ismail, I. H., Campbell, S. & Hendzel, M. J. Polycomb group proteins in the DNA damage response: a link between radiation resistance and "stemness". *Cell Cycle* **10**, 883-894 (2011).

156. Liu, J., Cao, L., Chen, J., Song, S., Lee, I. H., Quijano, C., Liu, H., Keyvanfar, K., Chen, H., Cao, L. Y., Ahn, B. H., Kumar, N. G., Rovira, II, Xu, X. L., van Lohuizen, M., Motoyama, N., Deng, C. X. & Finkel, T. Bmi1 regulates mitochondrial function and the DNA damage response pathway. *Nature* **459**, 387-392 (2009).
157. Pan, Y., Duncombe, T. A., Kellenberger, C. A., Hammond, M. C. & Herr, A. E. High-throughput electrophoretic mobility shift assays for quantitative analysis of molecular binding reactions. *Anal Chem* **86**, 10357-10364 (2014).
158. Chagraoui, J., Hebert, J., Girard, S. & Sauvageau, G. An anticlastogenic function for the Polycomb Group gene Bmi1. *Proc Natl Acad Sci U S A* **108**, 5284-5289 (2011).
159. Ginjala, V., Nacerddine, K., Kulkarni, A., Oza, J., Hill, S. J., Yao, M., Citterio, E., van Lohuizen, M. & Ganesan, S. BMI1 is recruited to DNA breaks and contributes to DNA damage-induced H2A ubiquitination and repair. *Mol Cell Biol* **31**, 1972-1982 (2011).
160. Ismail, I. H., Andrin, C., McDonald, D. & Hendzel, M. J. BMI1-mediated histone ubiquitylation promotes DNA double-strand break repair. *J Cell Biol* **191**, 45-60 (2010).
161. Pan, M. R., Peng, G., Hung, W. C. & Lin, S. Y. Monoubiquitination of H2AX protein regulates DNA damage response signaling. *J Biol Chem* **286**, 28599-28607 (2011).
162. Wang, E., Bhattacharyya, S., Szabolcs, A., Rodriguez-Aguayo, C., Jennings, N. B., Lopez-Berestein, G., Mukherjee, P., Sood, A. K. & Bhattacharya, R. Enhancing chemotherapy response with Bmi-1 silencing in ovarian cancer. *PLoS One* **6**, e17918 (2011).
163. Nakamura, S., Oshima, M., Yuan, J., Saraya, A., Miyagi, S., Konuma, T., Yamazaki, S., Osawa, M., Nakauchi, H., Koseki, H. & Iwama, A. Bmi1 confers resistance to oxidative stress on hematopoietic stem cells. *PLoS One* **7**, e36209 (2012).
164. Chato, W., Abdouh, M., David, J., Champagne, M. P., Ferreira, J., Rodier, F. & Bernier, G. The polycomb group gene Bmi1 regulates antioxidant defenses in neurons by repressing p53 pro-oxidant activity. *J Neurosci* **29**, 529-542 (2009).
165. Beck, B. & Blanpain, C. Unravelling cancer stem cell potential. *Nat Rev Cancer* **13**, 727-738 (2013).
166. Meacham, C. E. & Morrison, S. J. Tumour heterogeneity and cancer cell plasticity. *Nature* **501**, 328-337 (2013).
167. Lobo, N. A., Shimono, Y., Qian, D. & Clarke, M. F. The biology of cancer stem cells. *Annu Rev Cell Dev Biol* **23**, 675-699 (2007).
168. Mani, S. A., Guo, W., Liao, M. J., Eaton, E. N., Ayyanan, A., Zhou, A. Y., Brooks, M., Reinhard, F., Zhang, C. C., Shipitsin, M., Campbell, L. L., Polyak, K., Brisken, C., Yang, J. & Weinberg, R. A. The epithelial-mesenchymal transition generates cells with properties of stem cells. *Cell* **133**, 704-715 (2008).
169. Singh, A. & Settleman, J. EMT, cancer stem cells and drug resistance: an emerging axis of evil in the war on cancer. *Oncogene* **29**, 4741-4751 (2010).
170. Mitra, A., Mishra, L. & Li, S. EMT, CTCs and CSCs in tumor relapse and drug-resistance. *Oncotarget* **6**, 10697-10711 (2015).
171. Wicha, M. S. Targeting self-renewal, an Achilles' heel of cancer stem cells. *Nat Med* **20**, 14-15 (2014).
172. Cao, L., Bombard, J., Cintron, K., Sheedy, J., Weetall, M. L. & Davis, T. W. BMI1 as a novel target for drug discovery in cancer. *J Cell Biochem* **112**, 2729-2741 (2011).
173. Lessard, J. & Sauvageau, G. Polycomb group genes as epigenetic regulators of normal and leukemic hemopoiesis. *Exp Hematol* **31**, 567-585 (2003).

174. Natsume, A., Kinjo, S., Yuki, K., Kato, T., Ohno, M., Motomura, K., Iwami, K. & Wakabayashi, T. Glioma-initiating cells and molecular pathology: implications for therapy. *Brain Tumor Pathol* **28**, 1-12 (2011).
175. Abdouh, M., Facchino, S., Chatoo, W., Balasingam, V., Ferreira, J. & Bernier, G. BMI1 sustains human glioblastoma multiforme stem cell renewal. *J Neurosci* **29**, 8884-8896 (2009).
176. Chiba, T., Miyagi, S., Saraya, A., Aoki, R., Seki, A., Morita, Y., Yonemitsu, Y., Yokosuka, O., Taniguchi, H., Nakauchi, H. & Iwama, A. The polycomb gene product BMI1 contributes to the maintenance of tumor-initiating side population cells in hepatocellular carcinoma. *Cancer Res* **68**, 7742-7749 (2008).
177. Kreso, A., van Galen, P., Pedley, N. M., Lima-Fernandes, E., Frelin, C., Davis, T., Cao, L., Baiazitov, R., Du, W., Sydorenko, N., Moon, Y. C., Gibson, L., Wang, Y., Leung, C., Iscove, N. N., Arrowsmith, C. H., Szentgyorgyi, E., Gallinger, S., Dick, J. E. & O'Brien, C. A. Self-renewal as a therapeutic target in human colorectal cancer. *Nat Med* **20**, 29-36 (2014).
178. Takebe, N., Harris, P. J., Warren, R. Q. & Ivy, S. P. Targeting cancer stem cells by inhibiting Wnt, Notch, and Hedgehog pathways. *Nat Rev Clin Oncol* **8**, 97-106 (2011).
179. Vermeulen, L., De Sousa, E. M. F., van der Heijden, M., Cameron, K., de Jong, J. H., Borovski, T., Tuynman, J. B., Todaro, M., Merz, C., Rodermond, H., Sprick, M. R., Kemper, K., Richel, D. J., Stassi, G. & Medema, J. P. Wnt activity defines colon cancer stem cells and is regulated by the microenvironment. *Nat Cell Biol* **12**, 468-476 (2010).
180. Abramovich, C. & Humphries, R. K. Hox regulation of normal and leukemic hematopoietic stem cells. *Curr Opin Hematol* **12**, 210-216 (2005).
181. Argiropoulos, B. & Humphries, R. K. Hox genes in hematopoiesis and leukemogenesis. *Oncogene* **26**, 6766-6776 (2007).
182. Xiao, J. & Deng, C. Knockdown of Bmi-1 impairs growth and invasiveness of human gastric carcinoma cells. *Oncol Res* **17**, 613-620 (2009).
183. Qi, S., Li, B., Yang, T., Liu, Y., Cao, S., He, X., Zhang, P., Li, L. & Xu, C. Validation of Bmi1 as a therapeutic target of hepatocellular carcinoma in mice. *Int J Mol Sci* **15**, 20004-20021 (2014).
184. Shimono, Y., Zabala, M., Cho, R. W., Lobo, N., Dalerba, P., Qian, D., Diehn, M., Liu, H., Panula, S. P., Chiao, E., Dirbas, F. M., Somlo, G., Pera, R. A., Lao, K. & Clarke, M. F. Downregulation of miRNA-200c links breast cancer stem cells with normal stem cells. *Cell* **138**, 592-603 (2009).
185. Xu, Z., Liu, H., Lv, X., Liu, Y., Li, S. & Li, H. Knockdown of the Bmi-1 oncogene inhibits cell proliferation and induces cell apoptosis and is involved in the decrease of Akt phosphorylation in the human breast carcinoma cell line MCF-7. *Oncol Rep* **25**, 409-418 (2011).
186. Su, W. J., Fang, J. S., Cheng, F., Liu, C., Zhou, F. & Zhang, J. RNF2/Ring1b negatively regulates p53 expression in selective cancer cell types to promote tumor development. *Proc Natl Acad Sci U S A* **110**, 1720-1725 (2013).
187. Boukarabila, H., Saurin, A. J., Batsche, E., Mossadegh, N., van Lohuizen, M., Otte, A. P., Pradel, J., Muchardt, C., Sieweke, M. & Duprez, E. The PRC1 Polycomb group complex interacts with PLZF/RARA to mediate leukemic transformation. *Genes Dev* **23**, 1195-1206 (2009).

188. Jagani, Z., Wiederschain, D., Loo, A., He, D., Mosher, R., Fordjour, P., Monahan, J., Morrissey, M., Yao, Y. M., Lengauer, C., Warmuth, M., Sellers, W. R. & Dorsch, M. The Polycomb group protein Bmi-1 is essential for the growth of multiple myeloma cells. *Cancer Res* **70**, 5528-5538 (2010).
189. Smith, L. L., Yeung, J., Zeisig, B. B., Popov, N., Huijbers, I., Barnes, J., Wilson, A. J., Taskesen, E., Delwel, R., Gil, J., Van Lohuizen, M. & So, C. W. Functional crosstalk between Bmi1 and MLL/Hoxa9 axis in establishment of normal hematopoietic and leukemic stem cells. *Cell Stem Cell* **8**, 649-662 (2011).
190. Costello, R. T., Mallet, F., Gaugler, B., Sainy, D., Arnoulet, C., Gastaut, J. A. & Olive, D. Human acute myeloid leukemia CD34+/CD38- progenitor cells have decreased sensitivity to chemotherapy and Fas-induced apoptosis, reduced immunogenicity, and impaired dendritic cell transformation capacities. *Cancer Res* **60**, 4403-4411 (2000).
191. Bonnet, D. & Dick, J. E. Human acute myeloid leukemia is organized as a hierarchy that originates from a primitive hematopoietic cell. *Nat Med* **3**, 730-737 (1997).
192. Lapidot, T., Sirard, C., Vormoor, J., Murdoch, B., Hoang, T., Caceres-Cortes, J., Minden, M., Paterson, B., Caligiuri, M. A. & Dick, J. E. A cell initiating human acute myeloid leukaemia after transplantation into SCID mice. *Nature* **367**, 645-648 (1994).
193. Alchanati, I., Teicher, C., Cohen, G., Shemesh, V., Barr, H. M., Nakache, P., Ben-Avraham, D., Idelevich, A., Angel, I., Livnah, N., Tuvia, S., Reiss, Y., Taglicht, D. & Erez, O. The E3 ubiquitin-ligase Bmi1/Ring1A controls the proteasomal degradation of Top2alpha cleavage complex - a potentially new drug target. *PLoS One* **4**, e8104 (2009).
194. Ismail, I. H., McDonald, D., Strickfaden, H., Xu, Z. & Hendzel, M. J. A small molecule inhibitor of polycomb repressive complex 1 inhibits ubiquitin signaling at DNA double-strand breaks. *J Biol Chem* **288**, 26944-26954 (2013).
195. Rual, J. F., Venkatesan, K., Hao, T., Hirozane-Kishikawa, T., Dricot, A., Li, N., Berriz, G. F., Gibbons, F. D., Dreze, M., Ayivi-Guedehoussou, N., Klitgord, N., Simon, C., Boxem, M., Milstein, S., Rosenberg, J., Goldberg, D. S., Zhang, L. V., Wong, S. L., Franklin, G., Li, S., Albala, J. S., Lim, J., Fraughton, C., Llamas, E., Cevik, S., Bex, C., Lamesch, P., Sikorski, R. S., Vandenhaute, J., Zoghbi, H. Y., Smolyar, A., Bosak, S., Sequerra, R., Doucette-Stamm, L., Cusick, M. E., Hill, D. E., Roth, F. P. & Vidal, M. Towards a proteome-scale map of the human protein-protein interaction network. *Nature* **437**, 1173-1178 (2005).
196. Cierpicki, T. & Grembecka, J. Targeting protein-protein interactions in hematologic malignancies: still a challenge or a great opportunity for future therapies? *Immunol Rev* **263**, 279-301 (2015).
197. Makley, L. N. & Gestwicki, J. E. Expanding the number of 'druggable' targets: non-enzymes and protein-protein interactions. *Chem Biol Drug Des* **81**, 22-32 (2013).
198. Fry, D. C. Small-molecule inhibitors of protein-protein interactions: how to mimic a protein partner. *Curr Pharm Des* **18**, 4679-4684 (2012).
199. Thompson, A. D., Dugan, A., Gestwicki, J. E. & Mapp, A. K. Fine-tuning multiprotein complexes using small molecules. *ACS Chem Biol* **7**, 1311-1320 (2012).
200. Smith, M. C. & Gestwicki, J. E. Features of protein-protein interactions that translate into potent inhibitors: topology, surface area and affinity. *Expert Rev Mol Med* **14**, e16 (2012).
201. Arkin, M. R. & Wells, J. A. Small-molecule inhibitors of protein-protein interactions: progressing towards the dream. *Nat Rev Drug Discov* **3**, 301-317 (2004).

202. Mullard, A. Protein-protein interaction inhibitors get into the groove. *Nat Rev Drug Discov* **11**, 173-175 (2012).
203. Arkin, M. R., Tang, Y. & Wells, J. A. Small-molecule inhibitors of protein-protein interactions: progressing toward the reality. *Chem Biol* **21**, 1102-1114 (2014).
204. Wells, J. A. & McClendon, C. L. Reaching for high-hanging fruit in drug discovery at protein-protein interfaces. *Nature* **450**, 1001-1009 (2007).
205. Overington, J. P., Al-Lazikani, B. & Hopkins, A. L. How many drug targets are there? *Nat Rev Drug Discov* **5**, 993-996 (2006).
206. Imming, P., Sinning, C. & Meyer, A. Drugs, their targets and the nature and number of drug targets. *Nat Rev Drug Discov* **5**, 821-834 (2006).
207. Venkatesan, K., Rual, J. F., Vazquez, A., Stelzl, U., Lemmens, I., Hirozane-Kishikawa, T., Hao, T., Zenkner, M., Xin, X., Goh, K. I., Yildirim, M. A., Simonis, N., Heinzmann, K., Gebreab, F., Sahalie, J. M., Cevik, S., Simon, C., de Smet, A. S., Dann, E., Smolyar, A., Vinayagam, A., Yu, H., Szeto, D., Borick, H., Dricot, A., Klitgord, N., Murray, R. R., Lin, C., Lalowski, M., Timm, J., Rau, K., Boone, C., Braun, P., Cusick, M. E., Roth, F. P., Hill, D. E., Tavernier, J., Wanker, E. E., Barabasi, A. L. & Vidal, M. An empirical framework for binary interactome mapping. *Nat Methods* **6**, 83-90 (2009).
208. Stumpf, M. P., Thorne, T., de Silva, E., Stewart, R., An, H. J., Lappe, M. & Wiuf, C. Estimating the size of the human interactome. *Proc Natl Acad Sci U S A* **105**, 6959-6964 (2008).
209. Hwang, H., Vreven, T., Janin, J. & Weng, Z. Protein-protein docking benchmark version 4.0. *Proteins* **78**, 3111-3114 (2010).
210. Lo Conte, L., Chothia, C. & Janin, J. The atomic structure of protein-protein recognition sites. *J Mol Biol* **285**, 2177-2198 (1999).
211. Jones, S. & Thornton, J. M. Principles of protein-protein interactions. *Proc Natl Acad Sci U S A* **93**, 13-20 (1996).
212. DeLano, W. L. Unraveling hot spots in binding interfaces: progress and challenges. *Curr Opin Struct Biol* **12**, 14-20 (2002).
213. Bogan, A. A. & Thorn, K. S. Anatomy of hot spots in protein interfaces. *J Mol Biol* **280**, 1-9 (1998).
214. Ma, B., Elkayam, T., Wolfson, H. & Nussinov, R. Protein-protein interactions: structurally conserved residues distinguish between binding sites and exposed protein surfaces. *Proc Natl Acad Sci U S A* **100**, 5772-5777 (2003).
215. Goh, C. S., Milburn, D. & Gerstein, M. Conformational changes associated with protein-protein interactions. *Curr Opin Struct Biol* **14**, 104-109 (2004).
216. Colas, P. High-throughput screening assays to discover small-molecule inhibitors of protein interactions. *Curr Drug Discov Technol* **5**, 190-199 (2008).
217. Cochran, A. G. Antagonists of protein-protein interactions. *Chem Biol* **7**, R85-94 (2000).
218. Chang, L., Miyata, Y., Ung, P. M., Bertelsen, E. B., McQuade, T. J., Carlson, H. A., Zuiderweg, E. R. & Gestwicki, J. E. Chemical screens against a reconstituted multiprotein complex: myricetin blocks DnaJ regulation of DnaK through an allosteric mechanism. *Chem Biol* **18**, 210-221 (2011).
219. Higuero, A. P., Schreyer, A., Bickerton, G. R., Pitt, W. R., Groom, C. R. & Blundell, T. L. Atomic interactions and profile of small molecules disrupting protein-protein interfaces: the TIMBAL database. *Chem Biol Drug Des* **74**, 457-467 (2009).

220. Zhang, X., Betzi, S., Morelli, X. & Roche, P. Focused chemical libraries--design and enrichment: an example of protein-protein interaction chemical space. *Future Med Chem* **6**, 1291-1307 (2014).
221. Murray, C. W. & Rees, D. C. The rise of fragment-based drug discovery. *Nat Chem* **1**, 187-192 (2009).
222. Valkov, E., Sharpe, T., Marsh, M., Greive, S. & Hyvonen, M. Targeting protein-protein interactions and fragment-based drug discovery. *Top Curr Chem* **317**, 145-179 (2012).
223. Coyne, A. G., Scott, D. E. & Abell, C. Drugging challenging targets using fragment-based approaches. *Curr Opin Chem Biol* **14**, 299-307 (2010).
224. Wilson, C. G. & Arkin, M. R. Probing structural adaptivity at PPI interfaces with small molecules. *Drug Discov Today Technol* **10**, e501-508 (2013).
225. Fuller, J. C., Burgoyne, N. J. & Jackson, R. M. Predicting druggable binding sites at the protein-protein interface. *Drug Discov Today* **14**, 155-161 (2009).
226. Silvian, L., Enyedy, I. & Kumaravel, G. Inhibitors of protein-protein interactions: new methodologies to tackle this challenge. *Drug Discov Today Technol* **10**, e509-515 (2013).
227. Arrowsmith, C. H., Bountra, C., Fish, P. V., Lee, K. & Schapira, M. Epigenetic protein families: a new frontier for drug discovery. *Nat Rev Drug Discov* **11**, 384-400 (2012).
228. Finley, A. & Copeland, R. A. Small molecule control of chromatin remodeling. *Chem Biol* **21**, 1196-1210 (2014).
229. Helin, K. & Dhanak, D. Chromatin proteins and modifications as drug targets. *Nature* **502**, 480-488 (2013).
230. Filippakopoulos, P., Qi, J., Picaud, S., Shen, Y., Smith, W. B., Fedorov, O., Morse, E. M., Keates, T., Hickman, T. T., Felletar, I., Philpott, M., Munro, S., McKeown, M. R., Wang, Y., Christie, A. L., West, N., Cameron, M. J., Schwartz, B., Heightman, T. D., La Thangue, N., French, C. A., Wiest, O., Kung, A. L., Knapp, S. & Bradner, J. E. Selective inhibition of BET bromodomains. *Nature* **468**, 1067-1073 (2010).
231. Nicodeme, E., Jeffrey, K. L., Schaefer, U., Beinke, S., Dewell, S., Chung, C. W., Chandwani, R., Marazzi, I., Wilson, P., Coste, H., White, J., Kirilovsky, J., Rice, C. M., Lora, J. M., Prinjha, R. K., Lee, K. & Tarakhovskiy, A. Suppression of inflammation by a synthetic histone mimic. *Nature* **468**, 1119-1123 (2010).
232. Mirguet, O., Gosmini, R., Toum, J., Clement, C. A., Barnathan, M., Brusq, J. M., Mordaunt, J. E., Grimes, R. M., Crowe, M., Pineau, O., Ajakane, M., Daugan, A., Jeffrey, P., Cutler, L., Haynes, A. C., Smithers, N. N., Chung, C. W., Bamborough, P., Uings, I. J., Lewis, A., Witherington, J., Parr, N., Prinjha, R. K. & Nicodeme, E. Discovery of epigenetic regulator I-BET762: lead optimization to afford a clinical candidate inhibitor of the BET bromodomains. *J Med Chem* **56**, 7501-7515 (2013).
233. Herold, J. M., Wigle, T. J., Norris, J. L., Lam, R., Korboukh, V. K., Gao, C., Ingerman, L. A., Kireev, D. B., Senisterra, G., Vedadi, M., Tripathy, A., Brown, P. J., Arrowsmith, C. H., Jin, J., Janzen, W. P. & Frye, S. V. Small-molecule ligands of methyl-lysine binding proteins. *J Med Chem* **54**, 2504-2511 (2011).
234. James, L. I., Barsyte-Lovejoy, D., Zhong, N., Krichevsky, L., Korboukh, V. K., Herold, J. M., MacNevin, C. J., Norris, J. L., Sagum, C. A., Tempel, W., Marcon, E., Guo, H., Gao, C., Huang, X. P., Duan, S., Emili, A., Greenblatt, J. F., Kireev, D. B., Jin, J., Janzen, W. P., Brown, P. J., Bedford, M. T., Arrowsmith, C. H. & Frye, S. V. Discovery of a chemical probe for the L3MBTL3 methyllysine reader domain. *Nat Chem Biol* **9**, 184-191 (2013).

235. Fierz, B. & Muir, T. W. Chromatin as an expansive canvas for chemical biology. *Nat Chem Biol* **8**, 417-427 (2012).
236. Kong, X., Chen, L., Jiao, L., Jiang, X., Lian, F., Lu, J., Zhu, K., Du, D., Liu, J., Ding, H., Zhang, N., Shen, J., Zheng, M., Chen, K., Liu, X., Jiang, H. & Luo, C. Astemizole arrests the proliferation of cancer cells by disrupting the EZH2-EED interaction of polycomb repressive complex 2. *J Med Chem* **57**, 9512-9521 (2014).
237. Chen, H., Gao, S., Li, J., Liu, D., Sheng, C., Yao, C., Jiang, W., Wu, J., Chen, S. & Huang, W. Wedelolactone disrupts the interaction of EZH2-EED complex and inhibits PRC2-dependent cancer. *Oncotarget* (2015).
238. Borkin, D., He, S., Miao, H., Kempinska, K., Pollock, J., Chase, J., Purohit, T., Malik, B., Zhao, T., Wang, J., Wen, B., Zong, H., Jones, M., Danet-Desnoyers, G., Guzman, M. L., Talpaz, M., Bixby, D. L., Sun, D., Hess, J. L., Muntean, A. G., Maillard, I., Cierpicki, T. & Grembecka, J. Pharmacologic inhibition of the Menin-MLL interaction blocks progression of MLL leukemia in vivo. *Cancer Cell* **27**, 589-602 (2015).
239. Grembecka, J., He, S., Shi, A., Purohit, T., Muntean, A. G., Sorenson, R. J., Showalter, H. D., Murai, M. J., Belcher, A. M., Hartley, T., Hess, J. L. & Cierpicki, T. Menin-MLL inhibitors reverse oncogenic activity of MLL fusion proteins in leukemia. *Nat Chem Biol* **8**, 277-284 (2012).
240. He, S., Senter, T. J., Pollock, J., Han, C., Upadhyay, S. K., Purohit, T., Gogliotti, R. D., Lindsley, C. W., Cierpicki, T., Stauffer, S. R. & Grembecka, J. High-affinity small-molecule inhibitors of the menin-mixed lineage leukemia (MLL) interaction closely mimic a natural protein-protein interaction. *J Med Chem* **57**, 1543-1556 (2014).
241. McGinty, R. K., Henrici, R. C. & Tan, S. Crystal structure of the PRC1 ubiquitylation module bound to the nucleosome. *Nature* **514**, 591-596 (2014).
242. Makde, R. D., England, J. R., Yennawar, H. P. & Tan, S. Structure of RCC1 chromatin factor bound to the nucleosome core particle. *Nature* **467**, 562-566 (2010).

Chapter 2: Structural Characterization of BMI1 Protein-Protein Interaction Domain

A. Abstract

BMI1 is a subunit of the canonical polycomb repressive 1 (PRC1) complex which covalently modifies histones to repress transcription. To date the structural and mechanistic role of BMI1 in the PRC1 complex remains incompletely characterized. In this study we characterized the BMI1 interaction with PHC2 using biochemical and biophysical methods. Furthermore, we determined the high resolution structure of BMI1 in complex with its polycomb binding partner PHC2 using a hybrid approach involving X-ray crystallography and solution NMR. Structural details reveal that the protein-protein interaction domain of BMI1 adapts an ubiquitin-like fold. PHC2 adopts a β -hairpin conformation when in complex with BMI1 to form an intermolecular β -sheet. This BMI1 domain binds a short fragment of PHC2 with nanomolar affinity. Mutations at the BMI1-PHC2 interface disrupt this interaction *in vitro* and in cells leading to reduced U2OS cell proliferation. Additionally, we found that BMI1 can weakly self-associate *in vitro* and disruption of these oligomerization interfaces through mutagenesis impairs cellular proliferation. Together this data suggest a model of BMI1 as a central architectural protein of the PRC1 complex mediating multiple protein-protein interactions that contribute to regulation of cellular proliferation.

B. Background

BMI1 belongs to the polycomb group (PcG) family of proteins, an evolutionally conserved family of negative transcriptional regulators. In mammals PcG complexes target

hundreds of genes associated with cell fate and development.¹⁻³ As a component of polycomb repressive complex 1 BMI1 interacts directly with the Ring1B protein through a heterodimer involving N-terminal RING domains. Additionally, BMI1 has a C-terminal domain which previous studies suggested was critical for direct interactions with the polyhomeotic (PHC) proteins in the PRC1 complex.⁴⁻⁶ PHC proteins have two conserved domain, a C-terminal SAM domain that can form helical polymers⁷⁻⁹ and a homology domain 1 (HD1) which is suspected of mediating protein-protein interactions.^{4,10}

Other work demonstrated that the BMI1 C-terminus is critical for BMI1 function, suggesting an important role for these protein-protein interactions. The van Lohuizen group generated transgenic BMI1 domain deletion mice and observed that mice lacking the BMI1 C-terminus had skeletal deformations suggesting that the protein-protein interaction domain (“second domain”) is important for *Hox* gene regulation in mouse development.⁴ This domain was shown to be important for transcriptional silencing in *in vitro* luciferase assays.¹¹ This domain is critical in BMI1-mediated oncogenesis as deletion of the second domain of BMI1 reduces spontaneous lymphomagenesis in a mouse model for c-Myc mediated lymphoma.¹² Further, data suggest that this domain is critical for cell cycle regulation and a tumorigenic transcriptional program as deletion of the domain impairs human fibroblast and mammary epithelial cell immortalization.^{13,14} These studies demonstrate a critical role for this domain in BMI1 function, yet a complete structural characterization of this domain is missing. We were motivated to understand the molecular details of interactions mediated by the second domain to gain insight into the role of BMI1 in normal and cancer biology. Additionally, structural and biochemical characterization can support drug discovery efforts targeting activity of BMI1 in cancer.

C. Results

C.1. The second domain of BMI1 interacts directly with PHC2

Previous studies demonstrated that the C-terminus of BMI1 is required for binding PHC proteins.^{4,5} To confirm this we developed a cellular pull-down assay using the Avi-tag system. In this system coexpression of the BirA biotin ligase and Avi-tagged BMI1 results in biotinylation

of the Avi tag for pull-down with streptavidin beads (Figure 2.1.A).¹⁵ We used a construct of BMI1 lacking the N-terminal RING domain (Avi-BMI1 Δ Ring) allowing for detection of protein-protein interactions mediated by the C-terminus of BMI1, at the exclusion of non-direct interactions mediated by the RING domain. In this assay, myc-tagged PHC2 is pulled-down by wild type Avi-BMI1 Δ Ring from HEK293T cells (Figure 2.1.B). This demonstrates that the interaction between these proteins is mediated by the C-terminal fragment of BMI1 but not through the N-terminal RING domain. The constructs of BMI1 and PHC2 proteins used in this experiment are illustrated in Figure 2.1.C.

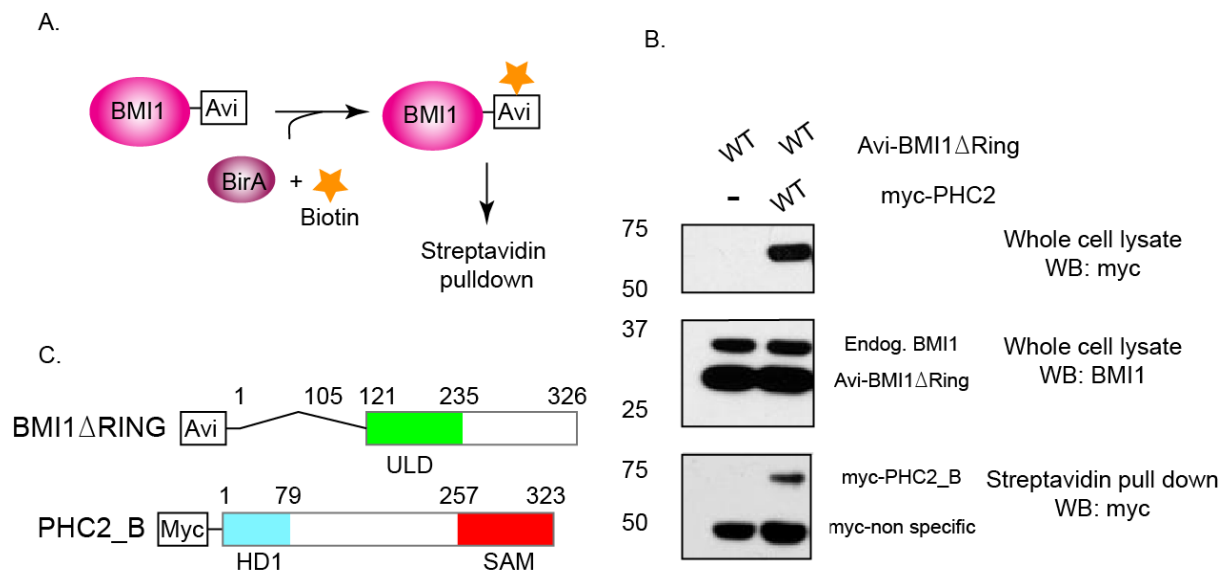


Figure 2.1. BMI1 C-terminus interacts directly with PHC2_B.

A. Schematic of streptavidin pull-down of Avi-tagged BMI1 from cell lysate from HEK293T cells co-expressing Avi-BMI1 Δ Ring and BirA. B. Pull-down of myc-PHC2_B from HEK293T cells using biotinylated-BMI1 Δ Ring. Western blots are probed as indicated and molecular weight marker is indicated on the left. C. Outline of constructs used in pull-down experiments; primary sequence numbers are provided for clarity.

C.2. Optimization of BMI1 second domain construct for structural studies

We next pursued *in vitro* characterization of the BMI1 second domain for detailed structural and functional characterization. To define the BMI1 protein-protein interaction domain and obtain soluble, folded protein we tested a construct of BMI1 encompassing residues 106-240. We measured ¹H-¹⁵N HSQC NMR spectra and found that this protein was folded in

solution. However, significant peak broadening for resonances with dispersed downfield and upfield chemical shifts and a large number of intense, sharp peaks in the random coil chemical shift region of the NMR spectrum indicated that this construct aggregated in solution, likely due to the presence of significant number of disordered residues (Figure 2.2.A). To improve this construct by removing such flexible regions we employed NMR spectroscopy. As described in Appendix A we developed a ^{13}C -detected NMR approach to detect disordered regions in globular proteins.¹⁶ 2D CACO spectrum for BMI1₁₀₆₋₂₄₀ showed ~34 peaks, and through sequential assignment process with CANCO and CBCACO experiments we identified BMI1 N- and C- terminal residues 106-120 and 236-240 as being highly flexible in solution (Figure 2.2.B). Deletion of these residues significantly improved the ^1H - ^{15}N HSQC spectrum (Figure 2.2.C), demonstrating that this is a well folded domain that behaves well in solution. Based on these studies we defined residues 121-235 as the BMI1 second domain and this construct is used for all subsequent studies.

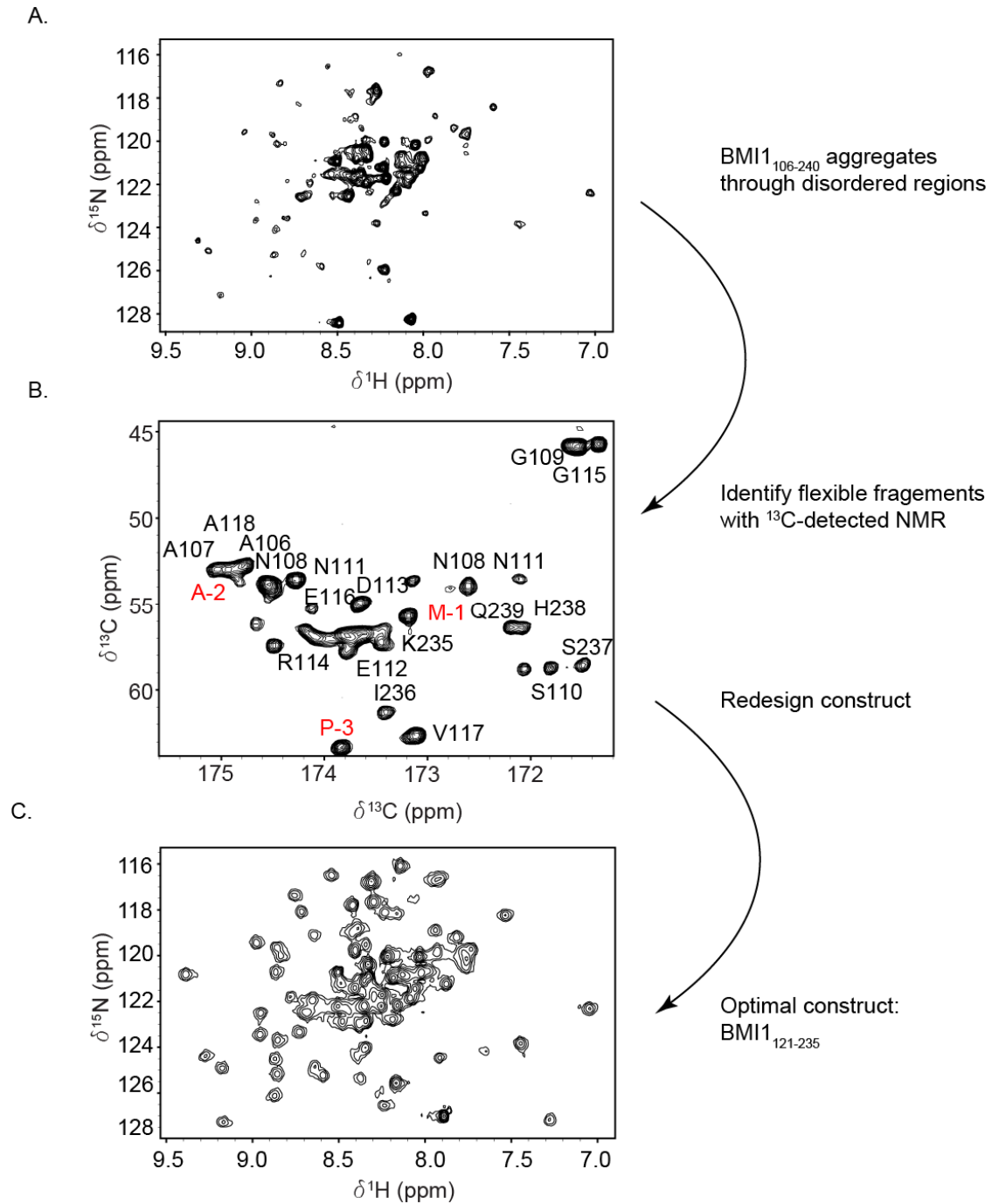


Figure 2.2. Optimization of BMI1 second domain construct using NMR

A. ^1H - ^{15}N HSQC spectrum of BMI1₁₀₆₋₂₄₀. B. ^{13}C -detected CBCACO spectrum for BMI1₁₀₆₋₂₄₀ with assignment for flexible regions. Residues remaining from tag cleavage during purification are colored in red. C. ^1H - ^{15}N HSQC spectrum of BMI1₁₂₁₋₂₃₅.

C.3. Characterization of BMI1-binding domain in PHC2

We next wanted to define the region of PHC2 that interacts directly with BMI1. It has been previously suggested that the N-terminal HD1 domain of PHC2 is involved in the BMI1 interaction but this interaction was not characterized before using *in vitro* methods.⁵ In order to confirm that the HD1 mediates the interaction with BMI1 we expressed a fragment of PHC2 including residues 1-79. To measure binding affinity between BMI1₁₂₁₋₂₃₅ and PHC2₁₋₇₉ we used isothermal titration calorimetry (ITC) which demonstrated mid-nanomolar affinity ($K_D = 398$ nM) with 1:1 stoichiometric ratio for this interaction (Figure 2.3).

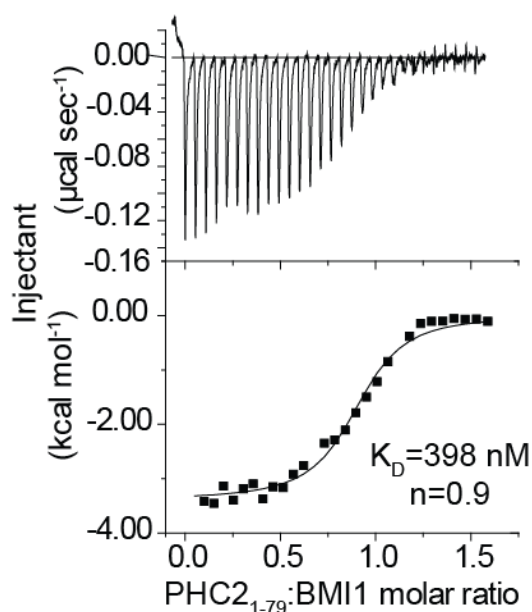


Figure 2.3. ITC determination of BMI1-PHC2 binding stoichiometry and affinity. Isothermal titration calorimetry experiment of BMI1₁₂₁₋₂₃₅ titrated with unlabeled PHC2₁₋₇₉.

To further characterize this interaction we turned to NMR spectroscopy. Poor peak dispersion on the ¹H-¹⁵N HSQC spectrum of PHC2₁₋₇₉ indicate that this protein is unstructured in solution (Figure 2.4.A). This heavy spectral overlap precluded backbone assignment using ¹H-detected triple resonance experiments, and thus to obtain a detailed characterization of this interaction we again applied carbon-detected NMR methodology. Using 2D CACO, CBCACO, CANCO experiments we assigned 90% of the unstructured PHC2₁₋₇₉. To identify the residues involved in BMI1 binding we recorded the same experiments for PHC2₁₋₇₉ saturated with BMI1.

Binding of BMI1 to PHC2₁₋₇₉ should result in significant peak broadening for involved residues due to the formation of a high molecular weight complex.¹⁷ Indeed, addition of BMI1 to PHC2₁₋₇₉ led to complete broadening of resonances on NMR spectra corresponding to residues 33-58 (Figure 2.4.B). These data confirm a structural transition from disordered to ordered for this motif and indicates that this region represents BMI1-binding motif.

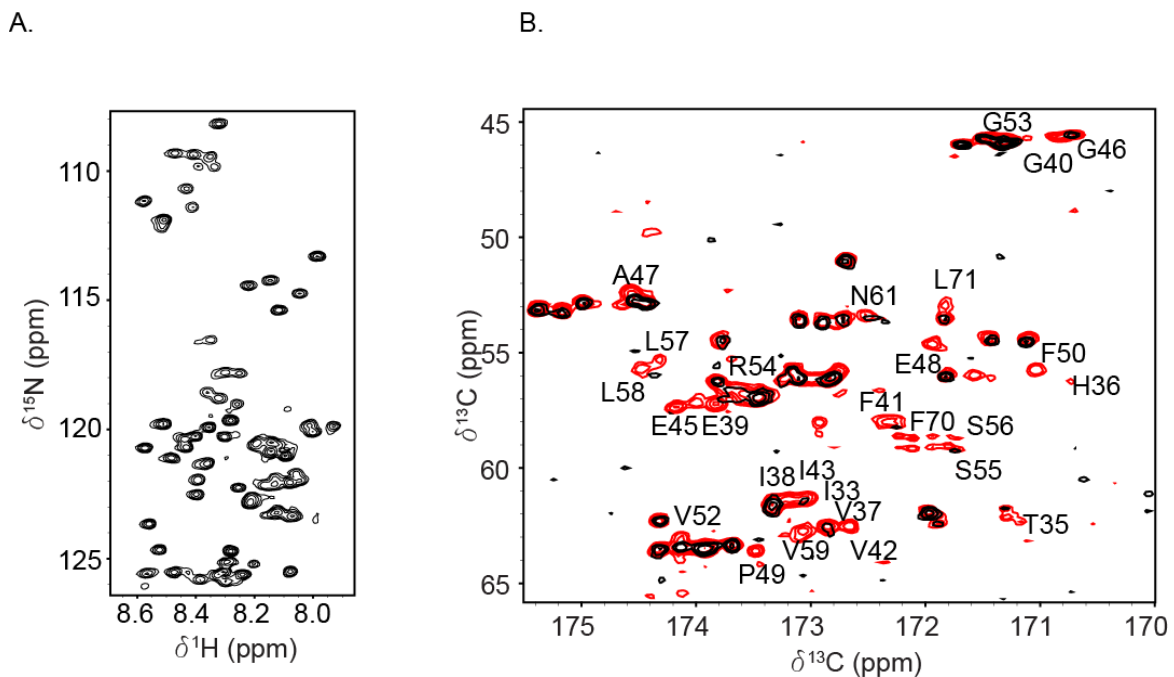


Figure 2.4. Mapping of BMI1-binding motif of PHC2 using ¹³C-detected NMR.

A. Selection of ¹H-¹⁵N HSQC spectrum for PHC2₁₋₇₉ showing poor dispersion of backbone amide resonances. Region of spectrum with sidechains is omitted for clarity. B. BMI1-binding motif mapping through carbon detected experiments. ¹³C-detected CACO spectrum for apo PHC2₁₋₇₉ (red) is overlaid with spectrum for PHC2₁₋₇₉ with equimolar unlabeled BMI1₁₂₁₋₂₃₅ (black). Residues broadened in the presence of BMI1₁₂₁₋₂₃₅ are labeled.

To test the role of this motif in mediating the direct BMI1-PHC2 interaction in cells we performed pull-down experiments from HEK293T cells overexpressing a construct of PHC2 lacking this fragment (PHC2Δ30-51) (Figure 2.5.A). The lack of observed interaction between BMI1 and PHC2Δ30-51 confirmed that in cells this section of PHC2 forms the critical interactions with BMI1. This region of PHC2 represents 52% of the conserved residues in the

HD1 domain (Figure 2.5.B). Isothermal titration calorimetry experiments with a synthetic PHC2₃₃₋₅₆ peptide showed similar affinity as the full length domain (PHC2₃₃₋₅₆ K_D = 413 nM) (Figure 2.5.C) suggesting that there are likely very few contacts outside of this motif.

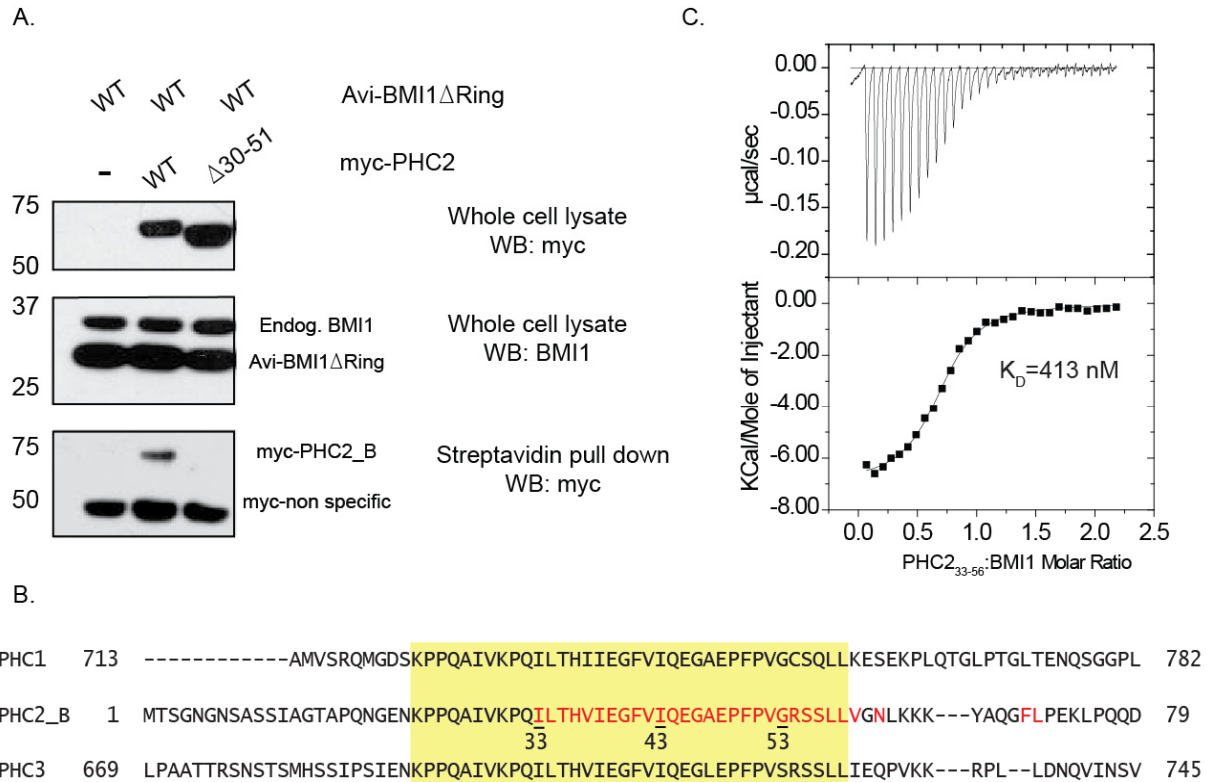


Figure 2.5. PHC2 interacts with BMI1 in cells through a conserved fragment.

A. Streptavidin pull-down of myc-tagged PHC2_B constructs from HEK293T cells transfected with Avi-BMI1 Δ Ring and BirA. Western blots are probed as indicated, molecular weight marker is labeled on the left side. B. Sequence alignment of the three human PHC proteins. Conserved residues are highlighted in yellow and residues of PHC2_B that are broadened in ¹³C-detected experiments upon BMI1 addition are colored in red. Numbering for PHC2_B sequence is provided for clarity. C. Characterization of the BMI1 – PHC2₃₃₋₅₆ interaction through isothermal titration calorimetry with BMI1 titrated with PHC2₃₃₋₅₆.

C.4. Structure determination of BMI1-PHC2 complex

Following mapping of the minimal interaction interfaces between BMI1 and PHC2 we pursued structural studies to determine the molecular basis of the complex. We have performed crystallization screens using both apo BMI1 and BMI1 in complex with different fragments of

PHC2. We obtained crystals of BMI1 co-crystallized in the presence of the PHC2₃₃₋₅₆ fragment in 100 mM MES, pH 6.5, 50 mM MgCl₂, 7% isopropanol, 6% PEG 4000 which diffracted to 2.5 Å. In order to determine the structure of the BMI1 second domain we purified seleno-methionine (Se-Met) labeled protein, however we were not able to reproduce the crystals obtained for the native protein. As an alternative approach, we performed bioinformatics analysis and found sequence similarity to the BMI1 ortholog, PCGF1 which has 29% identity to BMI1 second domain. Despite relatively low sequence similarity, we used a model derived from the PCGF1-BCOR structure¹⁸ and determined the crystal structure of the second domain of BMI1 using molecular replacement (Figure 2.6.A; Table 2.1- see Experimental Methods). As expected from sequence analysis, the structure of the BMI1 second domain has an ubiquitin-like fold and we named it the ubiquitin-like domain (ULD). There is one molecule per asymmetric unit and while we could refine the structure of BMI1 ULD, we were not able to model the PHC2 fragment into the remaining electron density which was found between two symmetry-related molecules of BMI1 ULD. The unmodeled electron density is found in a wide opening between the β 1 and α 1 elements, which has been previously found to be involved in protein-protein interactions for polycomb ubiquitin-like domains (Figure 2.6.B).¹⁸⁻²⁰

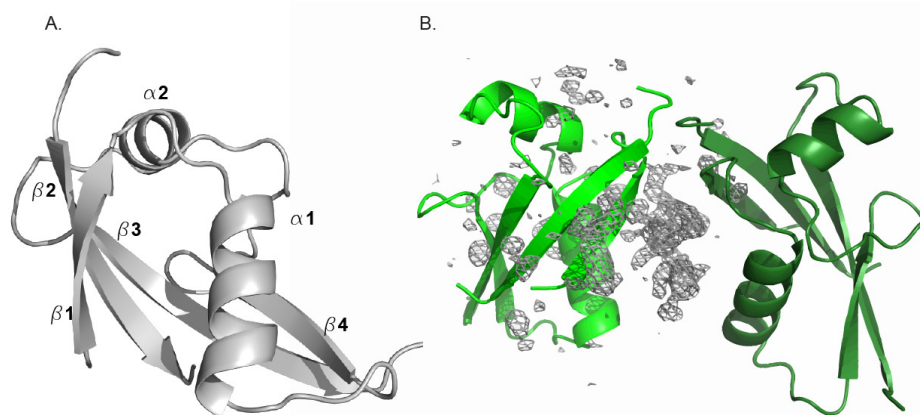


Figure 2.6. Crystal structure of BMI1 ULD.

A. Structure of BMI1 ULD as determined by X-ray crystallography from crystals obtained for BMI1 with PHC2₃₃₋₅₆ fragment. Secondary structure elements are labeled. B. 2Fo-Fc electron density map of PHC2 density (gray) with cartoon images of BMI1 ULD shows electron density found at crystallographic interface between two symmetry related BMI1 molecules (green).

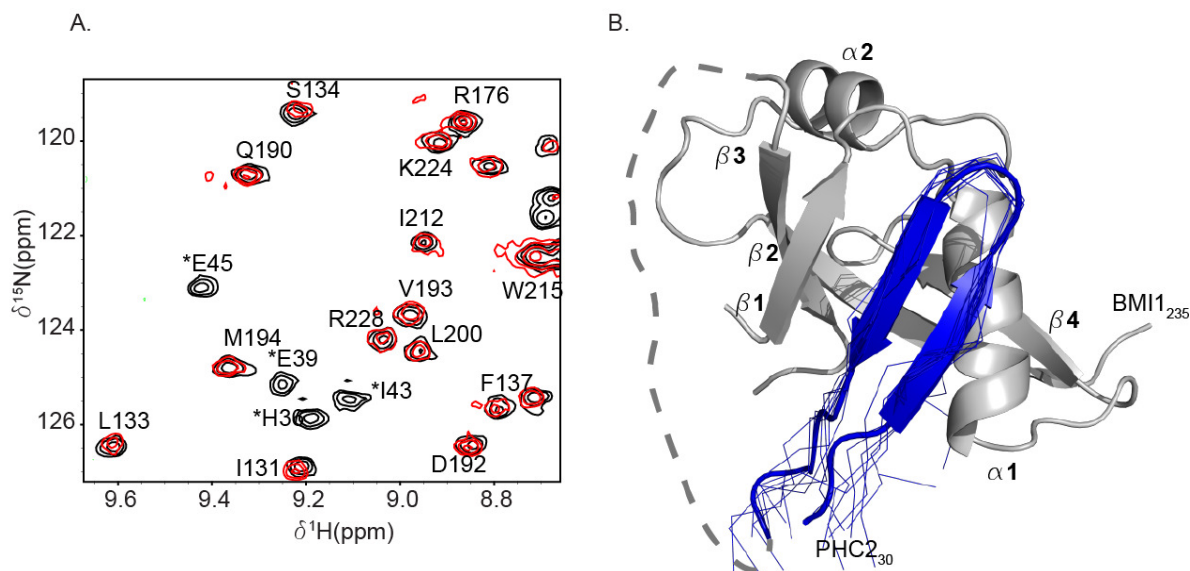


Figure 2.7. Hybrid solution NMR- x-ray crystal structure of BMI1-PhC2 complex.
 A. Selection of ^1H - ^{15}N HSQC spectra for BMI1 ULD saturated with PhC2₃₂₋₆₁-fragment (red) and for PhC2₃₀₋₆₃-BMI1 fusion protein (black). Resonance assignments are shown and resonances coming from the PhC2 portion of BMI1 fusion are indicated with asterisk. B. Hybrid X-ray crystal and solution NMR structure of the BMI1-PhC2 complex. The 10 lowest energy solution structures for PhC2 are shown in blue and the crystal structure of BMI1 is shown in gray.

While the structure of BMI1 ULD determined by X-ray crystallography provides insight into BMI1 structure when in complex with PhC2, the structure lacks the molecular details of the interaction. Therefore, to complete the structure of the BMI1-PhC2 complex, we turned to solution NMR. Because of limited stability of the BMI1-PhC2 complex we designed a fusion protein with PhC2 residues 30-63 fused to the N-terminal of BMI1. This construct was designed to include the intact BMI1 interacting motif in PhC2 as well as a short linker to insure proper folding of the complex. The ^1H - ^{15}N HSQC spectrum for this fusion protein is similar to that of ^{15}N BMI1 saturated with unlabeled PhC2 (Figure 2.7.A) demonstrating that the fusion protein accurately recapitulates the protein complex. We performed chemical shift assignment for the BMI1-PhC2 fusion protein based on triple resonance NMR spectra (Table 2.2- see Experimental Methods). We found that all BMI1 constructs with or without PhC2 fragments had a tendency to aggregate in solution. This resulted in the relatively poor quality of NMR spectra and prevented

structural determination of the BMI1-PHC2 complex solely based on NMR experiments. However, analysis of 3D ^{13}C -edited and 3D ^{15}N -edited NOESY spectra for the BMI1-PHC2 fusion protein allowed assignment of a large number of intra-PHC2 and inter-PHC2-BMI1 NOEs.

To circumvent the challenges in determining the structure of the complex solely using either X-ray crystallography or solution NMR methods we used a hybrid refinement method. In this approach we used the crystal structure of BMI1 and NMR restraints to define PHC2 structure and determine PHC2-BMI1 contacts. The initial structure was calculated in CYANA.²¹ Subsequently, we refined the structure using Rosetta²²⁻²⁴ employing coordinates for BMI1 derived from the crystal structure and NMR restraints (Table 2.2). The structure of the complex shows an antiparallel intermolecular β -sheet formed between the β -hairpin formed by PHC2 residues 33-47 and the β 1 strand of BMI1 (Figure 2.7.B). The hydrophobic core of the complex is made up of BMI1 Leu164, Leu175 and Phe178 packing with PHC2 Ala47, Ile43 and Ile38

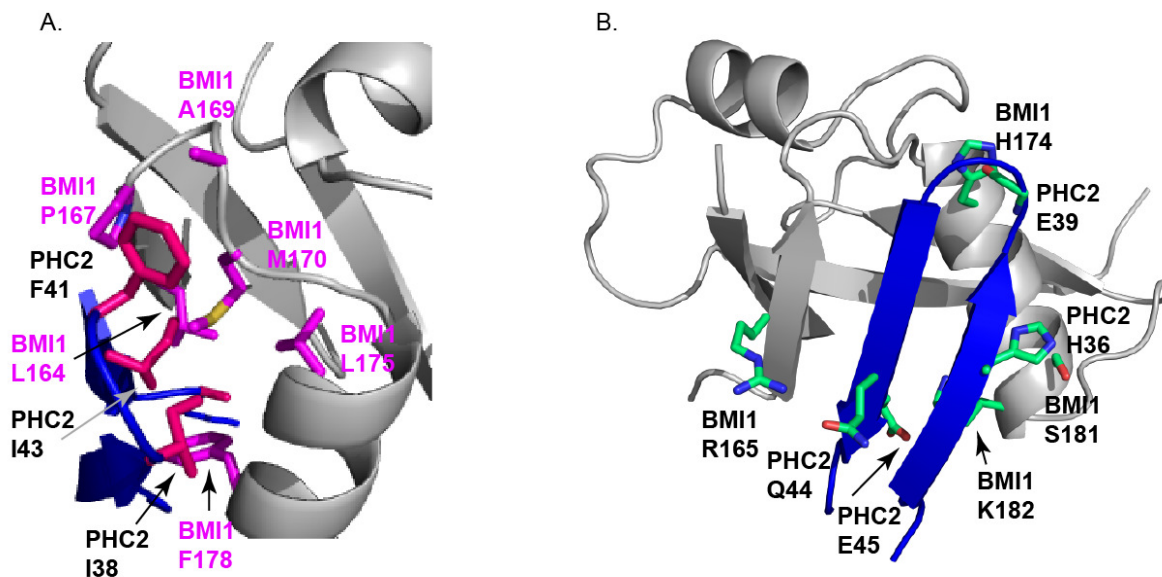


Figure 2.8. Structural details of the BMI1-PHC2 interaction.

A. Close-up view of the hydrophobic core of the BMI1-PHC2 interaction. Hydrophobic residues from BMI1 and PHC2 are shown in pink. Structure is rotated 90° relative to part B. B. Highlighted polar interactions in the BMI1-PHC2 interface. Key polar residues involved in the interaction are show in green.

(Figure 2.8.A). PHC2 Phe41 is sandwiched between BMI1 Pro167, Ala169 and Met170 at the top of the hairpin. One side of the PHC2 β -hairpin packs against the α 1 helix of BMI1 ULD with the sidechain of BMI1 His174 hydrogen bonding with the backbone carbonyl of PHC2 Glu39 in the β turn and PHC2 His36 forming polar contacts with BMI1 Ser181 (Figure 2.8.B). Polar contacts between the sidechains of BMI1 Arg165 and PHC2 Gln44 contribute to the affinity on the other side of the interface. Interestingly, we observed that PHC2 Glu45 is buried in the structure, extending towards BMI1 Lys182 which is also buried in the interface. Residues 47-63 of PHC2 are disordered in the structure, as supported by random coil chemical shifts and very few NOEs.

Re-examination of the crystal structure reveals that unmodeled electron density corresponds to PHC2 in the structure of a complex (Figure 2.9). PHC2 binds at the crystallographic interface between two BMI1 molecules. We concluded that in this crystal structure, PHC2 has to be bound in two different orientations with 50% occupancy for each PHC2 molecule. We cannot model the complex solely based on crystallographic data likely due to limited resolution (2.5 Å) and only partial occupancy of the PHC2 molecules.

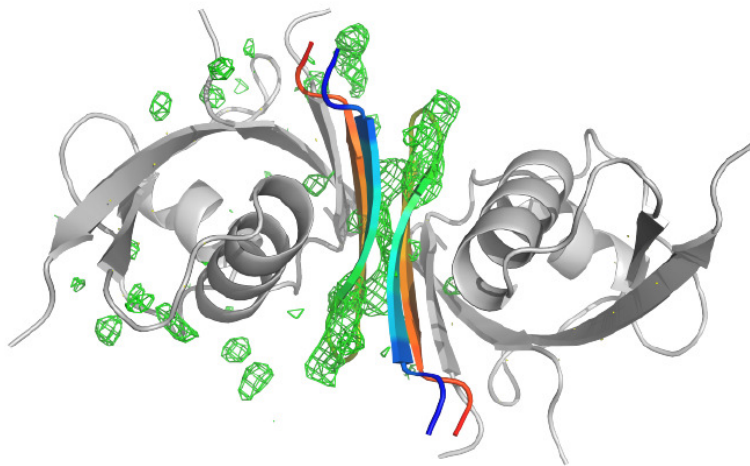


Figure 2.9. PHC2 electron density is at crystallographic interface in crystal structure. Overlay of the hybrid BMI1-PHC2 structure with the 2Fo-Fc electron density map from the BMI1 PHC2₃₃₋₅₆ crystal structure showing density identified for PHC2 peptide. The solution structure of PHC2 is colored from blue at the N-terminus to red at the C-terminus.

C.5. Design of mutations in BMI1 to disrupt BMI1-PHC2 interaction

We next wanted to probe the BMI1-PHC2 interaction through mutagenesis. Based on the structure of the complex we rationalized that mutation of Arg165 and His174 in BMI1 to glutamic acid would introduce significant electrostatic repulsion with PHC2 Glu45 and Glu39, respectively (Figure 2.8.A). We expressed these two BMI1 mutants (R165E and H174E) and verified that the proteins retained their secondary fold by NMR spectroscopy (data not shown). We tested these mutants in a fluorescence polarization assay with fluorescein-tagged PHC2₃₂₋₆₁ (FITC-PHC2) (Figure 2.10.A). In this assay wild type BMI1 binds PHC2₃₂₋₆₁ with $K_D = 0.215 \pm 0.016 \mu\text{M}$. The R165E mutation reduces the binding affinity by 30 fold ($K_D = 5.9 \pm 0.9 \mu\text{M}$) and H174E mutation results in a 100-fold loss in binding affinity ($K_D = 20.13 \pm 2.8 \mu\text{M}$). Introduction of the double R165E/H174E mutation into BMI1 completely abolished the interaction ($K_D > 50 \mu\text{M}$). In pull-down experiments from HEK293T cells, Avi-BMI1 Δ Ring R165E/H174E fails to pull-down myc-PHC2 demonstrating that disruption of this interaction through these point mutants is sufficient to disrupt the complex in cells (Figure 2.10.B).

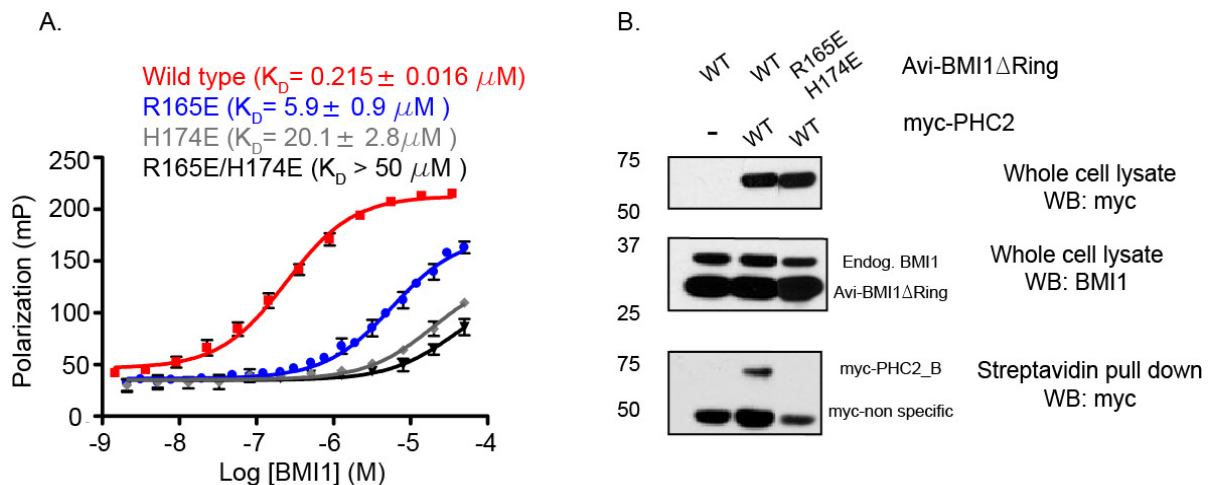


Figure 2.10. BMI1 mutations disrupt PHC2 interaction *in vitro* and in cells.

A. Fluorescence polarization experiments titrating FITC-PHC2 with wild type or mutant BMI1 ULD constructs. Experiments were performed in duplicates and error bars represent the standard deviation from three independent experiments; K_D is reported as average and standard deviation from three independent experiments. B. BMI1 ULD R165E/H174E cannot pull-down myc-PHC2_B from HEK293T cells. Western blots are probed as indicated and the molecular weight marker is shown on left.

C.6. BMI1 ULD can form higher order oligomers

Throughout the characterization of the BMI1-PHC2 interaction we made several observations that suggested that BMI1 may self-associate in solution. We observed that many resonances in the NMR spectra of BMI1 constructs showed various degrees of concentration-dependent peak broadening consistent with protein self-association in solution (Figure 2.11.A). Importantly, BMI1 self-association was not affected by the interaction with PHC2, as demonstrated by performing NMR binding experiments between BMI1 ULD and the PHC2-BMI1 fusion protein. This suggests that oligomerization may represent an intrinsic property of BMI1 ULD. To characterize BMI1 ULD self-association we employed analytical ultracentrifugation (AUC) experiments.²⁵ AUC is a biophysical method that measures the rate of particle sedimentation in solution in a gravitational field.²⁶ This allows determination of sedimentation coefficient which contains information on both size and shape of macromolecules in the experiment and is sensitive to heterogeneous distributions of species. Thus, from these experiments it is possible to assess if a protein or protein complex from multiple species in solution.

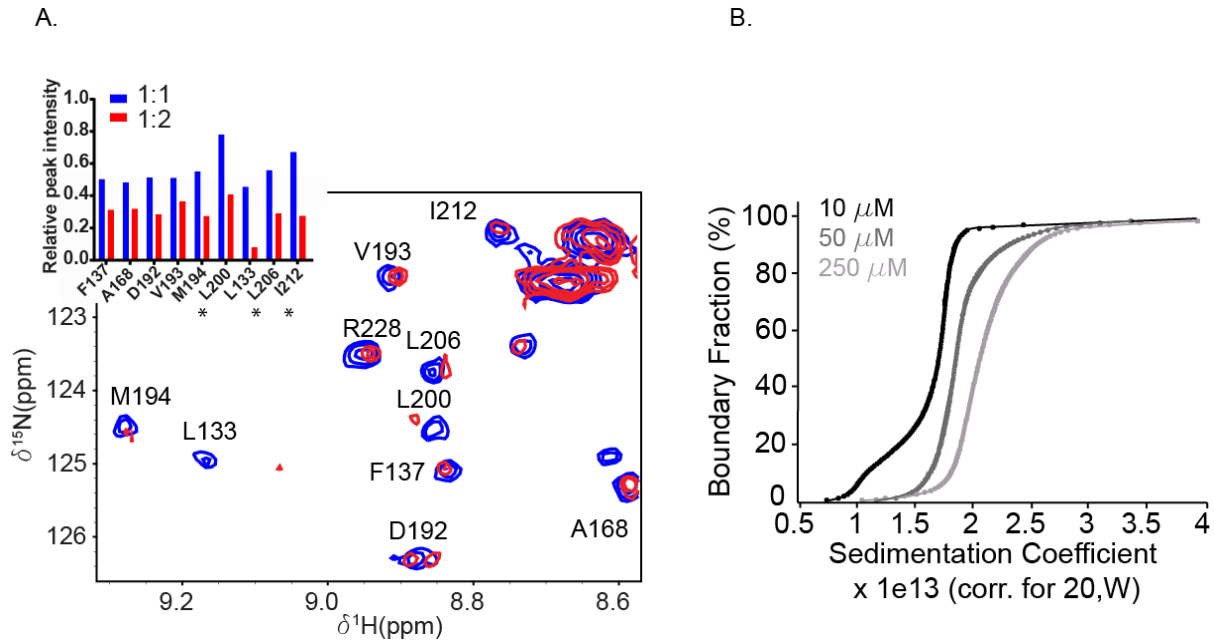


Figure 2.11. BMI1 self-associates in solution.

A. Portion of ^1H - ^{15}N HSQC spectrum for BMI1 ULN in the presence of 1 equivalent (blue) or 2 equivalents (red) of PHC2₃₀₋₇₉-BMI1 fusion protein. Inset shows relative peak intensity of select residues demonstrating broadening of signals with increased protein concentration from the formation of larger molecular species in solution. Residues at the second interface identified through X-ray crystallography are indicated. B. Van Holde-Weischet [G(s)] plots of sedimentation distributions for BMI1 ULN wild type-PHC2 complex at various concentrations shows concentration-dependent increase in particle size in solution

We first wanted to test solution sedimentation behavior of wild type BMI1 ULN. Because of limited solubility of BMI1 ULN, we measured sedimentation coefficients for the BMI1-PHC2 complex which has better solubility. AUC sedimentation velocity experiments were conducted at three concentrations of the complex and a concentration-dependent increase in sedimentation coefficient was determined with 10 μM , 50 μM and 250 μM samples having coefficients of 1.5, 1.8 and 2.0, respectively. This indicates a weak, yet significant formation of larger complexes at higher concentrations (Figure 2.11.B).

Inspection of the crystal packing in the structure of BMI1 with PHC2 shows two potential head-to-head homo-dimerization interfaces between BMI1 molecules (Figure 2.12.A). The first interface is predominantly hydrophobic and comprises residues Asp184-Phe189 and Tyr225-Thr230. The second interface is made up of polar interactions involving a hydrogen bonding

network between the sidechains of Tyr202, Tyr 203, Tyr 211 and Tyr 213 and backbone carbonyl groups. Ile212 is buried between two monomers at this interface. We observed that residues Leu200, Leu206 and Ile212 at this interface are particularly broadened at high concentrations of protein in solution NMR experiments (Figure 2.11.A). Examination of the interactions between symmetry-related BMI1 molecules in the crystal lattice reveal that through these head-to-head interactions BMI1 forms a hexamer in the crystal form (Figure 2.12.B).

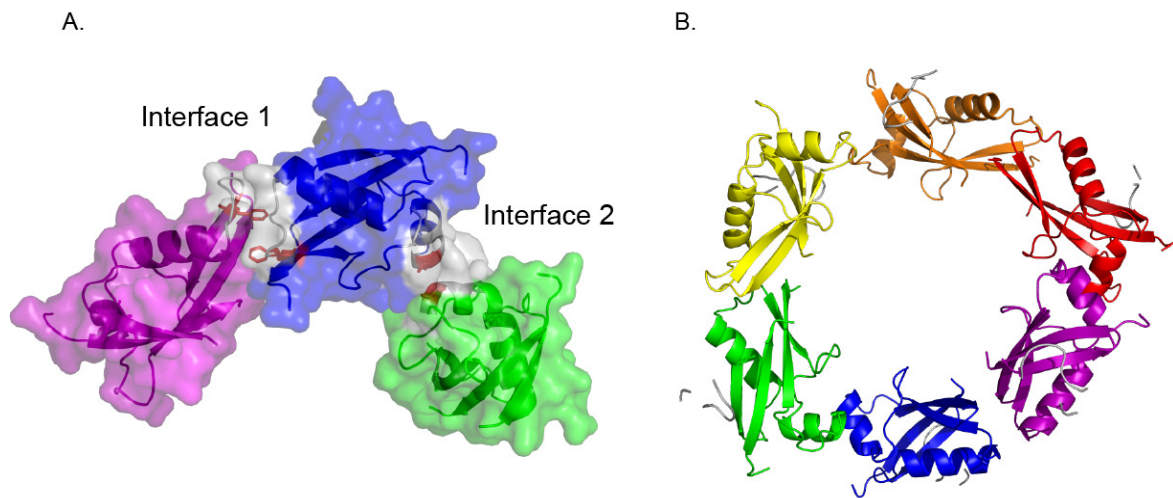


Figure 2.12. Crystal structure of BMI1 suggests two homodimer interfaces.

A. Crystal structure of BMI1 ULD shows two head-to-head interfaces colored in gray. Residues mutated in functional studies are colored red. B. Hexameric structure of BMI1 ULD formed through crystal contacts.

To test if interface mutants would disrupt BMI1 ULD oligomerization we designed point mutations F189Q and I212E to introduce polar groups in place of hydrophobic residues at these interfaces. Fluorescence polarization data indicate that these mutations do not very significantly affect binding of PHC2 *in vitro* (F189Q $K_D = 0.409 \pm 0.16 \mu\text{M}$ and I212E $K_D = 1.2 \pm 0.25 \mu\text{M}$) (Figure 2.13.A). Furthermore, NMR experiments show that both mutant proteins are folded in solution and bind PHC2 in a similar manner to wild type protein (data not shown). To assess whether these point mutations affect the propensity of BMI1 to self-associate we employed AUC (Figure 2.13.B). Sedimentation velocity experiments showed that the BMI1 F189Q –PHC2 complex has similar concentration dependent sedimentation behavior as wild type BMI1-PHC2

complex indicating that this mutant is still able to form higher order species. Interestingly, the I212E interface 2 mutant does not show concentration dependent effects, and sedimentation coefficients of 1.8 were calculated for I212E samples at both 10 μM and at 50 μM , indicating that this mutant has reduced tendency for oligomerization. These data support the conclusion that BMI1 self-associates through at least one interface identified through crystallographic studies.

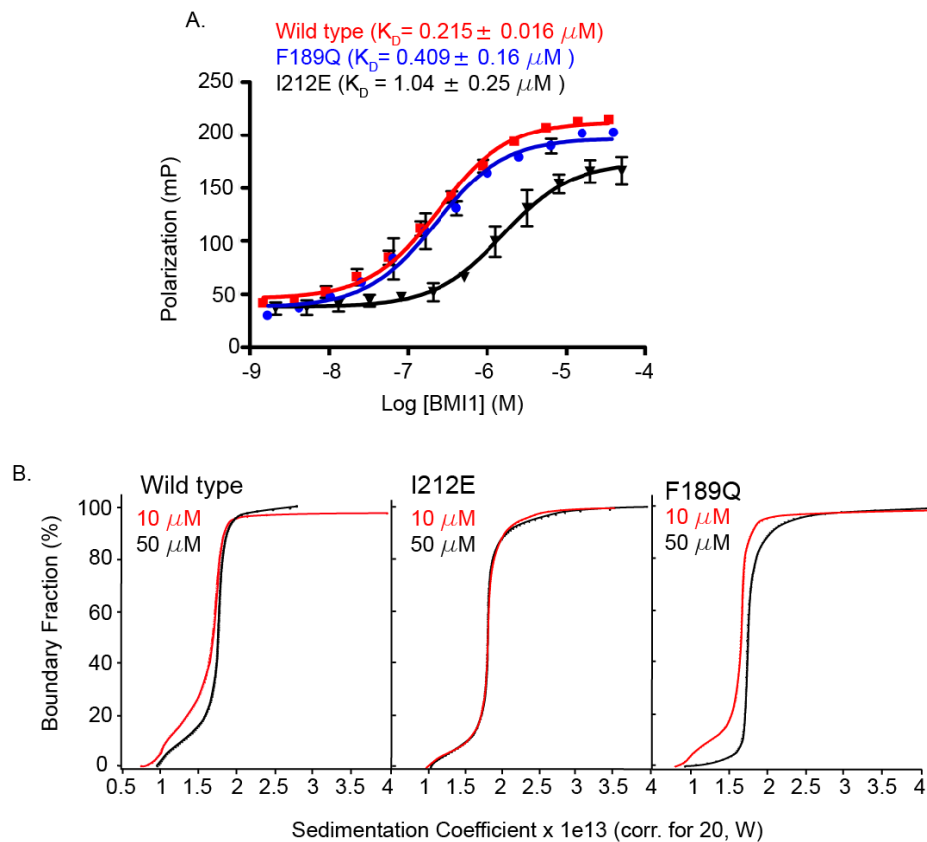


Figure 2.13. BMI1-BMI1 interface mutants disrupt self-association but not interaction with PHC2.

A. Fluorescence polarization experiments titrating FITC-PHC2 with wild type or mutant BMI1 ULD constructs. Experiments were performed in duplicates, error bars and K_D are standard deviation from three independent experiments; B. Van Holde-Weischet [G(s)] plots of sedimentation distributions for wild type or mutant BMI1 -PHC2 complexes at various concentrations.

C.7. Multiple BMI1 protein-protein interactions regulate cellular proliferation

We hypothesized that BMI1's multiple protein-protein interactions are critical to BMI1 function. To test this we developed a cellular proliferation assay in osteosarcoma U2OS cells, which express high levels of BMI1. We found that knockdown of BMI1 in U2OS using siRNA reduced proliferative capacity relative to cells transfected with non-targeting siRNA (Figure 2.14.A, black and blue curves). Following knockdown of endogenous BMI1 by siRNA, cellular proliferation could be rescued with overexpression of BMI1 wild type protein (Figure 2.13.A, green curve). However, overexpression of BMI1 mutants that disrupted either BMI1-PHC2 or BMI1-BMI1 interaction (BMI1 R165E/H174E or I212E mutants, respectively) failed to rescue cellular proliferation (Figure 2.14.A, purple and orange curves). This is further demonstrated by comparing cell numbers at the end of the five day experiment (Figure 2.14.B), demonstrating that both BMI1-PHC2 and BMI1-BMI1 interactions are critical for BMI1 function in regulating cellular proliferation. Expression of wild type or mutant BMI1 was analyzed by western blot showing complete loss of endogenous BMI1 by siRNA treatment and expression of all BMI1 constructs (Figure 2.14.C).

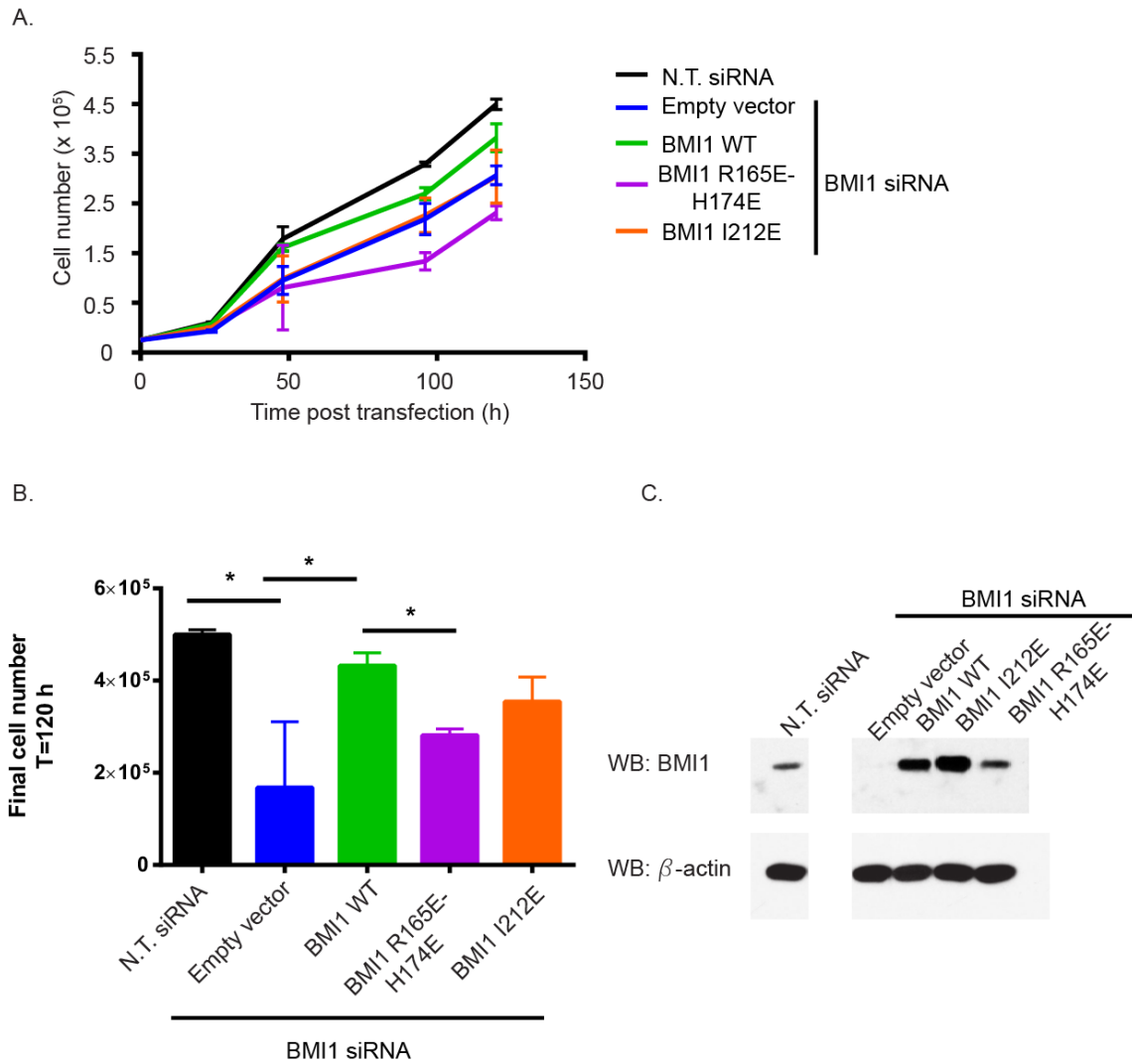


Figure 2.14. Multiple BMI1 protein-protein interactions regulate cellular proliferation.

A. Growth curve for U2OS cells that were transfected with either non-targeting (N.T) siRNA or BMI1 siRNA and transfected with vectors encoding for BMI1 wild type (WT) or mutant constructs. B. Quantification of cell number at end of proliferation experiment, t=120 hours. Mean and standard deviation of two technical replicates are shown and analyzed by the student's t-test. * $p < 0.05$. C. Western blot of cell lysate for cells quantified in part B.

D. Discussion and Conclusion

In the present study we have characterized the BMI1-PHC2 interaction using biophysical and structural biology methods. We found that the BMI1 protein-protein interaction domain forms a moderate affinity complex with the HD1 domain of PHC2 ($K_D= 398$ nM). Structural studies reveal that the BMI1 PPI domain has an ubiquitin-like fold, classifying this domain as an ubiquitin-like domain (ULD). The PHC2 HD1 domain is disordered in solution and forms an intermolecular β sheet upon binding to BMI1 ULD. Disruption of this interaction through mutagenesis impairs cellular proliferation of U2OS cells suggesting that the BMI1-PHC2 interaction plays a role in regulating cell proliferation. Additionally, we demonstrate through multiple approaches that BMI1 weakly self-associates *in vitro* potentially through two interfaces and that disruption of this homo-oligomerization through mutagenesis similarly reduces cellular proliferation. Together this data suggest a model of BMI1 function as a scaffold protein which participates in multiple protein-protein interactions to coordinate subunit organization within the canonical PRC1 complex.

D.1. Need for using multiple methodologies to determine structure of BMI1-PHC2 complex

In order to determine the structure of BMI1-PHC2 complex we needed to overcome multiple difficulties. We found that ^{13}C -detected NMR spectra (see Appendix A) were very helpful in designing BMI1 ULD and PHC2 constructs for structural studies. Previous attempts at structural studies of the BMI1 ULD were hampered by protein instability¹⁸ highlighting the utility of this methodology in designing constructs for structural biology.¹⁶ In the present studies we also used this methodology for the experimental determination of the minimal BMI1-binding motif in PHC proteins which corresponds to only a fraction of conserved residues in the natively unstructured HD1 domain. Identification of this motif through bioinformatics or sequential deletion studies would be time consuming. Additionally, as structural studies with larger fragments of unstructured proteins are challenging, this work reinforces the value of the ^{13}C -detected NMR methods in defining protein-protein interaction interfaces in natively unstructured proteins.¹⁷

We employed a hybrid X-ray crystallography and solution NMR approach to determine the structure of the BMI1-PHC2 complex. This approach incorporated data from both techniques to circumvent challenges encountered in structural studies with individual methods. Overall this tactic enabled solution of the 3D structure for a complete description of the BMI1-PHC2 complex. Hybrid structural biology approaches integrating multiple methodologies have been used to generate structures of individual proteins or protein complexes that may not be available through application of a single method. For example, combined use of solution NMR and cryo-EM revealed the structural details of regulation of the APC/C E3 ligase complex by the multi-domain protein EMI1.²⁷ Similarly, integration of solution NMR and small angle x-ray scattering (SAXS) have been used for a complete molecular depiction of multi-domain proteins or protein complexes.²⁸ These studies demonstrate the complementarity of structural biology methods with different resolutions to define both detailed molecular details as well as global architecture of macromolecular complexes. In contrast, the approach demonstrated here is novel in that it integrates data from two methods which were both challenged by the intrinsic behavior of the protein.

D.2. Structural studies of BMI1-PHC2 reveal basis for polycomb ULD specificity

In these studies we determined the structure of the BMI1-PHC2 complex, which was previously unknown. This structure demonstrates that a short fragment of PHC2 adopts a β hairpin structure when in complex with the BMI1 ubiquitin-like domain, which is a common architecture in other structures of polycomb protein complexes.^{18,19} The PPI domains of PCGF1 and Ring1B adopt an ubiquitin-like fold and interact with a short motifs of their respective binding partners, BCOR and CBX7 or RYBP through intermolecular β sheets (Figure 2.15.A). The stabilizing polar contacts between the sidechains of BMI1 Arg164 and PHC2 Gln44 are conserved between PCGF1 Arg193 and BCOR Glu1640 in the PCGF1-BCOR structure and mutation of PCGF1 Arg193 to Ala abolished the interaction in bacterial two-hybrid experiments.¹⁸ In the Ring1B-CBX7 complex Ring1B Arg246 and CBX7 Glu236 are the equivalent residues. Likewise, the critical histidine-backbone carbonyl hydrogen bond is conserved in all three structures, with PCGF1 His202 hydrogen bonding with the backbone carbonyl of BCOR Pro1746 and Ring1B His258 interacting with the backbone of both CBX7

Ala227 and Ile230 (Figure 2.14.A). The buried acidic residue observed in the BMI1-PHC2 structure is also conserved in the Ring1B-CBX7 or RYBP structures with CBX7 Glu236 and RYBP Asp172 extending towards Ring1B Arg266 through the core of the interface. This interaction is not conserved in the PCGF1-BCOR structure. The other PCGF orthologs (PCGF 2, 3, 5 and 6) are also predicted to have the same ubiquitin-like domain.²⁹ PCGF proteins have low overall sequence identity (average 36% between BMI1 and PCGF1,3,5,6; 62% between BMI1 and Mel-18/PCGF2) however the protein-protein interface identified by this study and previous structures is highly conserved among all polycomb proteins with this fold (Figure 2.15.B).¹⁸⁻²⁰

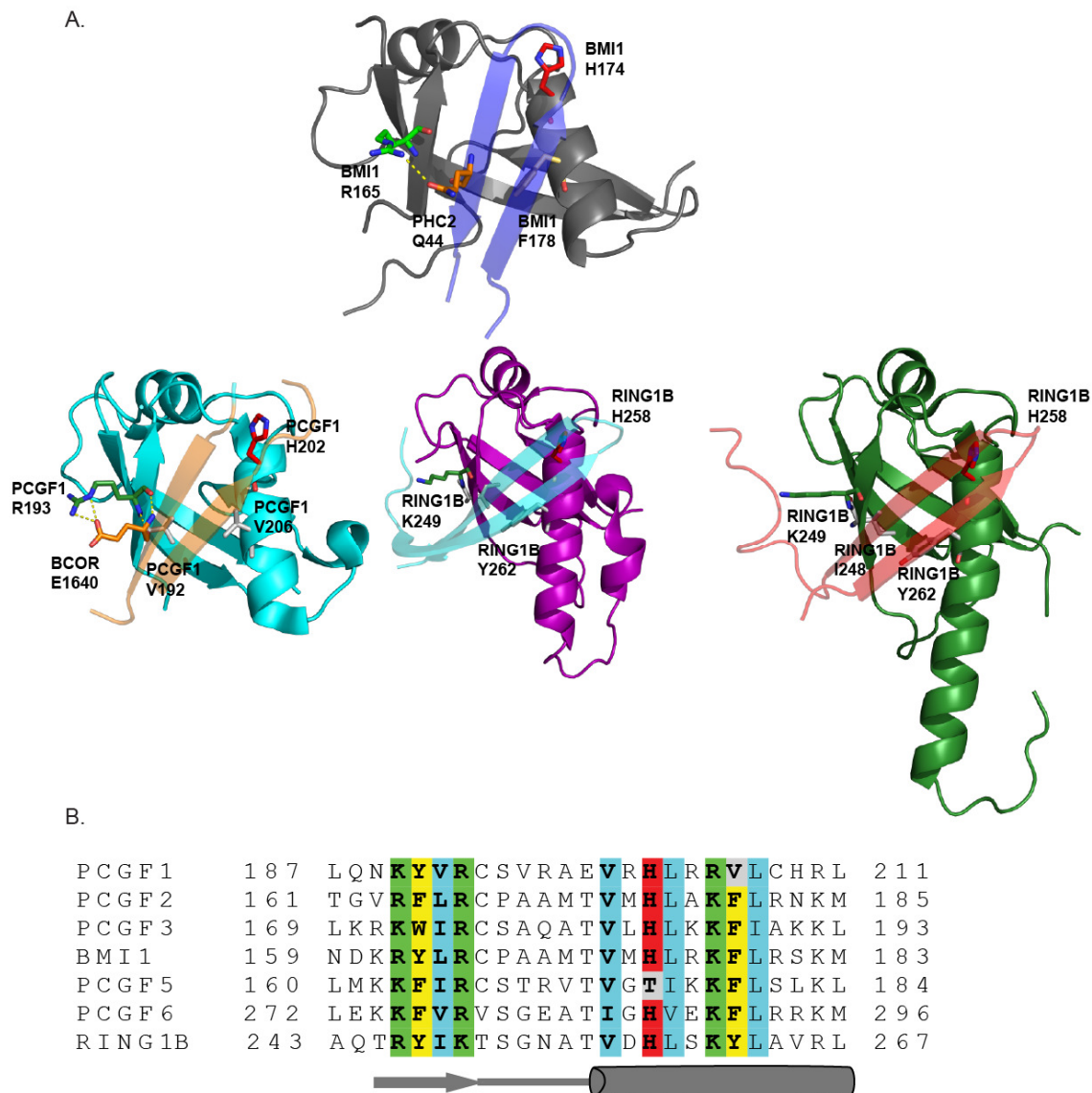


Figure 2.15. Basis for binding partner selectivity in ULD-domain polycomb proteins.

A. Structures of polycomb ULD complexes showing a conserved structure and similar biochemical contacts. Top: BMI1-Phc2 structure. Bottom Left: PCGF1-BCOR structure (PDB: 4HPL); center: Ring1B-RYBP structure (PDB: 3IXS); right: Ring1B-CBX7 structure (PDB: 3GS2). Critical interface residues are labeled. B. Sequence alignment of polycomb ULD domains. Conserved basic residues are shown in green, residues making up the hydrophobic core are in teal, the conserved His is shown in red and bulky residues are shown in yellow. Non-conserved residues are shown in gray. Secondary structure elements are illustrated below the sequence.

Despite such conserved structural and biochemical similarity there is no evidence for promiscuous binding among polycomb ULD-containing proteins. In bacterial- or yeast-two-hybrid experiments BMI1 does not interact with BCOR or CBX proteins,^{4,5} Ring1B does not interact with BCOR or PHC2 and similarly PCGF1 does not interact with PHC1 or CBX7.^{4,5,18} This lack of cross-interactions can be rationalized through residue mismatch at key positions. In the BMI1-PHC2 structure presented in this chapter, hydrophobic contacts between BMI1 Phe178 and PHC2 Ile43 contribute to the hydrophobic core of the structure. In the PCGF1-BCOR structure these residues are equivalent to the PCGF1 Val206 and BCOR Phe1641, respectively, and a likely steric clash between BMI1 Phe178 and BCOR Phe1641 would prevent this interaction. Binding selectivity between BMI1 and the Ring1B binding partners CBX7 or RYBP is likely due to the substitution of the key basic residue in PHC2 Gln44 with hydrophobic groups in the Ring1B binding partners (CBX7 Val238 and RYBP Leu165) thereby eliminating the sidechain polar interactions at this position that stabilize the BMI1-PHC2 interaction. This analysis demonstrates the molecular basis for polycomb complex heterogeneity through interactions of different ULD-containing subunits with distinct binding partners.

D.3. Implications of BMI1 protein-protein interactions on PRC1 architecture

Together, our studies suggest a central role for BMI1 in coordinating the architecture of the canonical PRC1 complex. Through interactions with Ring1B and PHC2 proteins BMI1 links the ubiquitin ligase Ring1B to the polymerizing PHC subunit (Figure 2.16.A). To our knowledge this model is the first of the overall complex based on structural data of all subunits and interactions. Further, we have identified that BMI1 ULD self-associates by two potential head-to-head interfaces. While these interactions are weak *in vitro* (mid-micromolar affinity), we can imagine that in a cellular context this interaction is enhanced by the high affinity interaction between the BMI1 ULD and the PHC HD1 domains and the polymerization of the PHC SAM domains (Figure 2.16.B). Complementary to PHC oligomerization, BMI1 ULD oligomerization may contribute to overall stability of PRC1 complex in cells. Together these studies provide detailed insight into BMI1's protein-protein interactions and set the stage for future work identifying other BMI1 potential binding partners and assessing the role of BMI1 self-association in PcG complex assembly and gene silencing mechanisms.

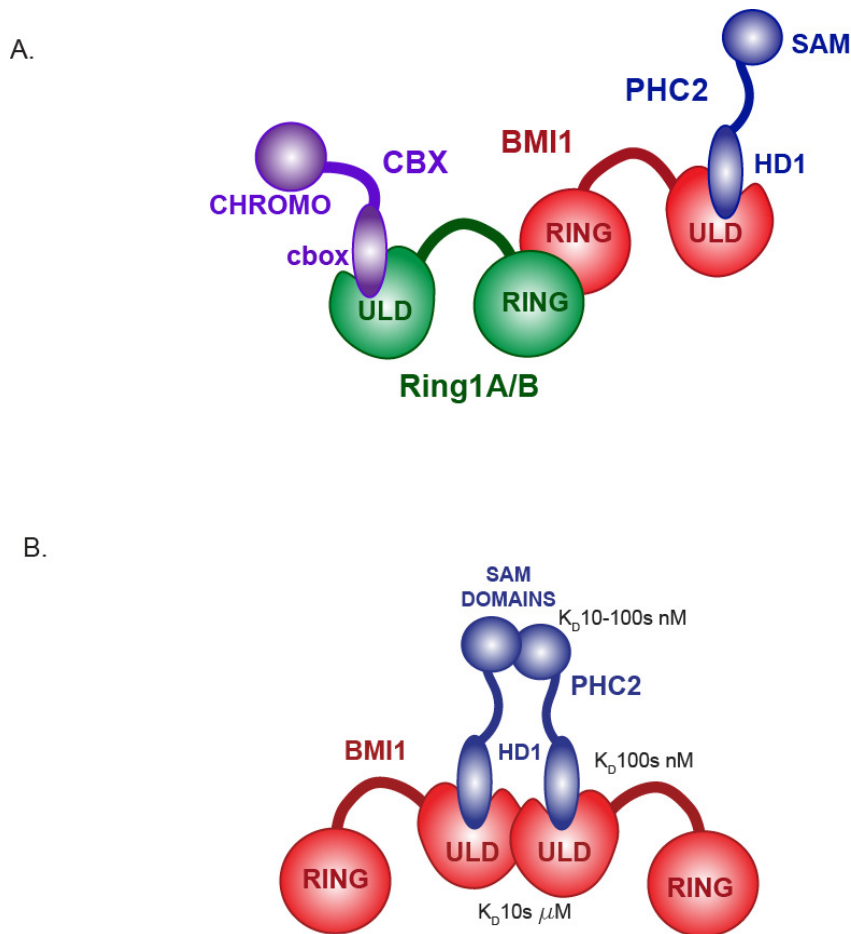


Figure 2.16. Models for PRC1 organization and oligomerization.

A. Model of canonical PRC1 complex based on structurally characterized protein-protein interactions. B. Model of PRC1 “oligomerization unit” showing affinities for different protein-protein interactions. Affinities for BMI1 ULD homo-oligomerization and BMI1-PHC2 HD1 heterodimerization are estimated or determined in this paper and SAM domain homo-oligomerization affinity is from previously published studies.^{7,8}

Acknowledgements

The work in this chapter was performed with assistance from various members of the Cierpicki and Grembecka labs. I thank previous members of the group for BMI1 ULD construct design, Dr. Hyoje Cho for helping solve the BMI1 ULD crystal structure, Dr. Bhavna Malik for help with cellular transfection and cell proliferation experiments.

E. Experimental methods

Plasmid construction

cDNA encoding full-length BMI1 and full-length PHC2_B was obtained from the Raul Lab (Department of Pathology, University of Michigan) and the desired constructs were subcloned into the pET32p bacterial expression vector. The pET32p vector is a modified pET32a vector (Novagen) that contains a Precision Protease recognition sequence between a thioredoxin (trx) solubility tag and the target protein. Linear DNA sequence for Avi-BMI1 Δ Ring and for the *E. Coli* BirA biotin ligase were ordered from Genscript. Plasmid encoding the synthetic gene for PHC2₃₀₋₆₃-BMI1 was ordered from Life Technologies and the gene coding sequence was subcloned into the pET32p vector. Mammalian expression plasmids were constructed using standard molecular biology techniques using the templates described above. Plasmids for mutant constructs were made using standard QuikChange mutagenesis protocol (Stratagene) and verified by Sanger sequencing.

Protein expression and purification

BMI1 constructs: All BMI1 proteins were expressed in Codon+ BL21 (DE3) *E. coli* cells (Sigma) with an N-terminal His₆- thioredoxin tag (trx). Cells subjected to ampicillin (100 μ g/mL) selection were grown in Luria broth (Fisher) or labeled M9 medium at 37°C (220 rpm) until OD₆₀₀ reached 0.6-0.8. Incubation temperature was lowered to 18°C for 1 hour then protein over-expression was induced with 0.5 mM (final concentration) Isopropyl β -D-1-thiogalactopyranoside (IPTG; Gold Bio) for 16-18 hours at 18°C. Cells were lysed in a buffer containing 50 mM Tris, pH 7.5 at 25°C, 250 mM NaCl, 20 mM Imidazole, 0.5 mM phenylmethylsulfonyl fluoride (PMSF), 1 mM β -mercaptoethanol (β -ME) using a French Press. Clarified lysate was applied to a HisTrap HP (GE Healthcare) column using Äkta Prime FPLC and eluted with lysis buffer containing 0.5 M imidazole. To remove the His₆- thioredoxin tag, the protein was cleaved with Precision protease and BMI1 ULD constructs were further purified using cation exchange chromatography. Purified protein solution was buffer exchanged into storage buffer (100 mM Bis-Tris, pH 6.5, 50 mM NaCl, 1 mM Tris(2-carboxyethyl)phosphine hydrochloride (TCEP-HCl)) using HiPrep Desalting Column (GE Healthcare). The PHC2₃₀₋₆₃-BMI1 fusion protein was expressed and purified as described above.

If the protein was to be used for crystallization, ion-exchange purified BMI1 ULD was saturated with PHC2₃₃₋₅₆ peptide (Genscript) and applied to a Superdex S75 gel filtration column (GE Healthcare) pre-equilibrated with storage buffer (100 mM Bis-Tris, pH 6.5, 50 mM NaCl, 1 mM TCEP). Protein concentration was measured using absorbance at 280 nm. Protein identity and purity was verified by sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE). Protein was flash frozen and stored at -80°C.

PHC2₁₋₇₉. PHC2₁₋₇₉ protein was expressed in BL21 (DE3) *E. coli* cells (Sigma) with an N-terminal His₆- thioredoxin tag. Cells were subjected to ampicillin (100 µg/mL) selection and grown in Luria broth (Fisher) or labeled M9 medium (Marley 2001) at 37°C (220 rpm) until OD₆₀₀ reached 0.6-0.8. Incubation temperature was lowered to 18°C for 1 hour then protein over expression was induced with 0.5 mM (final concentration) Isopropyl β-D-1-thiogalactopyranoside (IPTG; Gold Bio) for 16-18 hours at 18°C. Cells were lysed in a buffer containing 50 mM Tris, pH 7.5 at 25°C, 250 mM NaCl, 20 mM imidazole, 0.5 mM PMSF, 1 mM β-ME using a French Press. Clarified lysate was applied to a HisTrap HP (GE Healthcare) column using Äkta Prime FPLC and eluted with lysis buffer containing 0.5 mM imidazole. Trx-PHC2₁₋₇₉ protein was further purified by gel filtration on a Superdex S-75 column in buffer 50 mM phosphate, pH 6.5, 50 mM NaCl, 1 mM TCEP-HCl. To obtain untagged PHC2₁₋₇₉ protein, the thioredoxin tag was cleaved with Precision protease during dialysis against 100-fold excess dialysis buffer A (50 mM Tris, 50 mM NaCl, 1 mM TCEP-HCl, pH 7.5 at 25°C). The thioredoxin tag was separated from HD1 using a nickel affinity column to bind the thioredoxin protein. PHC2₁₋₇₉ was further purified by Superdex S75 gel filtration (GE Healthcare) pre-equilibrated with storage buffer (50 mM phosphate, pH 6.5, 50 mM NaCl, 1 mM TCEP) Protein concentration was measured using absorbance at 280 nm. Protein identity and purity was verified by sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE). Protein was flash frozen and stored at -80°C.

Isothermal Titration Calorimetry

BMI1 and trx- PHC2₁₋₇₉ were dialyzed extensively against ITC buffer (50 mM potassium phosphate, pH 6.5, 50 mM NaCl, 1 mM TCEP) and degassed prior to measurement. The titrations were performed using a VP-TIC titration calorimetric system (MicroCal) at 25°C. The calorimetric cell, containing BMI1 ULD (22 µM), was titrated with trx- PHC2₁₋₇₉ (164 µM)

injected in 10 μ L aliquots. Data were analyzed using Origin 7.0 (OriginLab) to obtain K_D and stoichiometry.

NMR Spectroscopy

Samples for BMI1 ULD- PHC2₁₋₇₉ interaction studies were made with ¹⁵N-labeled BMI1 ULD at 50 μ M in a buffer containing 50 mM Bis-Tris, pH 6.5, 50 mM NaCl, 1 mM TCEP and 5% D₂O. For PHC2₁₋₇₉ binding studies, samples were prepared with unlabeled PHC2₁₋₇₉ at 25 μ M, 50 μ M or 100 μ M, final concentration.

Samples for PHC2₁₋₇₉ backbone carbon assignment were made with 60 μ M ¹³C¹⁵N labeled protein prepared in a buffer containing 100 mM Bis-Tris, pH 6.5, 100 mM NaCl, 1 mM TCEP and 5% D₂O. ¹³C-detected CACO, CBCACO and CANCO experiments were used for carbon assignment^{30,31}. For BMI1 binding studies, equimolar concentrations of unlabeled BMI1 ULD was added to the labeled PHC2₁₋₇₉ and the same experiments were recorded.

All spectra were acquired at 30°C on a 600 MHz Bruker Advance III spectrometer equipped with cryoprobe, running Topspin version 2.1. Processing and spectral visualization was performed using NMRPipe³² and Sparky.³³

Crystallization and Crystal Structure Determination

Initial crystals were obtained through sitting drop screening of gel-filtration purified BMI1 ULD-PHC2₃₃₋₅₆ complex. Crystals were further optimized by hanging-drop vapor diffusion with equal volumes (1 μ L) of protein (9 mg/mL in 50 mM Bis-Tris, pH 6.5, 50 mM NaCl, 1 mM TCEP) and the precipitant solution (100 mM MES, pH 6.5, 50 mM MgCl₂, 7% isopropanol, 6% PEG 4000). Crystals formed within 7 days at 39°F. Crystals were cryoprotected using the precipitant solution containing 20% glycerol and flash frozen in liquid nitrogen. X-ray diffraction data of crystals were collected at a resolution of 2.5 Å at the Advanced Photon Source at LS-CAT beam line 21-ID-F. The data were then indexed, integrated, and scaled using the HKL2000 suite.³⁴ The crystals belonged to the space group P3₂12, with the unit cell parameters a=b=78.275, c=43.119 Å, $\alpha=\beta=90$, $\gamma=120^\circ$. With 1 molecule in the asymmetric unit, the crystal volume per unit of protein mass was 3.77 Å³ Da⁻¹, corresponding to a solvent content of 67.43%.³⁵ The structure was determined by molecular replacement method with the CCP4 version of MOLREP³⁶ using the polycomb group Ring finger protein complex structure (PDB code 4HPM B chain) as a

search model.¹⁹ Model building was performed manually using the program WinCoot³⁷ and the refinement was performed with CCP4 re mac5.³⁸ The data statistics are summarized in Table 2.1.

Table 2.1. Statistics of crystal structure of BMI1 ULD

Data Set	BMI1-PHC2
Experimental Data	
X-ray source	APS 21_ID-F
Wavelength (Å)	0.97872
Space group	P3 ₂ 12
Unit cell parameters (Å)	a=b=78.28, c=43.12, α=β=90°, γ=120°
Resolution limit (Å)	50 – 2.5 (2.54 – 2.50) ^a
Total reflections	52,849
Unique reflections	5,292
Redundancy	10.0 (8.3)
Completeness (%)	99.8 (99.6)
R _{sym} ^b	0.087 (0.477)
Average I/σ (I)	33.93 (3.28)
Refinement Details	
Resolutions (Å)	50 – 2.51
Reflections (working)	5,018
Reflections (test)	268
R _{work} /R _{free} (%) ^c	24.01/33.38
Number of water molecules	20
RMS deviation from ideal geometry	
bond length (Å)	0.012
bond angle (°)	1.737
Average B factors (Å ²)	57.91

^aThe numbers in parentheses describe the relevant value for the last resolution shell.

^b $R_{sym} = \sum |I_i - \langle I \rangle| / \sum I_i$ where I_i is the intensity of the i -th observation and $\langle I \rangle$ is the mean intensity of the reflections.

^c $R_{work} = \sum ||F_{obs}| - |F_{calc}|| / \sum |F_{obs}|$, crystallographic R factor, and $R_{free} = \sum ||F_{obs}| - |F_{calc}|| / \sum |F_{obs}|$ when all reflections belong to a test set of randomly selected data. The high R_{free} is due to the inability to model part of the density corresponding to PHC2, illustrated in Figure 2.9.

Measurement of BMI1-PHC2 distances using NMR

Protein samples for NMR structure determination contained 200 μM PHC2₃₀₋₆₃-BMI1 ULD fusion protein. Backbone and methyl sidechain assignment were made with ¹³C¹⁵N-labeled PHC2₃₀₋₆₃-BMI1 prepared in a buffer containing 50 mM Bis-Tris, pH 6.5, 50 mM NaCl, 1 mM TCEP and 5% D₂O. Backbone assignment was done using a series of triple-resonance

experiments including HNCACB, CBCA(CO)NH, HNCA, HN(CO)CA, HNCO and HN(CA)CO. Methyl side chain resonances were assigned using 3D ^{13}C - ^1H - ^1H HCCH-TOSCY. Nuclear Overhauser Effect (NOE) cross-peaks were identified using ^{15}N -separated NOESY-HSQC and ^{13}C -separated NOESY-HSQC spectra. Distance restraints were calculated with CYANA²¹ from ^1HN and methyl ^1H NOEs within PHC2 residues and between PHC2 and BMI1 ULD. The hybrid NMR- X-ray crystal structure of the BMI1 ULD-PHC2 complex was refined using Rosetta²²⁻²⁴ constrained with these distance restraints, PHC2 dihedral angles calculated by TALOS+³⁹ and BMI1 ULD coordinates from the crystal structure. The data statistics are summarized in table 2.2.

Table 2.2. Statistics for NMR Structure Determination.

<i>PHC2₃₀₋₆₃-BMI1 NMR assignment statistics</i>	
Assigned backbone amide resonances	91% (91% total BB amides)
Assigned methyl resonances	79% (57% of total methyl)
<i>NMR distance and dihedral constraints</i>	
Distance restraints	
Total number of NOE restraints	162
Intra-residue restraints (I=J)	46
Sequential restraints (I-J=1)	45
Backbone-backbone	24
Backbone-sidechain	8
Sidechain-sidechain	13
Medium-range restraints 1<(I-J)<5	10
Backbone-backbone	3
Backbone-sidechain	1
Sidechain-sidechain	6
Long-range restraints (I-J)>=5	61
Total Dihedral angle restraints	30
Phi	15
Psi	15

Fluorescence polarization assay

For binding experiments (K_D measurements) fluorescein-labeled PHC2₃₂₋₆₁ (FITC-PHC2; Genscript) at 20 nM was titrated with a range of BMI1 ULD concentrations in the FP buffer (50 mM Bis-Tris, pH 7.5, 50 mM NaCl, 1 mM TCEP, 0.01% BSA, 0.25% tween-20). After a 1 hour

incubation of the protein-peptide complexes, changes in fluorescence polarization and anisotropy were measured at 525 nm after excitation at 495 nm using PHERAstar microplate reader (BMG). Data fit with a sigmoidal dose response equation were used to assess BMI1 ULD construct binding with the Prism 4.0 (GraphPad) program.

Analytical Ultracentrifugation

Sedimentation velocity experiments were performed on a Beckman Optima XL-I at the Center for Analytical Ultracentrifugation of Macromolecular Assemblies (CAUMA) at the University of Texas Health Center at San Antonio. Calculations were performed with the UltraScan software⁴⁰ at the Texas Advanced Computing Center at the University of Texas at Austin and at the Bioinformatics Core Facility at the University of Texas Health Science Center at San Antonio.^{40,41} All samples were measured in a 50 mM potassium phosphate buffer, pH 6.5 containing 50 mM NaCl and 1 mM TCEP. All data were collected at 20°C and spun at 50 krpm. All data were first analyzed by two-dimensional spectrum analysis with simultaneous removal of time-invariant noise^{42,43} and then by enhanced van Holde-Weischet analysis and genetic algorithm refinement^{44,45} where applicable, followed by Monte Carlo analysis.⁴¹

Pull-down experiments

U2OS cells were transfected with BirA, myc-PHC2 and Avi-BMI1 Δ Ring constructs using Fugene 6 (Roche) transfection agent. Cells were harvested by centrifugation 48 hours after transfection and lysed through sonication in lysis buffer (50 mM HEPES, pH 7.0, 150 mM NaCl, 1 mM EDTA, 2.5 mM EGTA, 1 mM NEM, 1 mM NaF, 0.1 M Na₃VO₄, 10% glycerol, 0.1 mM β -glycerophosphate, 0.01% NP-40) with protease inhibitor cocktail (Sigma). Lysate was clarified by centrifugation and 10% total volume of whole cell lysate was taken as input. Streptavidin magnetic beads (Pierce) were added to each sample and incubated at 4°C with rotation for 16 hours. Beads were washed 4 times with wash buffer (20 mM Tris, pH 8.0, 300 mM KCl, 1 mM EDTA, 10% glycerol, 0.1% NP-40) with protease inhibitor cocktail (Sigma) and proteins were eluted with SDS in wash buffer. Samples were analyzed by SDS-Page and Western blotting probed with either myc antibody (Cell Signaling) or BMI1 antibody (Millipore).

Cellular proliferation

50,000 U2OS cells were transfected with 50nM of Control or BMI1 siRNA (Dharmacon) using lipofectamine RNAiMax (Life Technologies) and incubated for 48 hours. After incubation,

the medium was replaced with incomplete medium and BMI1 siRNA transfected cells were transfected with BMI1 WT and mutant constructs were transfected using Lipofectamine 2000 (Life Technologies). To count cell numbers at 24, 48, 96 and 120 hours after transfection, cells were washed, trypsinized and live cells were counted at different time intervals using Trypan Blue staining. At the end of the each time point whole cell lysates were harvested and protein levels visualized by Western blot.

F. References

1. Boyer, L. A., Plath, K., Zeitlinger, J., Brambrink, T., Medeiros, L. A., Lee, T. I., Levine, S. S., Wernig, M., Tajonar, A., Ray, M. K., Bell, G. W., Otte, A. P., Vidal, M., Gifford, D. K., Young, R. A. & Jaenisch, R. Polycomb complexes repress developmental regulators in murine embryonic stem cells. *Nature* **441**, 349-353 (2006).
2. Lee, T. I., Jenner, R. G., Boyer, L. A., Guenther, M. G., Levine, S. S., Kumar, R. M., Chevalier, B., Johnstone, S. E., Cole, M. F., Isono, K., Koseki, H., Fuchikami, T., Abe, K., Murray, H. L., Zucker, J. P., Yuan, B., Bell, G. W., Herbolsheimer, E., Hannett, N. M., Sun, K., Odom, D. T., Otte, A. P., Volkert, T. L., Bartel, D. P., Melton, D. A., Gifford, D. K., Jaenisch, R. & Young, R. A. Control of developmental regulators by Polycomb in human embryonic stem cells. *Cell* **125**, 301-313 (2006).
3. Bracken, A. P., Dietrich, N., Pasini, D., Hansen, K. H. & Helin, K. Genome-wide mapping of Polycomb target genes unravels their roles in cell fate transitions. *Genes Dev* **20**, 1123-1136 (2006).
4. Alkema, M. J., Bronk, M., Verhoeven, E., Otte, A., van 't Veer, L. J., Berns, A. & van Lohuizen, M. Identification of Bmi1-interacting proteins as constituents of a multimeric mammalian polycomb complex. *Genes Dev* **11**, 226-240 (1997).
5. Gunster, M. J., Satijn, D. P., Hamer, K. M., den Blaauwen, J. L., de Bruijn, D., Alkema, M. J., van Lohuizen, M., van Driel, R. & Otte, A. P. Identification and characterization of interactions between the vertebrate polycomb-group protein BMI1 and human homologs of polyhomeotic. *Mol Cell Biol* **17**, 2326-2335 (1997).
6. Kyba, M. & Brock, H. W. The Drosophila polycomb group protein Psc contacts ph and Pc through specific conserved domains. *Mol Cell Biol* **18**, 2712-2720 (1998).
7. Kim, C. A., Gingery, M., Pilpa, R. M. & Bowie, J. U. The SAM domain of polyhomeotic forms a helical polymer. *Nat Struct Biol* **9**, 453-457 (2002).
8. Robinson, A. K., Leal, B. Z., Nanyes, D. R., Kaur, Y., Ilangovan, U., Schirf, V., Hinck, A. P., Demeler, B. & Kim, C. A. Human polyhomeotic homolog 3 (PHC3) sterile alpha motif (SAM) linker allows open-ended polymerization of PHC3 SAM. *Biochemistry* **51**, 5379-5386 (2012).
9. Nanyes, D. R., Junco, S. E., Taylor, A. B., Robinson, A. K., Patterson, N. L., Shivarajpur, A., Halloran, J., Hale, S. M., Kaur, Y., Hart, P. J. & Kim, C. A. Multiple polymer architectures of human polyhomeotic homolog 3 sterile alpha motif. *Proteins* **82**, 2823-2830 (2014).

10. Levine, S. S., Weiss, A., Erdjument-Bromage, H., Shao, Z., Tempst, P. & Kingston, R. E. The core of the polycomb repressive complex is compositionally and functionally conserved in flies and humans. *Mol Cell Biol* **22**, 6070-6078 (2002).
11. Cohen, K. J., Hanna, J. S., Prescott, J. E. & Dang, C. V. Transformation by the Bmi-1 oncoprotein correlates with its subnuclear localization but not its transcriptional suppression activity. *Mol Cell Biol* **16**, 5527-5535 (1996).
12. Alkema, M. J., Jacobs, H., van Lohuizen, M. & Berns, A. Perturbation of B and T cell development and predisposition to lymphomagenesis in Emu Bmi1 transgenic mice require the Bmi1 RING finger. *Oncogene* **15**, 899-910 (1997).
13. Dimri, G. P., Martinez, J. L., Jacobs, J. J., Keblusek, P., Itahana, K., Van Lohuizen, M., Campisi, J., Wazer, D. E. & Band, V. The Bmi-1 oncogene induces telomerase activity and immortalizes human mammary epithelial cells. *Cancer Res* **62**, 4736-4745 (2002).
14. Itahana, K., Zou, Y., Itahana, Y., Martinez, J. L., Beausejour, C., Jacobs, J. J., Van Lohuizen, M., Band, V., Campisi, J. & Dimri, G. P. Control of the replicative life span of human fibroblasts by p16 and the polycomb protein Bmi-1. *Mol Cell Biol* **23**, 389-401 (2003).
15. de Boer, E., Rodriguez, P., Bonte, E., Krijgsveld, J., Katsantoni, E., Heck, A., Grosveld, F. & Strouboulis, J. Efficient biotinylation and single-step purification of tagged transcription factors in mammalian cells and transgenic mice. *Proc Natl Acad Sci U S A* **100**, 7480-7485 (2003).
16. Gray, F. L., Murai, M. J., Grembecka, J. & Cierpicki, T. Detection of disordered regions in globular proteins using (1)(3)C-detected NMR. *Protein Sci* **21**, 1954-1960 (2012).
17. Grembecka, J., Belcher, A. M., Hartley, T. & Cierpicki, T. Molecular basis of the mixed lineage leukemia-menin interaction: implications for targeting mixed lineage leukemias. *J Biol Chem* **285**, 40690-40698 (2010).
18. Junco, S. E., Wang, R., Gaipa, J. C., Taylor, A. B., Schirf, V., Gearhart, M. D., Bardwell, V. J., Demeler, B., Hart, P. J. & Kim, C. A. Structure of the polycomb group protein PCGF1 in complex with BCOR reveals basis for binding selectivity of PCGF homologs. *Structure* **21**, 665-671 (2013).
19. Wang, R., Taylor, A. B., Leal, B. Z., Chadwell, L. V., Ilangoan, U., Robinson, A. K., Schirf, V., Hart, P. J., Lafer, E. M., Demeler, B., Hinck, A. P., McEwen, D. G. & Kim, C. A. Polycomb group targeting through different binding partners of RING1B C-terminal domain. *Structure* **18**, 966-975 (2010).
20. Bezsonova, I., Walker, J. R., Bacik, J. P., Duan, S., Dhe-Paganon, S. & Arrowsmith, C. H. Ring1B contains a ubiquitin-like docking module for interaction with Cbx proteins. *Biochemistry* **48**, 10542-10548 (2009).
21. Guntert, P. Automated NMR structure calculation with CYANA. *Methods Mol Biol* **278**, 353-378 (2004).
22. Conway, P., Tyka, M. D., DiMaio, F., Kondering, D. E. & Baker, D. Relaxation of backbone bond geometry improves protein energy landscape modeling. *Protein Sci* **23**, 47-55 (2014).
23. Khatib, F., Cooper, S., Tyka, M. D., Xu, K., Makedon, I., Popovic, Z., Baker, D. & Players, F. Algorithm discovery by protein folding game players. *Proc Natl Acad Sci U S A* **108**, 18949-18953 (2011).

24. Tyka, M. D., Keedy, D. A., Andre, I., Dimaio, F., Song, Y., Richardson, D. C., Richardson, J. S. & Baker, D. Alternate states of proteins revealed by detailed energy landscape mapping. *J Mol Biol* **405**, 607-618 (2011).
25. Howlett, G. J., Minton, A. P. & Rivas, G. Analytical ultracentrifugation for the study of protein association and assembly. *Curr Opin Chem Biol* **10**, 430-436 (2006).
26. Cole, J. L., Lary, J. W., T, P. M. & Laue, T. M. Analytical ultracentrifugation: sedimentation velocity and sedimentation equilibrium. *Methods Cell Biol* **84**, 143-179 (2008).
27. Frye, J. J., Brown, N. G., Petzold, G., Watson, E. R., Grace, C. R., Nourse, A., Jarvis, M. A., Kriwacki, R. W., Peters, J. M., Stark, H. & Schulman, B. A. Electron microscopy structure of human APC/C(CDH1)-EMI1 reveals multimodal mechanism of E3 ligase shutdown. *Nat Struct Mol Biol* **20**, 827-835 (2013).
28. Grishaev, A., Wu, J., Trehwella, J. & Bax, A. Refinement of multidomain protein structures by combination of solution small-angle X-ray scattering and NMR data. *J Am Chem Soc* **127**, 16621-16628 (2005).
29. Sanchez-Pulido, L., Devos, D., Sung, Z. R. & Calonje, M. RAWUL: a new ubiquitin-like domain in PRC1 ring finger proteins that unveils putative plant and worm PRC1 orthologs. *BMC Genomics* **9**, 308 (2008).
30. Bermel, W., Bertini, I., Duma, L., Felli, I. C., Emsley, L., Pierattelli, R. & Vasos, P. R. Complete assignment of heteronuclear protein resonances by protonless NMR spectroscopy. *Angew Chem Int Ed Engl* **44**, 3089-3092 (2005).
31. Bermel, W., Bertini, I., Felli, I. C., Pierattelli, R. & Vasos, P. R. A selective experiment for the sequential protein backbone assignment from 3D heteronuclear spectra. *J Magn Reson* **172**, 324-328 (2005).
32. Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J. & Bax, A. NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J Biomol NMR* **6**, 277-293 (1995).
33. Goddard, T. D. & Kneller, D. G. *SPARKY 3* (University of California, San Francisco).
34. Otwinowski, Z. & Minor, W. Processing of X-ray Diffraction Data Collected in Oscillation Mode *Methods Enzymol* **276**, 307-326 (1997).
35. Matthews, B. W. Solvent content of protein crystals. *J Mol Biol* **33**, 491-497 (1968).
36. Barycki, J. J., O'Brien, L. K., Bratt, J. M., Zhang, R., Sanishvili, R., Strauss, A. W. & Banaszak, L. J. Biochemical characterization and crystal structure determination of human heart short chain L-3-hydroxyacyl-CoA dehydrogenase provide insights into catalytic mechanism. *Biochemistry* **38**, 5786-5798 (1999).
37. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* **60**, 2126-2132 (2004).
38. Murshudov, G. N., Vagin, A. A. & Dodson, E. J. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D Biol Crystallogr* **53**, 240-255 (1997).
39. Shen, Y., Delaglio, F., Cornilescu, G. & Bax, A. TALOS+: a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. *J Biomol NMR* **44**, 213-223 (2009).

40. Demeler, B. Methods for the design and analysis of sedimentation velocity and sedimentation equilibrium experiments with proteins. *Curr Protoc Protein Sci* **Chapter 7**, Unit 7 13 (2010).
41. Brookes, E. & Demeler, B. Parallel computational techniques for the analysis of sedimentation velocity experiments in UltraScan *Colloid Polym. Sci.* **286**, 139-148 (2008).
42. Brookes, E., Boppana, R. V. & Demeler, B. Computing large sparse multivariate optimization problems with an application in biophysics. *Supercomputing Proceedings of the ACM/IEEE*, 42 (2006).
43. Brookes, E., Demeler, B., Rosano, C. & Rocco, M. The implementation of SOMO (SOLution MOdeller) in the UltraScan analytical ultracentrifugation data analysis suite: enhanced capabilities allow the reliable hydrodynamic modeling of virtually any kind of biomacromolecule. *Eur Biophys J* **39**, 423-435 (2010).
44. Brookes, E. & Demeler, B. Genetic algorithm optimization for obtaining accurate molecular weight distributions from sedimentation velocity experiments *Prog. Colloid Polym. Sci.* **131**, 33-70 (2006).
45. Brookes, E. & Demeler, B. Parsimonious regularization using genetic algorithms applied to the analysis of analytical ultracentrifugation experiments *Proceedings of the 9th annual conference on Genetic and evolutionary computation* 361-368 (2007).

Chapter 3: High-throughput Screening to Identify Inhibitors of the BMI1-PHC2 Protein-Protein Interaction

This chapter contains proprietary data

A. Abstract

BMI1 has been identified as an oncogenic factor in many solid and blood cancers and is suggested to be a novel target for small molecule inhibitor development. BMI1 is best characterized as a member of the canonical polycomb repressive complex 1 where it serves as a scaffolding protein mediating multiple protein-protein interactions. Thus a challenge in inhibiting BMI1 is identifying an appropriate approach to target this protein. One opportunity is through disruption of the BMI1-PHC2 protein-protein interaction characterized in Chapter 2. Here we present the development of two biochemical assays monitoring this interaction that are suitable for high-throughput screening (HTS). These assays were employed in a HTS campaign with the University of Michigan Center for Chemical Genomics screening ~ 150,000 diverse small molecules. From these screening efforts we identified three classes of small molecules that bind directly to BMI1, paving the way towards development of inhibitors of the BMI1-PHC2 protein-protein interaction.

B. Background

B.1. Rationale for developing small molecule inhibitors of BMI1-PHC2 protein-protein interaction

As discussed in Chapter 1, polycomb proteins are important oncogenic factors in many cancers and as such have been identified as potential targets for therapy development.¹ BMI1 in particular represents a novel therapeutic target in many diseases. Previous studies using genetic inhibition demonstrated that knockdown of BMI1 reduces proliferation of many cancer cells including ex vivo models of gastric carcinoma,² hepatocellular carcinoma,³⁻⁵ breast cancer,⁶ colorectal cancer,⁷ multiple myeloma⁸ and acute myeloid leukemia.^{9,10} Further, loss of BMI1 reduces the tumorigenic potential of cancer initiating cells in models for acute myeloid leukemia^{9,10} and colorectal cancer⁷ demonstrating that targeting BMI1 may be a strategy to inhibit cancer cell populations that are resistant to current chemotherapeutics and that drive tumor relapse.¹¹

While these studies using BMI1 knockdown demonstrate the potential therapeutic benefit of inhibiting BMI1 they do not suggest the best strategy for targeting this protein. Indeed, the hematopoietic and skeletal abnormalities of BMI1 knockout mice demonstrate a critical role of BMI1 in regulating normal cellular processes,¹²⁻¹⁶ including normal hematopoiesis,⁹ demonstrate that targeted approaches to disrupt select BMI1 function are needed. As described in Chapter 2 BMI1 mediates multiple protein-protein interactions within the PRC1 complex providing multiple opportunities to selectively modulate BMI1 function. We further demonstrated in Chapter 2 that disruption of the BMI1-PHC2 interaction decreases cellular proliferation of U2OS cells. These data support a campaign to develop small molecule inhibitors of the BMI1 ULD by disrupting the BMI1-PHC2 protein-protein interaction as an approach to regulate BMI1 function.

B.2. Is the BMI1-PHC2 interaction a good target for disruption with small molecules?

We have provided a detailed structural and functional characterization of the BMI1-PHC2 interaction in Chapter 2. These studies provide valuable insight into the molecular basis for this interaction and suggest that this interaction is promising for inhibitor development for a number of reasons. First, the BMI1-PHC2 PPI is between the globular BMI1 ULD domain and a

peptide fragment of PHC2 which buries 1,338 Å² of accessible surface area (Figure 3.1). This category of PPI is recognized as amenable to orthosteric small molecule inhibition as there is one protein-protein interface and there may be potential small molecule binding pockets at the interface.¹⁷

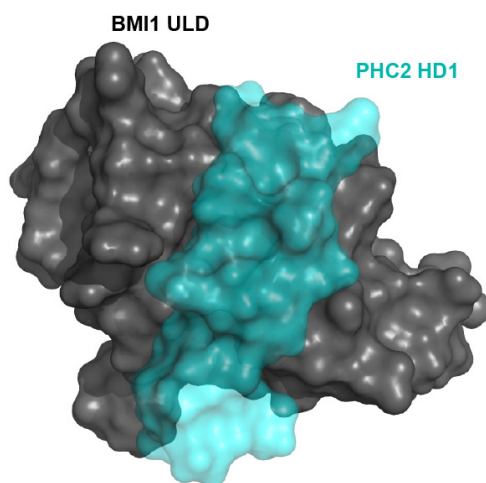


Figure 3.1. BMI1-PhC2 interface is an attractive target for small molecule inhibitors. Surface representation of BMI1-PhC2 complex.

Second, the BMI1-PhC2 interface is quite hydrophobic, a feature which generally supports high affinity small molecule ligand binding. The PhC2 β -hairpin encloses the hydrophobic core of BMI1 which forms a groove that may be amenable to small molecule binding (Figure 3.2). Third, with the exception of the BMI1 Arg165 and His174 residues which form critical polar interactions, the BMI1-PhC2 interface lacks charged residues (Figure 3.2). However, an intriguing aspect of the BMI1-PhC2 interaction is the presence of a buried glutamate in PhC2 (Glu45) which extends into a slightly basic surface formed by BMI1 Lys182. This feature suggests that molecules with acidic groups may appropriately mimic this interaction for potent inhibition. Finally, our solution NMR studies suggest that the apo BMI1 structure is perturbed upon binding of PhC2, indicating the presence of conformational dynamics for this domain. This is encouraging for inhibitor development as there may be protein conformations amenable to ligand binding that are not apparent in the crystal structure of the domain.

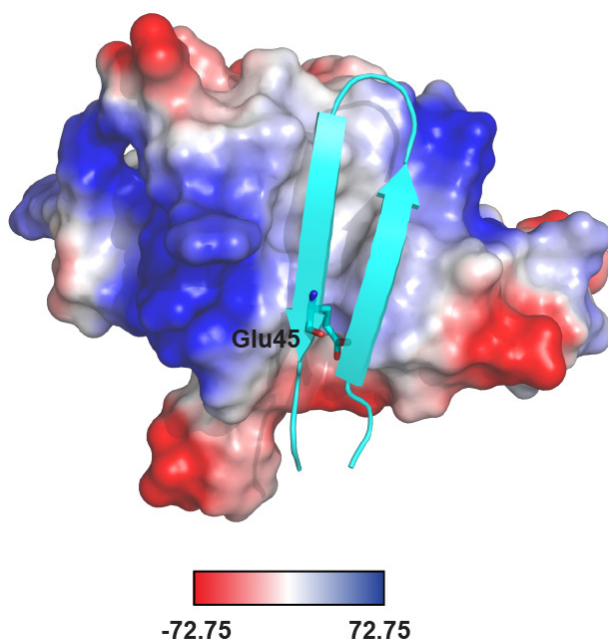


Figure 3.2. Features of BMI1-PHC2 interaction with relevance for drug discovery. Surface electrostatic potential representation of BMI1 with PHC2 shown as cyan cartoon and buried Glu45 labeled. The hydrophobic core of BMI1 is shown in white.

Together, this analysis suggests that the BMI1-PHC2 interaction is a favorable target for small molecule inhibitor development.

B.3. Biochemical assays and high-throughput screening strategy

Many recent reviews have highlighted the role of robust, solution phase biochemical assays as screening platforms for protein-protein interactions.¹⁸⁻²⁰ In a high-throughput screening (HTS) campaign these assays serve to quantify the protein-protein interaction in the presence of small molecules cataloged in compound libraries. PPI inhibitors are thus identified through change of assay signal compared to positive and negative control samples. To identify inhibitors of the BMI1-PHC2 interaction we designed two biochemical assays for the BMI1-PHC2 interaction that are amenable for HTS: fluorescence polarization and AlphaLISA.

Fluorescence polarization (FP) is a common bioassay for binary interactions.^{21,22} In this system one binding partner (protein, peptide, nucleotide or tracer molecule) is labeled with a

fluorophore which when free in solution tumbles rapidly and emission light is depolarized. Addition of a larger molecular weight binding partner slows this tumbling and emission light is polarized. The degree of light polarization is thus a direct readout of the interaction. This system has many advantages in a high-throughput screening setting: 1) the assay is not sensitive to environmental changes such as temperature and pH; 2) the degree of polarization is independent of dye concentration thus allowing good sensitivity with low concentrations of material; 3) modern instruments have very precise measurement of polarization thus FP assays tend to have good assay reproducibility and sensitivity and; 4) modern fluorophore labeling methods make development of the assay relatively straightforward and cost effective to develop and utilize.

To eliminate false-positive compounds that are potentially interfering with the primary screen we developed a secondary screening assay using an independent technology; the PerkinElmer AlphaLISA chemiluminescent assay.²³ In this assay tagged proteins of interest are mixed with antibody conjugated AlphaLISA beads that are themselves conjugated to light sensitive dyes. Excitation of the donor bead at 680 nm results in release of singlet oxygen and if the acceptor bead is within $\sim 1,000 \text{ \AA}$ the singlet oxygen is transferred to the acceptor bead resulting in AlphaLISA signal at 615 nm. Thus, AlphaLISA signal is a function of the proximity of the protein binding partners and disruption of this interaction results in signal decrease. While more sensitive to environmental conditions and assay interference by fluorescent compounds than the FP assay, positive aspects of the AlphaLISA system for high-throughput screening include: 1) homogeneous, no-wash set up which minimizes the number of assay steps leading to robust reproducibility; 2) large dynamic range from low bead concentrations therefore resulting in low background signal and; 3) applicability to low or moderate affinity interactions due to the conjugation of multiple antibodies to each donor or acceptor bead thus amplifying the signal.²⁴

Prior to investing in a HTS campaign it is necessary to evaluate the statistical robustness of proposed screening assays. To assess the quality of our assays we relied on three metrics: signal-to-noise (S/N), the coefficient of variance (CV)²⁵ and the Z-factor.²⁶ The S/N is calculated as the ratio of the raw signal of positive controls with all assay components to the signal of negative controls without protein targets. CV is a percentage measurement calculated by the ratio of the standard deviation to the mean and is interpreted as variation between measurements. Lower CVs thus signify lower variability. The final statistic reporting on assay robustness is the

Z-factor. This parameter is calculated from the S/N between positive and negative controls and the standard deviation of both controls (Equation 3.1). This term is reported from 0-1 and assays with Z-factors greater than 0.5 are considered acceptable for high-throughput screening.

$$\text{Z-factor} = 1 - \frac{3(\sigma_p - \sigma_n)}{|\mu_p - \mu_n|}$$

Equation 3.1. Z-factor calculation

Through application of these assays in an HTS campaign we aim to identify lead compounds that inhibit the BMI1-PHC2 interaction and can be optimized through medicinal chemistry to develop a quality chemical probe or clinical candidate.

C. Results

C.1. BMI1 -PHC2 Biochemical Assay Development

C.1.1 Fluorescence Polarization Assay

We developed a fluorescence-polarization assay to monitor the direct interaction between the BMI1 and PHC2 proteins in solution. A scheme of this assay is illustrated in Figure 3.3. In this system, recombinant BMI1 ULD was expressed with a thioredoxin (trx) tag. The trx tag was retained, essentially doubling the molecular weight of the protein, in order to achieve the greatest assay dynamic range.

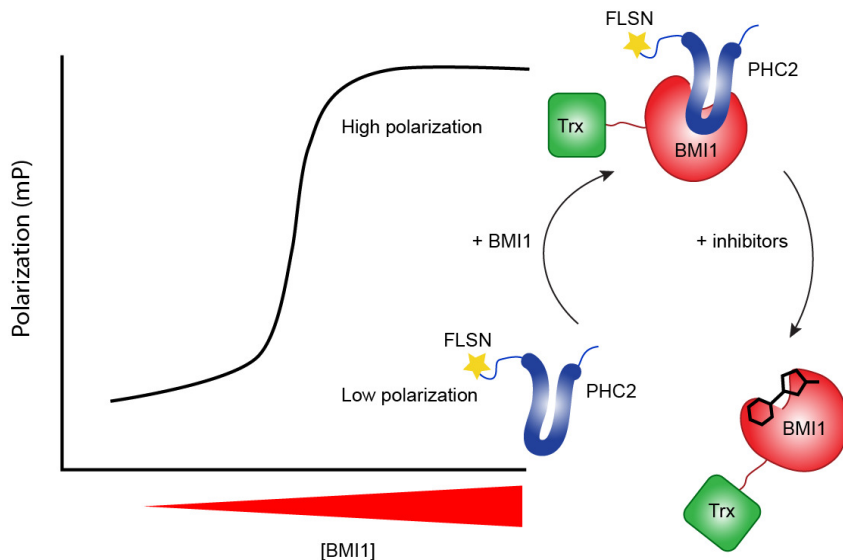


Figure 3.3. Scheme of fluorescence polarization assay for BMI1-PHC2 interaction.

In the absence of trx-BMI1 binding partner, FITC-PHC2 tumbles freely and emitted light is depolarized. Addition of trx-BMI1 causes dose-dependent increase in polarization. Competition of this interaction with small molecules subsequently decreases polarization.

First, we performed titration of BMI1 ULD constructs with the fluorescein-labelled PHC2₃₁₋₆₂ (FITC-PHC2) to determine the affinity of the interaction in this platform (Figure 3.4.A). The binding affinity of BMI1 ULD to PHC2 interaction in this assay is 216 nM, comparable to ITC measurements (Chapter 2). We also confirmed that the trx tag does not contribute to the affinity of the BMI1-PHC2 interaction but does increase the assay dynamic range by ~30% (Figure 3.4.A). For assay validation we performed a competition experiment titrating with untagged PHC2₃₂₋₆₁ competitor. The dose-dependent decrease in polarization following addition of the competitor (IC_{50} of 684 ± 106 nM), indicates that this assay is sensitive to disruption of the BMI1-PHC2 complex and is applicable for inhibitor screening (Figure 3.4.B).

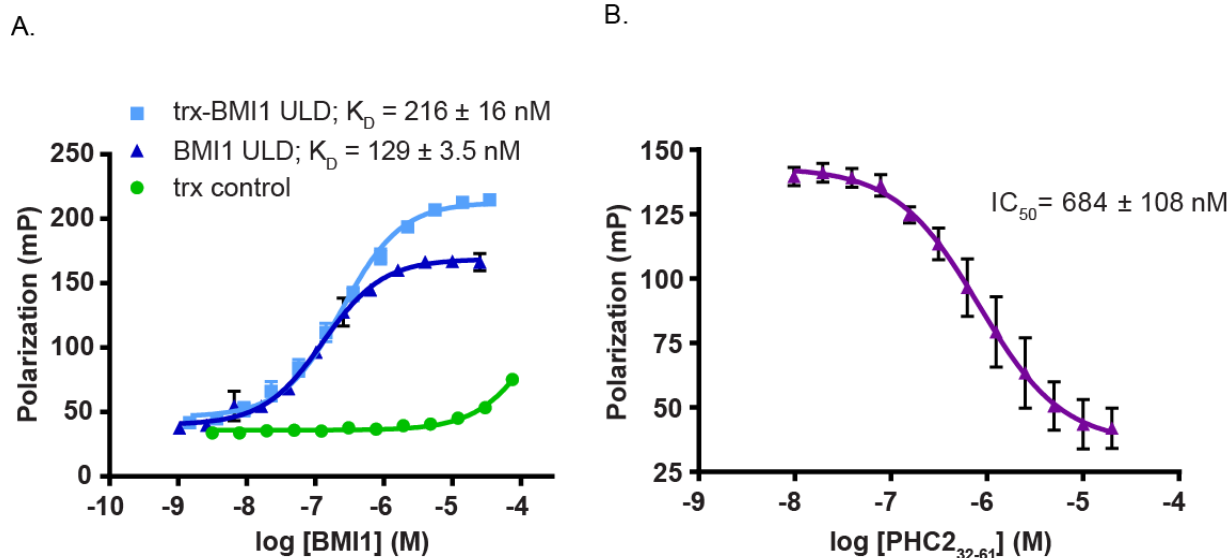


Figure 3.4. Development of fluorescence polarization assay for HTS.

A. Representative fluorescence polarization experiments titrating FITC-PHC2 with trx-tagged or untagged BMI1 ULD constructs or with trx-tag. B. Fluorescence polarization experiment titrating PHC2₃₂₋₆₁ competitor. In both graphs, error bars are standard deviation of two technical replicates. K_D and IC_{50} are reported as average and standard deviation from three independent experiments.

This assay was optimized in a low volume (15 μ L) format appropriate for high-throughput screening. We tested the performance and reproducibility of the assay using a multi-drop liquid dispenser and 384-well plate with multiple replicates of the positive control (100% inhibition- FITC-PHC2 alone) and of the negative control (no inhibition, trx- BMI1 ULD with FITC-PHC2). All controls contained a final DMSO concentration of 1.7%. This data were used to calculate the Z-factor (equation 3.1). Under optimized pre-HTS conditions the Z-factor was 0.83, indicative of a high quality assay. Compared to buffer controls, the assay had a 3-fold signal-to-noise, which is acceptable for fluorescence polarization assays. Statistics for this assay are summarized in Figure 3.5.A. Thus this fluorescence polarization assay is appropriate for high-throughput screening and we employed this as the primary assay in our HTS campaign (Figure 3.5.B).

A.

Signal-to-noise	3.2
Coefficient of variance (%)	1.7
Z-factor	0.83

B.

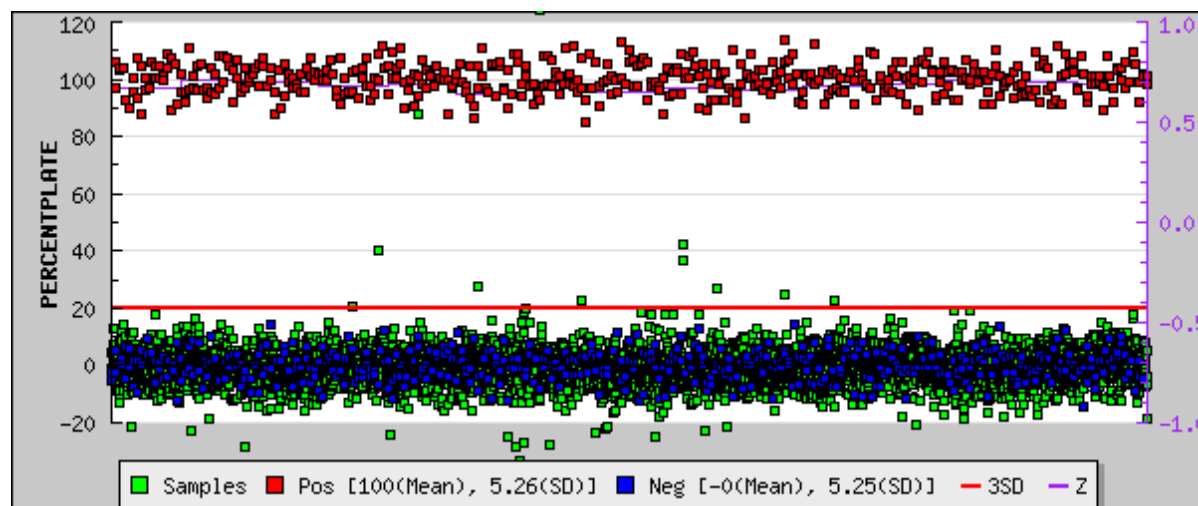


Figure 3.5. FP Assay Performance.

A. Pre-screening quality statistics for BMI1-PHC2 FP Assay. B. Scatter plot of FP assay data from one day of screening at the University of Michigan CCG (plates 37801-37815) showing positive controls (red), negative controls (blue) and compound wells (green). Z' is indicated in purple on the right axis.

C.1.2. AlphaLISA Assay Development

The AlphaLISA assay developed for the BMI1-PHC2 PPI was designed with N-terminally tagged His₆-trx-PHC2₁₋₇₉ (His-PHC2) and N-terminal FLAG tagged BMI1 ULD (FLAG-BMI1). When FLAG-BMI1 and His-PHC2 are in complex the Ni-chelate donor and anti-FLAG acceptor beads are in close proximity and maximum signal is achieved. However, inhibition of the FLAG-BMI1- His-PHC2 interaction by small molecules or protein competitors causes a dose-dependent decrease in Alpha signal. This assay is illustrated in Figure 3.6.

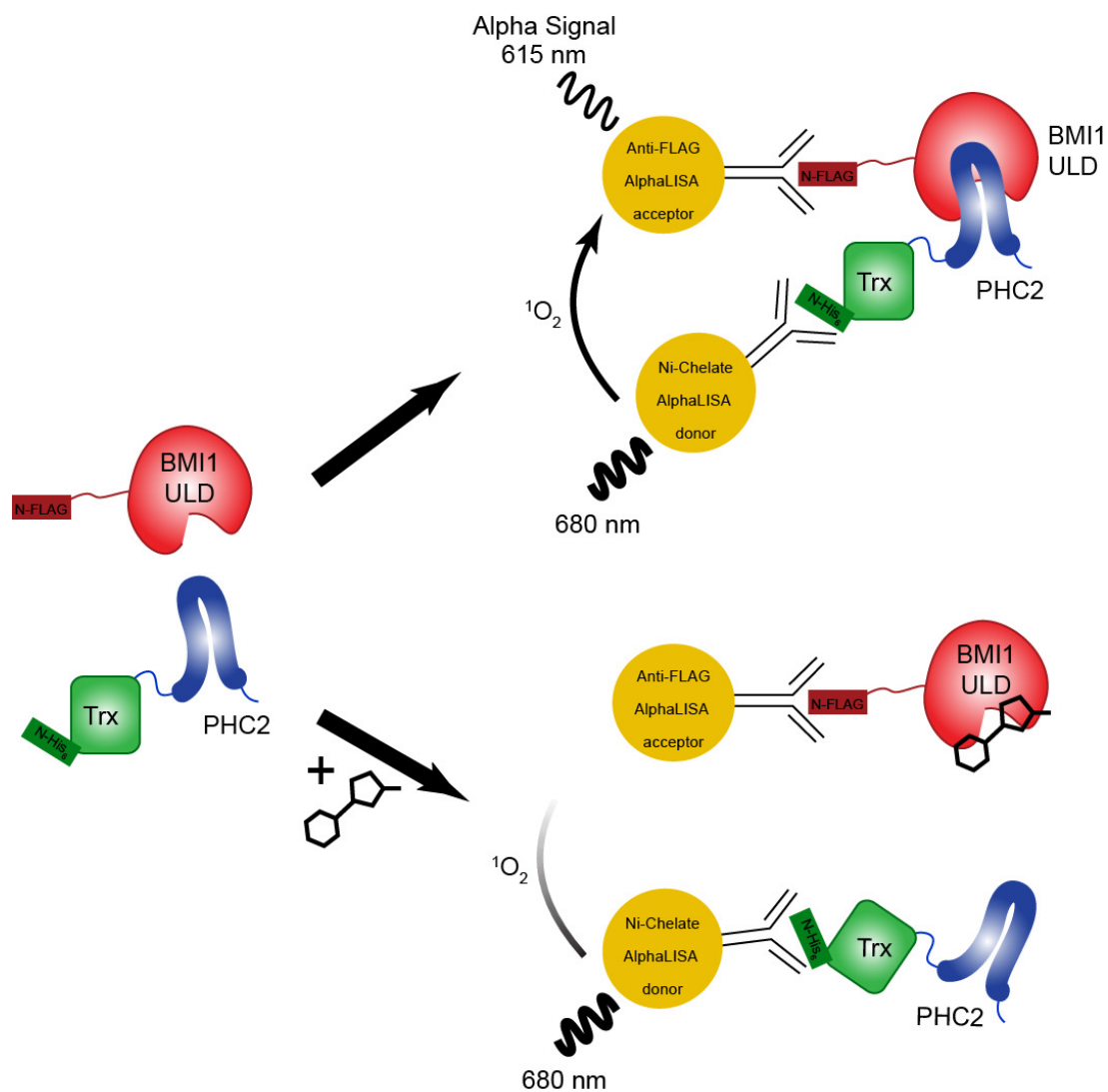


Figure 3.6. Scheme of AlphaLISA assay for BMI1-PhC2 interaction.

To validate this assay we tested competition with both untagged BMI1 ULD and with PhC2₃₂₋₆₁ peptide. Addition of these competitors to mixtures reduced Alpha signal in a dose-dependent manner where $IC_{50} = 354 \pm 77$ nM and 404 ± 132 nM for BMI1 ULD and PhC2 peptide respectively (Figure 3.7). This assay was optimized in a miniaturized 384-well format and the Z' for the assay was determined to be 0.87 (positive control: FLAG-BMI1 + anti-FLAG acceptor beads; negative control: all protein and bead components). All controls contained 1.7% DMSO. Compared to buffer controls, the assay had a signal-to-noise of 16. Statistics for this assay are summarized in Figure 3.8.A.

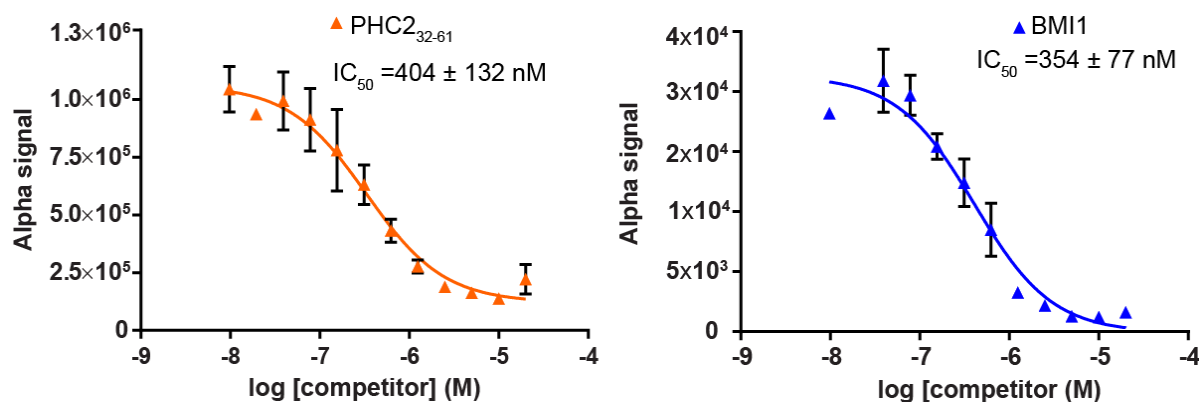


Figure 3.7. Competition experiments of AlphaLISA assay.

Representative AlphaLISA experiment titrating untagged PHC2₃₂₋₆₁ (left) or BMI1 (right) competitor. Error bars are standard deviation of two technical replicates and IC₅₀ is reported as average and standard deviation from two independent experiments performed in duplicate.

A.

Signal-to-noise	16
Coefficient of variance (%)	6.9
Z'	0.87

B.

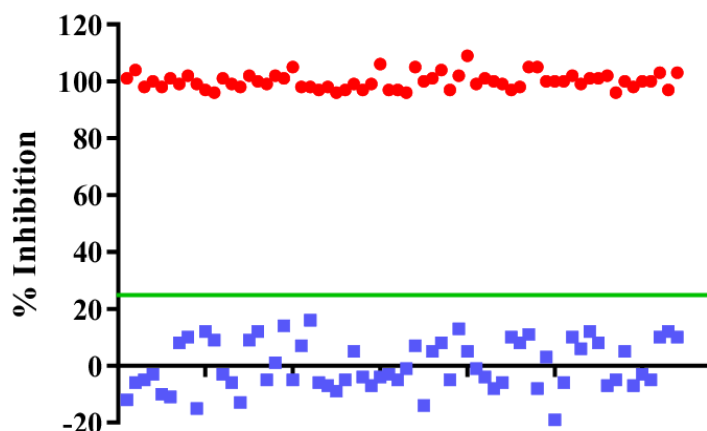


Figure 3.8. Quality assessment of BMI1-PHC2 AlphaLISA Assay

A. Pre-screening quality statistics for BMI1-PHC2 AlphaLISA Assay. B. Scatter plot of FP assay data from control plate showing positive controls (red) and negative controls (blue).

Based on the assays and metrics described we performed high-throughput screening to identify small molecule inhibitors of the BMI1-PHC2 interaction. HTS had three stages: primary screening with single point measurements, confirmation screening retesting active compounds in triplicate and dose-response screening with compound titration performed in two orthogonal

assays. Subsequently, active compounds were evaluated for direct protein binding, compound mechanism of action was assessed by NMR spectroscopy and mass spectrometry and limited SAR-by-catalog was performed.

C.2. Primary Screening

We performed HTS at the Center for Chemical Genomics (CCG) at the University of Michigan. For screening we selected commercial libraries from Chembridge, ChemDiv and Maybridge with most compounds meeting standard molecular weight, solubility and hydrogen bonding requirements for lead-like compounds. Some small libraries within the collection contained known pharmacologically active or FDA approved drugs. In total, 147,943 compounds were screened in the primary screen. Compounds were screened at an approximate concentration of 33 μ M with 1.7% DMSO delivered by pintool.

We defined hits as those with activity greater than 2.5 standard deviations from the mean of the negative control or had at least 12.5% inhibition. This selection criteria allowed us to identify even weak inhibitors of the interaction which could be used as scaffolds for future medicinal chemistry efforts to design more potent inhibitors. From this selection we removed compounds with molecular weight greater than 700 Da, with evident fluorescence interference and promiscuous activity in other biochemical screens performed at the CCG. Applying this criteria we identified 1,500 active compounds giving a ~1% overall hit ratio for the primary screen. These active compounds were then subjected to confirmation screening.

C.3. Confirmation Screening

To validate activity of hits identified in primary screening we tested the 1,500 active compounds in triplicate at 33 μ M with the FP assay. Active compounds were selected based on activity in both the primary and confirmation screen and we defined hits as those with greater than 12.5% inhibition or 2.5 times the standard deviation of the negative controls for at least 3 points. Applying this criteria and filtering out compounds with undesirable reactive groups a total of 360 compounds were selected for dose response, a 24% confirmation rate. Figure 3.9 shows activity correlation between the primary and confirmation screens.

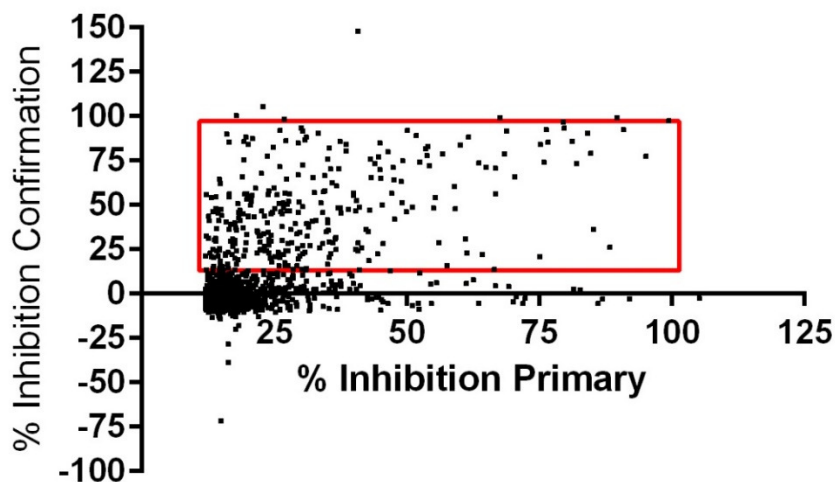


Figure 3.9. Comparing compound activity in confirmation screening. Correlation plot of compound activity in primary screen vs. confirmation screen. Compounds selected for dose-response screening are boxed in red box.

C.4. Dose Response Measurements

To test dose-dependent activity of the 360 selected compounds and determine IC_{50} values titration experiments were performed in an 8-point, 2-fold dilution series starting at 100 μM concentration, followed by sigmoidal curve fitting analysis. Active compounds were defined as those with IC_{50} less than 100 μM , minimal evidence of compound aggregation, precipitation or other assay interference. Applying these criteria, 180 compounds were selected for dose-dependent screening using the secondary AlphaLISA assay described above. In a similar titration experiment, 154 compounds (85%) showed dose-dependent activity. Active compounds were manually inspected and 60 compounds were selected for in-lab follow-up using fresh powder ordered directly from the vendor. A breakdown of the overall HTS screening campaign is shown in Figure 3.10.

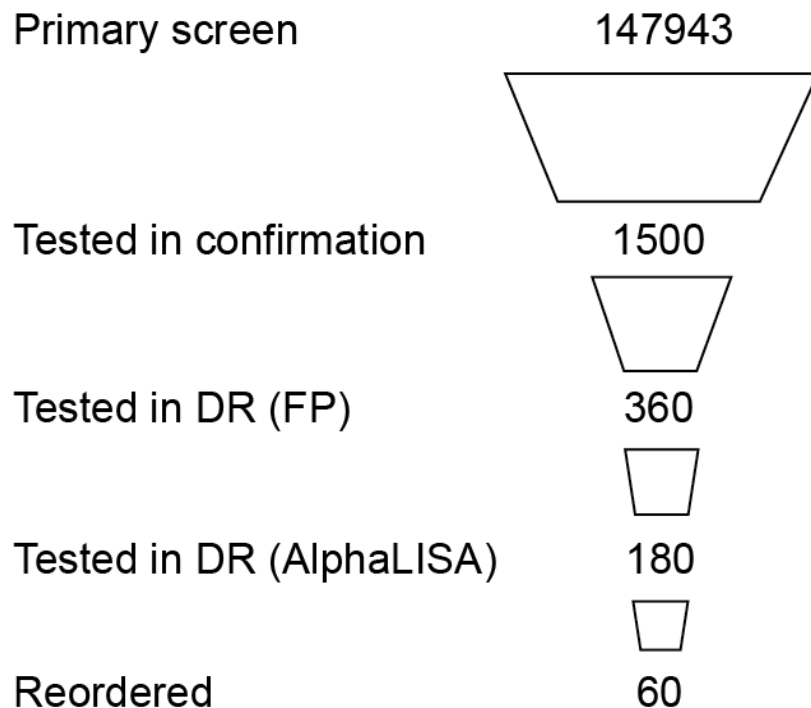


Figure 3.10. Summary of HTS.

D.R. = dose response.

C.5. Characterization of active compounds

Following the HTS campaign we selected 60 active compounds for follow up in in-lab experiments including testing freshly ordered powder in biochemical assays and validating direct binding to BMI1 by NMR spectroscopy. Through these efforts, we identified three classes of compounds (**BI-1**; **BI-2** and **BI-3**) that inhibit the BMI1-PHC2 interaction (Figure 3.11, Table 3.1). These fresh compounds exhibited low to high micromolar inhibition in the FP assay. Generally, the activity of fresh compounds did not correlate well with activity in the HTS campaign, suggesting that the compound age affected the potency of the inhibitors. However, these molecules are promising starting points for the development of optimized inhibitors.

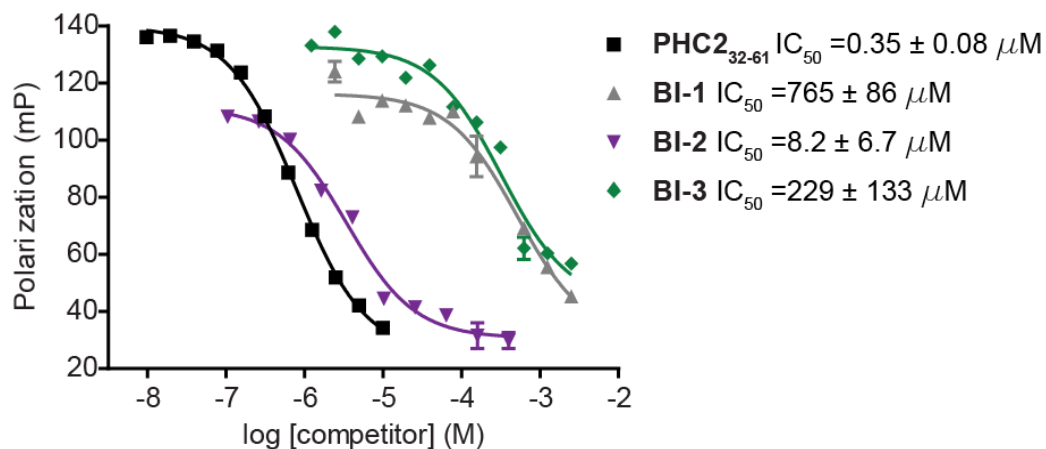


Figure 3.11. Validation of HTS hits from fresh powder.

Fluorescence polarization assay with trx-BMI1 and FITC-PHC2 titrated with active compounds from HTS that were reordered from commercial sources. Representative experiments are shown and error bars are the standard deviation of two technical replicates; IC_{50} is reported as average and standard deviation from at least two independent experiments.

Table 3.1. Confirmed hits from HTS campaign that bind directly to BMI1 ULD.

Compound ID	Structure	% Activity in Primary Screen	IC_{50} DRC1; FP Assay (μ M)	IC_{50} DRC2; AlphaLISA assay (μ M)	IC_{50} (Fresh); FP Assay (μ M)
BI-1-1 (CCG-14273)		21	11	60	765 ± 83
BI-2 (CCG-100325)		17	170	98	8.2 ± 6.7
BI-3 (CCG-104954)		23	16	0.77	229 ± 133

C5.1. Class 1 BMI1-PHC2 inhibitors

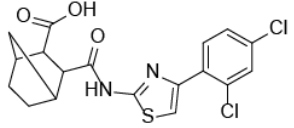
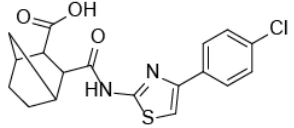
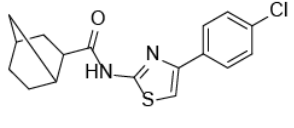
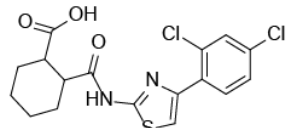
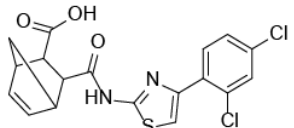
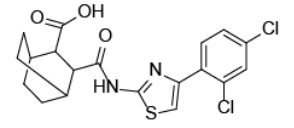
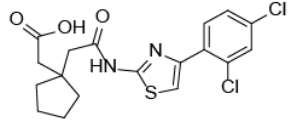
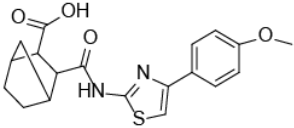
C.5.1.1. Structure Activity Relationship

Class 1 inhibitors feature an amino-thiazole core. The parent compound (**BI-1-1**) has high micromolar activity in the FP assay ($IC_{50} = 765 \pm 83 \mu M$) and structure-activity relationship studies were pursued through commercially available analogs. All analogs were tested for activity in the fluorescence polarization assay (Table 3.2).

The weak inhibition of **BI-1-2** and **BI-1-8** in the fluorescence polarization assay demonstrates the criticality of the 1,3-dichloro substitution on the phenyl ring. **BI-1-2** lacks chloro substitution at the 1 position and **BI-1-8** is mono-substituted at the 3 position with a methoxy group. From the greater than 2-fold reduced activity for these compounds we concluded that the hydrophobic chloro groups contribute greatly to compound activity and substitution with other similar groups may help drive potency. Compounds lacking the acidic moiety, **BI-1-3** and **BI-1-9**, were not soluble in buffer even at moderate concentrations (as determined by NMR). Therefore the role of this functionality in the activity of these compounds cannot be empirically determined.

Compounds **BI-1-4**, **BI-1-5**, **BI-1-6** and **BI-1-7** test the role of the rigid bicyclic group. **BI-1-4** which features a cyclohexyl group, and therefore has considerably greater conformational flexibility, does not have significantly different activity than the parent compound ($IC_{50} = 765 \mu M$ vs $734 \mu M$). **BI-1-5** has an unsaturated cyclohexene moiety which improves potency compared to a saturated cyclohexyl in **BI-1-1**, suggesting that ring conformation is not critical for potency. Interestingly, **BI-1-6** and **BI-1-7** which also have greater conformational flexibility than **BI-1-1** due to the addition of the extra carbon in the bicyclic ring (**BI-1-6**) or the five-membered ring with exocyclic aliphatic chain (**BI-1-7**), both have significantly improved activities ($IC_{50} = 37$ and $219 \mu M$, respectively) relative to the parent compound. The SAR reported here suggests that increased hydrophobicity in this region of the molecule improves potency and that ring conformation is not the most important considerations for this motif. The most potent compound, **BI-1-6** is 20-times more potent than the parent, demonstrating that future SAR studies may further improve these inhibitors.

Table 3.2. Structure-activity-relationship for BI-1 series of inhibitors.

Compound ID	Structure	IC ₅₀ (Fluorescence Polarization) (μM)
BI-1-1		765 ± 83
BI-1-2		1922 ± 976
BI-1-3		Not soluble
BI-1-4		739 ± 420
BI-1-5		296 ± 158
BI-1-6		37 ± 27
BI-1-7		219 ± 80
BI-1-8		No inhibition

C.5.1.2. Compound mechanism of action by NMR

To test direct binding of active compounds from HTS to BMI1 ULD we turned to NMR spectroscopy. Through the structural studies in Chapter 2, 82% of visible resonances for backbone amides of BMI1 ULD were assigned using triple resonance experiments. Many residues with missing backbone amide assignment are in the PHC2 binding site; in particular the β 1 strand which forms one side of the binding site is unassigned. However, there are 5 methyl-containing amino acids in the PHC2 binding site thus methyl groups can serve as probes for this site in the absence of complete backbone amide assignment. We used ^1H - ^{15}N and ^1H - ^{13}C -HSQC binding experiments to rapidly assess of direct binding of compounds to BMI1 ULD as well as to localize the compound binding site on the protein.

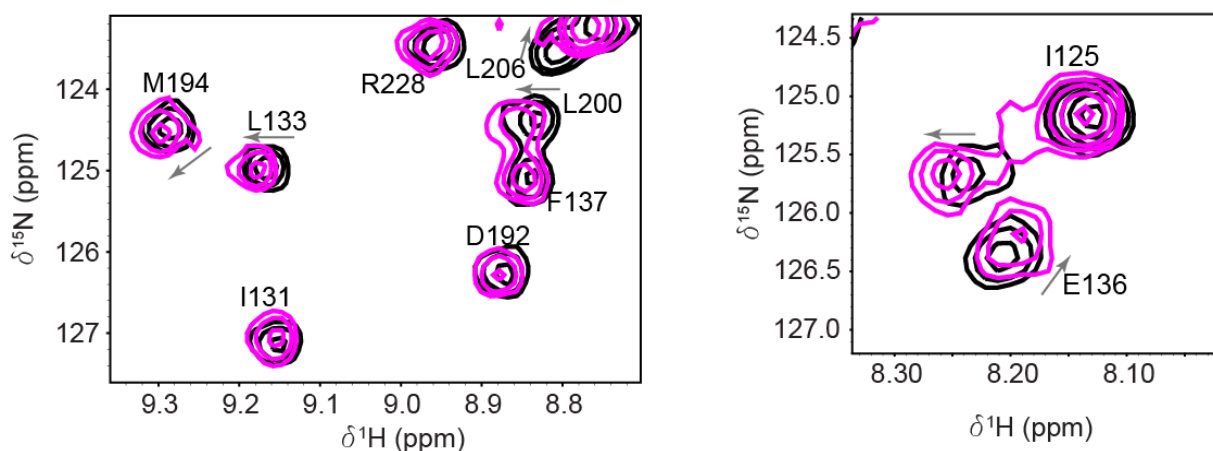


Figure 3.12. BI-1-1 binds directly to BMI1 ULD.

Selections of ^1H - ^{15}N HSQC spectra of apo BMI1 ULD (black) or with 100 μM BI-1-1 (magenta). Resonance assignments are labeled and directions of chemical shift perturbation are shown with arrows.

The ^1H - ^{15}N HSQC spectrum of BMI1 in the presence of **BI-1-1** showed a distinct set of chemical shift perturbations as illustrated in Figure 3.12. All active analogs in the BI-1 series showed similar chemical shift perturbations regardless of their activity in the fluorescence polarization assay. This led us to speculate that backbone amides in the protein may not be very

sensitive to compound binding in this case and that methyl groups might be more informative. Therefore ^1H - ^{13}C HSQC spectra were recorded for BMI1 in complex with **BI-1-7**. In this spectra, two leucine and methionine resonances are affected with compound binding (Figure 3.13.A).

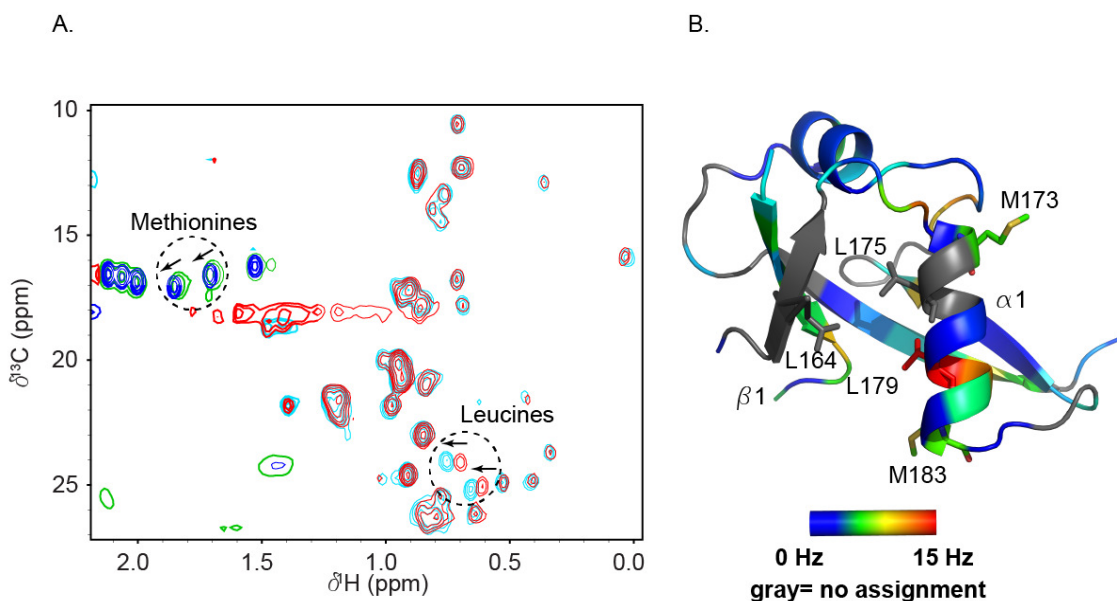


Figure 3.13. BI-1-7 binding site mapping by NMR spectroscopy.

A: Selection of ^1H - ^{13}C HSQC spectrum of apo BMI1 ULD (red/green) or with 200 μM BI-1-7 (dark blue/cyan). Resonances in the leucine and methionine regions of the spectrum that are affected are circled and direction of chemical shift perturbation are shown with arrows. B: Magnitude of chemical shift perturbations from ^1H - ^{15}N HSQC binding experiment with 0.5 mM equivalents of BI-1-7 are mapped onto the BMI1 ULD structure. The PHC2 binding site is between the $\beta 1$ and $\alpha 1$ structural elements. Methionine and leucine residues in the PHC2 binding site are labeled. The ^1H - ^{13}C NMR experiments were performed by Jon Pollock.

Using the amide backbone assignment of the BMI1 ULD we mapped the ^1H - ^{15}N HSQC chemical shift perturbations from **BI-1-7** onto the BMI1 ULD structure. This demonstrates that the BI-1 series compounds bind to the PHC2 binding site, as illustrated in Figure 3.13.B. Additionally, there are two methionines and three leucine residues in the PHC2 binding site, which may correspond to the resonances perturbed in ^1H - ^{13}C HSQC binding experiments. Together the data from two NMR methods demonstrates direct binding of this class of inhibitors to BMI1 ULD and suggests an orthostatic mechanism of inhibition through binding to the PHC2 binding site in BMI1.

C.5.2. Class 2 BMI1-PHC2 inhibitors

The **BI-2** class of BMI1 inhibitors features a benzamide core with a methylsulfone-substituted pyrimidine tail group. **BI-2** has low micromolar activity in the fluorescence polarization assay ($IC_{50} = 8.2 \pm 6.7 \mu M$). Eighteen **BI-2** analogs were purchased from commercial sources and most had similar low to mid micromolar IC_{50} (experiments performed by Jon Pollock, data not shown). NMR binding experiments with **BI-2** show significant chemical shift perturbations in slow exchange regime (Figure 3.14.A). Mapping of these perturbations onto the BMI1 ULD structure show the significant changes at the C-terminus of the protein (Figure 3.14.B). The strong effects of these compounds on BMI1 ULD spectra hinted that these compounds may be covalently modifying the protein.

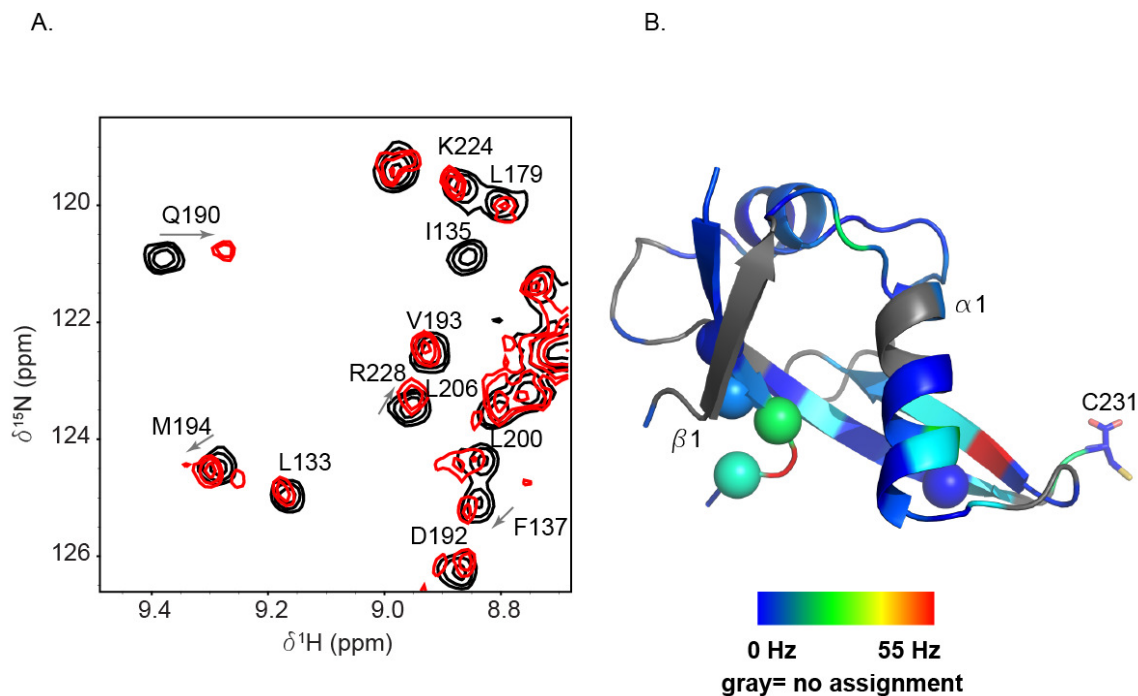


Figure 3.14. BI-2 binding site mapping by NMR spectroscopy.

A: Selection of 1H - ^{15}N HSQC spectrum of apo BMI1 ULD (black) or with 100 μM BI-2 (red). Perturbed residues are labeled and directions of chemical shift perturbation are shown with arrows. B: Magnitude of chemical shift perturbations from the experiment on the left are mapped onto the BMI1 ULD structure. Residues that are broadened with compound addition are shown as spheres. The covalently modified cysteine residue is labelled.

Mass spectrometry studies by my colleague Jon Pollock indeed showed covalent modification of BMI1 by this class of molecules with the addition of 394 Da in molecular weight corresponding to single site modification. We suspected that this molecule may be covalently modifying the protein at a cysteine residue. BMI1 has two cysteines, one is solvent-exposed at the C-terminus of the protein (Cys231) and the other is in the BMI1 binding site (Cys166). To identify which site was being modified we generated Cys231 to alanine mutant protein. By mass spectrometry and NMR there was little to no modification of the mutant protein suggesting that the C-terminal cysteine was the primary target (data not shown). Investigation into the mechanism of action for these compounds indicates that this reaction proceeds via nucleophilic aromatic substitution with the cysteine substituting for the methylsulfone leaving group in a reaction recently characterized by Guan et al.²⁷ Indeed, analogs lacking the methylsulfone pyrimidine did not show any inhibition in the FP assay (data not shown).

C.5.3. Class 3 BMI1-PHC2 inhibitors

The class 3 BMI1-PHC2 inhibitor is a cyclopentaquinoline fragment and compound has mid-micromolar activity ($IC_{50} = 229 \pm 133 \mu\text{M}$; Table 3.1). ^1H - ^{15}N HSQC NMR binding experiments suggest that this compound weakly interacts with the BMI1 ULD (Figure 3.15.A). Chemical shift mapping onto the BMI1 structure illustrates only very weak chemical shift perturbations with no localized cluster of perturbed resonances (Figure 3.15.B).

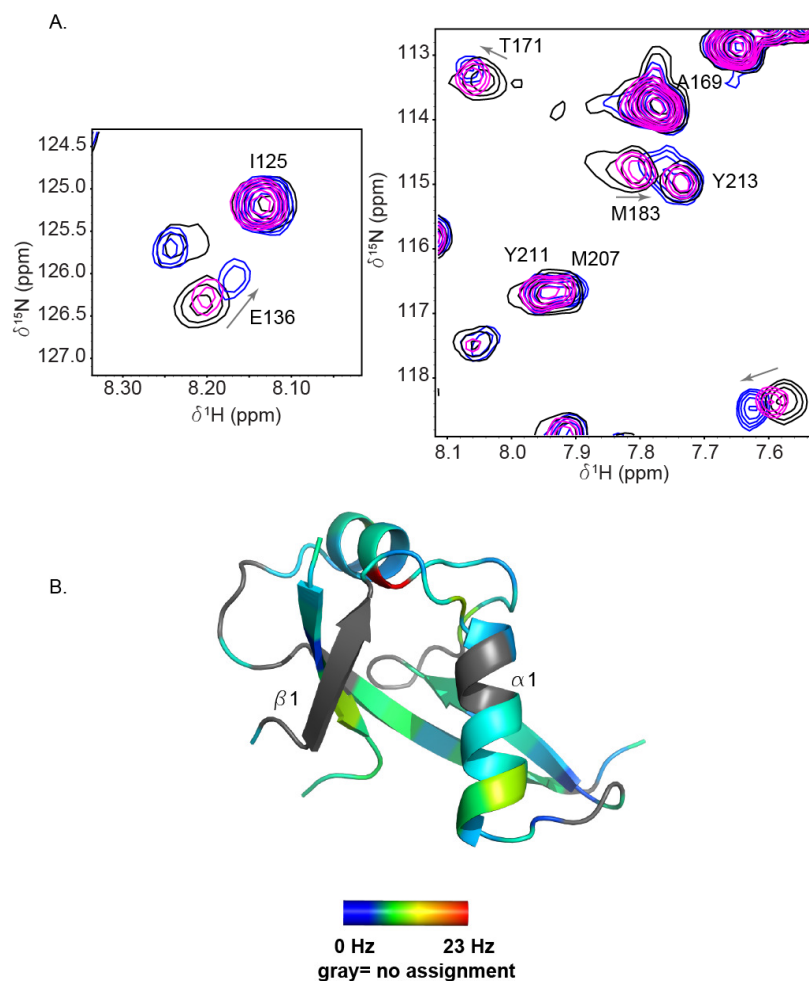


Figure 3.15. BI-3 binding site mapping by NMR spectroscopy.

A: Selections of ^1H - ^{15}N HSQC spectrum of apo BMI1 ULD (black) or with 0.5 mM or 2 mM equivalents of **BI-3** (magenta or dark blue, respectively). Perturbed residues are labeled and directions of chemical shift perturbation are shown with arrows. B: Magnitude of chemical shift perturbations with 0.5 mM of **BI-3** are mapped onto the BMI1 ULD structure. The PHC2 binding site is between the $\beta 1$ and $\alpha 1$ structural elements.

While the **BI-3** molecule had only weak activity, we are encouraged by the structural features of this molecule which may mimic the contacts between BMI1 and PHC2. The hydrophobic ring system may bind in the hydrophobic core of BMI1 ULD and similar to the **BI-1** series, the tail carboxylic acid may replace the buried PHC2 Glu45. Further, the direct interaction of **BI-3** with BMI1 observed in NMR experiments and inhibitory activity in the FP assay is promising as it suggests that BMI1 can bind low-molecular weight fragment-like

ligands. This supports future directions for the identification of BMI1 inhibitors using fragment-based approaches.

D. Discussion and Conclusions

In this Chapter we present our efforts to identify small molecule inhibitors of the BMI1-PHC2 interaction through high-throughput screening. We developed two orthogonal screening assays using the fluorescence polarization and AlphaLISA platforms, each of which showed good performance in pre-HTS testing with Z-factors greater than 0.8. Using the fluorescence polarization assay we screened almost 150,000 compounds from chemically diverse libraries. This assay had a hit rate of 1% which allowed follow up of even weak inhibitors in an effort to obtain chemical scaffolds that could be improved through medicinal chemistry. Use of the AlphaLISA assay in dose-response screening allowed the elimination of many false positives and the identification of the most promising compounds for follow up with fresh powder.

Through these efforts we identified three classes of small molecules that disrupt the BMI1-PHC2 PPI with micromolar activity and bind directly to BMI1 ULD. These compounds represent three different approaches to inhibiting this interaction. Structure-activity-relationship studies of the **BI-1** class of compounds identified **BI-1-6** which inhibits the BMI1-PHC2 *in vitro* interaction with an IC_{50} of 37 μ M, the most potent orthosteric inhibitor of the BMI1-PHC2 PPI identified thus far. NMR-based mapping studies suggest that this class of compounds bind to the BMI1 ULD domain at the PHC2 interface, demonstrating that this interface is amenable to small molecule binding. This series has a promising scaffold and future SAR studies may develop this series into potent inhibitors. Proposed modifications could test alkyl substitution of the phenyl ring, introduction of other bicyclic groups or replacement of the acid with different polar moieties such as sulfonamide or ester. Additionally, substantial scaffold modifications could be tested to determine the minimal pharmacophore and optimize the next set of ligands. Either crystallographic or solution structural studies with BMI1 ULD in complex with **BI-1** inhibitors could greatly enhance these SAR efforts.

The **BI-2** class disrupts the BMI1-PHC2 interaction through covalent modification of a C-terminal cysteine suggesting the presence of an allosteric regulatory network between distal sites of the protein and the PHC2 binding site. The **BI-3** class is a fragment-like ligand and direct

binding of this molecule to BMI1 ULD indicates that a fragment-based drug discovery approach yield other promising inhibitors for BMI1.

Acknowledgements

I am grateful for the technical support of Martha Larsen and Steve Vander Roest at the University of Michigan Center for Chemical Genomics during high-throughput screening. Additionally, we acknowledge the University of Michigan Cancer Center for funding this work.

E. Experimental methods

Plasmid construction.

cDNA encoding full-length BMI1 and full-length PHC2_B was obtained from the Raul Lab (Department of Pathology, University of Michigan) and the desired constructs were subcloned into the pET32p bacterial expression vector using standard molecular biology techniques. The pET32p vector is a modified pET32a vector (Novagen) that contains a Precision Protease recognition sequence between a thioredoxin (trx) solubility tag and the target protein.

Protein expression and purification

BMI1 constructs: All BMI1 proteins were expressed in Codon+ BL21 (DE3) *E. coli* cells (Sigma) with an N-terminal His₆- thioredoxin tag (trx). Cells subjected to ampicillin (100 µg/mL) selection were grown in Luria broth (Fisher) or labeled M9 medium (Marley 2001) at 37°C (220 rpm) until OD₆₀₀ reached 0.6-0.8. Incubation temperature was lowered to 18°C for 1 hour then protein over-expression was induced with 0.5 mM (final concentration) Isopropyl β-D-1-thiogalactopyranoside (IPTG; Gold Bio) for 16-18 hours at 18°C. Cells were lysed in a buffer containing 50 mM tris, pH 7.5 at 25°C, 250 mM NaCl, 20 mM Imidazole, 0.5 mM phenylmethylsulfonyl fluoride (PMSF), 1 mM β-mercaptoethanol (β-ME) using a French Press. Clarified lysate was applied to a HisTrap HP (GE Healthcare) column using Äkta Prime FPLC and eluted with lysis buffer containing 0.5 M imidazole. To remove the His₆- trx tag, the protein was cleaved with Precision protease and BMI1 ULD constructs were further purified using cation exchange chromatography. Purified protein solution was buffer exchanged into storage

buffer (100 mM bis tris, pH 6.5, 50 mM NaCl, 1 mM tris(2-carboxyethyl)phosphine hydrochloride (TCEP-HCl)) using HiPrep Desalting Column (GE Healthcare). Protein concentration was measured using absorbance at 280 nm. Protein identity and purity was verified by sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE). Protein was flash frozen and stored at -80°C.

¹⁵N-labeled BMI1 for NMR Spectroscopy: Protein was expressed as above with the exception of the use of the M9 minimal media supplemented with ¹⁵N ammonium sulfate (Cambridge Isotopes) and ¹⁵N Bioexpress (Cambridge Isotopes).²⁸

Table 3.3. Recipe for M9 minimal media for expression of ¹⁵N-labeled BMI1 ULD.

Reagent	Final Concentration
Sodium phosphate (dibasic)	28 mM
Potassium phosphate (monobasic)	14.7 mM
Sodium chloride	8.5 mM
Sodium sulfate	3 mM
Biotin	1 mg/L
Thiamin	1 mg/L
Magnesium sulfate	1 mM
Calcium chloride	3 mM
¹⁵ N ammonium sulfate	7.5 mM (1 g/L)
Glucose	10 g/L

PHC2₁₋₇₉. PHC2₁₋₇₉ protein was expressed in BL21 (DE3) *E. coli* cells (Sigma) with an N-terminal His₆-thioredoxin tag. Cells were subjected to ampicillin (100 µg/mL) selection and grown in Luria broth (Fisher) at 37°C (220 rpm) until OD₆₀₀ reached 0.6-0.8. Incubation temperature was lowered to 18°C for 1 hour then protein over expression was induced with 0.5 mM (final concentration) Isopropyl β-D-1-thiogalactopyranoside (IPTG; Gold Bio) for 16-18 hours at 18°C. Cells were lysed in a buffer containing 50 mM tris, pH 7.5 at 25°C, 250 mM NaCl, 20 mM imidazole, 0.5mM PMSF, 1 mM β-ME using a French Press. Clarified lysate was applied to a HisTrap HP (GE Healthcare) column using Äkta Prime FPLC and eluted with lysis buffer containing 0.5 mM imidazole. Trx-PHC2₁₋₇₉ protein was further purified by gel filtration on a Superdex S-75 column in buffer 50 mM phosphate, pH 6.5, 50 mM NaCl, 1 mM TCEP-HCl. PHC2 was further purified by Superdex S75 gel filtration (GE Healthcare) pre-equilibrated with storage buffer (50 mM phosphate, pH 6.5, 50 mM NaCl, 1 mM TCEP). Protein concentration was measured using absorbance at 280 nm. Protein identity and purity was verified

by sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE). Protein was flash frozen and stored at -80°C.

Table 3.4. Summary of high-throughput screening campaign.

Category	Parameter	Description
Primary Assay	Type of Assay	<i>In vitro</i> fluorescence polarization (FP) competition assay
	Target	BMI1 ULD (inhibition of the BMI1-PHC2 interaction)
	Primary measurement	Detection of fluorescence polarization signal
	Key reagents	His-trx-BMI1, FITC-PHC2 (PHC2 ₃₂₋₆₁) peptide, FP buffer: 100mM bis tris, 50mM NaCl, 1mM TCEP, 0.01% BSA, 0.025% tween-20, pH 7.25
	Assay protocol	Described in methods section
Library	Library size	147,943
	Library composition	Maybridge HitFinder, Chembridge custom collection of small molecules, ChemDiv custom collection of small molecules, Library of Pharmacologically Active Compounds (LOPAC- Sigma Aldrich), Prestwick Chemical Library, BioFocus NIH Clinical Collection, MicroSource Spectrum Collection
	Source	Center for Chemical Genomics (CCG), University of Michigan
Screen	Format	384-well, Corning 3676
	Concentration tested	33 μM, 1.7% DMSO
	Plate controls	NC: His-trx BMI1, FITC-PHC2; PC: FITC-PHC2
	Reagent/compound dispensing system	Biomek FX, Beckman
	Detection instrument and software	PERAstar, BMG
	Assay validation/QZ	Z'=0.83; Mean FP for NC=129 mP, SD=8 mP; Mean FP for PC=35 mP, SD=5 mP

	Normalization	$\% \text{ inhibition} = 100 \times (\text{average of FP for negative control} - \text{FP for sample}) / (\text{average of FP for negative control} - \text{average of FP for positive control})$
Post-HTS analysis	Hit criteria	Change in FP signal ≥ 2.5 standard deviations from mean of negative control OR 12.5% active
	Hit rate	1%
	Additional assay(s)	AlphaLISA with FLAG-BMI1 and His ₆ -trx-PHC ₂₁₋₇₉ , NMR ¹ H- ¹⁵ N HSQC to validate direct binding of compounds to BMI1 ULD
	Confirmation of hit purity and structure	Compounds were repurchased from Maybridge, Chembridge and ChemDiv and verified analytically

Fluorescence Polarization Assay: Development and High-Throughput Screening

Fluorescence Polarization Assay Development

Fluorescence Polarization Assay to Determine BMI1-PHC2 Affinity: The fluorescein-labeled peptide probe (FITC-PHC2) (FITC-PQILTHVIEGFVIQEGAEPFPVGRSSLLVGN- NH₂) was purchased from Genscript. The probe was dissolved in storage buffer (100 mM bis tris, pH 6.5, 50 mM NaCl, 1 mM TCEP) to a concentration of 1 mM and stored at -20°C. For the assay, the probe was diluted to 80 nM in FP assay buffer (100 mM bis tris, pH 7.25, 50 mM NaCl, 1 mM TCEP, 0.01% bovine serum albumin, 0.025% (v/v) Tween-20). Then 10 µL of probe solution was added to a series of 30 µL solutions of varying concentrations of His₆-trx BMI1 in FP assay buffer (100 mM bis tris, pH 7.25, 50 mM NaCl, 1 mM TCEP, 0.01% bovine serum albumin, 0.025% (v/v) Tween-20) to obtain a final concentration of up to 25 µM. Samples were incubated for 1 hour at room temperature in the dark before 15 µL samples were plated onto assay plate (Corning #3676). Changes in fluorescence polarization and anisotropy were measured at 525 nm after excitation at 495 nm using PHERAstar microplate reader (BMG). Data fit with a sigmoidal dose-response equation were used to assess BMI1 ULD construct binding with the Prism 4.0 (GraphPad) program.

Fluorescence Polarization Competition Assay with PHC2 peptide or compound: FITC-PHC2 was mixed with trx-BMI1 at 40 nM and 1 µM, respectively, in FP assay buffer (100 mM bis tris, pH 7.25, 50 mM NaCl, 1 mM TCEP, 0.01% bovine serum albumin, 0.025% (v/v) Tween-20). This protein-probe mix was then mixed 1:1 with competitor stocks of varying

concentrations for a final probe concentration of 20 nM and final protein concentration of 0.5 μ M and 5% DMSO in all reactions. Samples were incubated for 1 hour at room temperature in the dark before 15 μ L samples in duplicate were plated onto assay plate (Corning #3676).

Fluorescence polarization data was recorded and analyzed as above. Data fit with a one-site competition equation were used to assess inhibition with the Prism 4.0 (GraphPad) program.

Fluorescence Polarization Assay in High-Throughput Screening

Fluorescence polarization high-throughput primary and confirmation screening assay: 10 μ L of 0.75 μ M His₆-trx BMI1 in FP assay buffer (100 mM bis tris, pH 7.25, 50 mM NaCl, 1 mM TCEP, 0.01% bovine serum albumin, 0.025% (v/v) Tween-20) was dispensed to columns 1-22 of 384-well assay plate (Corning 3676). Reagents were dispensed with a multidrop liquid dispenser (Thermo Scientific). 10 μ L FP assay buffer was dispensed to columns 23, 24 of the plate.

Compounds were added to columns 3-22 and DMSO was added to columns 1, 2, 23 and 24 of the plate using Biomek FX dual head (Beckman) robot equipped with a pintool. 5 μ L of 60 nM FITC-PHC2 in FP assay buffer was added to all wells of the plate giving a final probe concentration of 20 nM and final protein concentration of 500 nM in a 15 μ L reaction. Plates were incubated for 1 hour at room temperature in the dark. The degree of fluorescence polarization was measured by a PHERAstar microplate reader (BMG).

Dose-response assay: 200 nL of compound with variable concentrations were pre-plated onto assay plates using a Mosquito X1 robot (TTP Labtech). His₆-trx BMI1 and FITC-PHC2 were diluted into FP assay buffer to a final concentration of 500 nM and 20 nM respectively. 15 μ L of the protein-probe mixture was dispensed to columns 1-22 of the assay plate. 15 μ L of 20 nM FITC-PHC2 in assay buffer was dispensed to columns 23-24 of the assay plate. Incubation and measurement was performed as described above.

AlphaLISA Assay: Development and High-Throughput Screening

AlphaLISA Assay Development

AlphaLISA Assay Development: FLAG-BMI1 and His₆-PHC2 were mixed varying ratios and concentrations up to 1.5 μ M in AlphaLISA assay buffer (100 mM bis tris, pH 7.25, 50 mM NaCl, 1 mM TCEP, 0.01% bovine serum albumin, 0.025% (v/v) Tween-20). AlphaLISA Ni-chelate donor beads and anti-FLAG acceptor beads (PerkinElmer) were each diluted 1:1,000 (v/v) in AlphaLISA buffer. 7.5 μ L of protein mixture was mixed with 7.5 μ L bead mixture in duplicates on PerkinElmer 384-well ProxiPlate PLUS for final protein concentrations up to 0.75 μ M and

final bead concentrations of 1:2,000 (v/v). All pipetting with beads was done in the dark. Reactions were incubated for 1 hour at room temperature in the dark. Alpha signal (europium emission at 607-623nm) was measured as with the PHERAStar plate reader (BMG). Data fit with a sigmoidal dose-response equation were used to quantify the PPI by the Prism 4.0 (GraphPad) program.

AlphaLISA Assay with BMI1 protein or PHC2₃₂₋₆₁ Peptide Competition: FLAG-BMI1 and His₆-PHC2 were mixed 1:1 at 1 μ M final concentration in AlphaLISA assay buffer (100 mM bis tris, pH 7.25, 50 mM NaCl, 1 mM TCEP, 0.01% bovine serum albumin, 0.025% (v/v) Tween-20). Protein mixture was incubated at room temperature for 30 minutes. AlphaLISA Ni-chelate donor beads and anti-FLAG acceptor beads (PerkinElmer) were each diluted 1:1,000 (v/v) in protein mixture. 40 μ L of protein-bead mixture was incubated with 40 μ L of varying concentrations of the PHC2₃₂₋₆₁ peptide or BMI1 protein to obtain concentrations of up to 20 μ M. Reactions were incubated for 1 hour at room temperature in the dark. Duplicates of 15 μ L per reaction were plated onto PerkinElmer 384-well ProxiPlate PLUS plate. Assay plates were quantified using excitation at 680 nm and europium emission at 615 nm as measured by Pherastar plate reader (BMG). Data were fit with a one-site competition equation were used to assess protein or peptide competition with the Prism 4.0 (GraphPad) program.

AlphaLISA High-Throughput Screening

High-throughput dose-response assay: 200 nL of compound with variable concentrations were pre-plated onto assay plates with a Mosquito X1 robot (TTP Labtech). FLAG- BMI1 and His-PHC2 were diluted into AlphaLISA buffer (100 mM bis tris, pH 7.25, 50 mM NaCl, 1 mM TCEP, 0.01% bovine serum albumin, 0.025% (v/v) Tween-20) to a final concentration of 500 nM each. Proteins were incubated on ice for 30 minutes. AlphaLISA Ni-chelate donor beads and anti-FLAG acceptor beads (PerkinElmer) were each diluted 1:2,000 (v/v) final concentration into protein mixture, protein-bead mixture was incubated on ice for 15 minutes. 15 μ L of the protein-bead mixture was dispensed to columns 1-22 of the assay plate (PerkinElmer 384-well ProxiPlate PLUS) using multidrop liquid dispenser (Thermo Scientific). 15 μ L of 500 nM FLAG-BMI1 with 1:2,000 (v/v) anti-FLAG AlphaLISA acceptor beads in assay buffer was dispensed to columns 23-24 of the assay plate. Plates were incubated for 1 hour at room temperature in the dark. Alpha signal (europium emission at 607-623 nm) was measured as by Envision plate reader (Perkin Elmer).

NMR experiments

All NMR experiments were performed on a Bruker Advance III 600-MHz spectrometer equipped with a 5 mM TCI cryogenic probe. All spectra were acquired at 30°C. Spectra were processed with NMRPipe²⁹ and analyzed with Sparky.³⁰

Small molecule binding experiments. To test direct binding of compounds to BMI1 ULD, samples were prepared with 50 μ M final ¹⁵N BMI1 concentration in NMR buffer (100 mM bis tris, 50 mM NaCl at pH 6.5, 1 mM TCEP) with 50, 100, or 500 μ M final compound concentration in 5% final DMSO. All samples contained 7% D₂O.

F. References

1. Cao, L., Bombard, J., Cintron, K., Sheedy, J., Weetall, M. L. & Davis, T. W. BMI1 as a novel target for drug discovery in cancer. *J Cell Biochem* **112**, 2729-2741 (2011).
2. Xiao, J. & Deng, C. Knockdown of Bmi-1 impairs growth and invasiveness of human gastric carcinoma cells. *Oncol Res* **17**, 613-620 (2009).
3. Chiba, T., Miyagi, S., Saraya, A., Aoki, R., Seki, A., Morita, Y., Yonemitsu, Y., Yokosuka, O., Taniguchi, H., Nakauchi, H. & Iwama, A. The polycomb gene product BMI1 contributes to the maintenance of tumor-initiating side population cells in hepatocellular carcinoma. *Cancer Res* **68**, 7742-7749 (2008).
4. Ruan, Z. P., Xu, R., Lv, Y., Tian, T., Wang, W. J., Guo, H. & Nan, K. J. Bmi1 knockdown inhibits hepatocarcinogenesis. *Int J Oncol* **42**, 261-268 (2013).
5. Qi, S., Li, B., Yang, T., Liu, Y., Cao, S., He, X., Zhang, P., Li, L. & Xu, C. Validation of Bmi1 as a therapeutic target of hepatocellular carcinoma in mice. *Int J Mol Sci* **15**, 20004-20021 (2014).
6. Douglas, D., Hsu, J. H., Hung, L., Cooper, A., Abdueva, D., van Doorninck, J., Peng, G., Shimada, H., Triche, T. J. & Lawlor, E. R. BMI-1 promotes ewing sarcoma tumorigenicity independent of CDKN2A repression. *Cancer Res* **68**, 6507-6515 (2008).
7. Kreso, A., van Galen, P., Pedley, N. M., Lima-Fernandes, E., Frelin, C., Davis, T., Cao, L., Baiazitov, R., Du, W., Sydorenko, N., Moon, Y. C., Gibson, L., Wang, Y., Leung, C., Iscove, N. N., Arrowsmith, C. H., Szentgyorgyi, E., Gallinger, S., Dick, J. E. & O'Brien, C. A. Self-renewal as a therapeutic target in human colorectal cancer. *Nat Med* **20**, 29-36 (2014).
8. Jagani, Z., Wiederschain, D., Loo, A., He, D., Mosher, R., Fordjour, P., Monahan, J., Morrissey, M., Yao, Y. M., Lengauer, C., Warmuth, M., Sellers, W. R. & Dorsch, M. The Polycomb group protein Bmi-1 is essential for the growth of multiple myeloma cells. *Cancer Res* **70**, 5528-5538 (2010).
9. Rizo, A., Olthof, S., Han, L., Vellenga, E., de Haan, G. & Schuringa, J. J. Repression of BMI1 in normal and leukemic human CD34(+) cells impairs self-renewal and induces apoptosis. *Blood* **114**, 1498-1505 (2009).
10. Lessard, J. & Sauvageau, G. Bmi-1 determines the proliferative capacity of normal and leukaemic stem cells. *Nature* **423**, 255-260 (2003).

11. Beck, B. & Blanpain, C. Unravelling cancer stem cell potential. *Nat Rev Cancer* **13**, 727-738 (2013).
12. Zencak, D., Lingbeek, M., Kostic, C., Tekaya, M., Tanger, E., Hornfeld, D., Jaquet, M., Munier, F. L., Schorderet, D. F., van Lohuizen, M. & Arsenijevic, Y. Bmi1 loss produces an increase in astroglial cells and a decrease in neural stem cell population and proliferation. *J Neurosci* **25**, 5774-5783 (2005).
13. Leung, C., Lingbeek, M., Shakhova, O., Liu, J., Tanger, E., Saremaslani, P., Van Lohuizen, M. & Marino, S. Bmi1 is essential for cerebellar development and is overexpressed in human medulloblastomas. *Nature* **428**, 337-341 (2004).
14. Lessard, J., Schumacher, A., Thorsteinsdottir, U., van Lohuizen, M., Magnuson, T. & Sauvageau, G. Functional antagonism of the Polycomb-Group genes *eed* and *Bmi1* in hemopoietic cell proliferation. *Genes Dev* **13**, 2691-2703 (1999).
15. van der Lugt, N. M., Domen, J., Linders, K., van Roon, M., Robanus-Maandag, E., te Riele, H., van der Valk, M., Deschamps, J., Sofroniew, M., van Lohuizen, M. & et al. Posterior transformation, neurological abnormalities, and severe hematopoietic defects in mice with a targeted deletion of the *bmi-1* proto-oncogene. *Genes Dev* **8**, 757-769 (1994).
16. Park, I. K., Qian, D., Kiel, M., Becker, M. W., Pihalja, M., Weissman, I. L., Morrison, S. J. & Clarke, M. F. Bmi-1 is required for maintenance of adult self-renewing haematopoietic stem cells. *Nature* **423**, 302-305 (2003).
17. Cierpicki, T. & Grembecka, J. Targeting protein-protein interactions in hematologic malignancies: still a challenge or a great opportunity for future therapies? *Immunol Rev* **263**, 279-301 (2015).
18. Nero, T. L., Morton, C. J., Holien, J. K., Wielens, J. & Parker, M. W. Oncogenic protein interfaces: small molecules, big challenges. *Nat Rev Cancer* **14**, 248-262 (2014).
19. Arkin, M. R., Tang, Y. & Wells, J. A. Small-molecule inhibitors of protein-protein interactions: progressing toward the reality. *Chem Biol* **21**, 1102-1114 (2014).
20. Gul, S. & Hadian, K. Protein-protein interaction modulator drug discovery: past efforts and future opportunities using a rich source of low- and high-throughput screening assays. *Expert Opin Drug Discov* **9**, 1393-1404 (2014).
21. Owicki, J. C. Fluorescence polarization and anisotropy in high throughput screening: perspectives and primer. *J Biomol Screen* **5**, 297-306 (2000).
22. Lea, W. A. & Simeonov, A. Fluorescence polarization assays in small molecule screening. *Expert Opin Drug Discov* **6**, 17-32 (2011).
23. Bielefeld-Sevigny, M. AlphaLISA immunoassay platform- the "no-wash" high-throughput alternative to ELISA. *Assay Drug Dev Technol* **7**, 90-92 (2009).
24. Eglen, R. M., Reisine, T., Roby, P., Rouleau, N., Illy, C., Bosse, R. & Bielefeld, M. The use of AlphaScreen technology in HTS: current status. *Curr Chem Genomics* **1**, 2-10 (2008).
25. Reed, G. F., Lynn, F. & Meade, B. D. Use of coefficient of variation in assessing variability of quantitative assays. *Clin Diagn Lab Immunol* **9**, 1235-1239 (2002).
26. Zhang, J. H., Chung, T. D. & 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).
27. Guan, Y., Wang, C., Wang, D., Dang, G., Chen, C., Zhou, H. & Zhao, X. Methylsulfone as a leaving group for synthesis of hyperbranched poly(arylene pyrimidine ether)s by nucleophilic aromatic substitution. *RSC Advances* **5**, 12821-12823 (2015).

28. Marley, J., Lu, M. & Bracken, C. A method for efficient isotopic labeling of recombinant proteins. *J Biomol NMR* **20**, 71-75 (2001).
29. Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J. & Bax, A. NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J Biomol NMR* **6**, 277-293 (1995).
30. Goddard, T. D. & Kneller, D. G. *SPARKY 3* (University of California, San Francisco).

Chapter 4: Development of Small Molecule Inhibitors of Ring1B/BMI1 E3 Ubiquitin Ligase

This chapter contains proprietary data

A. Abstract

The polycomb proteins BMI1 and Ring1B have been identified as important oncogenic factors in many cancers. Genetic studies demonstrate that knockdown of these proteins slows or abolishes cancer cell proliferation and tumor growth in a number of tumor types indicating that they may be promising targets for small molecule therapeutic development. BMI1 and Ring1B function as a heterodimeric E3 ubiquitin ligase for nucleosomal histone H2A, monoubiquitinating lysine 119- a chromatin modification associated with repressed gene transcription. Here we present a strategy to modulate the enzymatic activity of this complex through the development of small molecule inhibitors of the Ring1B/BMI1 E3 ubiquitin ligase. To identify ligands directly binding to the Ring1B/BMI1 RING domain heterodimer we performed a NMR-based fragment screen. The most potent hit from this screen showed specific binding to the complex and weak (mM) inhibition of E3 ligase activity. Through an extensive medicinal chemistry campaign we optimized the small molecule fragment ligand into to a potent inhibitor with low micromolar affinity and *in vitro* inhibitory activity of nucleosome ubiquitination by Ring1B/BMI1. Based on NMR chemical shift mapping and mutagenesis studies we identified the ligand binding site as the nucleosome binding interface of the Ring1B protein. Binding of the inhibitors results in significant conformational change in Ring1B. As a consequence, the ligand-bound conformation is incompatible with nucleosome substrate binding preventing H2A ubiquitination. In cellular studies potent Ring1B/BMI1 inhibitors decrease

global ubiquitinated H2A levels. Overall, we have developed a highly novel inhibitor for the polycomb E3 ubiquitin ligase complex which can be used as a chemical tool to explore the role of these proteins in normal and cancer biology.

B. Background

Chapter 1 details the role of BMI1 in cancer and describes studies linking Ring1B/BMI1-mediated H2A ubiquitination with transcriptional repression of key tumor suppressor genes.¹⁻³ It has been shown that increased levels of ubiquitinated H2A is correlated with poor prognosis for patients with pancreatic cancer.⁴ Further, genetic knockdown studies demonstrate that loss of BMI1 leads to reduced tumor growth in mouse models of many solid and hematologic malignancies.⁵⁻¹⁵ Inhibiting the ubiquitin ligase activity of Ring1B/BMI1 E3 ligase activity may therefore be a promising approach for therapeutic intervention.

B.2. Mechanism of protein ubiquitination

Protein ubiquitination is a signaling mechanism in many cellular pathways including protein degradation, nuclear transport, endocytosis, gene expression and DNA repair.¹⁶ Ubiquitin is a small, 8.5 kilo Dalton, protein that can be post-translationally attached to substrate proteins through the formation of an isopeptide bond between the C-terminus of ubiquitin and a target lysine. Mechanistically, the ubiquitination cascade begins with ubiquitin activation by the E1 enzyme which becomes covalently attached to ubiquitin at a catalytic cysteine through an ATP-dependent step (Figure 4.1). Ubiquitin is transferred from the E1 to a receptor cysteine on the E2 ubiquitin conjugating enzyme via a transthioesterification reaction.¹⁷ Ubiquitin is further transferred to substrate through interaction of E2 enzymes with E3 proteins which mediate substrate recognition. In the human genome there are only 9 E1 enzymes and 35 E2 enzymes, whereas there are hundreds of E3 enzymes;¹⁸ thus substrate specificity is achieved at the level of the E3.^{19,20}

There are two structural families of E3 enzymes, the HECT domain family and the RING domain family which have unique mechanisms of transferring ubiquitin to substrate. The smaller class of E3s, the HECT family, accepts the ubiquitin molecule from the E2 and then transfers the ubiquitin to substrate.^{21,22} The larger class of E3s, the RING family has hundreds of family members. This class of E3s does not have ubiquitin transferred directly to them. Instead these

proteins serve as scaffolding proteins mediating the interaction between E2 and substrate. Structural studies have revealed that through distal contacts with the substrate RING E3s help to position the E2 catalytic cysteine above the substrate lysine facilitating formation of the isopeptide bond.²³⁻²⁵ Mechanistic studies suggest that RING E3s may also allosterically activate the E2 enzyme^{26,27} through contacts with the ubiquitin molecule to rearrange the E2 active site for activation of the substrate lysine.^{28,29}

The overall ubiquitin ligase cascade is shown in Figure 4.1.

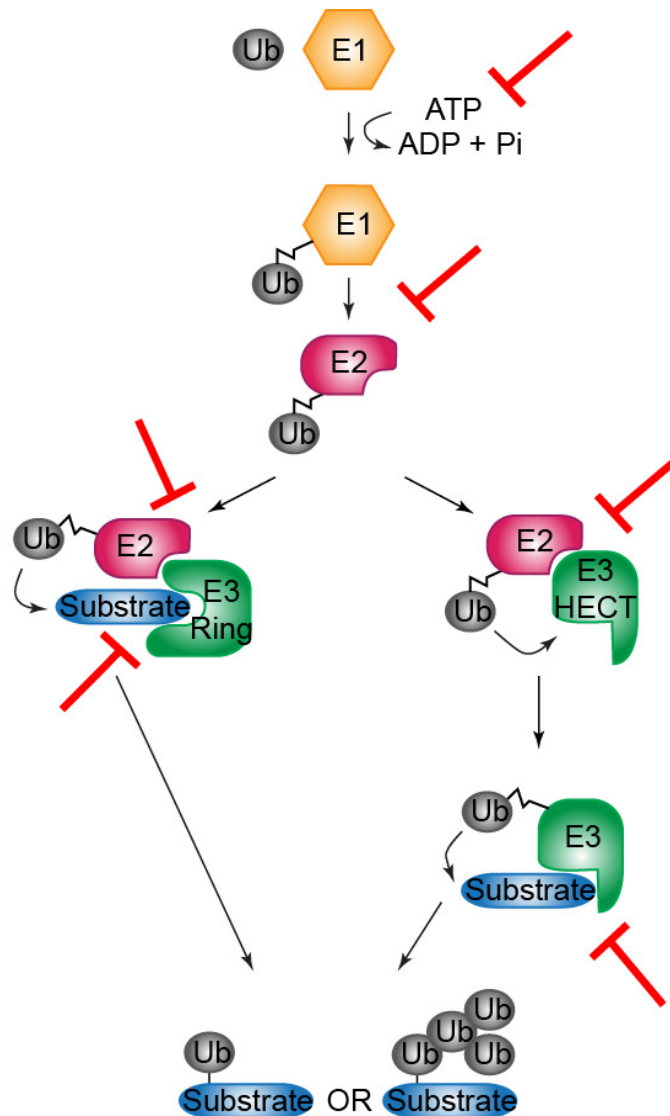


Figure 4.1. Overview of ubiquitination cascade.

Potential sites for small molecule inhibition are shown in red. Adapted from Di Fiore, P.P. et al *Nat. Rev. Mol. Cell Biol.* **2003**, *4*, 491-497.

B.3. Ring1B/BMI1 E3 ligase structure and substrate specificity

Within the PRC1 complex the Ring1B and BMI1 proteins each have N-terminal zinc-finger RING domains.^{30,31} Each domain coordinates two zinc molecules through C3HC4 coordination. The Ring1B RING domain is an active E3 ligase by itself, however the activity is enhanced through interaction with the BMI1 RING domain.³¹ The reverse is not true; the BMI1 RING domain has no intrinsic E3 ligase activity³¹ and the role of this domain is likely to stabilize Ring1B, maintain ubiquitin in a transfer-ready conformation^{28,29} and facilitate substrate recognition.

The Ring1B/BMI1 E3 ligase has a very narrow substrate profile; the canonical substrate is histone H2A lysine 119 in the context of the nucleosome. Ring1B/BMI1 lacks activity towards isolated histone octamer or H2A protein or peptide.³² The structural basis for this selectivity was revealed by McGinty and coworkers in a recently published crystal structure by of the Ring1B/BMI1-E2 “polycomb ubiquitination module” in complex with the nucleosome.²³ These studies identified the major E3-substrate interface as localized to a highly basic region on Ring1B and the H2A/H2B acidic patch. The “nucleosome binding loop” in Ring1B, from residues 86-98, features 8 arginine or lysine residues. Key to nucleosome binding is the “arginine anchor,” Arg98, which extends into an acidic pocket formed by H2A residues Glu61, Asp90 and Glu92. Additionally, Ring1B Lys97 interacts with H2A residues Glu61 and Glu64. These two Ring1B residues contribute significantly to nucleosome binding as single alanine mutations at either position reduces nucleosome binding affinity by at least 50 fold. Additional interactions between BMI1 and H3 and H4 histones and between the E2 (UbcH5c) and nucleosomal DNA further contribute to substrate binding. Through these structural studies it becomes clear that Ring1B/BMI1 would be unable to catalyze ubiquitination of histone octamer, isolated histone protein or peptide.

There is suggestion in the literature that the topoisomerase Top α ,³³ p53^{10,34,35} and the ribonucleotide reductase M1³⁶ may also be Ring1B/BMI1 substrates, however these are not well validated and direct binding between the E3 complex and these substrates has not been demonstrated.

B.4. E3 ligases as targets for small molecule inhibitors

B.4.1. Strategies to identify inhibitors of E3 ubiquitin ligases

Given the large number of cellular processes regulated by protein ubiquitination there has been significant interest in developing small molecule inhibitors of this pathway as potential therapeutics for a wide range of diseases. The ubiquitin cascade provides many avenues for inhibitor development targeting either enzymatic activity or protein-protein interactions (Figure 4.1).^{37,38} As there are very few E1 and E2 isozymes in the human genome E1 and E2 family members participate in many different ubiquitin pathways including the ubiquitin proteasome system.³⁹ Therefore, small molecule inhibitors of these enzymes have non-specific cellular activity and can be considered general proteasome inhibitors. Thus, from a drug discovery standpoint targeting the E3 enzyme represents the most promising avenue for inhibiting ubiquitination in a specific cellular pathway.

A challenge to identifying inhibitors of E3 ligases is the complexity of the biochemical process outlined above. Due to the multitude of enzymes, cofactors and protein-protein interactions required for this system, an activity-based screening campaign detecting ubiquitinated substrate would likely identify inhibitors of other steps in the cascade. Deconvolution of such hits to identify specific E3 inhibitors would be lengthy and likely unfruitful.

As RING domain E3s are not true enzymes and serve as scaffolding proteins, the tactic to inhibit these factors is through disruption of protein-protein interactions. Thus screening assays quantifying protein-protein interactions are required to identify inhibitors of these proteins in a high-throughput fashion. Förster resonance energy transfer (FRET) and fluorescence polarization (FP) assays have both been developed for E3 ubiquitin ligases.^{40,41} FRET-based assays sensitive to E3-ubiquitin, E3-substrate and ubiquitin-substrate interactions or the formation of poly-ubiquitin chains are representative approaches for these systems.⁴²⁻⁴⁵ These assays have been applied as screening platforms for inhibitors of p53 ubiquitination by MDM2,⁴⁶ and TRAF6 catalyzed poly-ubiquitin chain formation.⁴⁷ Similarly, a fluorescence polarization assay was used to screen inhibitors of the interaction between the SCF complex protein (Skp1-Cdc53/Cullin-F-box, E3 ligase complex) Cdc4 and a fluorescently-labeled peptide substrate.⁴⁸

B.4.2. Challenges identifying inhibitors of Ring1B/BMI1

While the technologies discussed above represent unique approaches to identify E3 ligase inhibitors, these assays are not fully applicable to the Ring1B/BMI1 system as: i) nucleosomal substrate is required for ligase activity, production of which may be limiting for HTS; ii) the interaction between Ring1B/BMI1 – isolated H2A protein is weak and has not been detected in direct binding experiments precluding adaption of substrate peptide-based assays such as FP (Cierpicki lab, unpublished); iii) given that the Ring1B/BMI1 heterodimer only catalyzes mono-ubiquitination assays detecting the creation of poly-ubiquitin chains are not applicable. In order to overcome these limitations of high-throughput screening systems and identify specific inhibitors of the Ring1B/BMI1 E3 ligase we turned to a fragment-based drug discovery approach.

B.4.3. Fragment-based drug discovery

The goal of fragment-based drug discovery is to identify small molecule ligands of a target protein with the aim of developing them into potent inhibitors through a medicinal chemistry campaign.^{49,50} In these screens, small libraries of compounds sampling diverse chemical space are screened using a biophysical method capable of detecting weak (mM) hits such as NMR spectroscopy, X-ray crystallography or isothermal titration calorimetry. Such libraries typically follow the “rule-of-three”⁵¹-guidelines defining promising fragments as those with molecular weights less than 300 Da, good solubility for screening at high concentration ($\text{clogP} \leq 3$), and limited chemical complexity with fewer than 3 hydrogen bond donors/acceptors and fewer than 3 rotatable bonds. Fragment-based screens typically have moderate hit rates often identifying multiple ligands which have high-quality interactions with the target protein. Based on the small size of fragment ligands these molecules generally have good ligand efficiency; a metric calculated as the ratio of the free energy of binding to the number of heteroatoms in the ligand.⁵⁰ Additionally, it has been demonstrated that the identification of protein ligands through a fragment screen is correlated with the success of lead discovery programs therefore fragment screening has been suggested indicate the “druggability” of a particular target.^{52,53} Through intensive medicinal chemistry campaigns these fragments can be developed into lead-like molecules with potent activity.

C. Results

C.1. Identification of small molecule fragment ligands of Ring1B/BMI1 fusion protein

In order to identify small molecule ligands of Ring1B/BMI1 we performed an NMR-based fragment screen of 1,000 compounds from a “rule of three” library. Compounds were screened in mixtures of 20 compounds at 0.25 mM final concentration and we analyzed binding of fragment compounds to Ring1B/BMI1 using NMR. Sixteen mixtures showed specific binding and deconvolution identified the active compounds, representing a 1.6% hit rate. The most promising compound from a potency and medicinal chemistry standpoint was compound **1**, 5-phenylthiophene-2-carboxylic acid (Figure 4.2.A). Compound **1** showed specific chemical shift perturbations clustering in a region of the Ring1B protein (binding site discussed below; Figure 4.2.B) and showed weak yet promising inhibitory activity in an *in vitro* ubiquitination assay,

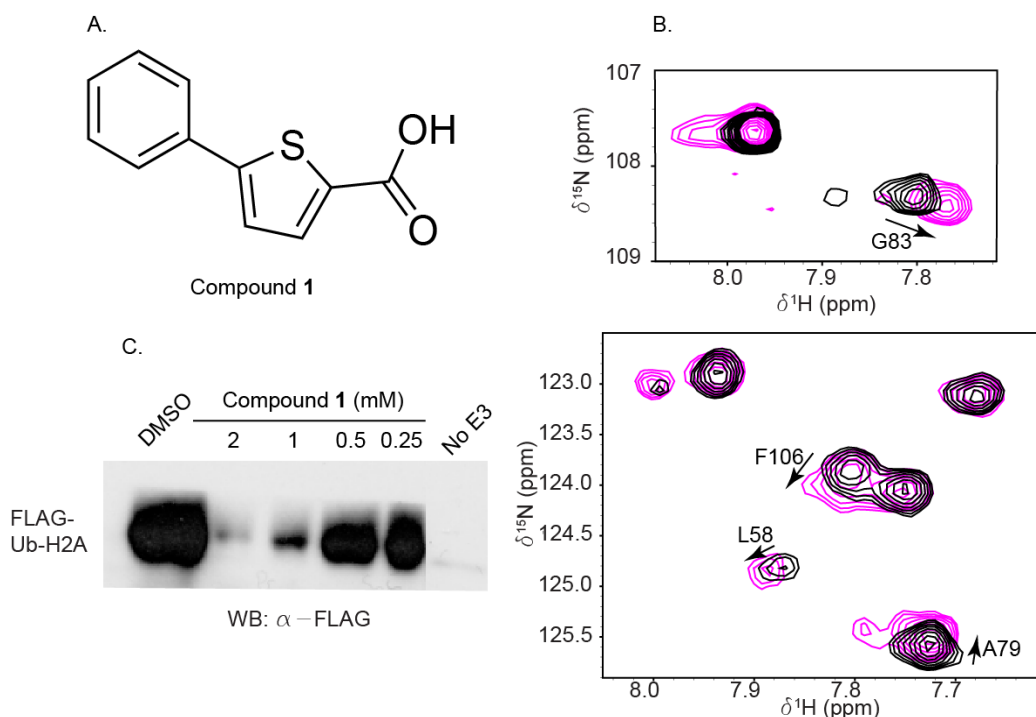


Figure 4.2. Characterization of Compound 1.

A. Structure of Compound **1**. B. Binding of compound **1** to Ring1B/BMI1. Selection of ^1H - ^{15}N HSQC spectra for Ring1B/BMI1 in the absence (black) or presence (magenta) of 0.5 mM compound **1**. C. Activity of compound **1** in ubiquitination assay. Western blot (α -FLAG detection) for Ring1B/BMI1 *in vitro* ubiquitin ligase assay with compound **1** or DMSO control.

inhibiting nucleosome H2A K119 ubiquitination by Ring1B/BMI1 (Figure 4.2.C). Through titration in NMR experiments the binding affinity of compound **1** for Ring1B/BMI1 was estimated as 7 mM.

C.2. NMR-guided SAR of Ring1B/BMI1 ligands

Using compound **1** as a scaffold we pursued a structure-activity-relationship study to grow the fragment into a more potent ligands. In the absence of structural information we proceeded with rationally exploring diverse chemical space and building on step-by-step improvements.

All medicinal chemistry efforts were performed in the Drs. Cierpicki and Grembecka labs.

C.2.1. Measuring affinity of compounds by NMR

To assess analog potency we used NMR spectroscopy. For the bulk of analogs tested, chemical exchange in the ^1H - ^{15}N HSQC spectra is in the fast exchange regime. To compare affinities of selected ligands we performed NMR titration experiments. Plots of the magnitude of the chemical shift perturbation (Equation 4.1- See Experimental Methods) vs. ligand concentration were fit with a saturation binding isotherm to calculate K_D and B_{max} (saturation chemical shift) (Figure 4.3.A). We observed in titration studies that many residues in the ligand binding site have approximately the same B_{max} (Figure 4.3.A). Thus we rationalized that the B_{max} obtained from titration studies could be used in equation 4.2 (see Experimental Methods) to fit chemical shift perturbations from experiments recorded at only a single concentrations. This is a rapid approach to quantify relative affinities of analogs based on just one binding experiment (Figure 4.3.B). This method has limitations due to the small number of data points analyzed, however it can be used to generally compare analog potency. Overall, this is an efficient approach to compare potency of ligands with millimolar to mid-micromolar affinity. We applied this method to estimate K_D for over one hundred analogs tested at a single concentration. For ease of protein production we expressed the Ring1B-BMI1 complex as a fusion protein (Ring1B/BMI1) with the C-terminus of Ring1B fused directly to the N-terminus of BMI1. This protein retains the biological activity of the non-covalent complex and therefore likely has a very similar structure.

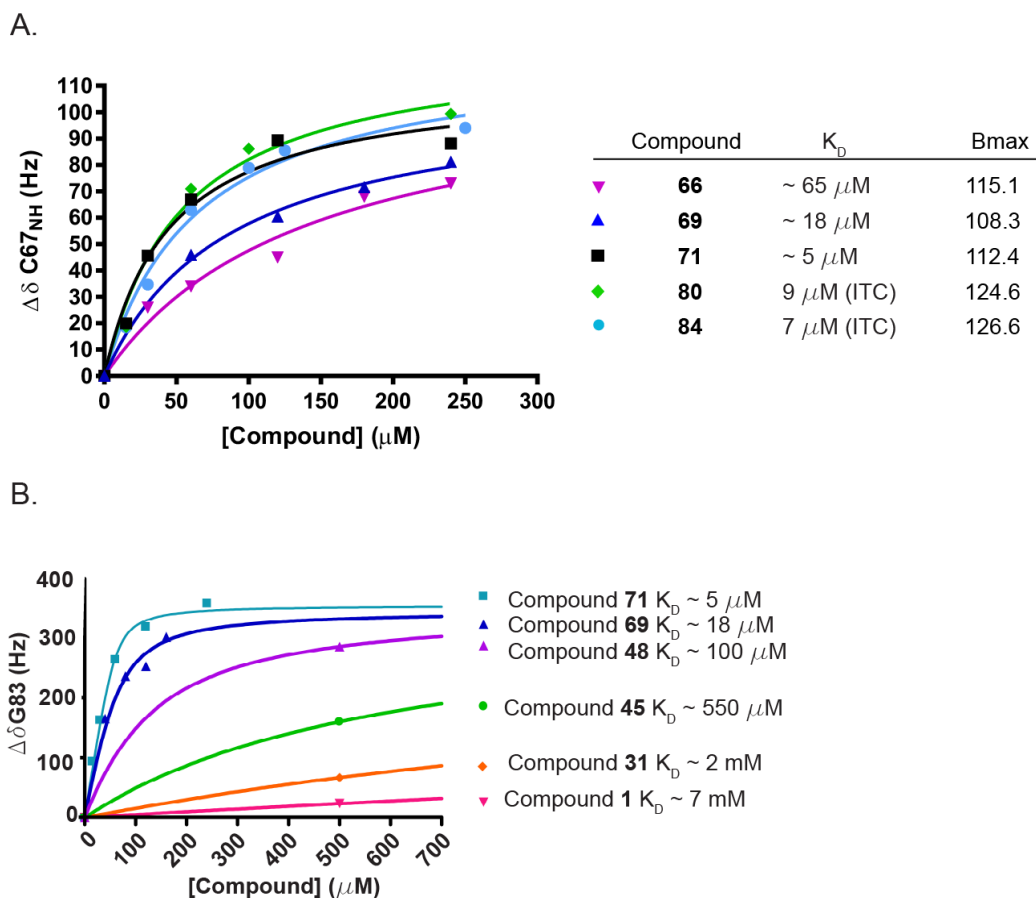


Figure 4.3. NMR estimation of binding affinity to guide SAR.

A. Example analysis of the chemical shift perturbation for a residue that shows saturation at the same chemical shift for ligands of different affinity. Left: curve fit for the Cys67 resonance with different ligands. Right: table with K_D and Bmax determined by the titration shown on the left. B. Analysis of Gly83 chemical shift perturbation for either titration or single point experiments for a selection of compound **1** analogs showing major improvements in affinity throughout the SAR studies.

For the majority of analogs tested the same set of Ring1B residues are perturbed and improved ligand potency did not cause more residues to be affected. Interestingly, for the most potent analogs ($K_D < 10 \mu\text{M}$) we observed appearance of new peaks likely due to ordering of several residues in the binding site. Additionally, for these more potent ligands we observed intermediate exchange phenomenon on NMR spectra which precluded quantitative chemical shift analysis using the method described above. Therefore, for the most potent compounds we determined binding affinity using ITC. All affinities reported in tables 4.1-4.6 are based on single point the NMR analysis protocol unless otherwise noted, (ie. NMR titration or ITC).

C2.1. Fragment growth approach with hydrophobic/hydrophilic group scanning

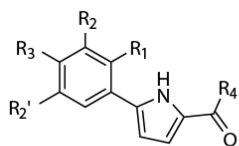
The first major improvement to the compound **1** was replacement of the thiophene core moiety with a pyrrole (compound **2**), which improved the affinity by approximately two-fold ($K_D \sim 3.6$ mM). All subsequent modifications were thus performed in the context of this scaffold which are detailed in Table 4.1. We tested replacement of the acidic group with an ester (compound **3**). This modification significantly reduced binding potency suggesting that the acidic group participates in critical interactions with the protein ($K_D \sim 20$ mM). Next we explored SAR around the phenyl ring. Compounds **4-9** explore substituents at the R_3 position. Substitution with hydrophobic methyl, chloro, trifluoromethyl or methoxy groups significantly reduced binding and substitution with polar sulfamide or amide group completely abolished binding.

In contrast to substitutions at the R_3 position which decreased binding affinity, substitutions at the R_2 position did not have a dramatic effect on affinity (compounds **10-20**). Alkyl or alkoxy groups of various size such as methyl, ethyl, methoxy, ethoxy, phenoxy and isopropoxy were well tolerated at R_2 (compounds **10-18**), neither enhancing nor diminishing binding. Interestingly, substitution with a hydroxyl group at this position (compound **19**) significantly decreased binding affinity whereas substitution with a primary amine (compound **20**; $K_D \sim 2.9$ mM) did not affect binding.

We tested substitutions at the R_1 position (compounds **21-27**) which followed a similar pattern where bulky hydrophobic groups such as phenoxy and ethoxy are well tolerated but do not improve binding affinity. Interestingly, addition of a hydroxyl group at R_1 slightly improved binding affinity whereas addition of a primary amine significantly reduced affinity ($K_D \sim 1.9$ and 8.9 mM, respectively), suggesting that the R_1 and R_2 occupy sites in the protein with different potential as hydrogen bond donors or acceptors. Together this scanning of R_1 and R_2 SAR suggests that careful arrangement of hydrogen bond donors and acceptors is critical at these positions.

Our studies with di-substitutions also did not significantly improve binding potency (compounds **28-30**). 2,4-dimethoxy (compound **28**) and 1-methoxy-4-methyl (compound **29**) substituted phenyl had similar affinity to the unsubstituted parent compound. The 1,4-dimethoxy

substituted compound (compound **30**) had ~4 fold reduced affinity relative to the parent compound. These studies indicate that careful orientation of substituents is required as two larger groups are not tolerated at the R₁ and R_{2'} positions yet can be tolerated at the R₂ and R_{2'} positions.









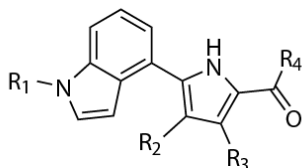
Compound number	R ₁	R ₂	R ₃	R ₄	K _D (from NMR fitting) (mM)
2	H	H	H	OH	3.6
3	H	H	H	OCH ₃	20.7
4	H	H	CH ₃	OH	5.4
5	H	H	Cl	OH	4.8
6	H	H	CF ₃	OH	9.7
7	H	H	OCH ₃	OH	9.8
8	H	H	SO ₂ NH ₂	OH	no binding
9	H	H	CONH ₂	OH	0
10	H	CH ₃	H	OH	6.5
11	H	OCH ₃	H	OH	2.3
12	H	OCH ₂ CH ₃	H	OH	6.3
13	H		H	OH	3.9
14	H		H	OH	3.8
15	H	CH ₃	H	OH	3
16	H		H	OH	4.3
17	H		H	OH	3.33
18	H	CH ₂ CH ₃	H	OH	4
19	H	OH	H	OH	14.3
20	H	NH ₂	H	OH	2.9
21	CH ₃	H	H	OH	16
22	OCH ₃	H	H	OH	3
23	OCH ₂ CH ₃	H	H	OH	4.7
24	NH ₂	H	H	OH	8.9
25	OH	H	H	OH	1.9
26		H	H	OH	3.8
27		H	H	OH	2.7
R₂ R₂'					
28	H	OCH ₃ OCH ₃	H	OH	3.8
29	OCH ₃	H CH ₃	H	OH	3.7
30	OCH ₃	H OCH ₃	H	OH	11.2

Table 4.1. SAR for 5-phenylpyrrole scaffold.

The biggest gain in affinity was through the addition of an indole group (compound **31**) which improved the potency by about two fold ($K_D \sim 2$ mM). Given the lack of other improvements thus far we designed all subsequent modifications in the context the indole ring, as detailed in Table 4.2.



Compound number	R ₁	R ₂	R ₃	R ₄	K _D (from NMR fitting unless noted (mM))
31	H	H	H	OH	2.0
32	H	H	H	NH ₂	5.9
33	CH ₃	H	H	OH	1.3
34	CH ₂ CH ₃	H	H	OH	3.4
35		H	H	OH	2.3
36		H	H	OH	3.8
37		H	H	OH	16.5
38		H	H	OH	6.7
39	H	CH ₃	H	OH	1.2
40	CH ₃	CH ₃	H	OH	1.5
41	H	H	CH ₃	OH	1.0
42	CH ₃	H	CH ₃	OH	1.4
43	H	CH ₂ CH ₃	CH ₂ CH ₃	OH	0.25
44	CH ₃	CH ₂ CH ₃	CH ₂ CH ₃	OH	0.19
45	H			OH	0.55
46	CH ₃			OH	0.43

Table 4.2. SAR of 5-indol-4-yl-pyrrole scaffold.

C.2.2 Optimization of 5-indol-4-yl-pyrrole scaffold

In an effort to determine the role of the R₄ acid in the context of the indole-pyrrole scaffold we synthesized compound **32** with an R₄ amide. This modification reduced potency by ~3 fold, leading us to conclude that a negative charge is required at this position. All further analogs were made in the context of the acidic moiety.

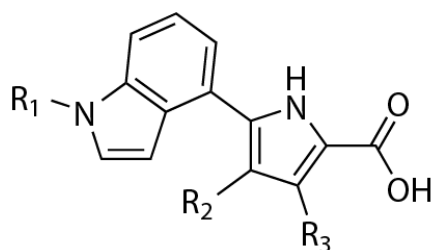
We designed compounds **33-38** to explore the role of the indole proton through substitution at the R₁ position with both hydrophobic and polar groups. Substitution with methyl (compound **33**) may slightly improve potency, although the improvement is likely within error of the analytical method ($K_D \sim 1.3$ mM). Substitution with larger groups such as ethyl and cyclopropyl reduced potency. Substitution with hydroxyethyl or aminoethyl reduced the affinity by 3.5 or 8 fold, respectively (compounds **37**, **38**). Overall, from these studies we concluded that substitution of the indole nitrogen is not favorable.

Next we explored substitution of the pyrrole ring in the context of either a methylated or unsubstituted indole nitrogen (compounds **39-48**). Regardless of indole nitrogen substituent, methyl substitution at either the R₂ or R₃ positions improved the potency by up to 2 fold ($K_D \sim 1.5-1.0$ mM) (compounds **39-42**). Significant improvement was attained through di-ethyl substitution, compounds **43** and **44**, with compound **44** having a 10-fold gain in affinity ($K_D \sim 190$ μ M) over unsubstituted pyrrole. We also introducing a fused cyclohexane ring in this position (compounds **45-46**) and found that these modification improved potency over the unsubstituted parent compound by ~4-fold ($K_D \sim 430$ μ M, compound **46**) but had reduced potency relative to diethyl substituted compounds (**43-44**). This indicated that these two positions likely occupy different sites in the protein and independent hydrophobic groups at these positions are beneficial.

We were encouraged by the large gain in affinity through ethyl and phenyl substitution at the R₂ and R₃ positions of the core pyrrole ring, respectively (compounds **47-48**). Table 4.3 describes SAR of hydrophobic groups at these positions. Compound **47** with 3-ethyl and 4-phenyl substituents on pyrrole had a measured of 30 μ M by NMR, a 200-fold improvement over the initial fragment hit. Importantly, switching the phenyl to the R₂ positions significantly reduced affinity (compounds **49-50**), with compound **50** having an affinity of 4.7 mM. Based on

the structural similarities yet significant differences in potency between the relatively low affinity compound **50** and high affinity compounds (such as compound **47**) we used compound **50** as a negative control in future studies. Additionally, from these studies we concluded that unsubstituted indole compounds have slightly improved potency compared to 1-methyl substituted indoles and for this reason, along with better solubility and ease of synthesis future analogs were generated with unsubstituted indole.

Based on the significant improvement of the phenyl substitution at R₃ we tested different six membered rings at this position (compounds **51-54**; Table 4.3). Replacement of phenyl with benzyl did not affect potency (compound **51** vs. **47**), however extending the group with a one carbon linker reduced the potency by ~ 3 fold (compound **52**). Changing phenyl to cyclohexyl



Compound number	R ₁	R ₂	R ₃	K _d from NMR fitting unless noted (μM)
47	H	CH ₂ CH ₃		30
48	CH ₃	CH ₂ CH ₃		100
49	H		H	3500
50	H		CH ₂ CH ₃	4700
51	H	CH ₂ CH ₃		27.5
52	H	CH ₂ CH ₃		87
53	H	CH ₂ CH ₃		150
54	H	CH ₂ CH ₃		35

Table 4.3. SAR of 5-indol-4-yl-pyrrole-2-carboxylic acid scaffold.

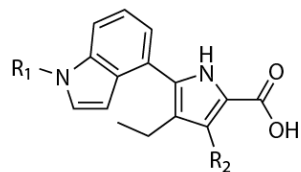
reduced the affinity by ~ 5-fold (compound **53**), however compound **54** with methylcyclohexyl has similar affinity as phenyl and benzyl substituents. Overall, these studies indicated that alkyl chain length and conformational flexibility are important for substituents at this position. Based on these results as well as the greater potential for chemical modification, we designed future compounds using the phenyl-substituted scaffold.

C.2.3 Further optimization of 5-indol-4-yl-pyrrole scaffold

In the next step we explored substitutions in the phenyl ring at R₂ in the context of 4-ethyl. These analogs were compared to the unsubstituted phenyl compound **47**, K_D ~ 30 μM. These studies are summarized in Table 4.4.

Substitution at the *para* position with either hydrophobic or polar groups resulted in a 2-5 fold loss of affinity (compounds **55-58**). The *meta* position is less sensitive to substitution as fluoro, methyl, or benzyloxy substitution resulted in only ~ 2 fold reduction in affinity (K_D ~ 40, 73 or 67 μM, compounds **59**, **60** and **62**, respectively). Interestingly, *meta*-methoxyphenyl (compound **61**) had approximately the same affinity as unsubstituted phenyl suggesting prospects for improving potency through this substitution. 3-methyl-4-fluorophenyl (compound **63**) had only 2-fold reduction in affinity (K_D ~ 72 μM) indicating that there is likely room in the binding pocket for multiple substituents on the phenyl ring and future exploration is needed to identify the optimal moieties to maximize hydrophobic or polar interactions.

We designed compounds **64-69** to explore *ortho*-substituents. In general substitutions at this position are well tolerated. Small hydrophobic groups such as methyl and trifluoromethyl do not improve or reduce binding affinity, while methoxy substitution reduces the affinity by ~ 2-fold (compounds **64**, **65** and **66**). 2-phenyl substitution (compound **67**) similarly reduced potency by ~ 2-fold suggesting that there is some room for small alkyl groups but not for larger groups. Importantly, compound **69**, featuring 2-chlorophenyl, had ~2-fold improved potency (K_D = 18 μM calculated through NMR titration).



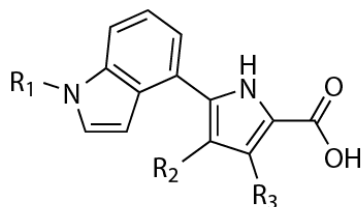
Compound Number	R ₁	R ₂	K _D from NMR fitting unless noted (μM)
47	H		30
55	H		76
56	H		50
57	CH ₃		109
58	H		262
59	H		40
60	H		73
61	H		36
62	H		67
63	CH ₃		72
64	H		33
65	H		41
66	H		65
67	H		75
68	CH ₃		30
69	H		18 (NMR titration)

Table 4.4. SAR of monosubstituted 5-indol-4-yl-pyrrole-2-carboxylic acid.

Encouraged by the improved potency of compound **69**, we tested analogs with di-substitutions in the context of 2-chloro. This series is summarized in table 4.5.

2,3- dichlorophenyl (compound **70**) had slightly reduced affinity, but 2-chloro-4-methoxyphenyl (compound **71**) significantly improved affinity ($K_D = 5 \mu\text{M}$ calculated through NMR titration). 2-chloro-4-ethoxyphenyl (compound **72**) had reduced affinity relative to compound **71** supporting the observations from mono substituted phenyl analogs that addition of a smaller alkyl group is tolerated but not a larger group. We designed compounds **73** and **74** to explore the addition of polar nitro or amine groups, neither of which improves nor reduces potency ($K_D \sim 20$ and $22 \mu\text{M}$, respectively). Based on the improved potency of compound **71**, we tested hydroxyethyl substitution at R_1 in an effort to retain potency while improving solubility for biochemical and cellular studies, however the binding affinity was decreased by ~ 70 fold (compound **75**). Finally, replacement of the chloro group with a larger bromo atom did not improve potency (compound **76**; $K_D \sim 7 \mu\text{M}$).

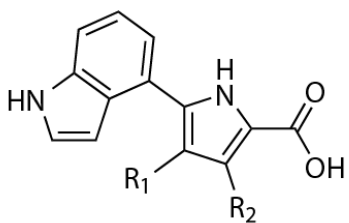
Following these studies, we explored substitution of alkyl groups at the R_2 position to gain affinity with the improved chloro substituted phenyl scaffold. Replacement of the ethyl group with methyl or propyl decreased activity (compounds **77**, **78**, $K_D \sim 67$ and $86 \mu\text{M}$, respectively). The butyl substituted compound **79** could not be tested due to solubility limitations. We were encouraged by significantly improved potency by isopropyl substitution (compound **80**; $K_D = 8.9 \mu\text{M}$ determined by ITC).



Compound number	R ₁	R ₂	R ₃	K _D from NMR fitting unless noted (μM)
70	H	CH ₂ CH ₃		30
71	H	CH ₂ CH ₃		5 (NMR titration)
72	H	CH ₂ CH ₃		16
73	H	CH ₂ CH ₃		20
74	H	CH ₂ CH ₃		22
75		CH ₂ CH ₃		344
76	H	CH ₂ CH ₃		7
77	H	CH ₃		67
78	H	CH ₂ CH ₂ CH ₃		86
79	H	CH ₂ CH ₂ CH ₂ CH ₃		N.D.*
80	H			ITC: 8.9

Table 4.5. SAR of disubstituted 3-phenyl-5-indol-4-yl-pyrrole-2-carboxylic acid.

*Affinity of compound **79** was not determined (N.D.) due to limited compound solubility.



Compound Number	R ₁	R ₂	K _D from NMR fitting unless noted (μM)
81	CH ₂ CH ₃		54
82	CH ₂ CH ₃		18
83	CH ₂ CH ₃		9 (NMR titration)
84	CH ₂ CH ₃		ITC: 6.9
85			ITC: 3.5

Table 4.6. SAR of heterocycle substituents of 5-indol-4-yl-pyrrole scaffold.

C.2.4. SAR of 5-indol-4-yl-pyrrole-2-carboxylic acid

Next, we tested different heterocycles at R₂; Table 4.6.

In the context of the ethyl at R₁ the addition of 4-chloropyridine at R₂ (compound **81**) reduced potency ~ 2 fold relative to 2-chlorophenyl (compound **69**). Substitution with either 6- or 4- indolyl (compounds **82**, **83**, respectively) significantly improved the potency, with compound **83** having an estimated affinity of 9 μM (NMR titration). Combining the gain in potency from indole and chloro modifications compound **84**, 5-chloro-4-indolyl, was constructed which had a slight gain in affinity (K_D = 6.9 μM by ITC). Ultimately, to maximize the effects of R₁ isopropyl compound **85** was designed with 5-chloro-4-indolyl at R₂, which had a two-fold gain in affinity (K_D = 3.5 μM as determined by ITC, Figure 4.4.A).

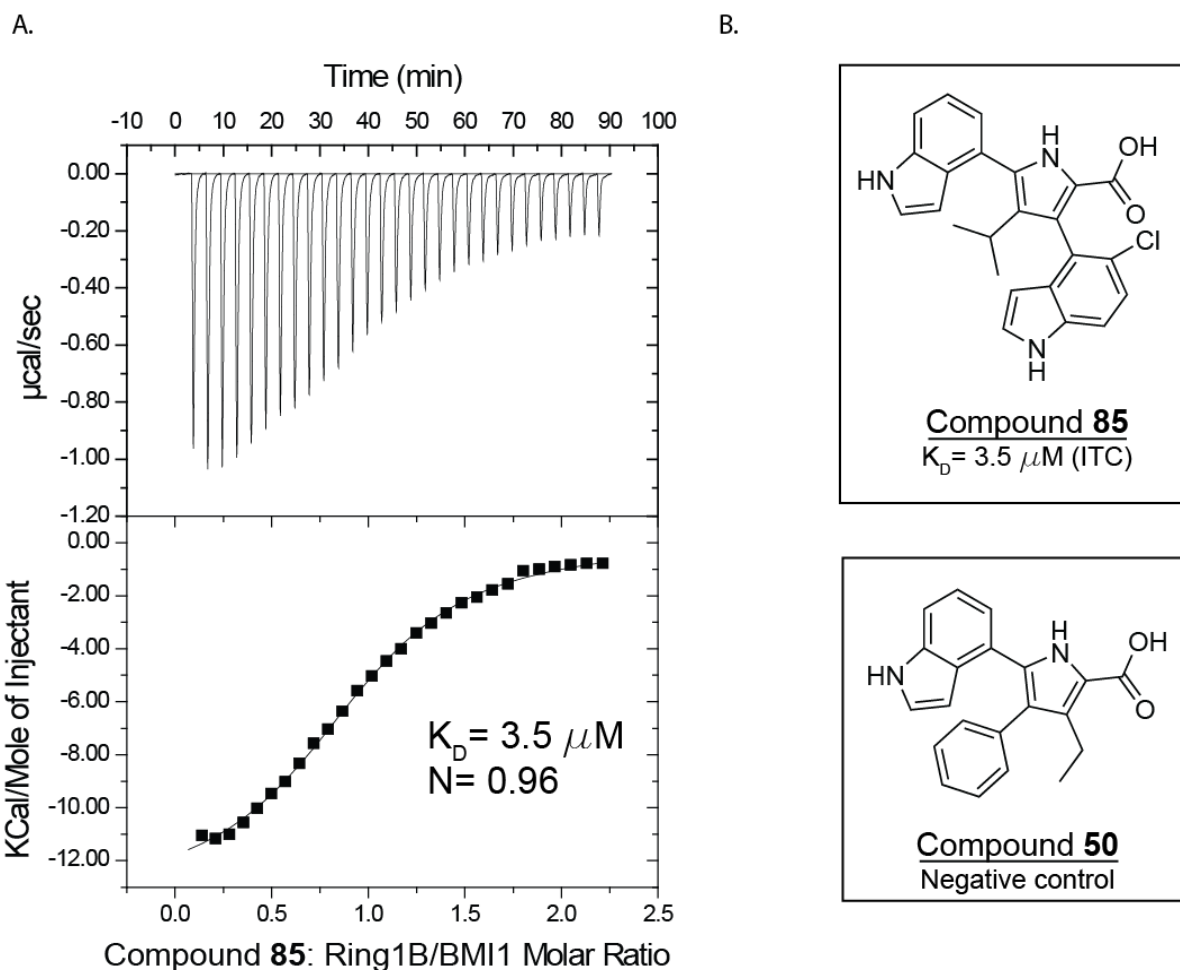


Figure 4.4. Characterization of most potent compound by ITC.

A. Isothermal titration calorimetry experiment of Ring1B/BMI1 titrated with compound **85**. B. Structures of most potent Ring1B/BMI1 inhibitor to date (compound **85**) and negative control compound (compound **50**).

Overall, through the extensive medicinal chemistry campaign a very weak fragment ligand of Ring1B/BMI1 was optimized into a potent ligand, improving the potency by three orders of magnitude. Compound **85** thus represents a novel chemical probe for Ring1B/BMI1 ubiquitin ligase activity with compound **50** serving as a structurally related negative control (Figure 4.4.B).

C.3. Ring1B/BMI1 ligands are potent and specific inhibitors of *in vitro* ubiquitin ligase assay activity

To test the inhibitory activity of Ring1B/BMI1 ligands we developed an *in vitro* ubiquitin ligase assay. In this assay the E1, UbcH5a (E2) enzymes and Ring1B/BMI1 fusion protein are

incubated with ATP, flag-ubiquitin and recombinant or commercial nucleosomes at 30° C. Ubiquitin ligase activity is visualized through western blot.

Throughout the medicinal chemistry campaign we observed a good correlation between improved binding affinity and *in vitro* inhibitory activity of Ring1B/BMI1 compounds (data not shown). The most potent ligand, compound **85**, has an IC₅₀ ~ 1.3 μM in the *in vitro* ubiquitination assay while no inhibitory activity was observed for compound **50** (negative control) (Figure 4.5). Significantly, this demonstrates that potent ligands of Ring1B/BMI1 are also inhibitors of ubiquitin ligase activity.

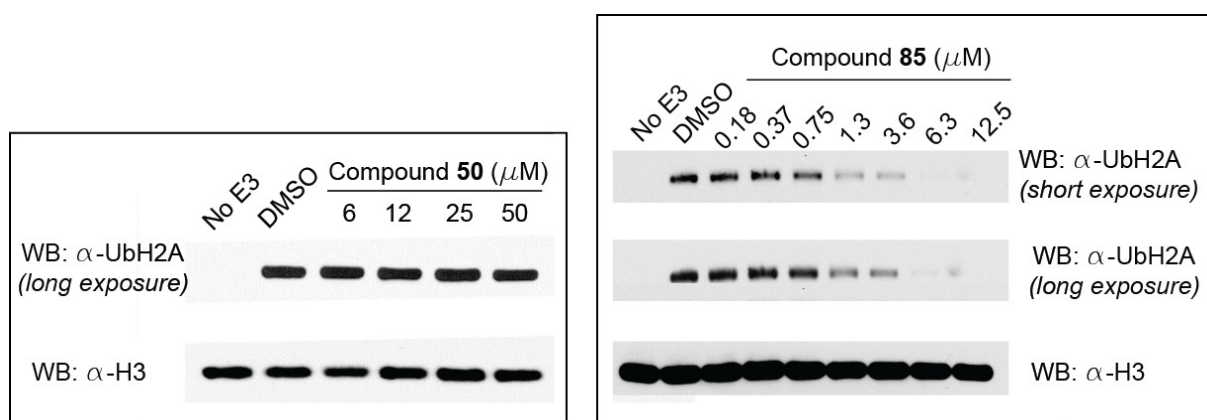


Figure 4.5. Optimized Ring1B/BMI1 ligand is inhibitor of *in vitro* ubiquitin ligase activity.

Western blots probed with antibody for UbH2A or H3 for *in vitro* ubiquitin ligase assay with Ring1B/BMI1 incubated with either negative control compound **50** (left) or the most potent ligand compound **85** (right).

To assess the potential of Ring1B/BMI1 inhibitors as chemical tools for biological experiments we wanted to determine their specificity against other RING E3 ubiquitin ligases. Recently two other E3 ligases, Brca1/Bard1 and Trim37 have been identified as mediating H2A ubiquitination in breast cancer.^{54,55} While these proteins have similar function they have low sequence identity to Ring1B (30% and 37% identity to Trim37 and Brca1, respectively) and thus Ring1B/BMI1 inhibitors are expected to have good selectivity over these other E3 ligases. To test this we performed ubiquitin ligase assays with these E3s. As expected compound **69** showed no inhibition of either Trim37 or Brca1 up to 500 μM, whereas the estimated IC₅₀ for this compound against Ring1B/BMI1 is 16 μM, showing at least 30-fold

selectivity (Figure 4.6). Additionally, no binding to the Brca1/Bard1 or Trim37 proteins was observed by NMR experiments with various Ring1B/BMI1 ligands we developed (data not shown). This demonstrates that we have developed potent and specific Ring1B/BMI1 inhibitors and supports their application as chemical tools.

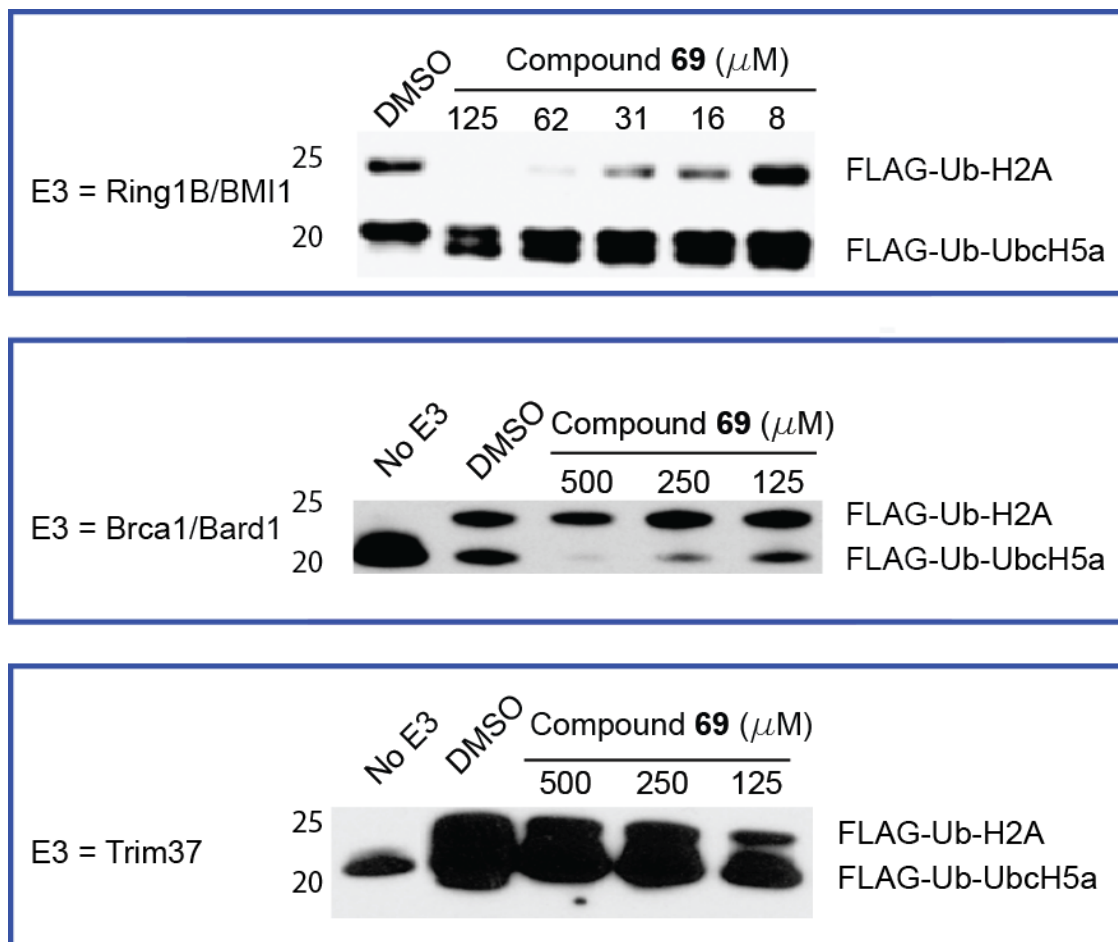


Figure 4.6. Inhibitors are specific for Ring1B/BMI1 E3 ubiquitin ligase.

Western blots probed with FLAG antibody for *in vitro* nucleosome ubiquitination assay with various E3 ligases incubated with compound **69** for Ring1B/BMI1 (top), Brca1/Bard1 (center) or Trim37 (bottom). Bands are labeled on right.

C.4. Ligand binding site and orientation probed by mutagenesis and NMR studies

To identify the ligand binding site on the Ring1B/BMI1 fusion protein, we mapped NMR chemical shift perturbations from compound **71** onto the crystal structure of the heterodimer complex (PDB: 2H0D).³⁰ This analysis shows that the ligand binds primarily to the Ring1B protein and that there are no interactions between BMI1 and the small molecule. Additionally,

chemical shift perturbations are localized to a region of Ring1B between the $\alpha\beta$ helix and the zinc binding loop (Figure 4.7.A). Inspection of the surface composition of the complex shows that there is a small hydrophobic cavity in this region but access to this pocket is occluded by the sidechain of Ring1B Lys97 (Figure 4.7.B).

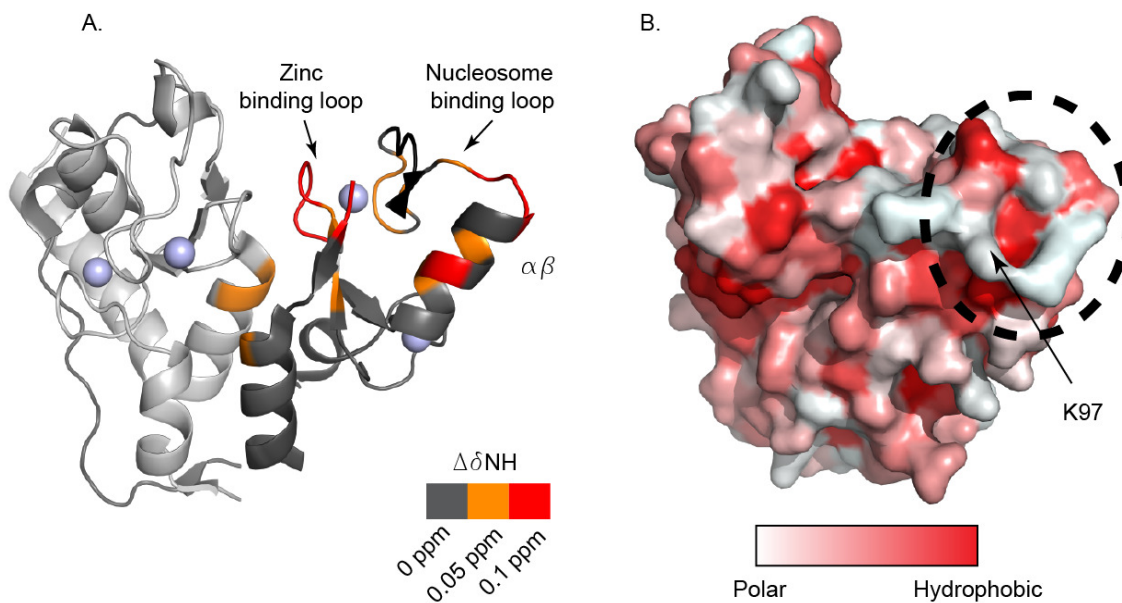


Figure 4.7. Mapping of NMR chemical shift perturbations onto Ring1B/BMI1 structure identifies ligand binding site.

A. Structure of Ring1B/BMI1 (PDB: 2H0D)³⁰ with magnitude of backbone NH chemical shift perturbations for binding of 500 μ M compound **71** colored as indicated. Ring1B protein is colored dark gray and BMI1 protein is colored light gray. Unassigned residues are colored black. Zinc molecules are shown as silver spheres. B. Surface representation of Ring1B/BMI1 structure colored by hydrophobicity of sidechains. Ligand binding site is indicated by dashed circle. Lys97 sidechain occluding binding site.

To validate this binding site, we designed mutants to perturb the conformation of this loop through charge repulsion and introduction of hydrophobic residues in the solvent exposed loop (Figure 4.8.A). We purified four mutant proteins and confirmed that the overall protein fold was not disrupted by NMR (data not shown). We tested ligand binding to these mutants by ¹H-¹⁵N NMR HSQC experiments (Figure 4.8.B). As expected, Ring1B Ile77Glu or Lys85Glu mutants had decreased affinity for the ligands. More significantly, the Leu94Ala mutation very dramatically reduced compound binding. Further, we designed Lys92Met mutant suspecting that replacement of the polar lysine with a larger hydrophobic methionine residue would cause significant structural perturbations of this loop. Unexpectedly, this mutation did not have an

effect on compound binding. We assayed these Ring1B mutants in the *in vitro* ubiquitin ligase assay and observed a strong correlation between reduced compound binding and reduced ubiquitin ligase activity (Figure 4.8.C).

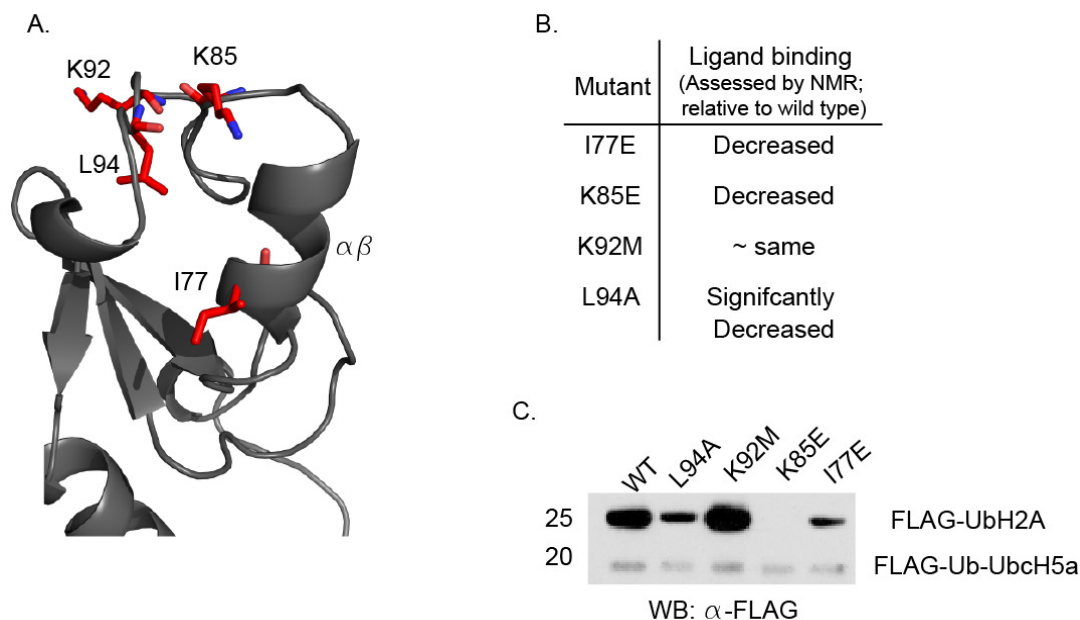


Figure 4.8. Probing compound binding site by mutagenesis.

A. Compound binding site with mutated residues indicated in red. B. Table of relative affinity of ligands for mutant Ring1B protein as assessed by NMR. C. Western blot probed with FLAG antibody for *in vitro* ubiquitin ligase activity for Ring1B mutants. Bands are labeled on right.

The reduced activity of the Lys85Glu and Ile77Glu mutants can be explained by the introduction of negative charges into this region of the protein that requires positively charged residues for interaction with the nucleosome. Inspection of the crystal structure suggests that the introduced methionine for Lys92 may be buried in the core of the loop region and thereby not affect protein activity. Leu94 mutation to alanine may disrupt the hydrophobic core of this region of the protein, and while the protein is still folded the decreased stability of this region may be enough to disrupt activity. Ring1B-nucleosome interface contacts are illustrated in Figure 4.10.

To gain insight into the ligand binding mode we turned to a paramagnetic relaxation enhancement (PRE). PRE is a NMR method used in structural biology to determine protein or protein complex structures and can be used to map the ligand binding.^{56,57} In this system a spin label with unpaired electrons is introduced into the protein or ligand of interest and magnetic

dipolar interactions between a physically close nuclei and the unpaired electrons in the spin label result in an increase in relaxation rate of the nucleus and thus loss of NMR signal.⁵⁸ This effect is visualized through the broadening of resonance signals for nuclei that are within a defined distance range of the spin label. The magnitude of the PRE is dependent on the electron-nucleus distance (r), with a proportionality of r^{-6} allowing both qualitative and quantitative assessment of the distances between the spin label and an observed nucleus.⁵⁸ It is estimated that PRE effects can be observed for nuclei within 15-25 Å of the spin label.⁵⁹

For PRE studies we designed compound **86**, which features the same basic chemical scaffold as compound **69** but is substituted with a pyrroline nitroxide at the 4 position connected to the core of the molecule by a short PEG linker (Figure 4.9.A). The SAR studies above suggest that this site is amenable to modification by large groups while still maintaining affinity for the protein. ¹H-¹⁵N HSQC experiments with compound **86** titration show that the compound still binds to the protein with $K_D \sim 250 \mu\text{M}$. In these experiments three effects are observed that allow estimation of the binding orientation of this ligand (Figure 4.9.B).

First, a number of resonances are completely broadened (eg. Asp74) upon addition of even low concentrations of compound suggesting they are within close proximity of the spin label. Second, some resonances show both fast exchange perturbations as well as dose-dependent peak broadening due to PRE (eg. Glu66) and these nuclei are likely close to the outer edge of the spin label effects (15-25 Å). Finally, a number resonances show fast exchange perturbations upon ligand binding indicating that they are further from the spin label yet still experience some change in chemical environment following ligand binding (eg. Thr64 and Thr89).

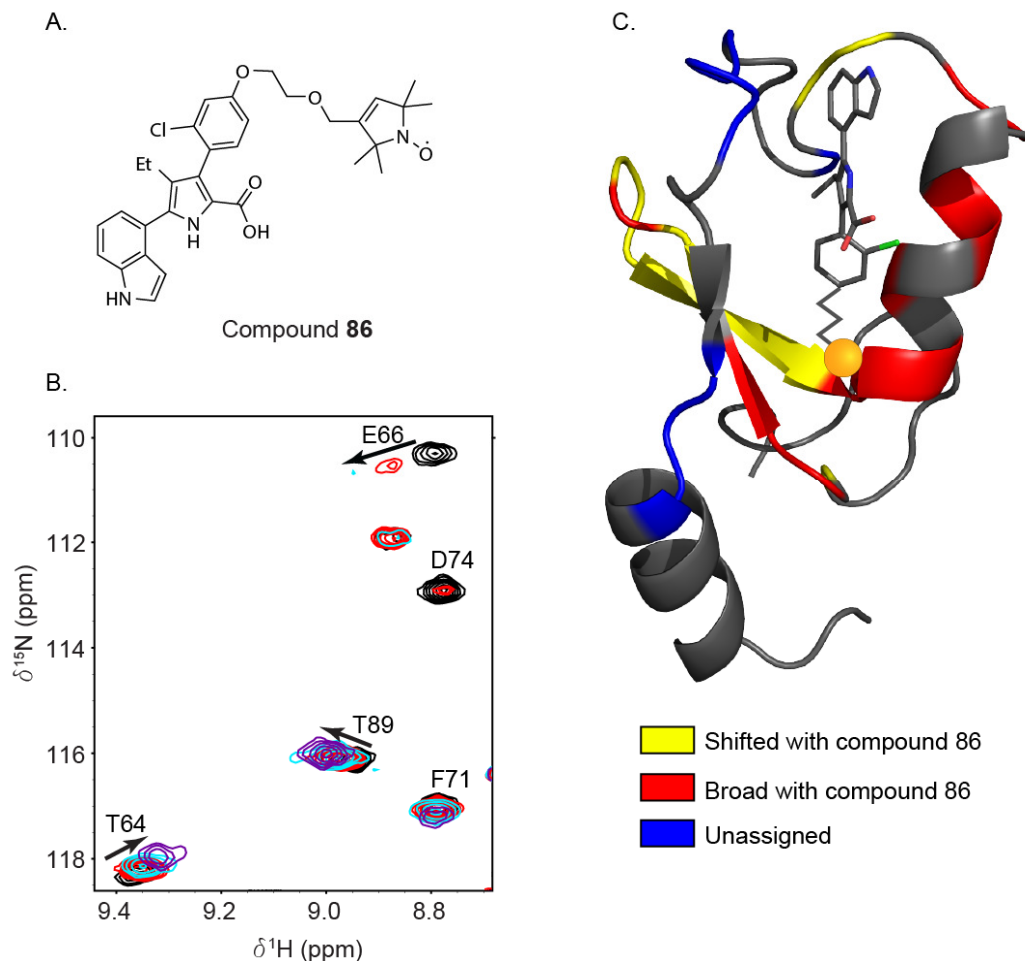


Figure 4.9. Paramagnetic relaxation enhancement (PRE) studies orient Ring1B/BMI1 ligands in binding site.

A. Structure of compound **86** with nitroxide spin label for PRE studies. B. Selection of ^1H - ^{15}N HSQC spectra for Ring1B/BMI1 titrated with compound **86**; DMSO= black, red= 0.5 equivalents, cyan= 1 equivalent, purple= 2 equivalents. C. Ring1B/BMI1 colored for the effects of compound **86** with residues that experience broadening effect indicated in red and residues that undergo chemical shift perturbations shown in yellow. Compound **86** is modeled into a previously published structure (PDB: 3RPG)³² with the nitroxide spin label shown as an orange sphere.

While multiple spin-labeled ligands would be required for the generation of a robust set of distance restraints from PRE studies and thus accurate positioning of the ligand in the binding site, the data from these experiments allow the determination of the orientation of the molecule with the protein. Figure 4.9.C illustrates the estimated orientation of compound **86** based on the analysis of ligand titration described above. The broadened residues in helix $\alpha\beta$ and at the base of the pocket suggest that the spin label is oriented towards these features. Residues that show

chemical shift perturbations but not PRE broadening are at the back of the ligand binding pocket indicating that the hydrophobic side of the ligand is likely orientated towards this region of the protein. Additionally, the residues that experience chemical shift perturbations but not PRE could be a result of protein conformational change as a result of ligand binding rather than interactions with the ligand.

C.5. Inhibitor binding results in conformational change in Ring1B

Analysis of the surface topology of the Ring1B/BMI1 complex does not suggest an obvious pocket for small molecule binding (Figure 4.7.A). However, robust ITC, mutagenesis and NMR binding data demonstrate that Ring1B/BMI1 inhibitors do bind directly to this protein at the nucleosome binding surface. This suggests that protein conformational change is required for ligand binding. Analysis of chemical shift perturbations, which are sensitive to both ligand binding and protein conformational change, validates this conjecture. For example, the magnitude of chemical shift perturbations, greater than 0.5 ppm for some resonances, suggesting these effects may be due to protein conformational change. Additionally, in binding studies with the most potent ligands we observed the appearance of new resonances in the ^1H - ^{15}N HSQC spectrum which we assigned to residues in the substrate binding site indicating that ligand binding results in stabilization of protein conformation in this region. Finally, the PRE studies detailed above demonstrate that there is a region of the ligand binding site that is further from the spin label but experiences a change in chemical environment following ligand binding, likely through protein conformational change.

The structural and biochemical characterization of the Ring1B/BMI1-nucleosome interaction by McGinty et al²³ demonstrated that several basic residues in the nucleosome binding site of Ring1B contribute to direct protein-nucleosome interaction and thus E3 ligase activity (Figure 4.10). On the basis of the observed conformational change in Ring1B with compound binding we suspected that the ligand-bound conformation of Ring1B/BMI1 is incompatible with nucleosome binding.

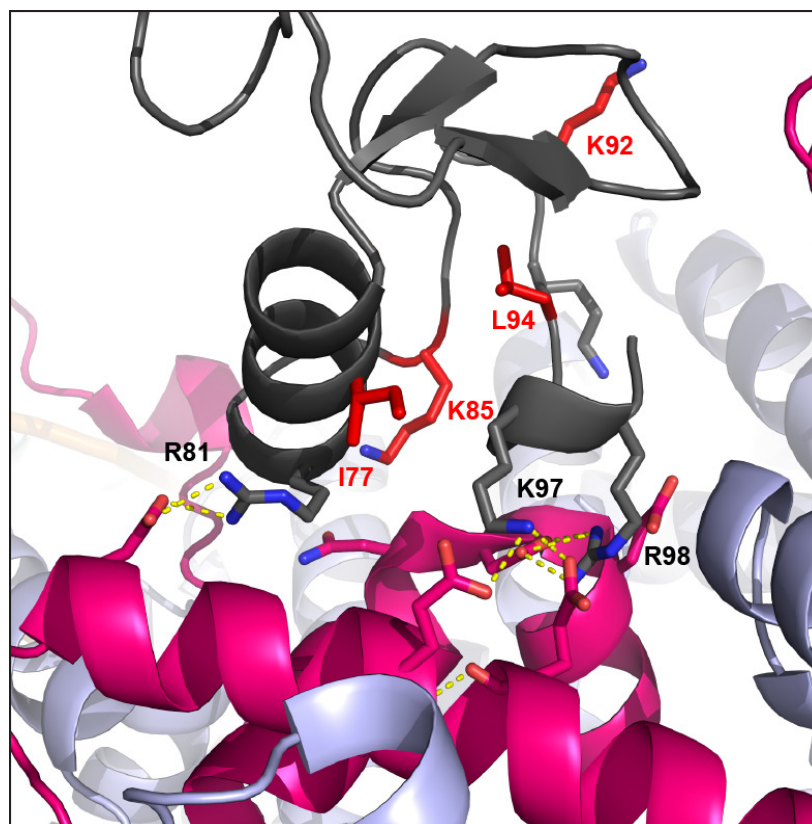


Figure 4.10. Ring1B-nucleosome interface.

Details of the Ring1B-nucleosome interface from PDB 4R8P²³ showing the requirement for careful orientation of basic sidechains in Ring1B (dark gray) to interact with acidic residues on the H2A (magenta) surface. Key basic residues are labeled. Other histones are colored in pale blue.

C.6. Ring1B/BMI1 inhibitors disrupt protein-nucleosome interaction

To experimentally validate that the ligand-bound conformation of Ring1B/BMI1 precludes binding to nucleosome substrate we developed a gel based EMSA assay (electrophoretic mobility shift assay) using nucleosomes reconstituted with Oregon Green-labeled H2B. Gel visualization by Typhoon imager allows detection of the mobility of nucleosomes in the presence of Ring1B/BMI1 protein. Experiments with pre-incubation of Ring1B/BMI1 with compounds allows analysis of the effect of compound binding on this interaction. We observed a clear change in mobility of nucleosomes following the addition of Ring1B/BMI1 (Figure 4.11.A). The potent E3 ligase inhibitors prevent this interaction as demonstrated by a dose dependent decrease in the bound nucleosome fraction with compound titration (Figure 4.11.B). The most potent inhibitor, compound **85**, has an $IC_{50} \sim 3 \mu M$ in this

assay, whereas the inactive compound **50** has no effect on nucleosome binding. Together these studies confirm that the inhibitor-bound conformation of Ring1B/BMI1 is incompatible with direct nucleosome binding thereby blocking its ubiquitination activity.

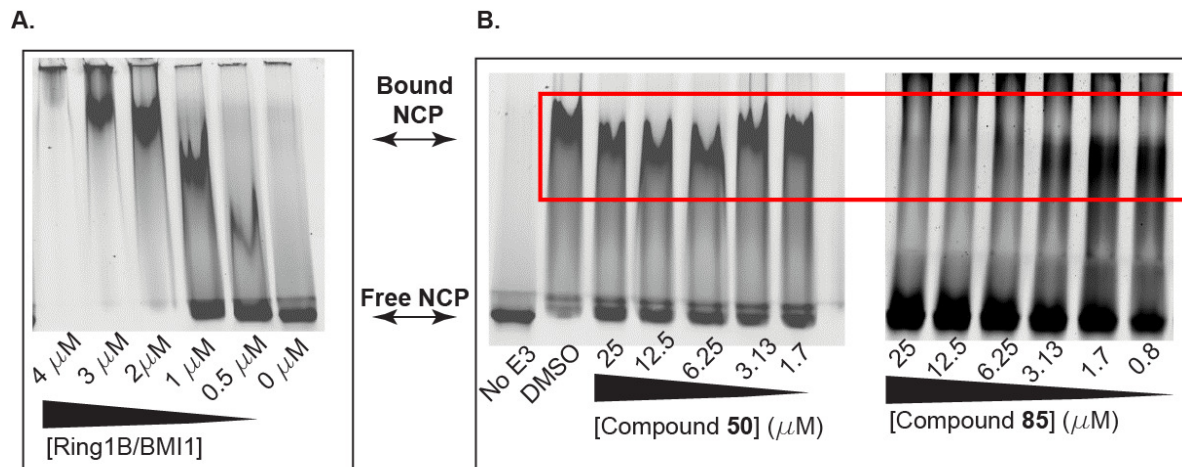


Figure 4.11. Ring1B/BMI1 ligand binding prevents interaction with nucleosome. A. EMSA assay for OregonGreen488 labeled nucleosomes titrated with Ring1B/BMI1. B. Competition EMSA experiments with Ring1B/BMI1 incubated with decreasing concentrations of either negative control compound **50** (left) or compound **85** (right). Shift positions of free and bound nucleosomes are labeled.

C.7. Preliminary cellular studies demonstrate utility of inhibitors as chemical probes

We next wanted to determine if the potent *in vitro* Ring1B/BMI1 inhibitors we developed have potential as novel chemical tools to address biological questions in a cellular context. To do this we assessed cell permeability and on-target activity by treating MCF10A cells with ligands with different *in vitro* potency. We selected MCF10A breast epithelial cells, a cell line in which Ring1B/BMI1 is the major H2A ubiquitin ligase.⁵⁴ Treatment of cells with compounds **69** and **85** showed dose-dependent decrease of global H2AK119Ub levels (Figure 4.12.A). Additionally, the activity of these compounds in cells correlates with their *in vitro* activity as the more potent compound, compound **85**, has $IC_{50} < 6.25 \mu\text{M}$ whereas the weaker compound **69** has $IC_{50} \sim 12.5 \mu\text{M}$ in these cellular studies. Together this data indicate that Ring1B/BMI1 inhibitors are cell permeable and have good on-target activity.

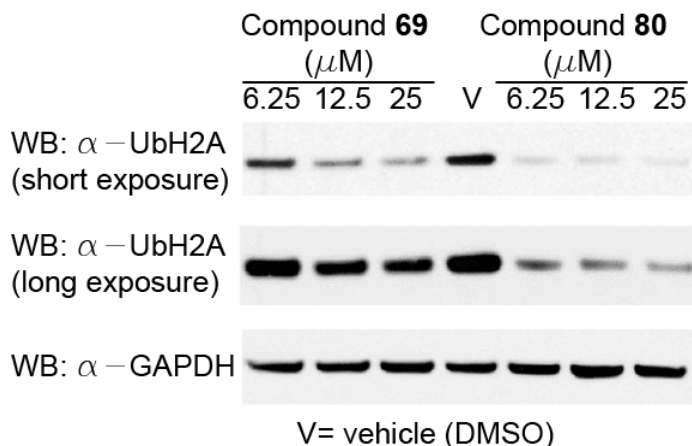


Figure 4.12. Ring1B/BMI1 inhibitors are cell permeable with on-target activity. Western blots of MCF10A cell lysate probed with antibodies for UbH2A and GAPDH as indicated. Cells were treated with compounds for 6 days.

D. Discussion and Conclusion

The work presented here describes the development of highly novel small molecule inhibitors of the Ring1B/BMI1 E3 ubiquitin ligase. This work employed fragment-based screening to identify small molecule ligands directly binding to the Ring1B/BMI1 E3 ubiquitin ligase. Optimized ligands are also potent and specific inhibitors of Ring1B/BMI1 E3 ligase activity both *in vitro* and in cells. Supporting mechanistic studies demonstrate that ligand binding causes significant structural perturbation in the substrate binding site of Ring1B preventing nucleosome binding and thus H2A ubiquitination. Therefore the optimized small molecule inhibitors of Ring1B/BMI1 presented here represent novel inhibitors of direct protein-nucleosome interaction. And finally, these molecules represent the first cell-permeable compounds that bind directly to Ring1B/BMI1 to block its E3 ligase activity setting the stage for use as chemical tools in polycomb biology and cancer biology.

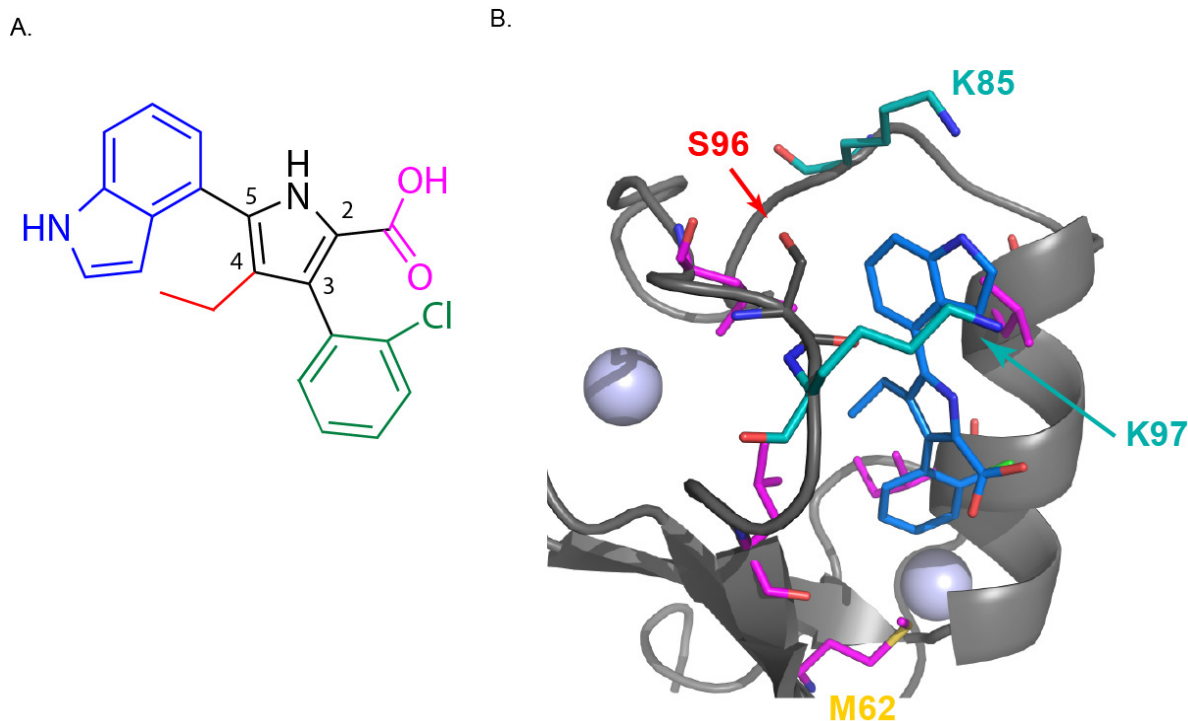


Figure 4.13. Docking model of Ring1B in complex with compound 69.

A. Structure of compound **69** with key features highlighted. B. Computational docking model of compound **69** in complex with Ring1B generated by Dr. George Lund. Hydrophobic residues in the binding site are colored magenta and key basic residues in the site are colored green. Zinc atoms are shown as gray spheres.

The SAR studies detailed in this chapter allow a comprehensive analysis of the chemical basis for potent inhibition of Ring1B/BMI1 through disruption of protein-nucleosome interaction. To design more potent compounds we have developed a structural model of Ring1B in complex with compound **69** based on computational docking studies supported by NMR-derived distance restraints (Figure 4.13.B). This model was generated by my colleague Dr. George Lund. Compound **69** features the pyrrole core substituted with 2-carboxylic acid, 3-chlorophenyl, 4-ethyl and 5-indole (Figure 4.13.A).

This model illustrates that the small molecule inhibitors insert between the $\alpha\beta$ helix and the zinc binding loop of Ring1B, pushing these elements apart and forming critical contacts with residues on both sides of this pocket. Key hydrophobic interactions between the indole and pyrrole ring systems and the hydrophobic core of the protein likely drive the affinity of this interaction. Critically, the 5-indole is bracketed by Leu80 and Leu94 providing a structure-based explanation for the loss of binding affinity to the Leu94A mutant. The sidechain of Ser96 is also

positioned in the ligand binding site and may contribute to some polar interactions. The 4-ethyl is positioned between Ile76 and Leu100, and while there is room in the structure to accommodate the larger isopropyl group featured in compound **85**, there is no room for a phenyl ring at this position supporting the observation that compound **50** has significantly reduced binding affinity. In the model, the 3-chlorophenyl is situated at the base of the binding site which is capped by Met62 and there may be hydrophobic interactions between these moieties. Additionally, the chloro group is oriented towards the $\alpha\beta$ helix and form a halogen bond with backbone carbonyl of Ile77. Finally, the 2-carboxylic acid is in close proximity to Lys97 and may form electrostatic interactions with this residue further contributing to compound activity. In summary, this structural model helps rationalize the steep SAR observed throughout the studies detailed in this chapter and provides a basis to design more potent compounds.

One such opportunity may be through exploiting the high density of basic residues in the ligand binding site. The structural model suggests that the sidechain of Lys85 is positioned near the 5-indole and future molecules designed with carbonyl or sulfonyl substituents on the indole could participate in polar interactions with this residue further contributing to binding affinity. Similarly, the sidechain of Lys97 may be in position to interact with the other side of this indole group. While the conformational flexibility of this region of the protein has thus far hindered structural studies of the protein-ligand complex by X-ray crystallography. Current efforts are focused on detailed structural characterization to drive structure-based design of more potent inhibitors.

Acknowledgements

Thank you to Dr. Shihan He who performed the fragment screen and taught me Ring1B/BMI1 purification. I thank Gireesh Reedy and Josh Abbott for help with large-scale protein purification. I thank Dr. Shirish Shukla for sharing his cellular data with me and Dr. George Lund for sharing the computational model of Ring1B/BMI1 and compound **69**. Finally, this work could not have been completed without the medicinal chemistry performed by Dr. Qingjie Zhao and Dr. Weijiang Yang. I think we make a great team.

E. Experimental methods

Plasmid construction

Ring1B/BMI1 fusion protein. A synthetic gene encoding Ring1B (res. 10-116) fused to BMI1 (res. 1-104) was ordered from Genscript and subcloned into the pET32a vector (Novagen).

Histone proteins. Ampicillin- resistant plasmids encoding *Xenopus* histones H2A, H2B, H3 and H4 in the pET3d vector were a kind gift from Dr. Yali Dou's lab (Department of Pathology, University of Michigan).

601 DNA sequence. Ampicillin- resistant plasmid encoding the 147 base pair Widom 601 sequence was a kind gift from Dr. Yali Dou's lab (Department of Pathology, University of Michigan). Plasmids for mutant constructs were made using standard QuikChange mutagenesis protocol (Stratagene) and verified by Sanger sequencing.

Protein expression and purification

Unlabeled Ring1B/Bmi1 wild type and mutants. Plasmids encoding RingBmi1 wild type or mutants were transformed into BL21 (DE3) *E. coli* cells T1R (Sigma) and selected with ampicillin (100 µg/mL). Seed cultures, 5 mL Luria broth (Fisher Scientific) cultures supplemented with ampicillin (100 µg/mL), were inoculated with a single colony and grown for 4 hours at 37°C (220 rpm). Seed cultures were diluted (200-fold) into 1 L cultures of Luria broth supplemented with ampicillin (100 µg/mL). Cultures were grown at 37°C (220 rpm) until OD₆₀₀ reached 0.6-0.8. Incubation temperature was lowered to 25°C and protein over expression was induced with 0.5 mM (final concentration) isopropyl β-D-1-thiogalactopyranoside (IPTG; Gold Bio) for 16-18 hours at 25°C. Cultures were harvested by centrifugation, resuspended in 30 mL lysis buffer (50 mM Tris, pH 7.5 at 25°C, 150 mM NaCl, 2 mM DTT, 100 µM ZnCl₂, 0.5 mM PMSF) and lysed using French press cell disrupter. Inclusion bodies were pelleted by centrifugation and washed three times in wash buffer A (lysis buffer A + 1% (v/v) triton-x100), washed once with lysis buffer and once with wash buffer B (50 mM Tris, pH 7.5 at 25°C, 150 mM NaCl, 100 µM ZnCl₂, 0.5 mM PMSF, 2 mM DTT, 1 M Guanidine HCl). Protein was solubilized in 6 M Guanidine HCl in lysis buffer by rocking at room temperature for 1 hour. Protein was refolded through serial dialysis into 100-fold excess refolding buffer (50 mM Tris, pH 7.5 at 25°C, 150 mM NaCl, 2 mM DTT, 100 µM ZnCl₂). To remove the His₆-thioredoxin tag, the protein was cleaved with Tobacco Etch Virus (TEV) protease. The His₆-thioredoxin protein was separated from RingBmi1 protein using cation exchange chromatography. The purified protein solution was buffer exchanged into storage buffer (50 mM HEPES, pH 7.2, 100

mM NaCl, 10 mM MgCl₂, 10 μM ZnCl₂, 1 mM DTT) through dialysis. Protein concentration was measured using absorbance at 280 nm. Protein identity and purity was verified by sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE). Protein was flash frozen and stored at -80°C.

¹⁵N-labeled Ring1B/Bmi1 for NMR Spectroscopy. Protein was expressed as above with the exception of the use of the M9 minimal media supplemented with ¹⁵N ammonium sulfate (Cambridge Isotopes) and ¹⁵N Bioexpress (Cambridge Isotopes).⁶⁰ Upon IPTG induction, media was supplemented with 100 μM ZnCl₂ (final concentration). Following protein refolding and tag cleavage, protein was dialyzed into 100-fold excess NMR buffer: 50 mM sodium phosphate, pH 6.8, 150 mM NaCl, 1 mM TCEP, 10 μM ZnCl₂.

Histone proteins. Plasmids encoding genes for the four histones were transformed into CodonPLUS BL21 (DE3) *E. coli* cells (SOURCE) and selected with ampicillin (100 μg/mL). Seed cultures, 5 mL Luria broth (Fisher Scientific) supplemented with ampicillin (100 μg/mL), were inoculated with a single colony and grown for 4 hours at 37°C (220 rpm). Seed cultures were diluted (200-fold) into 1 L cultures of Luria broth supplemented with ampicillin (100 μg/mL). Cultures were grown at 37°C (220 rpm) until OD₆₀₀ reached 0.5-0.7. Protein over expression was induced with 0.4 mM Isopropyl β-D-1-thiogalactopyranoside (IPTG; Gold Bio) (final concentration) for 4 hours at 37°C. Cultures were harvested by centrifugation. Cells were resuspended in 30 mL lysis buffer A (50 mM Tris, pH 7.5 at 25°C, 100 mM NaCl, 1 mM Na-EDTA, 1 mM PMSF, 1 mM β-ME) and lysed using French press cell disrupter. Inclusion bodies were pelleted by centrifugation and washed three times in wash buffer A (lysis buffer + 1% (v/v) triton-x100). Proteins were solubilized by mincing pellet in 0.5 mL DMSO and incubated at room temperature for 30 minutes. 5 mL unfolding buffer A (7 M Guanadine HCl, 20 mM sodium acetate, 10 mM DTT) was slowly added to pellet and solution was gently pipetted to facilitate resuspension. Protein solution was incubated with gentle stirring at room temperature for 4 hours before clarifying by centrifugation. Clarified, solubilized protein was dialyzed at room temperature into three changes of 100-fold excess dialysis buffer (7 M urea, 100 mM sodium acetate, 20 mM Tris, pH 7.5 at 25°C, 100 mM NaCl, 2 mM β-ME). Proteins were further purified using anion and cation exchange. Briefly, 4 mL gravity Q Sepharose (GE Healthcare) was equilibrated with dialysis buffer. Clarified protein sample was applied to the column and incubated for 10 minutes. Protein was eluted in 3 steps with Q elution buffer: Q100 (7 M urea,

20 mM sodium acetate, pH 5.2, 100 mM NaCl, 2 mM β -ME), Q150 (7 M urea, 20 mM sodium acetate, pH 5.2, 100 mM NaCl, 2 mM β -ME), Q200 (7 M urea, 20 mM sodium acetate, pH 5.2, 100 mM NaCl, 2 mM β -ME). Fractions containing protein were pooled and applied to 2 mL gravity SP Sepharose (GE Healthcare). Protein was eluted in 3 steps with S elution buffer: S300 (20 mM sodium acetate, pH 5.2, 300 mM NaCl, 7 M urea, 2 mM β -ME), SAU500 (100 mM sodium acetate, 20 mM Tris, pH 7.5 at 25°C, 500 mM NaCl, 7 M urea, 3 mM β -ME), SAU1000 (100 mM sodium acetate, 20 mM Tris, pH 7.5 at 25°C, 1 M NaCl, 7 M urea, 3 mM β -ME). Fractions containing purified protein was pooled and dialyzed against 3 changes of 4 L of MilliQ water with 2 mM β -ME at 4°C. Samples were clarified and lyophilized. Purified protein was stored at -80°C.

UbcH5c, Trim37 and Brca1/Bard1 were expressed and purified by other members of the Cierpicki lab.

NMR Spectroscopy

All NMR experiments were performed on a Bruker Advance III 600-MHz spectrometer equipped with a 5 mm TCI cryogenic probe. All spectra was acquired at 30°C. Spectra were processed with NMRPipe⁶¹ and analyzed with Sparky.⁶²

Compound stock preparation and concentration verification. Powdered compounds were resuspended in the appropriate amount of DMSO to make 100 mM stocks. Compound stock concentration was verified through ¹H NMR experiments compared to internal standard (BD-10). If necessary, stock concentrations were adjusted and re-verified though the same methodology.

SAR-by-NMR Experiments. For screening analogs, samples were prepared with 60 μ M final concentration ¹⁵N RingBMI1 in NMR buffer (50 mM sodium phosphate, pH 6.8, 150 mM NaCl, 1 mM TCEP, 10 μ M ZnCl₂) with compounds in 5% final DMSO. All samples contained 7% D₂O.

Chemical shift perturbation quantification. The magnitude of perturbation for fast-exchanging chemical shifts was calculated as the magnitude of change of both the ¹H and ¹⁵N chemical shifts in hertz.⁶³

Equation 4.1. Calculation of chemical shift perturbations

$$\Delta\delta_{NH} = \sqrt{((\delta N^{15} free - \delta N^{15} bound)^2 + (\delta H^1 free - \delta H^1 bound)^2)}$$

Determination of K_D using chemical shift perturbations of small molecule titration.

Chemical shift perturbations for small molecule titrations were fit in Prism GraphPad (reference) using a binding isotherm described by the equation 2. $[RingBmi1] = 60 \mu M$ for all titrations; $[Lig] =$ concentration of ligand. B_{max} is determined through fitting ligand titrations with a saturating binding isotherm.

Equation 4.2. Calculation of K_D and B_{max} for ligand titration.

$$\Delta\delta_{NH} = \frac{\left\{ \left[1 + \left(\frac{1}{K_D} + ([Lig] + [RingBmi1]) \right) \right] - \sqrt{\left[1 + \left(\frac{1}{K_D} + ([Lig] + [RingBmi1]) \right) \right]^2 - \left(\frac{4 \left(\frac{1}{K_D} \right) [RingBmi1]}{B_{max}} \right) * \left(B_{max} \frac{1}{K_D} * [Lig] \right)} \right\}}{\left\{ \frac{2 \left(\frac{1}{K_D} \right) [RingBmi1]}{B_{max}} \right\}}$$

Estimation of small molecule K_D using single point chemical shift perturbation. For rapid comparison of compound 1 analogs, K_D s were estimated in Prism GraphPad using equation 4.2 where $[RingBmi1] = 60 \mu M$. For each compound, equation 4.2 was solved using the magnitude of chemical shift perturbation for glycine 83 resonance at a single concentration and an estimated saturation chemical shift (B_{max}) of 360 Hz at 1 M ligand. Equation adapted from Ref. 63.

Nucleosome reconstitution (adapted from^{64,65})

Octamer reconstitution. Lyophilized histone proteins were resuspended in histone unfolding buffer (8 M guanidine HCl, 20 mM Tris pH, 7.5, 5 mM DTT). Proteins were rocked at room temperature for 1 hour. Proteins were mixed in a 1:1:1:1 molar ratio for a final protein concentration of 2 mg/mL. Histone mixture was dialyzed overnight into 2,000-fold excess octamer reconstitution buffer (2 M NaCl, 10 mM Tris, pH 7.5, 1 mM ETDA, 5 mM β -ME). Homogeneous octamer was purified via size exclusion chromatography using a Sephacryl S200 column (GE Healthcare) with octamer reconstitution buffer (2 M NaCl, 10 mM Tris, pH 7.5, 1 mM ETDA, 5 mM β -ME). Octamer quality was verified by sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) using 18% Tris-glycine-SDS gel (BioRad) and purified samples were flash frozen and stored at $-80^\circ C$. Concentration was measured using absorbance at 276 nm.

H2B^{OG488} Octamer Assembly. Reconstitution of labeled nucleosomes was performed as above with the exception that H2B S112C was labeled with OregonGreen 488 prior to assembly. Labeling was achieved by incubated H2B S112C with two equivalents of OregonGreen maleimide (Invitrogen) for 4 hours at 4°C with rotation in unfolding/labeling buffer (7 M guanidine HCl, 20 mM Tris, pH 7.5, 25 mM NaCl, 30 mM TCEP) before the reaction was spiked with an additional equivalent of dye. Excess dye was removed by dialyzing into two changes of 100-fold excess histone unfolding buffer. Labeling was verified by mass spectrometry.

601 DNA purification. The plasmid encoding the 23 copies of the 147 base pair Widom 601 sequence was amplified in DH5 α *E. Coli* cells and purified in milligram quantities with the GigaPrep kit (Qiagen). Purified DNA was digested at a concentration of 1 mg/mL with EcoRV (New England Bioscience) at a concentration of 30 units of EcoRV per nanomole EcoRV site in 1x Buffer #3 (New England Bioscience). Reactions were incubated at 37°C for 16 hours. Digestion progress was verified on a 10% agarose gel. If reaction was incomplete, reactions were supplemented with 50% more enzyme and incubated at 37°C for another 15 hours. The Widom 601 insert was separated from vector backbone by the addition of 0.192 volume equivalents of 4 M NaCl and 0.346 volume equivalents of 40% PEG 6000. Samples were incubated on ice for 1 hour before the vector was pelleted by centrifugation. The Widom 601 insert in the supernatant was further precipitated by the addition of 2.5 volume equivalents of cold 100% ethanol and incubated at -80°C for 1 hour. DNA was pelleted by centrifugation and air dried. Purified DNA was resuspended in 10:0.1 TE (10 mM Tris, 0.1 mM EDTA sodium salt, pH 8.5 at 25°C). 601 DNA purity was verified by 10% agarose gel stained in ethidium bromide and the concentration was measured by using absorbance at 260 nm.

Nucleosome reconstitution. Briefly, 601 DNA and histone octamer were mixed in varying molar ratios at 0.7 mg/mL final DNA concentration in high salt nucleosome reconstitution buffer (2 M NaCl, 20 mM Tris, pH 7.5, 1 mM EDTA, 1 mM DTT). Reactions were incubated on ice for 30 minutes before serially diluting the salt concentration (2 M NaCl, 0.85 M NaCl, 0.65 M NaCl, 0.2 M NaCl) in 1000-fold excess dialysis buffer (20 mM Tris, pH 7.5, 1 mM EDTA, 1 mM DTT) for a total of 20 hours at 4°C. Reconstituted nucleosomes were dialyzed into the final assay buffer (50 mM HEPES, pH 7.2, 100 mM NaCl, 10 mM MgCl₂, 1 μ M ZnCl₂, 1 mM DTT). Nucleosome quality was assessed by native-PAGE (polyacrylamide gel electrophoresis) using 7.5% Tris-glycine gel (BioRad) run under native conditions at 4°C and visualized with ethidium

bromide staining (20 µg/mL). Concentration was measured by absorbance at 260nm (DNA) and converted to total nucleosome concentration.

In vitro ubiquitin ligase assay

Reactions. ATP, E1, E2 and E3 enzymes were diluted in ubiquitination assay buffer (Ring1B/BMI1: 50 mM HEPES, pH 7.2, 100 mM NaCl, 10 mM MgCl₂, 1 µM ZnCl₂, 1 mM DTT; Brca1/Bard1: 50 mM Tris, pH 7.5, 5 mM MgCl₂, 15 mM KCl, 0.7 mM DTT, 2 mM NaF; Trim37: 50 mM Tris, pH 7.9, 5 mM MgCl₂, 2 mM NaF, 10 mM DTT) and incubated with varying concentrations of compound for 30 minutes at room temperatures. 30 µL final volume reactions were initiated with the addition of flag-ubiquitin and mono-nucleosomes (see table 4.7). Reactions were incubated at 30°C for 1.5 hours before being quenched with 10 µL 4xSDS buffer. Samples were boiled for 5 minutes.

Reactant	Amount per reaction	Source
E1	41 ng	BostonBioChem
E2 (UbcH5c)	876 ng	In house
E3 (RingBmi1 fusion wild type or mutant)	42 ng (50 nM)	In house
E3 (Brca1/Bard1)	55 ng (60 nM)	In house
E3 (Trim37)	1 µg (2.5 µM)	In house
ATP	3 µM	Calbiochem
FLAG-ubiquitin	8 µg	BostonBioChem
Mono-nucleosomes	250 ng	In house
Hela nucleosomes	250 ng	Reaction Biology

Table 4.7. Reaction conditions for ubiquitin-ligase assay.

Western blot. Samples were loaded onto 20 well 4-12% bis-Tris-glycine gel (Novex) and ran in 1x MES SDS running buffer (Invitrogen) at room temperature. Western transfer was completed using semi-try transfer apparatus (BioRad) to Immobilon-P PVDF membrane (Millipore). Membranes were blocked with either 5% milk or 5% BSA for 2 hours at room temperature. Membranes were probed with 1:8,000 anti-FLAG (Cell Signaling) or 1:3,000 anti-H2A (Abcam) overnight at 4°C. Membranes were washed for 1 hour with 5 changes in TBST (50 mM Tris-base, 150 mM sodium chloride, pH 7.6 at 25°C, 1% tween-20) at room temperature before being probed with 1:10,000 anti-rabbit IgG-HRP antibody (Cell Signaling) for 1 hour at room temperature. Blots were visualized with Amersham ECL Prime (GE Healthcare).

EMSA Assay

Ring1B/BMI1 was diluted to 1 μ M final concentration in EMSA buffer (50 mM HEPES, pH 7.2, 100 mM NaCl, 10 mM MgCl₂, 1 μ M ZnCl₂, 1 mM DTT, 5% glycerol, 0.01% BSA) and incubated with varying concentrations of compound at room temperature for 90 minutes. H2B_{OG488} Nucleosomes were added to a 10 ng/ μ L final concentration in 10 μ L reaction and reactions were incubated at room temperature for 30 minutes. 1 μ L DNA loading dye was added to each reaction (Life Technologies). Reaction species were separated by electrophoresis at 4° C under native conditions using 7.5% Tris-glycine acrylamide gel (BioRad) that was pre-run to equilibrate current and buffer. Gels were imaged with Typhoon Trio+ 9410 Variable Mode Imager (GE Healthcare) with the following parameters: power set to 600 PMT with Blue laser (488 nm) excitation and 520 nm bandpass filter and the focal plane at the platen surface. Images were analyzed with ImageQuant TL 7.0 (GE Healthcare).

ITC

Ring1B/BMI1 was exchanged into ITC buffer (50 mM phosphate, pH 6.8, 100 mM NaCl, 1 μ M ZnCl₂, 1 mM TCEP) by gel filtration and degassed prior to measurement. The titrations were performed using a VP-ITC titration calorimetric system (MicroCal) at 25°C. The calorimetric cell, containing Ring1B/BMI1 (26.6 μ M) with 5% DMSO, was titrated with compound **85** (270 μ M in 5% DMSO) injected in 10 μ L aliquots. Reference cell contained buffer with 5% DMSO. Data were analyzed using Origin 7.0 (OriginLab) to obtain K_D and stoichiometry.

F. References

1. Cao, R., Tsukada, Y. & Zhang, Y. Role of Bmi-1 and Ring1A in H2A ubiquitylation and Hox gene silencing. *Mol Cell* **20**, 845-854 (2005).
2. Stock, J. K., Giadrossi, S., Casanova, M., Brookes, E., Vidal, M., Koseki, H., Brockdorff, N., Fisher, A. G. & Pombo, A. Ring1-mediated ubiquitination of H2A restrains poised RNA polymerase II at bivalent genes in mouse ES cells. *Nat Cell Biol* **9**, 1428-1435 (2007).
3. Jacobs, J. J., Kieboom, K., Marino, S., DePinho, R. A. & van Lohuizen, M. The oncogene and Polycomb-group gene bmi-1 regulates cell proliferation and senescence through the ink4a locus. *Nature* **397**, 164-168 (1999).
4. Chen, H., Gao, S., Li, J., Liu, D., Sheng, C., Yao, C., Jiang, W., Wu, J., Chen, S. & Huang, W. Wedelolactone disrupts the interaction of EZH2-EED complex and inhibits PRC2-dependent cancer. *Oncotarget* (2015).
5. Shimono, Y., Zabala, M., Cho, R. W., Lobo, N., Dalerba, P., Qian, D., Diehn, M., Liu, H., Panula, S. P., Chiao, E., Dirbas, F. M., Somlo, G., Pera, R. A., Lao, K. & Clarke, M.

- F. Downregulation of miRNA-200c links breast cancer stem cells with normal stem cells. *Cell* **138**, 592-603 (2009).
6. Xu, C. R., Lee, S., Ho, C., Bommi, P., Huang, S. A., Cheung, S. T., Dimri, G. P. & Chen, X. Bmi1 functions as an oncogene independent of Ink4A/Arf repression in hepatic carcinogenesis. *Mol Cancer Res* **7**, 1937-1945 (2009).
 7. Chiba, T., Miyagi, S., Saraya, A., Aoki, R., Seki, A., Morita, Y., Yonemitsu, Y., Yokosuka, O., Taniguchi, H., Nakauchi, H. & Iwama, A. The polycomb gene product BMI1 contributes to the maintenance of tumor-initiating side population cells in hepatocellular carcinoma. *Cancer Res* **68**, 7742-7749 (2008).
 8. Qi, S., Li, B., Yang, T., Liu, Y., Cao, S., He, X., Zhang, P., Li, L. & Xu, C. Validation of Bmi1 as a therapeutic target of hepatocellular carcinoma in mice. *Int J Mol Sci* **15**, 20004-20021 (2014).
 9. Ruan, Z. P., Xu, R., Lv, Y., Tian, T., Wang, W. J., Guo, H. & Nan, K. J. Bmi1 knockdown inhibits hepatocarcinogenesis. *Int J Oncol* **42**, 261-268 (2013).
 10. Su, W. J., Fang, J. S., Cheng, F., Liu, C., Zhou, F. & Zhang, J. RNF2/Ring1b negatively regulates p53 expression in selective cancer cell types to promote tumor development. *Proc Natl Acad Sci U S A* **110**, 1720-1725 (2013).
 11. Douglas, D., Hsu, J. H., Hung, L., Cooper, A., Abdueva, D., van Doorninck, J., Peng, G., Shimada, H., Triche, T. J. & Lawlor, E. R. BMI-1 promotes ewing sarcoma tumorigenicity independent of CDKN2A repression. *Cancer Res* **68**, 6507-6515 (2008).
 12. Kreso, A., van Galen, P., Pedley, N. M., Lima-Fernandes, E., Frelin, C., Davis, T., Cao, L., Baiazitov, R., Du, W., Sydorenko, N., Moon, Y. C., Gibson, L., Wang, Y., Leung, C., Iscove, N. N., Arrowsmith, C. H., Szentgyorgyi, E., Gallinger, S., Dick, J. E. & O'Brien, C. A. Self-renewal as a therapeutic target in human colorectal cancer. *Nat Med* **20**, 29-36 (2014).
 13. Boukarabila, H., Saurin, A. J., Batsche, E., Mossadegh, N., van Lohuizen, M., Otte, A. P., Pradel, J., Muchardt, C., Sieweke, M. & Duprez, E. The PRC1 Polycomb group complex interacts with PLZF/RARA to mediate leukemic transformation. *Genes Dev* **23**, 1195-1206 (2009).
 14. Jagani, Z., Wiederschain, D., Loo, A., He, D., Mosher, R., Fordjour, P., Monahan, J., Morrissey, M., Yao, Y. M., Lengauer, C., Warmuth, M., Sellers, W. R. & Dorsch, M. The Polycomb group protein Bmi-1 is essential for the growth of multiple myeloma cells. *Cancer Res* **70**, 5528-5538 (2010).
 15. Smith, L. L., Yeung, J., Zeisig, B. B., Popov, N., Huijbers, I., Barnes, J., Wilson, A. J., Taskesen, E., Delwel, R., Gil, J., Van Lohuizen, M. & So, C. W. Functional crosstalk between Bmi1 and MLL/Hoxa9 axis in establishment of normal hematopoietic and leukemic stem cells. *Cell Stem Cell* **8**, 649-662 (2011).
 16. Woelk, T., Sigismund, S., Penengo, L. & Polo, S. The ubiquitination code: a signalling problem. *Cell Div* **2**, 11 (2007).
 17. Deshaies, R. J. & Joazeiro, C. A. RING domain E3 ubiquitin ligases. *Annu Rev Biochem* **78**, 399-434 (2009).
 18. Li, W., Bengtson, M. H., Ulbrich, A., Matsuda, A., Reddy, V. A., Orth, A., Chanda, S. K., Batalov, S. & Joazeiro, C. A. Genome-wide and functional annotation of human E3 ubiquitin ligases identifies MULAN, a mitochondrial E3 that regulates the organelle's dynamics and signaling. *PLoS One* **3**, e1487 (2008).

19. Schulman, B. A. & Harper, J. W. Ubiquitin-like protein activation by E1 enzymes: the apex for downstream signalling pathways. *Nat Rev Mol Cell Biol* **10**, 319-331 (2009).
20. Burroughs, A. M., Jaffee, M., Iyer, L. M. & Aravind, L. Anatomy of the E2 ligase fold: implications for enzymology and evolution of ubiquitin/Ub-like protein conjugation. *J Struct Biol* **162**, 205-218 (2008).
21. Huang, L., Kinnucan, E., Wang, G., Beaudenon, S., Howley, P. M., Huibregtse, J. M. & Pavletich, N. P. Structure of an E6AP-UbcH7 complex: insights into ubiquitination by the E2-E3 enzyme cascade. *Science* **286**, 1321-1326 (1999).
22. Metzger, M. B., Hristova, V. A. & Weissman, A. M. HECT and RING finger families of E3 ubiquitin ligases at a glance. *J Cell Sci* **125**, 531-537 (2012).
23. McGinty, R. K., Henrici, R. C. & Tan, S. Crystal structure of the PRC1 ubiquitylation module bound to the nucleosome. *Nature* **514**, 591-596 (2014).
24. Zheng, N., Schulman, B. A., Song, L., Miller, J. J., Jeffrey, P. D., Wang, P., Chu, C., Koepf, D. M., Elledge, S. J., Pagano, M., Conaway, R. C., Conaway, J. W., Harper, J. W. & Pavletich, N. P. Structure of the Cul1-Rbx1-Skp1-F boxSkp2 SCF ubiquitin ligase complex. *Nature* **416**, 703-709 (2002).
25. Berndsen, C. E. & Wolberger, C. New insights into ubiquitin E3 ligase mechanism. *Nat Struct Mol Biol* **21**, 301-307 (2014).
26. Ozkan, E., Yu, H. & Deisenhofer, J. Mechanistic insight into the allosteric activation of a ubiquitin-conjugating enzyme by RING-type ubiquitin ligases. *Proc Natl Acad Sci U S A* **102**, 18890-18895 (2005).
27. Pruneda, J. N., Littlefield, P. J., Soss, S. E., Nordquist, K. A., Chazin, W. J., Brzovic, P. S. & Klevit, R. E. Structure of an E3:E2~Ub complex reveals an allosteric mechanism shared among RING/U-box ligases. *Mol Cell* **47**, 933-942 (2012).
28. Plechanovova, A., Jaffray, E. G., Tatham, M. H., Naismith, J. H. & Hay, R. T. Structure of a RING E3 ligase and ubiquitin-loaded E2 primed for catalysis. *Nature* **489**, 115-120 (2012).
29. Dou, H., Buetow, L., Sibbet, G. J., Cameron, K. & Huang, D. T. BIRC7-E2 ubiquitin conjugate structure reveals the mechanism of ubiquitin transfer by a RING dimer. *Nat Struct Mol Biol* **19**, 876-883 (2012).
30. Li, Z., Cao, R., Wang, M., Myers, M. P., Zhang, Y. & Xu, R. M. Structure of a Bmi-1-Ring1B polycomb group ubiquitin ligase complex. *J Biol Chem* **281**, 20643-20649 (2006).
31. Buchwald, G., van der Stoop, P., Weichenrieder, O., Perrakis, A., van Lohuizen, M. & Sixma, T. K. Structure and E3-ligase activity of the Ring-Ring complex of polycomb proteins Bmi1 and Ring1b. *EMBO J* **25**, 2465-2474 (2006).
32. Bentley, M. L., Corn, J. E., Dong, K. C., Phung, Q., Cheung, T. K. & Cochran, A. G. Recognition of UbcH5c and the nucleosome by the Bmi1/Ring1b ubiquitin ligase complex. *EMBO J* **30**, 3285-3297 (2011).
33. Alchanati, I., Teicher, C., Cohen, G., Shemesh, V., Barr, H. M., Nakache, P., Ben-Avraham, D., Idelevich, A., Angel, I., Livnah, N., Tuvia, S., Reiss, Y., Taglicht, D. & Erez, O. The E3 ubiquitin-ligase Bmi1/Ring1A controls the proteasomal degradation of Top2alpha cleavage complex - a potentially new drug target. *PLoS One* **4**, e8104 (2009).
34. Wen, W., Peng, C., Kim, M. O., Ho Jeong, C., Zhu, F., Yao, K., Zykova, T., Ma, W., Carper, A., Langfald, A., Bode, A. M. & Dong, Z. Knockdown of RNF2 induces apoptosis by regulating MDM2 and p53 stability. *Oncogene* **33**, 421-428 (2014).

35. Calao, M., Sekyere, E. O., Cui, H. J., Cheung, B. B., Thomas, W. D., Keating, J., Chen, J. B., Raif, A., Jankowski, K., Davies, N. P., Bekkum, M. V., Chen, B., Tan, O., Ellis, T., Norris, M. D., Haber, M., Kim, E. S., Shohet, J. M., Trahair, T. N., Liu, T., Wainwright, B. J., Ding, H. F. & Marshall, G. M. Direct effects of Bmi1 on p53 protein stability inactivates oncoprotein stress responses in embryonal cancer precursor cells at tumor initiation. *Oncogene* **32**, 3616-3626 (2013).
36. Zhang, Y., Li, X., Chen, Z. & Bepler, G. Ubiquitination and degradation of ribonucleotide reductase M1 by the polycomb group proteins RNF2 and Bmi1 and cellular response to gemcitabine. *PLoS One* **9**, e91186 (2014).
37. Zhang, W. & Sidhu, S. S. Development of inhibitors in the ubiquitination cascade. *FEBS Lett* **588**, 356-367 (2014).
38. Landre, V., Rotblat, B., Melino, S., Bernassola, F. & Melino, G. Screening for E3-ubiquitin ligase inhibitors: challenges and opportunities. *Oncotarget* **5**, 7988-8013 (2014).
39. Markson, G., Kiel, C., Hyde, R., Brown, S., Charalabous, P., Bremm, A., Semple, J., Woodsmith, J., Duley, S., Salehi-Ashtiani, K., Vidal, M., Komander, D., Serrano, L., Lehner, P. & Sanderson, C. M. Analysis of the human E2 ubiquitin conjugating enzyme protein interaction network. *Genome Res* **19**, 1905-1911 (2009).
40. Sun, Y. Overview of approaches for screening for ubiquitin ligase inhibitors. *Methods Enzymol* **399**, 654-663 (2005).
41. Goldenberg, S. J., Marblestone, J. G., Mattern, M. R. & Nicholson, B. Strategies for the identification of ubiquitin ligase inhibitors. *Biochem Soc Trans* **38**, 132-136 (2010).
42. Gururaja, T. L., Pray, T. R., Lowe, R., Dong, G., Huang, J., Daniel-Issakani, S. & Payan, D. G. A homogeneous FRET assay system for multiubiquitin chain assembly and disassembly. *Methods Enzymol* **399**, 663-682 (2005).
43. Boisclair, M. D., McClure, C., Josiah, S., Glass, S., Bottomley, S., Kamerkar, S. & Hemmila, I. Development of a ubiquitin transfer assay for high throughput screening by fluorescence resonance energy transfer. *J Biomol Screen* **5**, 319-328 (2000).
44. Ungermannova, D., Lee, J., Zhang, G., Dallmann, H. G., McHenry, C. S. & Liu, X. High-throughput screening AlphaScreen assay for identification of small-molecule inhibitors of ubiquitin E3 ligase SCFSkp2-Cks1. *J Biomol Screen* **18**, 910-920 (2013).
45. Davydov, I. V., Woods, D., Safiran, Y. J., Oberoi, P., Fearnhead, H. O., Fang, S., Jensen, J. P., Weissman, A. M., Kenten, J. H. & Vousden, K. H. Assay for ubiquitin ligase activity: high-throughput screen for inhibitors of HDM2. *J Biomol Screen* **9**, 695-703 (2004).
46. Murray, M. F., Jurewicz, A. J., Martin, J. D., Ho, T. F., Zhang, H., Johanson, K. O., Kirkpatrick, R. B., Ma, J., Lor, L. A., Thrall, S. H. & Schwartz, B. A high-throughput screen measuring ubiquitination of p53 by human mdm2. *J Biomol Screen* **12**, 1050-1058 (2007).
47. Hong, C. A., Swearingen, E., Mallari, R., Gao, X., Cao, Z., North, A., Young, S. W. & Huang, S. G. Development of a high throughput time-resolved fluorescence resonance energy transfer assay for TRAF6 ubiquitin polymerization. *Assay Drug Dev Technol* **1**, 175-180 (2003).
48. Orlicky, S., Tang, X., Neduva, V., Elowe, N., Brown, E. D., Sicheri, F. & Tyers, M. An allosteric inhibitor of substrate recognition by the SCF(Cdc4) ubiquitin ligase. *Nat Biotechnol* **28**, 733-737 (2010).

49. Valkov, E., Sharpe, T., Marsh, M., Greive, S. & Hyvonen, M. Targeting protein-protein interactions and fragment-based drug discovery. *Top Curr Chem* **317**, 145-179 (2012).
50. Murray, C. W. & Rees, D. C. The rise of fragment-based drug discovery. *Nat Chem* **1**, 187-192 (2009).
51. Congreve, M., Carr, R., Murray, C. & Jhoti, H. A 'rule of three' for fragment-based lead discovery? *Drug Discov Today* **8**, 876-877 (2003).
52. Hajduk, P. J., Huth, J. R. & Fesik, S. W. Druggability indices for protein targets derived from NMR-based screening data. *J Med Chem* **48**, 2518-2525 (2005).
53. Edfeldt, F. N., Folmer, R. H. & Breeze, A. L. Fragment screening to predict druggability (ligandability) and lead discovery success. *Drug Discov Today* **16**, 284-287 (2011).
54. Bhatnagar, S., Gazin, C., Chamberlain, L., Ou, J., Zhu, X., Tushir, J. S., Virbasius, C. M., Lin, L., Zhu, L. J., Wajapeyee, N. & Green, M. R. TRIM37 is a new histone H2A ubiquitin ligase and breast cancer oncoprotein. *Nature* **516**, 116-120 (2014).
55. Kalb, R., Mallery, D. L., Larkin, C., Huang, J. T. & Hiom, K. BRCA1 is a histone-H2A-specific ubiquitin ligase. *Cell Rep* **8**, 999-1005 (2014).
56. Dedmon, M. M., Lindorff-Larsen, K., Christodoulou, J., Vendruscolo, M. & Dobson, C. M. Mapping long-range interactions in alpha-synuclein using spin-label NMR and ensemble molecular dynamics simulations. *J Am Chem Soc* **127**, 476-477 (2005).
57. Battiste, J. L. & Wagner, G. Utilization of site-directed spin labeling and high-resolution heteronuclear nuclear magnetic resonance for global fold determination of large proteins with limited nuclear overhauser effect data. *Biochemistry* **39**, 5355-5365 (2000).
58. Clore, G. M. & Iwahara, J. Theory, practice, and applications of paramagnetic relaxation enhancement for the characterization of transient low-population states of biological macromolecules and their complexes. *Chem Rev* **109**, 4108-4139 (2009).
59. Gottstein, D., Reckel, S., Dotsch, V. & Guntert, P. Requirements on paramagnetic relaxation enhancement data for membrane protein structure determination by NMR. *Structure* **20**, 1019-1027 (2012).
60. Marley, J., Lu, M. & Bracken, C. A method for efficient isotopic labeling of recombinant proteins. *J Biomol NMR* **20**, 71-75 (2001).
61. Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J. & Bax, A. NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J Biomol NMR* **6**, 277-293 (1995).
62. Goddard, T. D. & Kneller, D. G. *SPARKY 3* (University of California, San Francisco).
63. Tugarinov, V. & Kay, L. E. Quantitative NMR studies of high molecular weight proteins: application to domain orientation and ligand binding in the 723 residue enzyme malate synthase G. *J Mol Biol* **327**, 1121-1133 (2003).
64. Luger, K., Rechsteiner, T. J. & Richmond, T. J. Expression and purification of recombinant histones and nucleosome reconstitution. *Methods Mol Biol* **119**, 1-16 (1999).
65. Luger, K., Rechsteiner, T. J. & Richmond, T. J. Preparation of nucleosome core particle from recombinant histones. *Methods Enzymol* **304**, 3-19 (1999).

Chapter 5. Discussion

A. Conclusions

As outlined in Chapter 1, we were motivated to undertake the work in this dissertation to obtain a more complete understanding of BMI1's role in cancer and to develop strategies to inhibit BMI1 with small molecule inhibitors. The results presented here have expanded our knowledge of the molecular details of BMI1's protein-protein interactions and demonstrated that distinct interactions are important for BMI1 function in regulating cellular proliferation. Further, we hypothesized that targeted disruption of BMI1's protein-protein interactions is an attractive approach to inhibit this protein. Through pursuit of complementary approaches to inhibit BMI1 either through disruption of protein-protein interactions within the PRC1 complex or targeting the E3 ligase we established a number of chemical approaches to advance the understanding of BMI1 function in cancer.

A.1. Structural insights into BMI1 function in PRC1 complex

In Chapter 2 we characterized the BMI1-PHC2 interaction using biophysical methods and determined the 3D structure of the BMI1-PHC2 complex. Through this characterization we identified that BMI1 ULD is also involved in oligomerization through self-association. The structural insights suggest a number of hypotheses about the function of the canonical PRC1 complex. Although a complete understanding of polycomb gene targeting and silencing mechanisms remains elusive^{1,2} it has been demonstrated that vertebrate polycomb PcG proteins are capable of condensing chromatin *in vitro* and in cells suggesting that chromatin compaction is a polycomb transcriptional repression mechanism.^{3,4,5} In this mechanism chromatin

compaction is believed to be a result of polycomb complexes bridging neighboring or distant nucleosome via inter-complex subunit interactions.³ With this in mind we propose that BMI1 self-association may contribute to polycomb complex spreading via direct interactions between PRC1 complexes leading to chromatin compaction and robust gene silencing (Figure 5.1.). Bioinformatics analysis of the published ULD-domain containing polycomb structures by the PISA (Proteins, Interfaces, Surfaces and Assemblies)⁶ server and examination of the sequences of other PCGF orthologs suggests that self-association may be limited to BMI1 and its close ortholog MEL-18/PCGF2. Therefore chromatin compaction may be restricted to the canonical PRC1 complex.

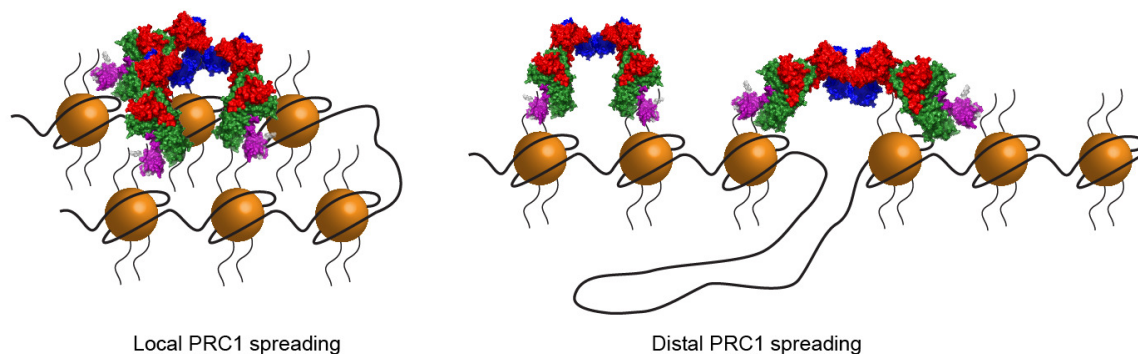


Figure 5.1. Model for role of BMI1-PHC2 oligomerization in PRC1 spreading.

Model for PRC1 spreading over either local (left) or distal (right) chromatin domains through the BMI1-PHC2 oligomerization unit. Subunits are colored as in Figure 5.2.

Recent studies have highlighted the relevance of PRC1 subunit diversity in PcG gene targeting and silencing mechanisms.⁷⁻¹⁰ We speculate that BMI1 and PHC oligomerization could serve to incorporate a variety of PcG proteins into a larger assembly. The so-called variant-PRC1 complexes are defined by the incorporation of other PcG orthologs such as the Ring1B ortholog Ring1A or the BMI1/PCGF4 orthologs, PCGF1-6, with conserved ULD domains.¹¹ Further, PCGF proteins form direct interactions with distinct binding partners, including BCOR, E2F6, KDM2B and L3MBPL2, thereby contributing to complex heterogeneity.^{7,12-14} Interestingly, mass spectrometry studies indicate that all PRC1-like complexes contain both Ring1A and Ring1B orthologs although the functional significance and molecular mechanisms behind this subunit redundancy have not been established.^{4,9,15,16} BMI1 multimerization could facilitate

integration of multiple RING orthologs in the same complex where BMI1-Ring1A and BMI1-Ring1B heterodimers are linked by the BMI1-PhC2 oligomeric unit (Figure 5.2).

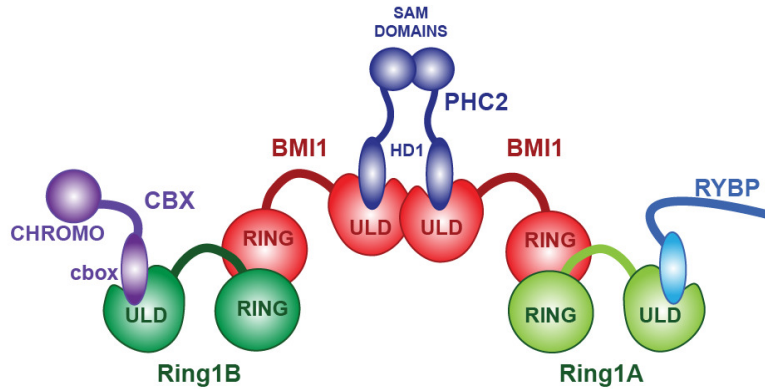


Figure 5.2. Model for role of BMI1-PhC2 oligomerization in PRC1 subunit heterogeneity.

Model for the contribution of BMI1 self-association to PRC1 heterogeneity where complexes incorporating different Ring orthologs and various binding partners are linked through the BMI1-PhC2 oligomerization unit.

Similarly, recent mass spectrometry studies using BMI1 as bait identified both the Ring1B binding partners RYBP and CBX as associated proteins.^{8,9} Structural studies demonstrated that RYBP and CBX interactions with Ring1B are physically mutually exclusive.¹⁷ It is therefore puzzling that in functional studies it was observed that RYBP and CBX can co-localize on the same loci.^{7,10} These observations could be explained by the contribution of homodimerized BMI1 molecules in complex with Ring1B (or Ring1A) molecules that are themselves bound by different binding partners thereby amplifying subunit heterogeneity within a larger complex (Figure 5.2). This analysis supports the hypothesis that the ULD-containing PCGF proteins define the subunit diversity of PRC1 complexes through protein-protein interactions with distinct binding partners.^{7,12}

A.2. Multiple approaches to inhibit BMI1 with small molecules

In Chapters 3 and 4 we demonstrate two approaches to inhibit BMI1 with small molecules: inhibition of BMI1 ULD protein-protein interactions and blocking the E3 ligase activity of the RING domains of Ring1B/BMI1 (Figure 5.3). The small molecules described in these chapters therefore represent chemical tools that can be used to address specific questions about BMI1 function.

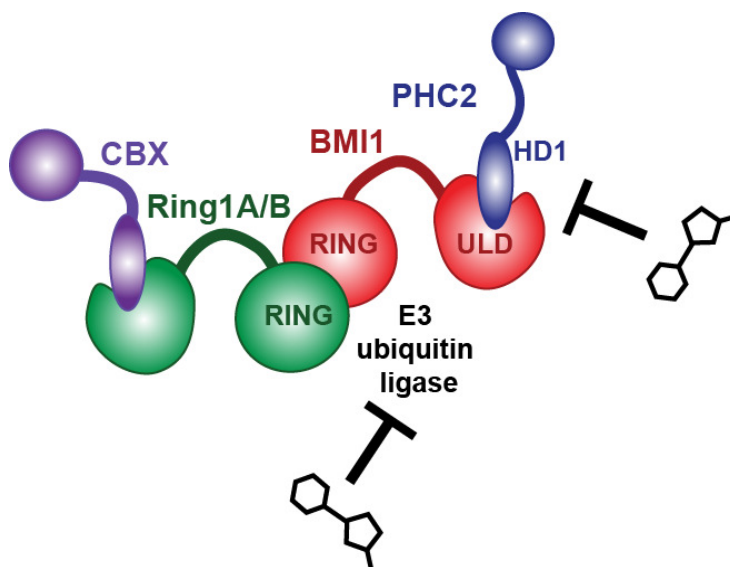


Figure 5.3. Two approaches to inhibit BMI1 with small molecules.

In Chapter 3 through high-throughput screening we identified three classes of compounds that bind directly to the BMI1 ULD. The **BI-1** class of compounds binds to BMI1 at the BMI1-PHC2 interface to orthosterically disrupt this interaction. The **BI-2** class covalently modifies the C-terminal Cys231 of BMI1 ULD. Cys231 is distant from the PHC2 binding site and this suggests an allosteric mechanism to inhibit this interaction. Finally, the **BI-3** fragment demonstrates that smaller ligands can disrupt BMI1 ULD- PHC2 protein-protein interaction. Overall, these results demonstrate the feasibility of identifying small molecules by high-throughput screening to block the BMI1-PHC2 PPI. In cells disruption of this interaction would likely perturb the overall architecture of the PRC1 complex; it remains to be determined if this is an appropriate strategy to inhibit BMI1.

As an alternative approach to inhibit BMI1, in Chapter 4 we focused on developing small molecule inhibitors of the Ring1B/BMI1 E3 ubiquitin ligase. We identified a fragment ligand and optimized it into low micromolar inhibitors of both *in vitro* and cellular Ring1B/BMI1 E3 ligase activity. To our knowledge these molecules represent the first directly binding inhibitors of this complex. We demonstrated a novel mechanism of action where ligands insert into the core of the protein by opening the nucleosome binding loop in Ring1B and thus preventing substrate recognition. These novel molecules demonstrates the feasibility of direct disruption of protein-nucleosome interactions with small molecules. Further, this work supports the development of small molecule modulators of other RING E3 ligases through fragment screening as an opportunity to develop inhibitors of select ubiquitination pathways and thus specific cellular processes.

B. Future Directions

B.1. BMI1 function in protein complexes

B.1.1. BMI1 function within the PRC1 complex

The structural characterization of the BMI1 ULD in Chapter 2 clarifies the role of BMI1 in mediating the interaction between PHC2 and the Ring1B and CBX subunits of the canonical PRC1 complex. A remaining question about BMI1 function is what effect BMI1-PHC2 or BMI1-BMI1 interactions have on the E3 ligase capacity of the complex. It was previously observed that while the close BMI1 ortholog, MEL-18 (PCGF2), can be incorporated into the canonical PRC1 complex, this protein does not contribute to E3 ligase activity of the PRC1 complex.¹⁸ Further, mice deficient in *Mel-18* have a different phenotype than those lacking *Bmi1*^{19,20} and in contrast to BMI1, MEL-18 is suggested to be a tumor suppressor.²¹⁻²³ It remains to be determined if MEL-18 can self-associate in a similar manner as BMI1, although we hypothesize that this may be a unique characteristic of BMI1 and may explain its particular function within the PRC1 complex. Future biochemical or cellular experiments with BMI1 protein-protein interaction mutants or small molecule inhibitors could be used to address this question. Additional work dissecting the spatial arrangements of the PRC1 subunits by electron microscopy (EM), small angle X-ray scattering (SAXS) or mass spectrometry can provide insights into the mechanisms and allostery of intra-complex regulation. Additionally, BMI1 self-association characterized in this work suggests a role for BMI1 in chromatin compaction or

contributing to PRC1 subunit heterogeneity. Future work clarifying the functional role of this interaction by *in vitro*²⁴ or in cell²⁵ chromatin compaction assays will further illuminate BMI1 function within the PRC1 complex.

B.1.2. BMI1 beyond the PRC1 complex

It is tempting to speculate that BMI1 may have other protein binding partners outside the PRC1 complex which contribute to its oncogenic function. While there is evidence from coimmunoprecipitation (Co-IP) and yeast-two-hybrid experiments that BMI1 interacts with the transcription factors E4F1,²⁶ PLZF²⁷ and Zfp277²⁸ there has been no demonstrated direct binding between these proteins by biophysical methods. Therefore, it remains to be robustly evaluated if BMI1 has non-PRC1 binding partners. The 20 amino acid BMI1-interaction motif in PHC proteins identified in Chapter 2 can be used in a bioinformatics search to identify such partners based on sequence identity or similarity. A challenge to this tactic is that the structural characterization other polycomb ULD-binding partner complexes demonstrated that the intermolecular β sheet in these interactions can be formed by disparate parts of the binding partner protein which are not necessarily contiguous in primary sequence. For example, in the structure of PCGF1 and BCOR, BCOR interacts with PCGF1 through an intermolecular β sheet formed by two strands of non-contiguous residues; one strand is from residues 1594 to 1601 and the other strand from 1703 –1707.¹² Therefore, a BLAST search to identify potential BMI1 binding partners is complicated by the requirement for two potential motifs that may not be linear in sequence yet in the tertiary fold can form interactions at the same site.

Alternatively, an experimental proteomics approach could be used to identify BMI1 binding partners in cells. However, a challenge of this method is that due to multiple protein-protein interactions within the PRC1 complex, the majority of BMI1 molecules are likely buried within this complex and pulldown studies may not identify new binding partners. An alternative approach to identify BMI1 protein binding partners is through covalent crosslinking coupled with downstream mass-spectrometry experiments.^{29,30} These techniques are particularly well suited to capture transient interactions, which may be advantageous in identifying non-PRC1 BMI1 binding partners.

B.2. Future efforts to inhibit BMI1's protein-protein interactions with small molecules

The compounds identified by HTS in Chapter 3 demonstrate that the BMI1-PHC2 interaction is amenable to inhibition by small molecules. Going forward we are interested in developing more potent small molecule inhibitors of this interaction by exploring new strategies. First, the covalent modification of Cys231 by the **BI-2** class opens the door for covalent approaches to disrupt this PPI. In addition to the C-terminal cysteine, BMI1 ULD has only one other cysteine which is in the PHC2 binding site (Cys166). As such, screening alkylating agents against the C231A mutant protein may allow identification of covalent ligands that bind in the PHC2 binding site. My colleague Jon Pollock has confirmed by mass spectrometry that Cys166 can be modified by a maleimide-containing compound (data not shown). This proof-of principle experiment demonstrates the viability of this approach and supports future efforts screening covalent ligands targeting this site.

Second, the low hit rate from HTS may be due to the general unsuitability of many current commercial high-throughput screening libraries for identification of protein-protein interaction inhibitors.^{31,32} Applying the robust biochemical assays developed in Chapter 3 to screen natural product or more diverse libraries may yield more promising lead compounds. Finally, as demonstrated by the direct binding of the fragment-like compound **BI-3** to BMI1 ULD a fragment-based drug discovery approach may represent a promising avenue to develop inhibitors of this target.

B.3. Use of BMI1 inhibitors as chemical tools

The molecules developed in Chapter 4 have exciting prospects as chemical tools and future generations of ligands with sub-micromolar activity can be applied to address important biological questions.

B.3.1. Chemical tool for polycomb silencing mechanisms

We have demonstrated *in vitro* that the Ring1B/BMI1 inhibitors developed through this work disrupt nucleosome binding by this E3 ubiquitin ligase. However, in a cellular context Ring1B and BMI1 associate with other proteins that contribute to chromatin targeting, such as CBX7. Thus, in the complex cellular environment it is unlikely that Ring1B would be displaced from chromatin following treatment with these inhibitors. This presents an excellent opportunity

to use these novel chemical probes to address important questions about polycomb gene silencing.

Of particular interest is clarifying the role of H2A monoubiquitination in transcriptional repression and to dissect the molecular mechanisms of different PRC1 complexes. It has been demonstrated that in embryonic stem cells only some polycomb target genes are regulated by histone ubiquitination expression of other genes is repressed by the association of PcG proteins on chromatin.^{5,33} Other recent studies have suggested that some PRC1 complexes do not stimulate H2A ubiquitination,⁹ suggesting that distinct Ring1B binding partners determine the silencing mechanisms at different target loci. The inhibitors developed in this work could be used to explore these heterogeneous systems.

First, chem-seq experiments could be used to clarify precisely which genes or pathways are sensitive to H2A ubiquitination as opposed to Ring1B chromatin association and other protein-protein interactions.^{34,35} These experiments would require the development of a biotinylated inhibitor which can be used as an affinity tag for Ring1B preceding sequencing of cross-linked genes and qRT-PCR analysis of gene transcription. Based on the modeling studies described in Chapter 4 the biotin group could be attached at either the 1 or the 2-position on the unsubstituted indole which are likely solvent exposed.

Second, a biotinylated inhibitor enables proteomics studies by streptavidin pull-down of Ring1B/BMI1- inhibitor complexes to identify Ring1B binding partners. Combined with ChIP-seq methodologies this may allow identification of distinct PRC1 complexes responsible for repression of individual genes through H2Aub-dependent and independent mechanisms. Performing these experiments in multiple cell lines or developmental states could shed light into various transcriptional repression programs in different contexts.

B.3.2. Chemical tool in cancer biology

As discussed in Chapter 1 the polycomb proteins Ring1B and BMI1 are attractive therapeutic targets in many tumor types based on the reduced *in vitro* proliferative capacity and reduced tumor growth observed with BMI1 knockdown.^{27,36-45} However, BMI1 knockout mice have hematopoietic and skeletal abnormalities demonstrating that complete loss of BMI1 through genetic ablation may have toxic effects.^{20,37-39} However, inhibiting BMI1's protein-

protein interaction may provide good opportunities for targeted therapeutic intervention. Further it remains to be determined to what extent BMI1's oncogenic function is dependent on its E3 ligase activity. The inhibitors generated in Chapter 4 could therefore be used as tool compounds to establish if inhibition of the polycomb E3 ligase complex provides a therapeutic window to reduce tumor burden without generating adverse effects.

Additionally, targeting so-called cancer initiating cells has emerged as an exciting area for therapy development to treat treatment-resistant tumors.⁴⁰⁻⁴³ BMI1 has been identified as an important factor in regulating self-renewal and maintenance of these cells in a number of tumor types^{36-39,44} and has thus been proposed as a therapeutic target to inhibit these cancer initiating cell populations.⁴⁵ The compounds reported in Chapter 4 could therefore be used as tools to test the hypothesis that small molecule inhibitors of the Ring1B/BMI1 E3 ubiquitin ligase represent a valid strategy to prevent tumor growth or metastasis by blocking cancer initiating cells.

B.4. Investigation into RING E3 ligase mechanisms

Through the biophysical and biochemical studies in Chapter 4 we demonstrated that potent inhibitors of the Ring1B/BMI1 complex induce significant conformational change in Ring1B to prevent nucleosome binding. This suggests that Ring1B is capable of adopting multiple conformations with different functional consequences; a closed "active" conformation and an "open" inactive conformation. What is not clear is if this conformational change is a consequence of inhibitor binding (induced-fit) or the trapping of an inactive conformation that is natively sampled (conformational selection).⁴⁶ NMR dynamics studies could provide insight into these questions. Our group has obtained robust assignment of NMR spectra of Ring1B and preliminary relaxation dispersion and R1/R2 relaxation data from our lab suggests that the nucleosome binding loop is dynamic with motions on the micro to nanosecond timescale, suggesting that conformational flexibility might contribute to protein regulation. Future studies sorting protein dynamics associated with ligand binding from intrinsic protein motion could provide novel insight into a potential auto-regulatory mechanism for RING E3 ligases.

B.5. Inhibitors of RING E3 ligases

As discussed in Chapter 4, RING E3 ubiquitin ligases represent attractive, yet challenging, targets for inhibitor development to block many cellular pathways. Given the

complicated biochemical assay for ubiquitination, the fragment-based approach used in Chapter 4 to identify ligands of the Ring1B/BMI1 complex likely represents one of the most promising avenues to identify ligands of specific RING domains. Further, as the core of RING domain proteins featuring zinc-coordinating motifs is highly conserved,⁴⁷ targeting the substrate-binding site is the most promising avenue to block activity of specific E3s. It is attractive to speculate that the conformational flexibility observed in the Ring1B protein to regulate substrate binding and prevent E3 ligase activity may be a conserved feature of RING domain E3s. If indeed this is a conserved mechanism, it may be possible to develop small molecule ligands that modulate these conformations to inhibit protein function in a similar manner as the inhibitors developed for Ring1B. Such inhibitors would therefore be specific inhibitors of different cellular ubiquitination pathways and provide novel approaches to regulate cellular processes.

In summary, the work presented here provides insight into BMI1 protein-protein interactions and describes various strategies to inhibit these interactions with small molecules. Together this supports future efforts deploying these chemical tools to study BMI1 in cancer and their development into new potential therapeutics.

C. References

1. Simon, J. A. & Kingston, R. E. Occupying chromatin: Polycomb mechanisms for getting to genomic targets, stopping transcriptional traffic, and staying put. *Mol Cell* **49**, 808-824 (2013).
2. Simon, J. A. & Kingston, R. E. Mechanisms of polycomb gene silencing: knowns and unknowns. *Nat Rev Mol Cell Biol* **10**, 697-708 (2009).
3. Francis, N. J. & Kingston, R. E. Mechanisms of transcriptional memory. *Nat Rev Mol Cell Biol* **2**, 409-421 (2001).
4. Levine, S. S., Weiss, A., Erdjument-Bromage, H., Shao, Z., Tempst, P. & Kingston, R. E. The core of the polycomb repressive complex is compositionally and functionally conserved in flies and humans. *Mol Cell Biol* **22**, 6070-6078 (2002).
5. Eskeland, R., Leeb, M., Grimes, G. R., Kress, C., Boyle, S., Sproul, D., Gilbert, N., Fan, Y., Skoultchi, A. I., Wutz, A. & Bickmore, W. A. Ring1B compacts chromatin structure and represses gene expression independent of histone ubiquitination. *Mol Cell* **38**, 452-464 (2010).
6. Krissinel, E. & Henrick, K. Inference of macromolecular assemblies from crystalline state. *J Mol Biol* **372**, 774-797 (2007).
7. Gao, Z., Zhang, J., Bonasio, R., Strino, F., Sawai, A., Parisi, F., Kluger, Y. & Reinberg, D. PCGF homologs, CBX proteins, and RYBP define functionally distinct PRC1 family complexes. *Mol Cell* **45**, 344-356 (2012).
8. Tavares, L., Dimitrova, E., Oxley, D., Webster, J., Poot, R., Demmers, J., Bezstarosti, K., Taylor, S., Ura, H., Koide, H., Wutz, A., Vidal, M., Elderkin, S. & Brockdorff, N. RYBP-

- PRC1 complexes mediate H2A ubiquitylation at polycomb target sites independently of PRC2 and H3K27me3. *Cell* **148**, 664-678 (2012).
9. Blackledge, N. P., Farcas, A. M., Kondo, T., King, H. W., McGouran, J. F., Hanssen, L. L., Ito, S., Cooper, S., Kondo, K., Koseki, Y., Ishikura, T., Long, H. K., Sheahan, T. W., Brockdorff, N., Kessler, B. M., Koseki, H. & Klose, R. J. Variant PRC1 complex-dependent H2A ubiquitylation drives PRC2 recruitment and polycomb domain formation. *Cell* **157**, 1445-1459 (2014).
 10. Morey, L., Pascual, G., Cozzuto, L., Roma, G., Wutz, A., Benitah, S. A. & Di Croce, L. Nonoverlapping functions of the Polycomb group Cbx family of proteins in embryonic stem cells. *Cell Stem Cell* **10**, 47-62 (2012).
 11. Bezsonova, I., Walker, J. R., Bacik, J. P., Duan, S., Dhe-Paganon, S. & Arrowsmith, C. H. Ring1B contains a ubiquitin-like docking module for interaction with Cbx proteins. *Biochemistry* **48**, 10542-10548 (2009).
 12. Junco, S. E., Wang, R., Gaipa, J. C., Taylor, A. B., Schirf, V., Gearhart, M. D., Bardwell, V. J., Demeler, B., Hart, P. J. & Kim, C. A. Structure of the polycomb group protein PCGF1 in complex with BCOR reveals basis for binding selectivity of PCGF homologs. *Structure* **21**, 665-671 (2013).
 13. Gearhart, M. D., Corcoran, C. M., Wamstad, J. A. & Bardwell, V. J. Polycomb group and SCF ubiquitin ligases are found in a novel BCOR complex that is recruited to BCL6 targets. *Mol Cell Biol* **26**, 6880-6889 (2006).
 14. Trojer, P., Cao, A. R., Gao, Z., Li, Y., Zhang, J., Xu, X., Li, G., Losson, R., Erdjument-Bromage, H., Tempst, P., Farnham, P. J. & Reinberg, D. L3MBTL2 protein acts in concert with PcG protein-mediated monoubiquitination of H2A to establish a repressive chromatin structure. *Mol Cell* **42**, 438-450 (2011).
 15. Elderkin, S., Maertens, G. N., Endoh, M., Mallery, D. L., Morrice, N., Koseki, H., Peters, G., Brockdorff, N. & Hiom, K. A phosphorylated form of Mel-18 targets the Ring1B histone H2A ubiquitin ligase to chromatin. *Mol Cell* **28**, 107-120 (2007).
 16. Wang, H., Wang, L., Erdjument-Bromage, H., Vidal, M., Tempst, P., Jones, R. S. & Zhang, Y. Role of histone H2A ubiquitination in Polycomb silencing. *Nature* **431**, 873-878 (2004).
 17. Wang, R., Taylor, A. B., Leal, B. Z., Chadwell, L. V., Ilangoan, U., Robinson, A. K., Schirf, V., Hart, P. J., Lafer, E. M., Demeler, B., Hinck, A. P., McEwen, D. G. & Kim, C. A. Polycomb group targeting through different binding partners of RING1B C-terminal domain. *Structure* **18**, 966-975 (2010).
 18. Cao, R., Tsukada, Y. & Zhang, Y. Role of Bmi-1 and Ring1A in H2A ubiquitylation and Hox gene silencing. *Mol Cell* **20**, 845-854 (2005).
 19. Akasaka, T., van Lohuizen, M., van der Lugt, N., Mizutani-Koseki, Y., Kanno, M., Taniguchi, M., Vidal, M., Alkema, M., Berns, A. & Koseki, H. Mice doubly deficient for the Polycomb Group genes Mel18 and Bmi1 reveal synergy and requirement for maintenance but not initiation of Hox gene expression. *Development* **128**, 1587-1597 (2001).
 20. van der Lugt, N. M., Domen, J., Linders, K., van Roon, M., Robanus-Maandag, E., te Riele, H., van der Valk, M., Deschamps, J., Sofroniew, M., van Lohuizen, M. & et al. Posterior transformation, neurological abnormalities, and severe hematopoietic defects in mice with a targeted deletion of the bmi-1 proto-oncogene. *Genes Dev* **8**, 757-769 (1994).

21. Guo, W. J., Datta, S., Band, V. & Dimri, G. P. Mel-18, a polycomb group protein, regulates cell proliferation and senescence via transcriptional repression of Bmi-1 and c-Myc oncoproteins. *Mol Biol Cell* **18**, 536-546 (2007).
22. Guo, W. J., Zeng, M. S., Yadav, A., Song, L. B., Guo, B. H., Band, V. & Dimri, G. P. Mel-18 acts as a tumor suppressor by repressing Bmi-1 expression and down-regulating Akt activity in breast cancer cells. *Cancer Res* **67**, 5083-5089 (2007).
23. Lee, J. Y., Park, M. K., Park, J. H., Lee, H. J., Shin, D. H., Kang, Y., Lee, C. H. & Kong, G. Loss of the polycomb protein Mel-18 enhances the epithelial-mesenchymal transition by ZEB1 and ZEB2 expression through the downregulation of miR-205 in breast cancer. *Oncogene* **33**, 1325-1335 (2014).
24. Li, G., Margueron, R., Hu, G., Stokes, D., Wang, Y. H. & Reinberg, D. Highly compacted chromatin formed in vitro reflects the dynamics of transcription activation in vivo. *Mol Cell* **38**, 41-53 (2010).
25. Winter, S. L., Wong, P. & Alexandrow, M. G. In vivo chromatin decondensation assays: molecular genetic analysis of chromatin unfolding characteristics of selected proteins. *Methods Mol Biol* **523**, 27-40 (2009).
26. Chagraoui, J., Hebert, J., Girard, S. & Sauvageau, G. An anticlastogenic function for the Polycomb Group gene Bmi1. *Proc Natl Acad Sci U S A* **108**, 5284-5289 (2011).
27. Boukarabila, H., Saurin, A. J., Batsche, E., Mossadegh, N., van Lohuizen, M., Otte, A. P., Pradel, J., Muchardt, C., Sieweke, M. & Duprez, E. The PRC1 Polycomb group complex interacts with PLZF/RARA to mediate leukemic transformation. *Genes Dev* **23**, 1195-1206 (2009).
28. Negishi, M., Saraya, A., Mochizuki, S., Helin, K., Koseki, H. & Iwama, A. A novel zinc finger protein Zfp277 mediates transcriptional repression of the Ink4a/arf locus through polycomb repressive complex 1. *PLoS One* **5**, e12373 (2010).
29. Pham, N. D., Parker, R. B. & Kohler, J. J. Photocrosslinking approaches to interactome mapping. *Curr Opin Chem Biol* **17**, 90-101 (2013).
30. Suchanek, M., Radzikowska, A. & Thiele, C. Photo-leucine and photo-methionine allow identification of protein-protein interactions in living cells. *Nat Methods* **2**, 261-267 (2005).
31. Sperandio, O., Reynes, C. H., Camproux, A. C. & Villoutreix, B. O. Rationalizing the chemical space of protein-protein interaction inhibitors. *Drug Discov Today* **15**, 220-229 (2010).
32. Pagliaro, L., Felding, J., Audouze, K., Nielsen, S. J., Terry, R. B., Krog-Jensen, C. & Butcher, S. Emerging classes of protein-protein interaction inhibitors and new tools for their development. *Curr Opin Chem Biol* **8**, 442-449 (2004).
33. Endoh, M., Endo, T. A., Endoh, T., Isono, K., Sharif, J., Ohara, O., Toyoda, T., Ito, T., Eskeland, R., Bickmore, W. A., Vidal, M., Bernstein, B. E. & Koseki, H. Histone H2A mono-ubiquitination is a crucial step to mediate PRC1-dependent repression of developmental genes to maintain ES cell identity. *PLoS Genet* **8**, e1002774 (2012).
34. Anders, L., Guenther, M. G., Qi, J., Fan, Z. P., Marineau, J. J., Rahl, P. B., Loven, J., Sigova, A. A., Smith, W. B., Lee, T. I., Bradner, J. E. & Young, R. A. Genome-wide localization of small molecules. *Nat Biotechnol* **32**, 92-96 (2014).
35. Rodriguez, R. & Miller, K. M. Unravelling the genomic targets of small molecules using high-throughput sequencing. *Nat Rev Genet* **15**, 783-796 (2014).
36. Kreso, A., van Galen, P., Pedley, N. M., Lima-Fernandes, E., Frelin, C., Davis, T., Cao, L., Baiazitov, R., Du, W., Sydorenko, N., Moon, Y. C., Gibson, L., Wang, Y., Leung, C.,

- Iscove, N. N., Arrowsmith, C. H., Szentgyorgyi, E., Gallinger, S., Dick, J. E. & O'Brien, C. A. Self-renewal as a therapeutic target in human colorectal cancer. *Nat Med* **20**, 29-36 (2014).
37. Park, I. K., Qian, D., Kiel, M., Becker, M. W., Pihalja, M., Weissman, I. L., Morrison, S. J. & Clarke, M. F. Bmi-1 is required for maintenance of adult self-renewing haematopoietic stem cells. *Nature* **423**, 302-305 (2003).
38. Lessard, J. & Sauvageau, G. Bmi-1 determines the proliferative capacity of normal and leukaemic stem cells. *Nature* **423**, 255-260 (2003).
39. Rizo, A., Olthof, S., Han, L., Vellenga, E., de Haan, G. & Schuringa, J. J. Repression of BMI1 in normal and leukemic human CD34(+) cells impairs self-renewal and induces apoptosis. *Blood* **114**, 1498-1505 (2009).
40. Ischenko, I., Seeliger, H., Schaffer, M., Jauch, K. W. & Bruns, C. J. Cancer stem cells: how can we target them? *Curr Med Chem* **15**, 3171-3184 (2008).
41. McCubrey, J. A., Steelman, L. S., Abrams, S. L., Misaghian, N., Chappell, W. H., Basecke, J., Nicoletti, F., Libra, M., Ligresti, G., Stivala, F., Maksimovic-Ivanic, D., Mijatovic, S., Montalto, G., Cervello, M., Laidler, P., Bonati, A., Evangelisti, C., Cocco, L. & Martelli, A. M. Targeting the cancer initiating cell: the ultimate target for cancer therapy. *Curr Pharm Des* **18**, 1784-1795 (2012).
42. Raggi, C., Mousa, H. S., Correnti, M., Sica, A. & Invernizzi, P. Cancer stem cells and tumor-associated macrophages: a roadmap for multitargeting strategies. *Oncogene* (2015).
43. Beck, B. & Blanpain, C. Unravelling cancer stem cell potential. *Nat Rev Cancer* **13**, 727-738 (2013).
44. Liu, S., Dontu, G., Mantle, I. D., Patel, S., Ahn, N. S., Jackson, K. W., Suri, P. & Wicha, M. S. Hedgehog signaling and Bmi-1 regulate self-renewal of normal and malignant human mammary stem cells. *Cancer Res* **66**, 6063-6071 (2006).
45. Wicha, M. S. Targeting self-renewal, an Achilles' heel of cancer stem cells. *Nat Med* **20**, 14-15 (2014).
46. Boehr, D. D., Nussinov, R. & Wright, P. E. The role of dynamic conformational ensembles in biomolecular recognition. *Nat Chem Biol* **5**, 789-796 (2009).
47. Deshaies, R. J. & Joazeiro, C. A. RING domain E3 ubiquitin ligases. *Annu Rev Biochem* **78**, 399-434 (2009).

Appendix A

*The text and data presented here are adapted from the following manuscript: Gray, F.L.V.; Murai, M.; Grembecka, J.; Cierpicki, T. “Detection of Disordered Regions in Globular Proteins Using ^{13}C -Detected NMR.” *Protein Science*, **2012**, *21*, 1954-60.

A. Abstract

Characterization of disordered regions in globular proteins constitutes a significant challenge. Here, we report development of an approach based on ^{13}C -detected NMR experiments for the identification and assignment of disordered regions in large proteins. Using this method we show that disordered fragments can be accurately identified in menin, a globular protein with a molecular weight over 50 kDa. Our work demonstrates an efficient way to characterize disordered fragments in globular proteins for structural biology applications.

B. Introduction

B.1. Function of disordered regions in proteins

Disordered regions in proteins can play important roles in protein function. These regions are frequently involved in cell signal transduction, transcriptional regulation, molecular recognition and protein regulation through post-translational modification.^{1,2} In particular, unstructured regions in proteins have become recognized as a feature facilitating promiscuous interactions with many protein binding partners through the adoption of different secondary conformations through coupled binding-and-folding.^{3,4} For example, the disordered N and C-termini of p53 have been identified as having over 40 different protein binding partners and structural studies have shown that these sequences can attain varying structures when in complex with different binding partners.⁵ Conversely, some proteins have multiple binding motifs within

a larger disordered region that form interactions with different sites in a globular protein. This is illustrated by the MLL-menin interaction where a natively unstructured region of MLL interacts with the menin protein through two short motifs connected by a flexible linker.^{6,7} Methods to rapidly define the boundaries of regions involved in molecular recognition and to analyze their structural transitions are valuable to advance functional characterization by disordered proteins of these interactions and to facilitate protein-protein interaction inhibitor development.

Additionally, the identification and characterization of disordered regions in proteins has become an important task for computational protein structure prediction and for structural biology.⁸⁻¹¹ Disordered fragments can be predicted using various bioinformatics methods,^{8,11,12} however high resolution experimental validation and biophysical characterization of these regions remains challenging. Accurate methods of identifying disordered fragments in globular proteins is of significant interest to the structural biology community as the presence of flexible protein segments may interfere with production of diffraction quality crystals. Consequently, extensive protein engineering can be required to remove these flexible regions to enable crystallization or improve the quality of protein crystals. For example, the recent X-ray structure of the *Drosophila* effector caspase drICE required the deletion of a highly flexible internal fragment.¹³ Additionally, deletion of internal flexible regions in the GluR2 receptor ligand binding domain resulted in improved diffraction of protein crystals from 2.5 Å to 1.5 Å.^{14,15} Experimental identification of internal disordered regions is commonly based on rapid hydrogen-deuterium exchange rates for solvent exposed amide protons which can be detected using mass spectrometry.^{16,17} Nevertheless, accurate identification of disordered residues remains difficult and an efficient strategy to experimentally detect such fragments in globular proteins would significantly facilitate the design of protein constructs suitable for crystallization.

B.2. ¹³C-detected NMR

NMR is a valuable experimental technique uniquely suited for high resolution studies of disorder in proteins.¹⁸ However, amide proton-detected NMR experiments commonly used for protein studies are hindered in the characterization of intrinsically disordered proteins due to poor resonance dispersion and fast exchange of amides with water limiting the observation of complete set of resonances. To the contrary, NMR experiments directly detecting ¹³C overcome these limitations as random coil carbon chemical shifts have greater dispersion than proton

chemical shifts and observation of ^{13}C is not affected by exchange of amide protons with water.¹⁹ Additional advantages of carbon detected experiments include the observation of resonances corresponding to the backbone $\text{C}\alpha$ and C' carbons allowing for detection of all amino acids, including proline, which is frequently found in disordered regions.²⁰ Finally, carbon detected experiments are less sensitive to environmental conditions such as pH and temperature which can limit the experimental applications of proton detected experiments.²¹

B.3. Model system: the menin protein

As a model system to test a ^{13}C -detected NMR approach to identify both short and long disordered regions that inhibit structural studies we used the menin protein. Menin is a tumor suppressor protein which controls cellular growth in endocrine tissues²² and also functions as an oncogenic cofactor required for leukemogenesis.²³ Structural studies were recently undertaken and while full length menin proved recalcitrant to crystallization experiments, successful crystallization of the protein was achieved through deletion of internal disordered fragments.^{24,25} The first menin to be crystallized was the homolog from *Nematostella vectensis* and crystallization required truncation of the C-terminus and deletion of one internal disordered fragment.²⁴ In this study we evaluated whether this internal disordered fragment could be identified through ^{13}C -detected NMR experiments. As a model protein we chose C-terminally truncated constructs of *Nematostella* menin (*N_menin* Δ C corresponding to residues 1-468).

C. Results

Sequence analysis using the DISOPRED2 server⁸ revealed that all menin homologs have multiple internal regions predicted to be disordered. We first tested whether ^{13}C NMR experiments could be used to identify these internal disordered fragments in *Nematostella* menin. The CACO spectrum of $^{13}\text{C},^{15}\text{N}$ labeled *N_menin* Δ C revealed the presence of approximately 27 resonances (Figure A.1.A). Given the significant molecular weight of the protein (55 kDa), we expect that all observable signals correspond to the most disordered residues. Slow tumbling of the protein molecule leads to very strong broadening of resonances for structured fragments and acts as an efficient filter leaving observable signals only for highly mobile residues. To assign these observed resonances we also collected CBCACO and CANCO experiments.^{26,27} A feature of the ^{13}C -detected experiments employed here is the acquisition of 2D spectra which facilitates complete backbone carbon assignment through the straightforward analysis of the 2D CANCO

and CACO experiments. In each spectra $C\alpha$ chemical shifts are on the y-axis and C' chemical shifts are on the x-axis. In the CANCO spectrum resonances are seen for correlations between the $C'_{i-1}-C\alpha_{i+1}$ chemical shifts and connecting the resonance daisy chain allows for facile assignment by even novice spectroscopists. An overview of this simple assignment strategy is illustrated in Figure A.1.A. The CBCACO experiment was employed as analysis of $C\beta-C'$ correlations was essential for unambiguous assignment due to significantly less peak overlap in this region and because the $C\beta$ chemical shifts allow for the identification of the amino acid type.

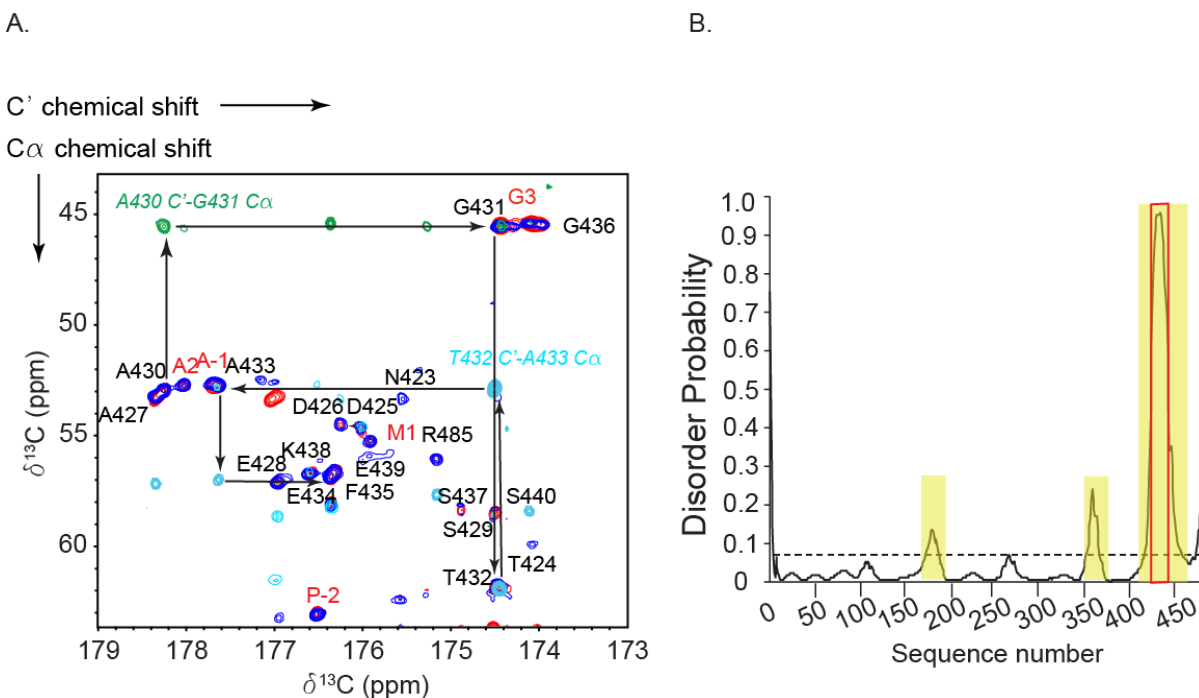


Figure A.1. Assignment of disordered residues in *Nematostella menin*.

A. 2D ^{13}C CACO (blue), CBCACO (red) and CANCO (green/cyan) spectra for *N_menin* ΔC . Arrows illustrate process of sequential assignment for a selection of residues. Residues retained from the vector following TEV cleavage are labeled in red. B. Disorder prediction for *N_menin* ΔC as predicted by DISOPRED2 server.⁸ Dashed line shows the threshold for predicted disordered regions (highlighted in yellow). Red box indicates the experimentally observed disordered fragment of the protein.

Through the sequential assignment procedure outlined above we found that the majority of CACO signals correspond to an internal fragment (residues 423-440), 5 N-terminal residues and 2 C-terminal residues (Figure A.1.A). Several remaining peaks were unassigned due to

reduced intensities. Most likely these peaks correspond to the shorter and less disordered loops. The assigned residues 423-440 yield strong resonances, clearly indicating that this fragment is disordered in solution. Consistently, deletion of residues 426-442 in *Nematostella* menin was necessary to obtain diffraction quality crystals and to determine the X-ray structure of the protein.²⁴

The internal disordered fragment in *Nematostella* menin is fairly short and its assignment based on ¹³C-detected experiments was relatively straightforward. However, unambiguous assignment for more complex proteins with multiple disordered fragments would be more difficult due to increased peak overlap and complexity of 2D spectra. Therefore, we assessed whether assignment of disordered regions based on ¹³C experiments could be facilitated by combining bioinformatics methods for disorder prediction and chemical shift calculation. To test this, we first employed the program DISOPRED2⁸ for the prediction of internal regions of increased disorder in *N_menin*ΔC. Based on this method, three possible internal disordered regions were identified: residues 177-187, 356-367 and 418-456 (Figure A.1.B). We next assumed that these disordered fragments would have chemical shifts consistent with random-coil values, which can be predicted with high accuracy.²⁸⁻³⁰ Thus, we used the ncIDP program²⁸ to generate predicted chemical shifts for these regions. These predicted chemical shifts were used to simulate spectra with Cβ-C', Cα-C' and C'_i-Cα_{i+1} correlations and compared to experimental data for *N_menin*ΔC. Using this approach we found that observed resonances correspond to residues 423-440, consistent with the manual assignment. Overall, this analysis validated the use of chemical shift prediction as a very efficient strategy to aid in completing the assignment of disordered regions in large proteins.

D. Discussion

In summary, we have developed a simple method for the identification and assignment of disordered regions in large proteins based on carbon-detected NMR experiments. We have demonstrated that this approach allows for identification of disordered residues in the 50 kDa menin protein. The assignment of relatively complex spectra can be rapidly achieved through the combination of experimental data with chemical shift calculation. Importantly, the NMR experiments allowed for highly accurate identification of even relatively short disordered

fragments (~10 amino acid long), which are more difficult to predict using bioinformatics methods.¹¹

We applied this method in Chapter 2 to optimize BMI1 ULD constructs for structural biology studies. Further, we used this method to map the protein-protein interaction motif of the intrinsically disordered protein PHC2 which undergoes disorder-to-order transitions when in complex with BMI1. Thus, we demonstrated that the highly sensitive ^{13}C -detected experiments are valuable tools to study protein mechanisms.

The overall approach of this method and its applications is summarized in Figure A.2.

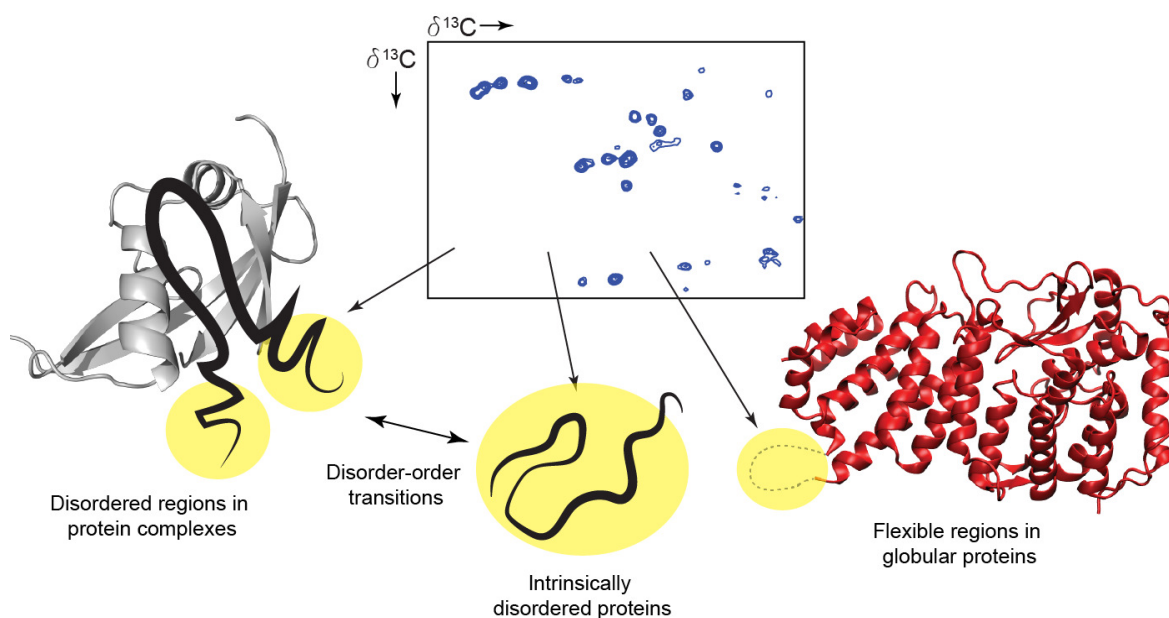


Figure A.2. Applications of ^{13}C -detected NMR approach for rapid characterization of disordered regions of proteins.

E. Materials and Methods:

Protein purification: The synthetic construct encoding *Nematostella* menin was ordered from Genscript and cloned into the pET32a vector. The truncation after residue 487 led to generation of the *N_menin* ΔC which was used for NMR experiments. The $^{13}\text{C},^{15}\text{N}$ labeled *N_menin* ΔC protein was expressed by growing bacterial cells in isotopically enriched M9 minimal media. The purification was carried out following previously described protocol.²⁴

NMR Spectroscopy: For NMR experiments the ^{13}C , ^{15}N -labeled *N_menin* ΔC sample was prepared at a final concentration of 100 μM in 50 mM Tris buffer, pH 7.5, 150 mM NaCl, 1 mM TCEP with 10% D_2O . NMR measurements were performed using a Bruker Advance III 600-MHz spectrometer equipped with 5 mm TCI cryogenic probe. The following parameters were used for ^{13}C -detected experiments: 2D CACO²⁶ data size: 64 (t_1) x 512 (t_2) complex points, $t_{1\text{max}}(^{13}\text{C}) = 16$ ms, $t_{2\text{max}}(^{13}\text{C}) = 85.2$ ms; 2D CBCACO²⁶ data size: 70 (t_1) x 512 (t_2) complex points, $t_{1\text{max}}(^{13}\text{C}) = 7.8$ ms, $t_{2\text{max}}(^{13}\text{C}) = 85.2$ ms; 2D CANCO²⁷ data size 50 (t_1) x 512 (t_2) complex points, $t_{1\text{max}}(^{13}\text{C}) = 7.4$ ms, $t_{2\text{max}}(^{13}\text{C}) = 85.2$ ms. All ^{13}C detected experiments were recorded with ^1H excitation in order to increase the sensitivity and processed with the IPAP scheme for decoupling. These experiments were recorded with 1 second relaxation delay and 32, 64, 448 scans per increment, respectively. This lead to total acquisition times of 2.5, 6 and 31 hours. All experiments were collected at 25° C. Spectra were processed with NMRPipe³¹ and analyzed with Sparky.³²

F. References

1. Dyson, H. J. & Wright, P. E. Intrinsically unstructured proteins and their functions. *Nat Rev Mol Cell Biol* **6**, 197-208 (2005).
2. Dunker, A. K., Brown, C. J., Lawson, J. D., Iakoucheva, L. M. & Obradovic, Z. Intrinsic disorder and protein function. *Biochemistry* **41**, 6573-6582 (2002).
3. Wright, P. E. & Dyson, H. J. Linking folding and binding. *Curr Opin Struct Biol* **19**, 31-38 (2009).
4. Turoverov, K. K., Kuznetsova, I. M. & Uversky, V. N. The protein kingdom extended: ordered and intrinsically disordered proteins, their folding, supramolecular complex formation, and aggregation. *Prog Biophys Mol Biol* **102**, 73-84 (2010).
5. Oldfield, C. J. & Dunker, A. K. Intrinsically disordered proteins and intrinsically disordered protein regions. *Annu Rev Biochem* **83**, 553-584 (2014).
6. Grembecka, J., Belcher, A. M., Hartley, T. & Cierpicki, T. Molecular basis of the mixed lineage leukemia-menin interaction: implications for targeting mixed lineage leukemias. *J Biol Chem* **285**, 40690-40698 (2010).
7. Shi, A., Murai, M. J., He, S., Lund, G., Hartley, T., Purohit, T., Reddy, G., Chruszcz, M., Grembecka, J. & Cierpicki, T. Structural insights into inhibition of the bivalent menin-MLL interaction by small molecules in leukemia. *Blood* **120**, 4461-4469 (2012).
8. Ward, J. J., Sodhi, J. S., McGuffin, L. J., Buxton, B. F. & Jones, D. T. Prediction and functional analysis of native disorder in proteins from the three kingdoms of life. *J Mol Biol* **337**, 635-645 (2004).
9. Wang, R. Y., Han, Y., Krassovsky, K., Sheffler, W., Tyka, M. & Baker, D. Modeling disordered regions in proteins using Rosetta. *PLoS One* **6**, e22060 (2011).

10. Bordoli, L., Kiefer, F. & Schwede, T. Assessment of disorder predictions in CASP7. *Proteins* **69 Suppl 8**, 129-136 (2007).
11. He, B., Wang, K., Liu, Y., Xue, B., Uversky, V. N. & Dunker, A. K. Predicting intrinsic disorder in proteins: an overview. *Cell Res* **19**, 929-949 (2009).
12. Deng, X., Eickholt, J. & Cheng, J. A comprehensive overview of computational protein disorder prediction methods. *Mol Biosyst* **8**, 114-121 (2012).
13. Li, X., Wang, J. & Shi, Y. Structural mechanisms of DIAP1 auto-inhibition and DIAP1-mediated inhibition of drICE. *Nat Commun* **2**, 408 (2011).
14. Armstrong, N., Sun, Y., Chen, G. Q. & Gouaux, E. Structure of a glutamate-receptor ligand-binding core in complex with kainate. *Nature* **395**, 913-917 (1998).
15. Chen, G. Q., Sun, Y., Jin, R. & Gouaux, E. Probing the ligand binding domain of the GluR2 receptor by proteolysis and deletion mutagenesis defines domain boundaries and yields a crystallizable construct. *Protein Sci* **7**, 2623-2630 (1998).
16. Pantazatos, D., Kim, J. S., Klock, H. E., Stevens, R. C., Wilson, I. A., Lesley, S. A. & Woods, V. L., Jr. Rapid refinement of crystallographic protein construct definition employing enhanced hydrogen/deuterium exchange MS. *Proc Natl Acad Sci U S A* **101**, 751-756 (2004).
17. Sharma, S., Zheng, H., Huang, Y. J., Ertekin, A., Hamuro, Y., Rossi, P., Tejero, R., Acton, T. B., Xiao, R., Jiang, M., Zhao, L., Ma, L. C., Swapna, G. V., Aramini, J. M. & Montelione, G. T. Construct optimization for protein NMR structure analysis using amide hydrogen/deuterium exchange mass spectrometry. *Proteins* **76**, 882-894 (2009).
18. Dyson, H. J. & Wright, P. E. Unfolded proteins and protein folding studied by NMR. *Chem Rev* **104**, 3607-3622 (2004).
19. Felli, I. C. & Pierattelli, R. Recent progress in NMR spectroscopy: toward the study of intrinsically disordered proteins of increasing size and complexity. *IUBMB Life* **64**, 473-481 (2012).
20. Bermel, W., Bertini, I., Felli, I. C., Lee, Y. M., Luchinat, C. & Pierattelli, R. Protonless NMR experiments for sequence-specific assignment of backbone nuclei in unfolded proteins. *J Am Chem Soc* **128**, 3918-3919 (2006).
21. Gil, S., Hosek, T., Solyom, Z., Kummerle, R., Brutscher, B., Pierattelli, R. & Felli, I. C. NMR spectroscopic studies of intrinsically disordered proteins at near-physiological conditions. *Angew Chem Int Ed Engl* **52**, 11808-11812 (2013).
22. Chandrasekhara, S. C., Guru, S. C., Manickam, P., Olufemi, S. E., Collins, F. S., Emmert-Buck, M. R., Debelenko, L. V., Zhuang, Z., Lubensky, I. A., Liotta, L. A., Crabtree, J. S., Wang, Y., Roe, B. A., Weisemann, J., Boguski, M. S., Agarwal, S. K., Kester, M. B., Kim, Y. S., Heppner, C., Dong, Q., Spiegel, A. M., Burns, A. L. & Marx, S. J. Positional cloning of the gene for multiple endocrine neoplasia-type 1. *Science* **276**, 404-407 (1997).
23. Yokoyama, A., Somerville, T. C., Smith, K. S., Rozenblatt-Rosen, O., Meyerson, M. & Cleary, M. L. The menin tumor suppressor protein is an essential oncogenic cofactor for MLL-associated leukemogenesis. *Cell* **123**, 207-218 (2005).
24. Murai, M. J., Chruszcz, M., Reddy, G., Grembecka, J. & Cierpicki, T. Crystal structure of menin reveals binding site for mixed lineage leukemia (MLL) protein. *J Biol Chem* **286**, 31742-31748 (2011).

25. Huang, J., Gurung, B., Wan, B., Matkar, S., Veniaminova, N. A., Wan, K., Merchant, J. L., Hua, X. & Lei, M. The same pocket in menin binds both MLL and JUND but has opposite effects on transcription. *Nature* **482**, 542-546 (2012).
26. Bermel, W., Bertini, I., Duma, L., Felli, I. C., Emsley, L., Pierattelli, R. & Vasos, P. R. Complete assignment of heteronuclear protein resonances by protonless NMR spectroscopy. *Angew Chem Int Ed Engl* **44**, 3089-3092 (2005).
27. Bermel, W., Bertini, I., Felli, I. C., Pierattelli, R. & Vasos, P. R. A selective experiment for the sequential protein backbone assignment from 3D heteronuclear spectra. *J Magn Reson* **172**, 324-328 (2005).
28. Tamiola, K., Acar, B. & Mulder, F. A. Sequence-specific random coil chemical shifts of intrinsically disordered proteins. *J Am Chem Soc* **132**, 18000-18003 (2010).
29. Schwarzingers, S., Kroon, G. J., Foss, T. R., Chung, J., Wright, P. E. & Dyson, H. J. Sequence-dependent correction of random coil NMR chemical shifts. *J Am Chem Soc* **123**, 2970-2978 (2001).
30. Wang, Y. & Jardetzky, O. Probability-based protein secondary structure identification using combined NMR chemical-shift data. *Protein Sci* **11**, 852-861 (2002).
31. Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J. & Bax, A. NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J Biomol NMR* **6**, 277-293 (1995).
32. Goddard, T. D. & Kneller, D. G. *SPARKY 3* (University of California, San Francisco).