Biophysical Approaches to Target Protein-Protein Interactions of Epigenetic Complexes by Small-Molecule Inhibitors

by

Brian Linhares

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

Doctoral Committee:

Associate Professor Tomasz Cierpicki, Chair
Professor Charles L. Brooks, III
Associate Professor Jolanta Grembecka
Professor John J. G. Tesmer, Purdue University

Brian Matthew Linhares

linhares@umich.edu

ORCID ID: 0000-0002-0906-6191

DEDICATION

To my father and mother, Rod and Lisa Linhares

iii

knowledgeable than before (or so I'd like to think). Lastly, I would like to give a special shout-out to my lab bay partner of the past two years, graduate student Christina Howard. Though not on any of my projects, she has been a valued comrade and sounding board, and has made the many hours spent in the lab – including the *occasional* weekend or holiday – much more enjoyable. Over one year later, I will still never cease to be amazed by the amount of work done, with machinelike efficiency, on the Fourth of July, 2018.

I must thank my dissertation committee, Profs. Charles L. Brooks III, Jolanta Grembecka, and John J. G. Tesmer. In the three years following my qualifying exams, my committee has been vital to my progression through my doctoral training, and I am profoundly grateful for letters of recommendation on my behalf for my post-doctoral fellowship applications. In addition, as Prof. Grembecka is a Principal Investigator of the Cierpicki and Grembecka Labs, I have to thank her for her mentorship, insight, and financial support on various projects over the years.

I will always be indebted to my mentor, Prof. Tomasz Cierpicki, for his scientific mentorship and training; for bestowing guidance through the long, arduous process of doctoral training; and for support over the years, financial, professional, or otherwise. I never hesitate to take for granted the good fortune by which he accepted me as his student, and I am proud of what we have accomplished over the past half-decade. Moreover, I see myself as incredibly privileged for having had the opportunity to work alongside the talented, dedicated, diligent members of the Cierpicki and Grembecka Labs, and to be just one part, however small, of the remarkable science that is done in this group on a routine basis.

I would like to conclude with an expression of gratitude to my friends and family. I am thankful for the support from my friends from my undergraduate studies, who to this day remain

my close friends, Tyler Boyd, Ronald Bruntz, Ph.D., Zach Heath, Adam Kerbis, Alex Kurtz, and Dan Nechamkin; from my gap years, the "In-N-Out crew" comprised of Grayson Kochi, Danny Leong, and Jaime Tawney; and from my graduate studies, who include (but certainly are not limited to) fellow then-doctoral students Tyler Beyett, Amy Fraley, Jared Houghtaling, and Meredith Skiba, and summer undergraduate trainee Megan Hoinville. To all who have assisted on my journey yet whose names elude me at this moment, please accept my apologies, for it is the fallibility of my own memory, as well as the constraints on statement length inherent to this medium, that preclude recognition – not your contributions.

I extend my most sincere appreciation to my extended family: my maternal grandparents, Ann and Fred; my late paternal grandparents, Rod and Josephine; my aunts, Amy and Julie; my uncles, Vince and Michael; my great-aunt, Leslie; and my cousins, Nick and Kim and Wesley, the latter of whom is celebrating his third birthday on the very day of this writing, and also of whom I hope to count as a friend in the upcoming years. Last but certainly not in the least, I must give special thanks to my sisters, Hannah and Sarah, for their camaraderie and kinship. To all of my family: I am forever grateful for the collective decency, loyalty, support, friendship, advice, and companionship, plus any number of additional offerings that I am surely forgetting.

Finally, I will conclude with the dedication: to my mother, Lisa, and father, Rod. Neither can the extent of their contributions, nor can the depths of my gratitude, be adequately put into words. And, try as I might, my efforts would undoubtedly be in vain. Begrudgingly, I must defer to a phrase from my mother, a personal favorite of hers: "Talk is cheap". Point taken.

And so, without further ado, it is you, mom and dad, to whom this work is dedicated.

Ann Arbor, Michigan

July 2019

Table of Contents

List of Figures

List of Tables

List of Schemes

List of Equations

List of Abbreviations

6PA –   6 Peak Addition

ABC –   Activated B-Cell

AF9 –   ALL-Fused gene from chromosome 9 protein

ALL –   Acute Lymphoid Leukemia

Alpha –   amplified luminescent proximity homogeneous assay (AlphaLisa/AlphaScreen assays)

AML –   Acute Myeloid Leukemia

B-ALL –   BCR-ABL-positive B-cell precursor Acute Lymphoblastic Leukemia

BBD –   BCL6 Binding Domain (of BCL6 co-repressors SMRT and BCoR)

BCL6 –   B-Cell Lymphoma protein 6

BCoR –   BCL-6 Co-Repressor

BLI –   bio-layer interferometry

BP-CML –   Blast-Phase Chronic Myeloid Leukemia

BPI –   BCL6 Peptide Inhibitor

BTB –   Bric-*à*-brac, Tamtrack, Broad complex domain

C_2H_2 –   Cysteine$_2$-Histidine$_2$ (zinc finger motif)

CC –   Coiled-coil (domain)

CD –   circular dichroism

ChIP –   Chromatin Immunoprecpitation

CSP –   chemical shift perturbation

DLBCL –   Diffuse Large B-Cell Lymphoma

DLG –   Dymytrii Listunov GAS41 (scaffold nomenclature)

DNA –   deoxyribonucleic acid

ENL –   Eleven-Nineteen-Leukemia protein

EFS –   event-free survival rate

FBLD – fragment-based lead discovery

FBS – fragment-based screening

FL – Follicular Lymphoma

FP – fluorescence polarization assay

GAS41– Glioma-Amplified Sequence 41

GC – Germinal Center (of B-Cell maturation)

GI$_{50}$ – half-maximal growth-inhibition concentration

GPCR – G-protein coupled-receptor

GST – Glutathione S-Transferase

HA – heavy (non-hydrogen) atoms

HDAC – histone deacetylase

HDCH – Histidine14-Aspartic acid (D)17-Cysteine53-Histidine116 (pocket at BCL6^{BTB} LG, SMRT/BCoR binding site)

HEK 293T – human embryonic kidney 293 cells, containing SV40 T-antigen

HEPES – 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid

HSQC – heteronuclear single quantum coherence (two-dimensional NMR experiment)

HTS – high-throughput screening

IC$_{50}$ – half-maximal inhibitory concentration

IPTG – isopropyl 1-thio-β-D-galactopyranoside

ITC – isothermal titration calorimetry

K_D – dissociation constant

LAC – Lung Adenocarcinoma

LB – Luria Broth (*E. Coli* growth medium)

LCC – large cell carcinoma

LE – ligand efficiency

LG – lateral groove (motif of BCL6^{BTB})

Lg(α%) – log10 value of remaining AlphaLisa signal (in percentage) in presence of inhibitor at 50 μM

LRF – Leukemia/Lymphoma-Related Factor protein

LSCC – lung squamous cell carcinoma

M9 – Miller 9 (isotope-labeled mimimal *E. Coli* growth medium)

MBM1/2 – Menin Binding Motif 1 or 2 of MLL (*see below*)

MI – Menin Inhibitor (scaffold nomenclature)

MIZ1 – Myc-Interacting Zinc finger protein 1

MLL – Mixed Lineage Leukemia 1 protein

MST – microscale thermophoresis

MW – molecular weight

NanoBit – NanoLuc Binary Technology

NCoR – Nuclear-receptor Co-Repressor

Ni-NTA – nickel-nitrilotriacetic acid (resin)

NMR – nuclear magnetic resonance

NSCLC – Non-Small Cell Lung Cancer

OD_{600} – optical density at 600 nm

PCR – polymerase chain reaction

PDB – Protein Data Bank

PPI – protein-protein interaction

SAR – structure-activity relationship

SASA – solvent-accessible surface area

SD – standard deviation

shRNA – short hairpin RNA

SMRT – Silencing Mediator of Retinoic Acid and Thyroid Hormone

TCEP – tris(2-carboxyethyl)phosphine

TEV – Tobacco Etch Virus protease

TKI – Tyrosine kinase inhibitors

T_m – melting temperature

TNBC – triple-negative breast cancer (cell lines)

YEATS – Yaf9, Enl, Af9, Taf14, Sas5 (epigenetic reader domain)

YEATS2 – YEATS protein 2

Abstract

Innumerable biological systems rely on Protein-Protein Interactions (PPIs) to govern basal physiological processes. Deregulation of PPIs is implicated in various human diseases, and PPIs have thus been frequently promoted as pharmacological targets. Yet, physiochemical and structural characteristics of binding interfaces, as well as modest affinities for binding partners, have yielded tremendous challenges in targeting PPIs with small-molecule inhibitors. Here we demonstrate biophysics-guided methodologies to reveal the mechanistic and structural bases of a tumorigenic PPI, and to identify and optimize small-molecule inhibitors of PPIs implicated in oncogeneses by Fragment-based lead discovery (FBLD) campaigns. We targeted GAS41 and BCL6, implicated as oncogenic drivers of non-small cell lung cancers and diffuse large B-cell lymphomas, respectively. First, we present biophysical approaches to characterize a bivalent mode of recognition of site-specific acetylated-Histone H3 by histone reader protein GAS41. We demonstrate that the bivalent mode of recognition confers an improvement in affinity by an order of magnitude, and further solve the crystal structure of complex to reveal the molecular details of Histone H3 acetyl-lysine binding GAS41 YEATS (Yaf9, ENL, Af9, Taf14, Sas5) domain. Subsequently, we report the development of small-molecule inhibitors of GAS41 histone reader function, and of BCL6-co-repressor interaction. We highlight a rational design strategy in which the molecular details of macromolecular recognition are incorporated into optimization of small-molecule inhibitors. In addition, we develop NMR-based methodology to quantify and rank binding of compounds in early-stage FBLD projects. We used protein-NMR-based fragment screens against GAS41 YEATS and BCL6 BTB domains to identify hits, and we applied protein-NMR, x-ray

crystallography, biophysical studies, and biochemical assays to lead design of compounds targeting GAS41 and BCL6 with nanomolar and mid-micromolar inhibitory activities, respectively, *in vitro*. In total, our work presents an integrated, iterative strategy to characterize and block PPIs that may be deemed intractable to conventional small-molecule inhibitor development approaches.

Chapter 1

Introduction

A. Motivation

Protein-protein interactions (PPIs) are ubiquitous in biological systems and pathways, governing a myriad of physiological processes, such as chromatin remodeling[1], transcriptional activation or repression[2], signal transduction[3], cell growth and division[4], and cell metabolism[5]. Furthermore, PPIs and multi-protein complexes comprise epigenetic complexes, responsible for one of either recognition of epigenetic markings on histone lysine residues via reader proteins; addition of markings to histone lysine residues via writer proteins; or removal of markings from histone lysine residues via eraser proteins[6]. The orchestrated function of histone readers, writers, and erasers figures prominently in regulation of chromatin structure and gene activation.

By extension, deregulation of epigenetic PPIs is implicated in numerous human diseases, given the basal roles that histone acetyl-lysine mediated PPIs play in integral processes spanning various biological functions[7,8]. Particularly, given the significance of such in gene regulation and in chromatin remodeling, many human cancers present oncogenes dependent on epigenetic PPIs to drive oncogenic transformation[2,9,10]. For this reason, we have identified several PPIs as pharmacological targets for therapeutic intervention against varying types of cancers.

1

Yet, PPIs possess biophysical, physiochemical, and structural properties that pose

challenges in design of small-molecule inhibitors. Specifically, binding interfaces present

large, hydrophobic, flat surfaces seen in classically defined "undruggable" or challenging

targets[11], and moreover may often display modest affinities for endogenous binding partners

(K_Ds >10 μM)[12,13,14], which begets multiple questions, including what methods are appropriate

to quantify binding, how much affinity is *to be had* at such binding sites, and how to develop

weakly-binding hits into potent small-molecule inhibitors. Here we present a two-pronged

biophysics- and structural biology-based approach to target difficult-to-prosecute PPIs. The

first is characterization of the biophysical, mechanistic, and structural bases of a molecular

recognition event of histone acetyl-lysine by an epigenetic histone reader protein that is

implicated in oncogenesis of a sub-type of non-small cell lung cancer. The second is

identification and optimization of small-molecule inhibitors in two cases of epigenetic PPIs,

validated as indispensible to oncogenic transformations in their respective cancers, by

fragment-based lead discovery, guided by biophysics-, chemical biology-, and structural

biology-based methodologies.

B. Epigenetic protein-protein interactions (PPIs) as pharmacological targets in various

human cancers

B.1. PPIs as regulators of epigenetic systems

Epigenetic systems regulate gene expression independent of genetic sequence

(reviewed extensively in[6]). This principally occurs via the addition of epigenetic markings to

either DNA or Histone proteins, which are comprised of modifications of any number of

functional groups, including acetyl[15], butyryl[16], crotonyl[17], methyl[18], propionyl[16], and

succinyl[19]. Lysine residues on histone proteins are commonly targeted for epigenetic

markings, which exert demonstrable effects on chromatin remodeling and condensation[15].

Three distinct classes of epigenetic proteins are responsible for regulating histone lysine modification: histone reader proteins, which recognize or "read" modifications; histone writer proteins, which add modifications; and histone eraser proteins, which remove modifications[6].

Epigenetic reader proteins act in concert with writer or eraser proteins, and moreover often act within multi-protein complexes that regulate transcription[8]. As well as reader, writer, and eraser proteins, scaffolding proteins play fundamental roles in the function of gene activation or repression, via recruitment of transcriptional activator or repressor proteins within the context of multi-protein complexes[20].

B.2. Epigenetic PPIs implicated in oncogeneses of several human cancers

Extensive *in cell*, *in vivo*, and translational studies have yielded insight into the roles of epigenetic PPIs in oncogenic transformations of various cancers (reviewed extensively in [2,7,9,10]). In particular, we have selected two PPIs for inhibitor development campaigns, B-Cell Lymphoma protein 6 (BCL6)-co-repressor interactions and Glioma-Amplified Sequence 41 (GAS41)-acetylated-Histone H3 interactions, identified as oncogenic drivers in diffuse large B-cell lymphoma (DLBCL)[21] and non-small cell lung cancer (NSCLC)[22], respectively.

B.2.1. BCL6 is constitutively overexpressed in diffuse large B-cell lymphomas (DLBCLs)

An estimated ~30 – 40% of all DLBCLs and ~5 – 15% of all follicular lymphomas present chromosomal translocations targeting the *BCL6* gene[23], primarily localized at a 5-prime non-coding exon[24,25]. This disrupts a negative regulatory feedback loop by which BCL6 expression is auto-regulated[26]. Constitutive overexpression of BCL6 has been demonstrated to recapitulate lymphomagenesis in a murine model, consistent with DLBCLs

seen in humans as characterized by immunohistochemical and genetic techniques[27]. In otherwise normal B-Cell development, BCL6 governs affinity antibody maturation in Germinal Center B-Cells (GC B-Cells), by regulation of gene pathways involved in DNA damage response, cell cycle arrest, and cellular differentiation, among other cellular functions[28]. Affinity antibody maturation is relatively quite genotoxic among physiological processes, as the rapid generation of antibodies with maximal clonal diversity mandates somatic hypermutation and class-switch recombination[29]. GC B-Cells must permit these events for affinity antibody maturation[29]. BCL6's dual functions as transcriptional repressor and scaffolding protein regulate the formation of transcriptional repressor complexes that include eraser proteins Histone Deacetlyases (HDACs)[30]. Such complexes regulate expression of numerous downstream target genes that include differentiation factors, cyclin-dependent kinases, and tumor suppressor pathways[31,32,33,34,35].

BCL6 co-repressors include silencing mediator of retinoic acid and thyroid hormone (SMRT)[36], and BCL6 co-repressor (BCoR)[37]. SMRT has been reported to interact with histone eraser proteins HDAC3, which is responsible for Histone H3K27 de-acetylation, and further is associated with control of transcriptional programming in normal and malignant B-Cells[38]. BCL6 recruits co-repressors SMRT and BCoR via its N-terminal BTB domain (BCL6^{BTB}), which bind BCL6-Binding Domains of SMRT[12] and BCoR[13], SMRT^{BBD} and BCoR^{BBD}, respectively. BCoR^{BBD} presents approximately 10-fold higher affinity for BCL6^{BTB} over SMRT^{BBD}, as assessed by Isothermal Titration Calorimetry (ITC) experiment (BCL6^{BTB}-BCoR^{BBD} K_D = 1.3 μM, whereas BCL6^{BTB}-SMRT^{BBD} K_D = 11.4 μM)[12,13].

B.2.2. GAS41 is a pharmacological target in non-small cell lung cancers (NSCLCs)

GAS41 is a histone reader protein responsible for recognition of Histone H3 acetyl-lysine residues, and further regulates formation of multi-protein complexes responsible for H2A.Z deposition[47]. Recent work has demonstrated that GAS41 binds specific acetylated-[14,22,47], crotonylated-[14], and succinylated-Histone H3[48], via its N-terminal YEATS domain (GAS41 YEATS). Biophysical studies presented generally modest affinities of GAS41 YEATS for Histone H3 peptides, with K_Ds ≥ 10 μM[14,22,48].

GAS41 is amplified in NSCLC sub-type lung adenocarcinoma (LAC), and GAS41 knockdown has been demonstrated to attenuate proliferation and tumor growth in LAC cell lines[22]. In addition, GAS41 YEATS has been implicated directly in tumorigenesis, as transfection of *wild-type (wt)*-GAS41 was able to restore the oncogenic phenotype in GAS41-dependent LAC cells, whereas GAS41-mutants Y74A and W93A were not[22]. Structural studies have revealed that Tyr74 and Trp93 comprise the acetyl-lysine binding pocket of GAS41, and therefore are integral to acetylated-Histone H3 recognition by GAS41[14,22]. This result demonstrates that epigenetic reader function of GAS41 YEATS via recognition of acetyl-lysine is necessary for oncogenic transformation in GAS41-dependent NSCLCs. As of this writing, there are no reported inhibitors of GAS41.

C. Challenges in design of small-molecule inhibitors targeting PPIs

C.1. Structural and physiochemical properties of challenging drug targets

PPI interfaces have demonstrated structural and physiochemical properties that present challenges in development of small-molecule inhibitors. Their binding sites are characterized by interfaces that have large, flat, "feature-less" topologies[11,49]. These surfaces starkly contrast with deep, well-defined binding pockets that are characteristic of small molecule-binding sites on proteins[50,51]. Moreover, compared to binding pockets typically seen on

protein targets for endogenous small molecule ligands, which often present only ~300 – 1000

$Å^2$ [52,53] of solvent-accessible surface area (SASA)[54,55] for contact, PPIs' binding interfaces

often possesses several-fold greater SASA, ~1000 – 3000 $Å^2$ [56,57,58]. For example, the LG

interface formed by BCL6[BTB] dimer, responsible for macromolecular recognition of SMRT,

presents ~1270 $Å^2$ SASA[59].

Challenging PPI interface	Conventional target
More solvent-exposed surface area (~1000 – 3000 $Å^2$)	*Less solvent-exposed surface area (~300 – 1000 $Å^2$)*
Topology generally large, flat, "featureless" character	*Well-defined, "ligandable" binding pocket*
Scarcity of well-defined pockets	*Examples include GPCRs, kinases*
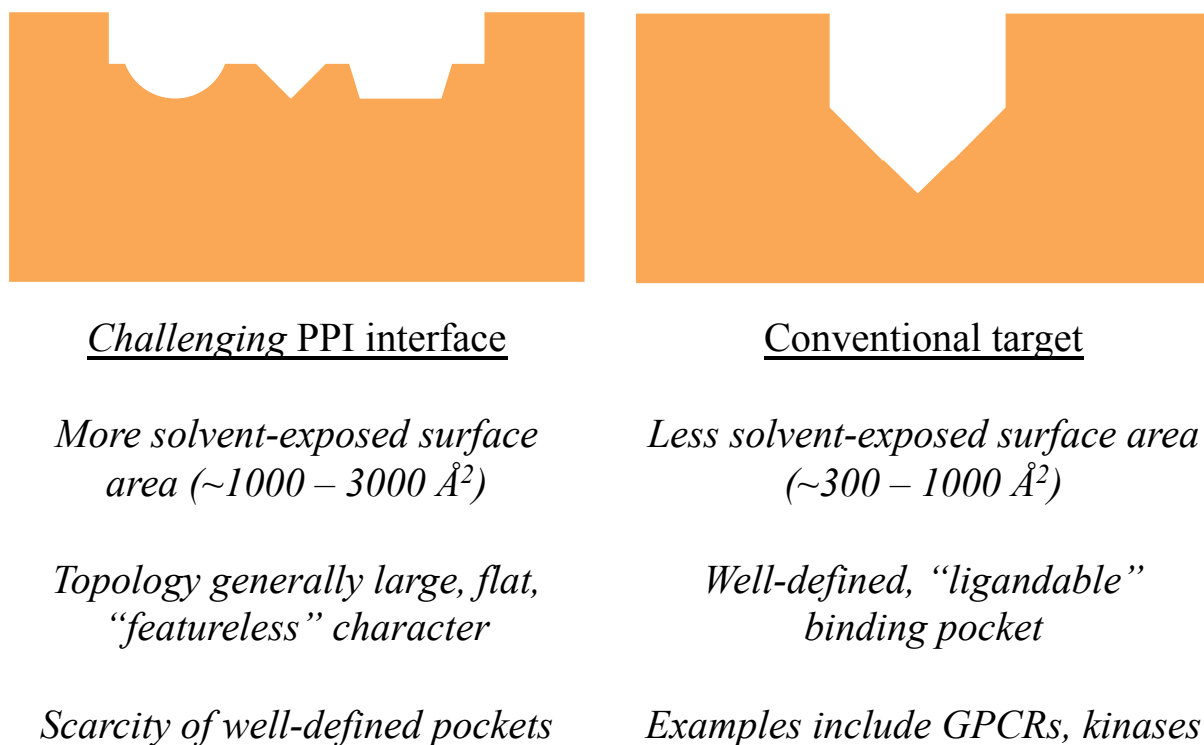
Figure 1.1. Contrast in structural and physiochemical properties of difficult-to-prosecute PPI surfaces, and of classically defined "druggable" targets (*e.g.* GPCRs and kinases)

The greater SASA and unique topology of PPIs manifest themselves as significant

hurdles in the design of inhibitors to block PPIs, which are not readily apparent in targeting

conventionally "druggable" targets amendable to binding low-molecular weight (MW) small

molecules, such as kinases, G-Protein-Coupled Receptors (GPCRs), or ion channels[60]. Small molecules must compete with binding partners whose free energy (ΔG) of molecular recognition is distributed over a larger contact area[61], and furthermore at PPI interfaces typically not possessing discrete groove- or pocket-like features which would be readily occupied by low-MW small molecule ligands[59]. Structural and physical considerations notwithstanding, the challenges from a biological perspective lie in development of small molecules to block a PPI at a binding interface that never evolved to bind small molecules, but rather larger, more complex protein partners[62].

C.2. Development of potent small-molecule inhibitors of PPIs that possess modest affinities

Affinities of PPIs span multiple orders of magnitude, presenting K_Ds ranging from as strong as mid-fM (on the order of 10^{-14} M)[63], to as weak as high-μM[64]. Several epigenetic PPIs implicated in oncogeneses that we have identified for targeting by small molecules have demonstrated modest affinities, towards the weaker end of the spectrum of PPI potency: BCL6^{BTB}-SMRT^{BBD} and BCL6^{BTB}-BCoR^{BBD} K_Ds = 11.4 and 1.3 μM[12,13], respectively; GAS41-monoacetylated Histone H3 and GAS41-diacetylated-Histone H3 K_Ds = ~30 – 80 μM and 3 – 5 μM[14], respectively.

Such affinities beget questions of just how much affinity is *to be had* at PPI interfaces that present modest K_Ds for endogenous ligands. In other words, if the K_D of a given PPI is measured to be on the orders of 10^{-5} – 10^{-6} M, can small-molecule inhibitors with stronger potencies by several orders of magnitude (on the order of 10^{-9} and tighter, which has been suggested as appropriate for lead compounds[65]) that bind at such interface and block the interaction be achieved? Computational efforts by Cheng, *et al.*, published in 1999,

developed a mathematical model to predict the maximum possible binding energy for a drug-like small molecule, dependent on ligand and target desolvation energies (derived in part from solvent-accessible surface area), and on electrostatic and van der Waals interactions[52]. Correlation of druggability with radius of curvature of binding site suggests that difficult-to-drug targets possess less radius of curvature (*i.e.* flatter and/or "feature-less") binding sites[52]. Over a decade later, a 2012 review by Thompson, *et al*., categorized PPIs reportedly targeted by small-molecule inhibitors: by small or large surface area, less or greater than 1,800 $Å^2$, respectively; and by strong or weak affinities, more or less potent than 1 μM K_D, respectively[66]. The authors found that 68% of reported small-molecule inhibitors of PPIs have been developed for PPIs with relatively smaller surface areas *and* stronger affinities, whereas only 10% of reported small-molecule inhibitors target PPIs with relatively larger surface areas *and* weaker affinities[66]. Small molecules targeting PPIs with smaller surface areas yet weaker affinities, or with larger surface areas yet stronger affinities, are comparably scarce[66]. Specifically regarding surface area, this finding is consistent with a recent review by Cierpicki and Grembecka, which concluded that protein-peptide fragment (approximately six to eight residues) interactions are least challenging to target with small molecules, followed by, in order of increasing difficulty, protein-peptides of longer length (12 – 30 residues, roughly), protein-domain interactions, and finally protein-intrinsically disordered protein interactions[59]. Altogether, computational models and multi-year drug discovery campaigns suggest that larger PPI interfaces with fewer well-defined pockets and weaker affinities are less tractable for small-molecule inhibitor development.

D. Biophysics- and structural biology-based approaches to identify and optimize inhibitors

D.1. Fragment-based lead discovery to target PPIs

D.1.1. The case for fragment-based lead discovery

Fragment-based lead discovery (FBLD) relies on the identification of one or more fragment-like small molecules (referred to herein as "fragments") that binds a protein target. Fragments are generally classified by adherence to the "rule of three", which mandates that compounds have MWs under 300 Da, fewer than three hydrogen bond donors and acceptors, fewer than three rotatable bonds, and *cLogP* value of ≤ 3[67]. In more recent years, the criterion that has figured most prominently is MW, and several preeminent practitioners of FBLD have defined the upper limit of fragment sizes at approximately 17 – 20 heavy (non-hydrogen) atoms[68,69].

Fragment libraries are generally ~1,000 – 5,000 compounds in size, which should be sufficient to yield a reasonable amount of hits, yet is not too great in number beyond what Fragment-Based Screening (FBS) methodologies can reasonably handle[68]. In addition, libraries comprised of ~10^3 fragments sample chemical space far more efficiently than high-throughput screening (HTS) libraries[68]. Given the larger MWs of compounds in HTS libraries, which are typically up to 500 Da in size, estimates place the number of possible drug-like small molecules that can exist in chemical space on the order of 10^{63} compounds[70]. By comparison, an estimate of chemical space of fragment-like small molecules (up to 17 heavy atoms in size) places the possible number of compounds at 166 billion[68]. Though HTS libraries in industrial research groups range from 0.5 to 3 million compounds in size[71], a ~1,000-member fragment library samples chemical space far more efficiently, by orders of magnitude, than a ~1,000,000-member HTS library. In fact, the library of drug-sized small molecules (MWs $\leq$ 500 Daltons) would have to be 10 trillion compounds in size to sample

chemical space as efficiently as the 1,000-member fragment library[72]. Strikingly, FBS results have been demonstrated to be predictive of both druggability and non-druggability of targets, and moreover of favorable outcomes in lead discovery projects, with remarkable success[72].

In addition to relatively efficient sampling of chemical space, another advantage of FBLD is that fragment hits must form high-quality interactions with protein targets, and therefore may make good starting points for optimization[68,73]. Given the upper-limit of size imposed on fragment libraries, it is likely that fewer interactions will be made with targets than would be made by HTS hits. And, though fragment hits tend to bind with modest affinities (possessing K_Ds for targets typically in the range of ~100 μM to low-mM[74]), the few contacts that are made should each confer significant binding energy in complex formation; this is requisite for molecular recognition[75,76]. Furthermore, given the sizable entropic penalty of fragment binding, these interactions (as few as one or two in some cases) must form favorable, strong contacts with target protein[77]. The metric Ligand Efficiency (LE) has been developed to assess binding of fragments,

Equation 1.1.

$$LE = \Delta G/HA$$

where ΔG is the Gibbs free energy of binding, and HA is the number of heavy (non-hydrogen atoms)[78,79,80]. To this point, the LE of a hypothetical drug-like small molecule that is 38 heavy atoms in size (corresponding to a MW of approximately 500 Daltons) and binds a protein target with 10 nM K_D has an LE of ~0.3[81]. This is consistent with the LE of a fragment with 10 heavy atoms that binds a protein target with 1 mM K_D, as with the LE of a fragment with 20 heavy atoms that binds a protein target with 1 μM K_D[82]. Despite variations in affinities that span five orders of magnitude, all three compounds bind with equal

efficiency, or binding energy per atom. The propensity for hits to make high-quality

interactions, synthetic tractability, and feasible growth vectors (determined on the basis of

compound binding mode in complex with target) all contribute to the sentiment that fragments

may be good starting points for compound optimization in small-molecule inhibitor

development campaigns[68].
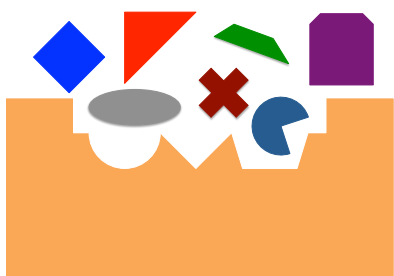
Screen ~1000-member fragment library
against target by protein-NMR

Identify hit by chemical shift perturbations
on NMR spectra
$K_Ds \sim 100\ \mu M$ to low-mM

Grow inhibitors by biophysics- and
structural biology-guided approaches
Improve potency by orders of magnitude

Figure 1.2. Fragment-Based Lead Discovery (FBLD) methodology to identify and

optimize small-molecule inhibitors of PPIs.

D.1.2. FBLD addresses challenges PPI present as drug targets

It has been suggested that HTS campaigns targeting PPIs are not routinely fruitful[49]. Our observations are consistent with this impression, as ~150,000-member compound library screened against GAS41 YEATS, performed at the Center for Chemical Genomics at the Life Sciences Institute at the University of Michigan, did not yield any hits. Particularly, PPIs whose molecular recognition events rely on large, shallow interfaces with weak affinities may not be tractable to probing by drug-like small molecules found in HTS chemical libraries[83]. Notably, 20% of all PPI inhibitors have been reported since 2013, which has been at least partly attributed to advancements in FBLD libraries and methodologies[83]. I argue that FBLD offers advantages that other approaches, of which HTS is most prominent, do not. First, the lower MWs of fragments (in accordance with the "rule of three") facilitate the identification of hits at challenging-to-drug interfaces: there is a relatively limited number of interactions that fragments make with a target protein, which therefore reduces the complexity of the protein-ligand system and in turn lessens the constrain on the fragment as it explores the target protein's surface[68,69]. Moreover, given what is known about the surfaces of PPIs and how this affects druggability, specifically the general paucity of well-defined, ligandable pockets, it would stand to reason that fragments may feasibly produce more hits than would HTS libraries, on the bases of fragments' lesser molecular complexities. In addition, given the high-quality molecular interactions that fragment hits form with protein targets[68], it is further not unreasonable that fragment hits would be good starting points for inhibitor optimization. To the latter point, this can be empirically tested, and comprehensive analysis of later generations of fragment-based inhibitors' binding modes in FBLD campaigns has revealed that fragment hits' binding modes are often conserved[73]. And yet, the investment in synthetic medicinal chemistry that grows fragment hits of low-MWs and modest affinities into potent

and selective small-molecule inhibitors is non-trivial. Within the context of FBLD, it is typically guided by biophysics- and structural biology-based approaches, which we will entertain in the following sections.

D.2. Biophysical and structural techniques to characterize targets and develop small-molecule inhibitors

D.2.1. Characterization of the biophysical, mechanistic, and structural bases of molecular recognition events of PPIs with weak affinities

We have identified two PPIs as pharmacological targets for inhibitor development campaigns (discussed in Section B.2.). First, I worked on transcriptional repressor BCL6, which is overexpressed in a number of DLBCL subtypes, and has been shown to regulate oncogenic transcriptional programming to drive lymphomagenesis. BCL6 recruits co-repressors SMRT[BBD] and BCoR[BBD] via its N-terminal BTB domain, and the previous undertaking of biophysical characterization by ITC presents K_Ds in the range of ~1 – 12 μM[12,13]. Second, I focused on epigenetic reader protein GAS41, amplified in several different sub-types of NSCLCs. GAS41 YEATS, specifically, has been directly implicated in oncogenic transformation. Recent studies have shown that GAS41 recognizes acetylated-Histone H3 peptides with low- to mid-micromolar affinities (K_Ds = ~9 – 33 μM)[22,47]. GAS41 is one of four YEATS domain epigenetic readers proteins in humans (the remaining members of the family are ENL, AF9, and YEATS2), all of which have emerged as oncogenic drivers of various human cancers[2,22,84,85] over the past half-decade (further discussed in Chapter 4). Comprehensive studies to reveal the molecular bases of human YEATS domains' recognition of acylated-Histone H3 peptides have presented K_Ds as potent as 3 μM, for AF9, to ~30 – 120 μM, for ENL and YEATS2[85,86,87,88]. Structure-based approaches have yielded compounds

targeting ENL and AF9, from which we have gleaned insight into the structural basis of inhibition of YEATS domains[89,90,91,92] but GAS41 and YEATS2 have remained elusive as of this writing. As we begun our campaign to identify acetylated-Histone H3 selectivity by GAS41, we therefore enlisted biophysics-based approaches that would be sensitive for modest interactions. In summary, the PPIs we have identified for targeting by small molecules do not demonstrate potent (K_Ds <1 μM) binding affinities.

To the contrary, PPIs possessing K_Ds >1 μM have been classified as both weak and transient[93,94]. Albeit, these properties are not mutually exclusive, as binding affinity is extracted from on- and off-rates. Weak, transient interactions pose challenges in characterization, for techniques have to be sensitive to mid- to high-μM concentrations of protein. This is requisite to extract K_Ds of weak PPIs, such as human YEATS domains' recognition of acylated-histone H3 peptides[22,47]. NMR, in particular, has proven tremendously valuable for investigation of weak PPIs, revealing a wealth of insight into the structural and biophysical bases of their molecular recognition events[95,96,97,98]. Guided by protein-observed solution-Nuclear Magnetic Resonance spectroscopy (referred to herein as protein-NMR) to both validate and quantify binding, we contend that complementary studies for structure determination to high-resolution by x-ray crystallography, and biophysical characterization to derive binding affinity, specifically kinetic and thermodynamic parameters by Bio-Layer Interferometry (BLI)[99] and ITC, respectively, yields a comprehensive picture of macromolecular recognition events that drive weak PPIs that may not be sensitive to other techniques.

Such binding events, in particular, are largely driven by "hot spots", which are sites on PPI interfaces that contribute significantly to binding energy[100]. Further analysis across

numerous PPIs indicated that binding energy tends to be not uniformly distributed throughout, but rather highly localized in a small subset of residues at PPI interfaces[101]. Strikingly, tryptophan, tyrosine, and arginine are heavily overrepresented at hot spots[101]. This extensive mechanistic and biophysical characterization can provide insight into the structural basis of PPI inhibition, as hot spots present themselves as druggable sites on PPI interfaces[102]. Moreover, hot spots are remarkably amendable to FBS approaches, as fragment hit binding sites frequently overlap with hot spots[103].

D.2.2. Biophysics- and structural biology-based approaches to identify and optimize small-molecule inhibitors of PPIs

I contend that protein-NMR is uniquely appropriate for FBS, because it allows for detection of weak-binding fragment hits (typically possessing mid-μM to low-mM K_Ds for target proteins) and subsequent quantification of their affinities; may identify fragment hits' binding sites given spectrum assignment; and is performed at near-physiological conditions (relative to other screening techniques), in which conformational flexibility and dynamics of target proteins are permitted[104,105,106]. Regarding difficult-to-prosecute PPIs, protein-NMR is a technique routinely used in FBLD that is viable for hit validation, and versatile for hit and hit analogues' characterization[77]. Following hit identification, protein-NMR can be further advantageous in exploration of modifications during early stages of optimization, as chemical shift perturbations on NMR experiment ^{1}H-^{15}N Heteronuclear Nuclear Single Coherence (HSQC) spectra in presence of fragment hit and of hit analogues can be used to quantify, and thusly rank, compound binding[107].
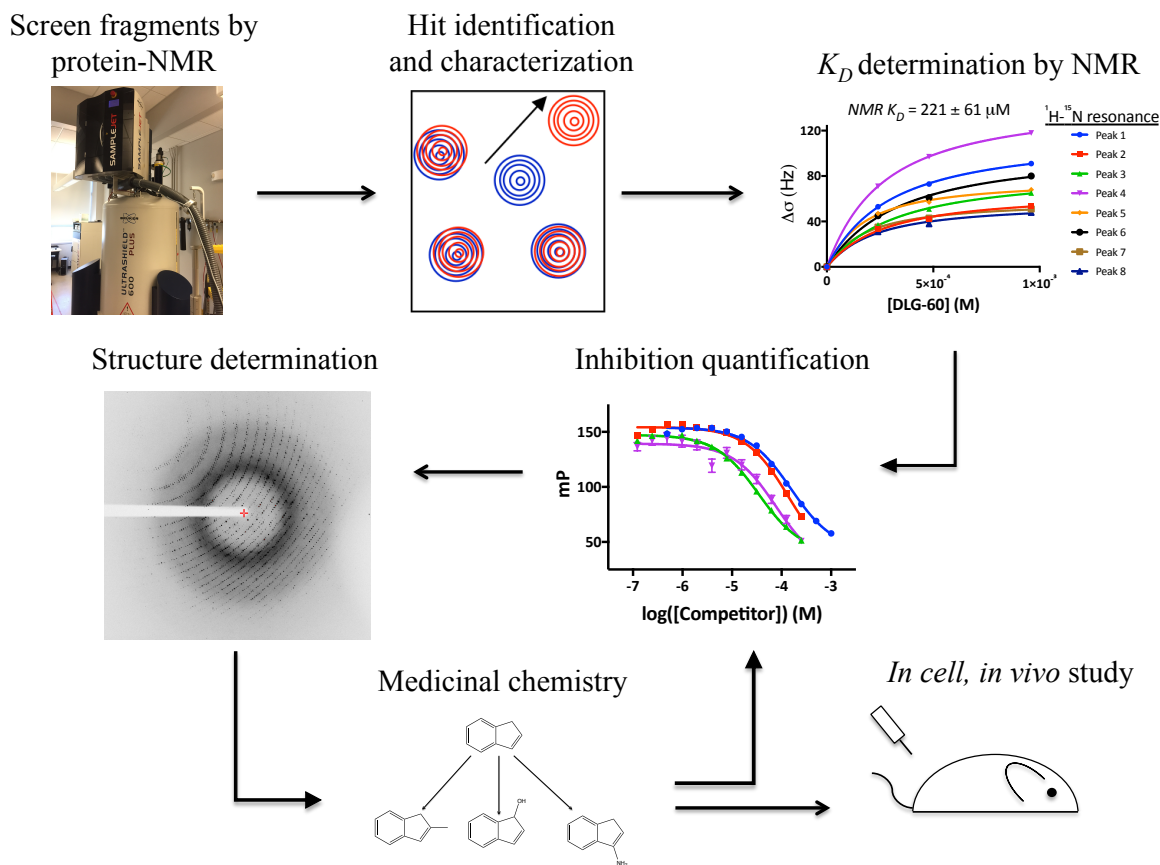
Figure 1.3. FBLD approach integrated with biophysics-guided methodologies to develop small-molecule inhibitors of PPIs. Screen in-house fragment library against target by protein-NMR. Identify and characterize hit and hit analogues by quantification of chemical shift perturbations (CSPs) and NMR-titration experiments. Develop assay to characterize inhibition. Determine crystal structure in complex with inhibitor to reveal binding mode. Inform design of new compounds. Process is iterative. Cartoon showing CSPs courtesy Christina Howard.

In tandem with protein-NMR, structure determination of the fragment hit in complex with target protein to high-resolution by x-ray crystallography is quite profitable in inhibitor design. The crystal structure of protein-fragment hit complex reveals the binding mode of the

fragment hit and molecular details of complex formation, and furthermore allows for the identification of growth vectors for optimization[68,75,108]. Determination of the crystal structure of hit in complex with protein is necessary for a structure-guided inhibitor development program, at which point biophysical and chemical biological methods can be incorporated into an iterative process to guide the discovery of small-molecule inhibitors. Altogether, biophysical and structural techniques have been adeptly assimilated in multi-pronged campaigns that have yielded potent and selective lead compounds[109,110,111,112].

E. Thesis summary

First, in Chapter 2, adapted from Cho, *et al*., 2018, we demonstrate biophysics- and structural biology-guided approaches, in parallel with *in cell* studies, to report the molecular basis of Histone H3 acetyl-lysine recognition by histone reader GAS41, via its YEATS domain. We identify specific sites of Histone H3 acetylation in GAS41-Histone H3 complex formation, and quantify their affinities in the mid-micromolar range (K_Ds $\sim 30 - 100\ \mu$M)[14]. Further, we present a model of bivalent recognition of di-acetylated Histone H3 by GAS41, which confers an order of magnitude improvement of affinity, into the low-micromolar range (K_Ds $\sim 3 - 5\ \mu$M)[14]. Finally, we report the crystal structure of di-acetylated Histone H3 peptide in complex with GAS41 YEATS domain, which reveals the molecular details of acetyl-lysine recognition by GAS41.

In Chapter 3, we detail the development of small-molecules inhibitors targeting di-acetylated Histone H3 recognition by histone reader GAS41. Recent studies have demonstrated that GAS41 is an oncogenic driver of certain sub-types of NSCLCs, and moreover that the YEATS domain specifically is required for oncogenic transformation[22]. As of this writing, there are no inhibitors of GAS41 reported in the literature. We screened an in-

house library of fragment-like small-molecules against GAS41 YEATS domain by protein-NMR, and identified a hit that presents high-μM inhibitory activity in blocking GAS41 YEATS-di-acetyl-lysine Histone H3 peptide complex formation. We incorporated insight into the molecular basis of Histone H3 acetyl-lysine recognition by GAS41 YEATS domain, discussed in Chapter 2 and reported in Cho, *et al*., to design inhibitors. We combined this insight with protein-NMR, x-ray crystallography, affinity determination by ITC, and Fluorescence Polarization and AlphaScreen Competition Assay development, which yielded small-molecule inhibitors with low-nanomolar activity *in vitro* for GAS41. This represents an over-10,000-fold improvement in potency. Studies to demonstrate on-target activity and anti-proliferative effects in GAS41-dependent NSCLC cell lines are ongoing in our lab.

In Chapter 4, we report a fragment-based approach to develop small-molecule inhibitors of BCL6-co-repressor interactions, adapted from Cheng and Linhares, *et al*., 2018. BCL6 is transcription factor that regulates B-Cell maturation, and governs formation of epigenetic complexes responsible for gene repression via recruitment of co-repressors SMRT and BCoR. Constitutively overexpressed in several sub-types of DLBCLs and FLs[21,113], BCL6 is a demonstrably validated pharmacological target (discussed both earlier in this chapter and in Chapter 4). We screened an in-house library of fragment-like small molecules against BCL6[BTB] by protein-NMR, and identified a hit scaffold that is comprised of a previously unreported ligand for BCL6[BTB]. We solved the crystal structure of hit in complex with BCL6[BTB], and subsequently introduced novel methodology to rank binding of weak ligands by protein-NMR[114]. We used protein-NMR, x-ray crystallography, and computational chemistry to guide an investment in synthetic medicinal chemistry. This improved low-mM fragment hit approximately 100-fold to yield mid-μM small-molecule inhibitors[114]. We

solved crystal structures of our most potent inhibitors in complex with BCL6[BTB], which reveal

insights into the structural basis of inhibition of BCL6-co-repressor complexes.

F. References

1   Mujtaba, S., Zeng, L. & Zhou, M. M. Structure and acetyl-lysine recognition of the bromodomain. *Oncogene* **26**, 5521-5527, doi:10.1038/sj.onc.1210618 (2007).
2   Zhao, D., Li, Y., Xiong, X., Chen, Z. & Li, H. YEATS Domain-A Histone Acylation Reader in Health and Disease. *J Mol Biol* **429**, 1994-2002, doi:10.1016/j.jmb.2017.03.010 (2017).
3   Mayer, B. J. Protein-protein interactions in signaling cascades. *Mol Biotechnol* **13**, 201-213, doi:10.1385/MB:13:3:201 (1999).
4   Kaake, R. M., Milenkovic, T., Przulj, N., Kaiser, P. & Huang, L. Characterization of cell cycle specific protein interaction networks of the yeast 26S proteasome complex by the QTAX strategy. *J Proteome Res* **9**, 2016-2029, doi:10.1021/pr1000175 (2010).
5   Durek, P. & Walther, D. The integrated analysis of metabolic and protein interaction networks reveals novel molecular organizing principles. *BMC Syst Biol* **2**, 100, doi:10.1186/1752-0509-2-100 (2008).
6   Allis, C. D. & Jenuwein, T. The molecular hallmarks of epigenetic control. *Nat Rev Genet* **17**, 487-500, doi:10.1038/nrg.2016.59 (2016).
7   Rodriguez-Paredes, M. & Esteller, M. Cancer epigenetics reaches mainstream oncology. *Nat Med* **17**, 330-339, doi:10.1038/nm.2305 (2011).
8   Diaz-Eufracio, B. I., Naveja, J. J. & Medina-Franco, J. L. Protein-Protein Interaction Modulators for Epigenetic Therapies. *Adv Protein Chem Struct Biol* **110**, 65-84, doi:10.1016/bs.apcsb.2017.06.002 (2018).
9   Flavahan, W. A., Gaskell, E. & Bernstein, B. E. Epigenetic plasticity and the hallmarks of cancer. *Science* **357**, doi:10.1126/science.aal2380 (2017).
10   Smith, S. G. & Zhou, M. M. The Bromodomain: A New Target in Emerging Epigenetic Medicine. *ACS Chem Biol* **11**, 598-608, doi:10.1021/acschembio.5b00831 (2016).
11   Guo, W., Wisniewski, J. A. & Ji, H. Hot spot-based design of small-molecule inhibitors for protein-protein interactions. *Bioorg Med Chem Lett* **24**, 2546-2554, doi:10.1016/j.bmcl.2014.03.095 (2014).
12   Ahmad, K. F. *et al.* Mechanism of SMRT corepressor recruitment by the BCL6 BTB domain. *Mol Cell* **12**, 1551-1564 (2003).
13   Ghetu, A. F. *et al.* Structure of a BCOR corepressor peptide in complex with the BCL6 BTB domain dimer. *Mol Cell* **29**, 384-391, doi:10.1016/j.molcel.2007.12.026 (2008).
14   Cho, H. J. *et al.* GAS41 Recognizes Diacetylated Histone H3 through a Bivalent Binding Mode. *ACS Chem Biol* **13**, 2739-2746, doi:10.1021/acschembio.8b00674 (2018).
15   Bannister, A. J. & Kouzarides, T. Regulation of chromatin by histone modifications. *Cell Res* **21**, 381-395, doi:10.1038/cr.2011.22 (2011).
16   Chen, Y. *et al.* Lysine propionylation and butyrylation are novel post-translational modifications in histones. *Mol Cell Proteomics* **6**, 812-819, doi:10.1074/mcp.M700021-MCP200 (2007).

17   Tan, M. *et al.* Identification of 67 histone marks and histone lysine crotonylation as a new type of histone modification. *Cell* **146**, 1016-1028, doi:10.1016/j.cell.2011.08.008 (2011).

18   Hyun, K., Jeon, J., Park, K. & Kim, J. Writing, erasing and reading histone lysine methylations. *Exp Mol Med* **49**, e324, doi:10.1038/emm.2017.11 (2017).

19   Xie, Z. *et al.* Lysine succinylation and lysine malonylation in histones. *Mol Cell Proteomics* **11**, 100-107, doi:10.1074/mcp.M111.015875 (2012).

20   Ehrlich, L. *et al.* A Review of the Scaffold Protein Menin and its Role in Hepatobiliary Pathology. *Gene Expr* **17**, 251-263, doi:10.3727/105221617X695744 (2017).

21   Hatzi, K. & Melnick, A. Breaking bad in the germinal center: how deregulation of BCL6 contributes to lymphomagenesis. *Trends Mol Med* **20**, 343-352, doi:10.1016/j.molmed.2014.03.001 (2014).

22   Hsu, C. C. *et al.* Recognition of histone acetylation by the GAS41 YEATS domain promotes H2A.Z deposition in non-small cell lung cancer. *Genes Dev* **32**, 58-69, doi:10.1101/gad.303784.117 (2018).

23   Ye, B. H. *et al.* Chromosomal translocations cause deregulated BCL6 expression by promoter substitution in B cell lymphoma. *EMBO J* **14**, 6209-6217 (1995).

24   Baron, B. W. *et al.* The human BCL6 transgene promotes the development of lymphomas in the mouse. *Proc Natl Acad Sci U S A* **101**, 14198-14203, doi:10.1073/pnas.0406138101 (2004).

25   Ye, B. H. *et al.* Alterations of a zinc finger-encoding gene, BCL-6, in diffuse large-cell lymphoma. *Science* **262**, 747-750, doi:10.1126/science.8235596 (1993).

26   Pasqualucci, L. *et al.* Mutations of the BCL6 proto-oncogene disrupt its negative autoregulation in diffuse large B-cell lymphoma. *Blood* **101**, 2914-2923, doi:10.1182/blood-2002-11-3387 (2003).

27   Cattoretti, G. *et al.* Deregulated BCL6 expression recapitulates the pathogenesis of human diffuse large B cell lymphomas in mice. *Cancer Cell* **7**, 445-455, doi:10.1016/j.ccr.2005.03.037 (2005).

28   Basso, K. & Dalla-Favera, R. BCL6: master regulator of the germinal center reaction and key oncogene in B cell lymphomagenesis. *Adv Immunol* **105**, 193-210, doi:10.1016/S0065-2776(10)05007-8 (2010).

29   De Silva, N. S. & Klein, U. Dynamics of B cells in germinal centres. *Nat Rev Immunol* **15**, 137-148, doi:10.1038/nri3804 (2015).

30   Polo, J. M., Ci, W., Licht, J. D. & Melnick, A. Reversible disruption of BCL6 repression complexes by CD40 signaling in normal and malignant B cells. *Blood* **112**, 644-651, doi:10.1182/blood-2008-01-131813 (2008).

31   Wagner, S. D., Ahearne, M. & Ko Ferrigno, P. The role of BCL6 in lymphomas and routes to therapy. *Br J Haematol* **152**, 3-12, doi:10.1111/j.1365-2141.2010.08420.x (2011).

32   Tunyaplin, C. *et al.* Direct repression of prdm1 by Bcl-6 inhibits plasmacytic differentiation. *J Immunol* **173**, 1158-1165, doi:10.4049/jimmunol.173.2.1158 (2004).

33   Phan, R. T. & Dalla-Favera, R. The BCL6 proto-oncogene suppresses p53 expression in germinal-centre B cells. *Nature* **432**, 635-639, doi:10.1038/nature03147 (2004).

34   Phan, R. T., Saito, M., Basso, K., Niu, H. & Dalla-Favera, R. BCL6 interacts with the transcription factor Miz-1 to suppress the cyclin-dependent kinase inhibitor p21 and

cell cycle arrest in germinal center B cells. *Nat Immunol* **6**, 1054-1060, doi:10.1038/ni1245 (2005).

35 Shaffer, A. L. *et al.* BCL-6 represses genes that function in lymphocyte differentiation, inflammation, and cell cycle control. *Immunity* **13**, 199-212 (2000).

36 Huynh, K. D. & Bardwell, V. J. The BCL-6 POZ domain and other POZ domains interact with the co-repressors N-CoR and SMRT. *Oncogene* **17**, 2473-2484, doi:10.1038/sj.onc.1202197 (1998).

37 Huynh, K. D., Fischle, W., Verdin, E. & Bardwell, V. J. BCoR, a novel corepressor involved in BCL-6 repression. *Genes Dev* **14**, 1810-1823 (2000).

38 Hatzi, K. *et al.* A hybrid mechanism of action for BCL6 in B cells defined by formation of functionally distinct complexes at enhancers and promoters. *Cell Rep* **4**, 578-588, doi:10.1016/j.celrep.2013.06.016 (2013).

39 Polo, J. M. *et al.* Specific peptide interference reveals BCL6 transcriptional and oncogenic mechanisms in B-cell lymphoma cells. *Nat Med* **10**, 1329-1335, doi:10.1038/nm1134 (2004).

40 Cerchietti, L. C. *et al.* A small-molecule inhibitor of BCL6 kills DLBCL cells in vitro and in vivo. *Cancer Cell* **17**, 400-411, doi:10.1016/j.ccr.2009.12.050 (2010).

41 Mendgen, T., Steuer, C. & Klein, C. D. Privileged scaffolds or promiscuous binders: a comparative study on rhodanines and related heterocycles in medicinal chemistry. *J Med Chem* **55**, 743-753, doi:10.1021/jm201243p (2012).

42 Cardenas, M. G. *et al.* Rationally designed BCL6 inhibitors target activated B cell diffuse large B cell lymphoma. *J Clin Invest* **126**, 3351-3362, doi:10.1172/JCI85795 (2016).

43 McCoull, W. *et al.* Discovery of Pyrazolo[1,5-a]pyrimidine B-Cell Lymphoma 6 (BCL6) Binders and Optimization to High Affinity Macrocyclic Inhibitors. *J Med Chem* **60**, 4386-4402, doi:10.1021/acs.jmedchem.7b00359 (2017).

44 Yasui, T. *et al.* Discovery of a novel B-cell lymphoma 6 (BCL6)-corepressor interaction inhibitor by utilizing structure-based drug design. *Bioorg Med Chem* **25**, 4876-4886, doi:10.1016/j.bmc.2017.07.037 (2017).

45 Sameshima, T. *et al.* Discovery of an Irreversible and Cell-Active BCL6 Inhibitor Selectively Targeting Cys53 Located at the Protein-Protein Interaction Interface. *Biochemistry* **57**, 1369-1379, doi:10.1021/acs.biochem.7b00732 (2018).

46 Kerres, N. *et al.* Chemically Induced Degradation of the Oncogenic Transcription Factor BCL6. *Cell Rep* **20**, 2860-2875, doi:10.1016/j.celrep.2017.08.081 (2017).

47 Hsu, C. C. *et al.* Gas41 links histone acetylation to H2A.Z deposition and maintenance of embryonic stem cell identity. *Cell Discov* **4**, 28, doi:10.1038/s41421-018-0027-0 (2018).

48 Wang, Y. *et al.* Identification of the YEATS domain of GAS41 as a pH-dependent reader of histone succinylation. *Proc Natl Acad Sci U S A* **115**, 2365-2370, doi:10.1073/pnas.1717664115 (2018).

49 Wells, J. A. & McClendon, C. L. Reaching for high-hanging fruit in drug discovery at protein-protein interfaces. *Nature* **450**, 1001-1009, doi:10.1038/nature06526 (2007).

50 Arkin, M. R., Tang, Y. & Wells, J. A. Small-molecule inhibitors of protein-protein interactions: progressing toward the reality. *Chem Biol* **21**, 1102-1114, doi:10.1016/j.chembiol.2014.09.001 (2014).

51   Fuller, J. C., Burgoyne, N. J. & Jackson, R. M. Predicting druggable binding sites at the protein-protein interface. *Drug Discov Today* **14**, 155-161, doi:10.1016/j.drudis.2008.10.009 (2009).

52   Cheng, A. C. *et al.* Structure-based maximal affinity model predicts small-molecule druggability. *Nat Biotechnol* **25**, 71-75, doi:10.1038/nbt1273 (2007).

53   Smith, R. D. *et al.* Exploring protein-ligand recognition with Binding MOAD. *J Mol Graph Model* **24**, 414-425, doi:10.1016/j.jmgm.2005.08.002 (2006).

54   Lee, B. & Richards, F. M. The interpretation of protein structures: estimation of static accessibility. *J Mol Biol* **55**, 379-400, doi:10.1016/0022-2836(71)90324-x (1971).

55   Street, A. G. & Mayo, S. L. Pairwise calculation of protein solvent-accessible surface areas. *Fold Des* **3**, 253-258, doi:10.1016/S1359-0278(98)00036-4 (1998).

56   Jones, S. & Thornton, J. M. Principles of protein-protein interactions. *Proc Natl Acad Sci U S A* **93**, 13-20, doi:10.1073/pnas.93.1.13 (1996).

57   Lo Conte, L., Chothia, C. & Janin, J. The atomic structure of protein-protein recognition sites. *J Mol Biol* **285**, 2177-2198, doi:10.1006/jmbi.1998.2439 (1999).

58   Hwang, H., Vreven, T., Janin, J. & Weng, Z. Protein-protein docking benchmark version 4.0. *Proteins* **78**, 3111-3114, doi:10.1002/prot.22830 (2010).

59   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, doi:10.1111/imr.12244 (2015).

60   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, doi:10.1111/cbdd.12066 (2013).

61   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, doi:10.1017/erm.2012.10 (2012).

62   Arkin, M. R. & Wells, J. A. Small-molecule inhibitors of protein-protein interactions: progressing towards the dream. *Nat Rev Drug Discov* **3**, 301-317, doi:10.1038/nrd1343 (2004).

63   Kobe, B. & Deisenhofer, J. A structural basis of the interactions between leucine-rich repeats and protein ligands. *Nature* **374**, 183-186, doi:10.1038/374183a0 (1995).

64   Davis, S. J., Davies, E. A., Tucknott, M. G., Jones, E. Y. & van der Merwe, P. A. The role of charged residues mediating low affinity protein-protein recognition at the cell surface by CD2. *Proc Natl Acad Sci U S A* **95**, 5490-5494, doi:10.1073/pnas.95.10.5490 (1998).

65   Hefti, F. F. Requirements for a lead compound to become a clinical candidate. *BMC Neurosci* **9 Suppl 3**, S7, doi:10.1186/1471-2202-9-S3-S7 (2008).

66   Thompson, A. D., Dugan, A., Gestwicki, J. E. & Mapp, A. K. Fine-tuning multiprotein complexes using small molecules. *ACS Chem Biol* **7**, 1311-1320, doi:10.1021/cb300255p (2012).

67   Congreve, M., Carr, R., Murray, C. & Jhoti, H. A 'rule of three' for fragment-based lead discovery? *Drug Discov Today* **8**, 876-877 (2003).

68   Erlanson, D. A., Fesik, S. W., Hubbard, R. E., Jahnke, W. & Jhoti, H. Twenty years on: the impact of fragments on drug discovery. *Nat Rev Drug Discov* **15**, 605-619, doi:10.1038/nrd.2016.109 (2016).

69   Jhoti, H., Williams, G., Rees, D. C. & Murray, C. W. The 'rule of three' for fragment-based drug discovery: where are we now? *Nat Rev Drug Discov* **12**, 644-645, doi:10.1038/nrd3926-c1 (2013).

70   Bohacek, R. S., McMartin, C. & Guida, W. C. The art and practice of structure-based drug design: a molecular modeling perspective. *Med Res Rev* **16**, 3-50, doi:10.1002/(SICI)1098-1128(199601)16:1<3::AID-MED1>3.0.CO;2-6 (1996).

71   Macarron, R. *et al.* Impact of high-throughput screening in biomedical research. *Nat Rev Drug Discov* **10**, 188-195, doi:10.1038/nrd3368 (2011).

72   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, doi:10.1016/j.drudis.2011.02.002 (2011).

73   Murray, C. W., Verdonk, M. L. & Rees, D. C. Experiences in fragment-based drug discovery. *Trends Pharmacol Sci* **33**, 224-232, doi:10.1016/j.tips.2012.02.006 (2012).

74   Lamoree, B. & Hubbard, R. E. Current perspectives in fragment-based lead discovery (FBLD). *Essays Biochem* **61**, 453-464, doi:10.1042/EBC20170028 (2017).

75   Murray, C. W. & Rees, D. C. The rise of fragment-based drug discovery. *Nat Chem* **1**, 187-192, doi:10.1038/nchem.217 (2009).

76   Kashyap, A., Singh, P. K. & Silakari, O. Counting on Fragment Based Drug Design Approach for Drug Discovery. *Curr Top Med Chem* **18**, 2284-2293, doi:10.2174/1568026619666181130134250 (2018).

77   Price, A. J., Howard, S. & Cons, B. D. Fragment-based drug discovery and its application to challenging drug targets. *Essays Biochem* **61**, 475-484, doi:10.1042/EBC20170029 (2017).

78   Carr, R. A., Congreve, M., Murray, C. W. & Rees, D. C. Fragment-based lead discovery: leads by design. *Drug Discov Today* **10**, 987-992, doi:10.1016/S1359-6446(05)03511-7 (2005).

79   Hopkins, A. L., Groom, C. R. & Alex, A. Ligand efficiency: a useful metric for lead selection. *Drug Discov Today* **9**, 430-431, doi:10.1016/S1359-6446(04)03069-7 (2004).

80   Murray, C. W. *et al.* Validity of ligand efficiency metrics. *ACS Med Chem Lett* **5**, 616-618, doi:10.1021/ml500146d (2014).

81   Hopkins, A. L., Keseru, G. M., Leeson, P. D., Rees, D. C. & Reynolds, C. H. The role of ligand efficiency metrics in drug discovery. *Nat Rev Drug Discov* **13**, 105-121, doi:10.1038/nrd4163 (2014).

82   Kenny, P. W., Leitao, A. & Montanari, C. A. Ligand efficiency metrics considered harmful. *J Comput Aided Mol Des* **28**, 699-710, doi:10.1007/s10822-014-9757-8 (2014).

83   Taylor, I. R. *et al.* High-throughput screen for inhibitors of protein-protein interactions in a reconstituted heat shock protein 70 (Hsp70) complex. *J Biol Chem* **293**, 4014-4025, doi:10.1074/jbc.RA117.001575 (2018).

84   Zhou, J., Ng, Y. & Chng, W. J. ENL: structure, function, and roles in hematopoiesis and acute myeloid leukemia. *Cell Mol Life Sci* **75**, 3931-3941, doi:10.1007/s00018-018-2895-8 (2018).

85   Mi, W. *et al.* YEATS2 links histone acetylation to tumorigenesis of non-small cell lung cancer. *Nat Commun* **8**, 1088, doi:10.1038/s41467-017-01173-4 (2017).

86   Li, Y. *et al.* AF9 YEATS domain links histone acetylation to DOT1L-mediated H3K79 methylation. *Cell* **159**, 558-571, doi:10.1016/j.cell.2014.09.049 (2014).

87   Erb, M. A. *et al.* Transcription control by the ENL YEATS domain in acute leukaemia. *Nature* **543**, 270-274, doi:10.1038/nature21688 (2017).

88   Wan, L. *et al.* ENL links histone acetylation to oncogenic gene expression in acute myeloid leukaemia. *Nature* **543**, 265-269, doi:10.1038/nature21687 (2017).

89   Heidenreich, D. *et al.* Structure-Based Approach toward Identification of Inhibitory Fragments for Eleven-Nineteen-Leukemia Protein (ENL). *J Med Chem* **61**, 10929-10934, doi:10.1021/acs.jmedchem.8b01457 (2018).

90   Moustakim, M. *et al.* Discovery of an MLLT1/3 YEATS Domain Chemical Probe. *Angew Chem Int Ed Engl* **57**, 16302-16307, doi:10.1002/anie.201810617 (2018).

91   Li, X. *et al.* Structure-guided development of YEATS domain inhibitors by targeting pi-pi-pi stacking. *Nat Chem Biol* **14**, 1140-1149, doi:10.1038/s41589-018-0144-y (2018).

92   Klein, B. J. *et al.* Structural insights into the pi-pi-pi stacking mechanism and DNA-binding activity of the YEATS domain. *Nat Commun* **9**, 4574, doi:10.1038/s41467-018-07072-6 (2018).

93   Perkins, J. R., Diboun, I., Dessailly, B. H., Lees, J. G. & Orengo, C. Transient protein-protein interactions: structural, functional, and network properties. *Structure* **18**, 1233-1243, doi:10.1016/j.str.2010.08.007 (2010).

94   Acuner Ozbabacan, S. E., Engin, H. B., Gursoy, A. & Keskin, O. Transient protein-protein interactions. *Protein Eng Des Sel* **24**, 635-648, doi:10.1093/protein/gzr025 (2011).

95   Qin, J., Vinogradova, O. & Gronenborn, A. M. Protein-protein interactions probed by nuclear magnetic resonance spectroscopy. *Methods Enzymol* **339**, 377-389, doi:10.1016/s0076-6879(01)39323-0 (2001).

96   Vaynberg, J. & Qin, J. Weak protein-protein interactions as probed by NMR spectroscopy. *Trends Biotechnol* **24**, 22-27, doi:10.1016/j.tibtech.2005.09.006 (2006).

97   Lian, L. Y. NMR studies of weak protein-protein interactions. *Prog Nucl Magn Reson Spectrosc* **71**, 59-72, doi:10.1016/j.pnmrs.2012.11.002 (2013).

98   Zuiderweg, E. R. Mapping protein-protein interactions in solution by NMR spectroscopy. *Biochemistry* **41**, 1-7, doi:10.1021/bi011870b (2002).

99   Sultana, A. & Lee, J. E. Measuring protein-protein and protein-nucleic Acid interactions by biolayer interferometry. *Curr Protoc Protein Sci* **79**, 19 25 11-26, doi:10.1002/0471140864.ps1925s79 (2015).

100   Clackson, T. & Wells, J. A. A hot spot of binding energy in a hormone-receptor interface. *Science* **267**, 383-386, doi:10.1126/science.7529940 (1995).

101   Bogan, A. A. & Thorn, K. S. Anatomy of hot spots in protein interfaces. *J Mol Biol* **280**, 1-9, doi:10.1006/jmbi.1998.1843 (1998).

102   Cukuroglu, E., Engin, H. B., Gursoy, A. & Keskin, O. Hot spots in protein-protein interfaces: towards drug discovery. *Prog Biophys Mol Biol* **116**, 165-173, doi:10.1016/j.pbiomolbio.2014.06.003 (2014).

103   Zerbe, B. S., Hall, D. R., Vajda, S., Whitty, A. & Kozakov, D. Relationship between hot spot residues and ligand binding hot spots in protein-protein interfaces. *J Chem Inf Model* **52**, 2236-2244, doi:10.1021/ci300175u (2012).

104   Pfaff, S. J., Chimenti, M. S., Kelly, M. J. & Arkin, M. R. Biophysical methods for identifying fragment-based inhibitors of protein-protein interactions. *Methods Mol Biol* **1278**, 587-613, doi:10.1007/978-1-4939-2425-7_39 (2015).

105   Shuker, S. B., Hajduk, P. J., Meadows, R. P. & Fesik, S. W. Discovering high-affinity ligands for proteins: SAR by NMR. *Science* **274**, 1531-1534, doi:10.1126/science.274.5292.1531 (1996).

106   Harner, M. J., Frank, A. O. & Fesik, S. W. Fragment-based drug discovery using NMR spectroscopy. *J Biomol NMR* **56**, 65-75, doi:10.1007/s10858-013-9740-z (2013).

107   Lund, G. *et al.* Inhibition of CDC25B phosphatase through disruption of protein-protein interaction. *ACS Chem Biol* **10**, 390-394, doi:10.1021/cb500883h (2015).

108   Hartshorn, M. J. *et al.* Fragment-based lead discovery using X-ray crystallography. *J Med Chem* **48**, 403-413, doi:10.1021/jm0495778 (2005).

109   Friberg, A. *et al.* Discovery of potent myeloid cell leukemia 1 (Mcl-1) inhibitors using fragment-based methods and structure-based design. *J Med Chem* **56**, 15-30, doi:10.1021/jm301448p (2013).

110   Wyss, D. F. *et al.* Combining NMR and X-ray crystallography in fragment-based drug discovery: discovery of highly potent and selective BACE-1 inhibitors. *Top Curr Chem* **317**, 83-114, doi:10.1007/128_2011_183 (2012).

111   Maurer, T. Advancing fragment binders to lead-like compounds using ligand and protein-based NMR spectroscopy. *Methods Enzymol* **493**, 469-485, doi:10.1016/B978-0-12-381274-2.00018-2 (2011).

112   Jhoti, H., Cleasby, A., Verdonk, M. & Williams, G. Fragment-based screening using X-ray crystallography and NMR spectroscopy. *Curr Opin Chem Biol* **11**, 485-493, doi:10.1016/j.cbpa.2007.07.010 (2007).

113   Parekh, S., Prive, G. & Melnick, A. Therapeutic targeting of the BCL6 oncogene for diffuse large B-cell lymphomas. *Leuk Lymphoma* **49**, 874-882, doi:10.1080/10428190801895345 (2008).

114   Cheng, H. *et al.* Identification of Thiourea-Based Inhibitors of the B-Cell Lymphoma 6 BTB Domain via NMR-Based Fragment Screening and Computer-Aided Drug Design. *J Med Chem* **61**, 7573-7588, doi:10.1021/acs.jmedchem.8b00040 (2018).

Chapter 2

GAS41 Recognizes Diacetylated Histone H3 through a Bivalent Binding Mode

*The text and data presented here are adapted from the following manuscript:

Cho, H. J., Li, H., Linhares, B. M., Kim, E., Ndoj, J., Miao, H., Grembecka, J., and T. Cierpicki. "GAS41 Recognized Diacetylated Histone H3 through a Bivalent Binding Mode." *ACS Chemical Biology*. **13**, 2739 – 46 (2018).

A. Abstract

GAS41 is a chromatin-associated protein that belongs to the YEATS family and is involved in the recognition of acetyl-lysine in histone proteins. A unique feature of GAS41 is the presence of a C-terminal coiled-coil domain, which is responsible for protein dimerization. Here, we characterized specificity of the GAS41 YEATS domain, and found that it preferentially binds to acetylated H3K18 and H3K27 peptides. Interestingly, we found that full-length, dimeric GAS41 binds to di-acetylated H3 peptides with an enhanced affinity when compared to mono-acetylated peptides, through a bivalent binding mode. We determined the crystal structure of the GAS41 YEATS domain with H3K23acK27ac to visualize the molecular basis of di-acetylated histone binding. Our results suggest a unique binding mode in which full-length GAS41 is a reader of di-acetylated histones.

B. Introduction

GAS41, also known as YEATS4, is a nuclear protein encoded by the *GAS41* (glioma-amplified sequence 41) gene that was identified in a glioblastoma cell line[1]. It is a member of the YEATS family of proteins, and is implicated in chromatin remodeling and transcriptional regulation[2-3]. GAS41 interacts with chromatin-modifying complexes, including TIP60 and SRCAP[4-5], and promotes deposition of histone H2A.Z[6]. Functional studies implicate GAS41 in cell growth and survival[7-8], and in the maintenance of embryonic stem cell identity[9]. Acetylation of histone lysine residues is abundant in cells, and a high level of acetylated histones typically correlates with actively transcribed genes[10]. Proteins involved in transcriptional regulation recognize acetylated lysine residues utilizing two major families of reader domains: bromodomains[11] and YEATS domains[12]. There are four YEATS domain-containing proteins in humans: ENL, YEATS2, AF9, and GAS41. Both AF9 and ENL bind histone H3 peptides with acetylated K9, K18, and K27, with low- to mid-micromolar affinities[13-14]. Recently, GAS41 has been shown to interact with H3K14ac and H3K27ac peptides with affinities ranging from 9.3 to 32.7 μM[6,9], and with H3K122 succinylated peptide at low pH (6.0)[15]. Notably, YEATS domains have been observed to bind histone peptides with crotonylated lysine with higher affinities[16-19]. This is particularly evident for YEATS2, which favors crotonylated lysine with a 7-fold stronger binding affinity when compared to acetylated lysine[16].

In this study, we report that the YEATS domain of GAS41 binds acetylated histone peptides with moderate, mid-micromolar affinities, and preferentially recognizes acetylated H3K18 and H3K27. GAS41 contains a C-terminal coiled-coil domain, and we demonstrate that full-length GAS41 is dimeric in HEK293T cells. We found that full-length GAS41 binds

with an enhanced affinity towards di-acetylated over mono-acetylated histone H3, *in vitro* and in pull-down experiments. We determined the crystal structure of the YEATS domain in a complex with H3K23acK27ac, illustrating a bivalent mechanism of di-acetylated histone recognition by GAS41. Our findings suggest a unique recognition mode of acetylated histone by GAS41 through higher-order interactions.
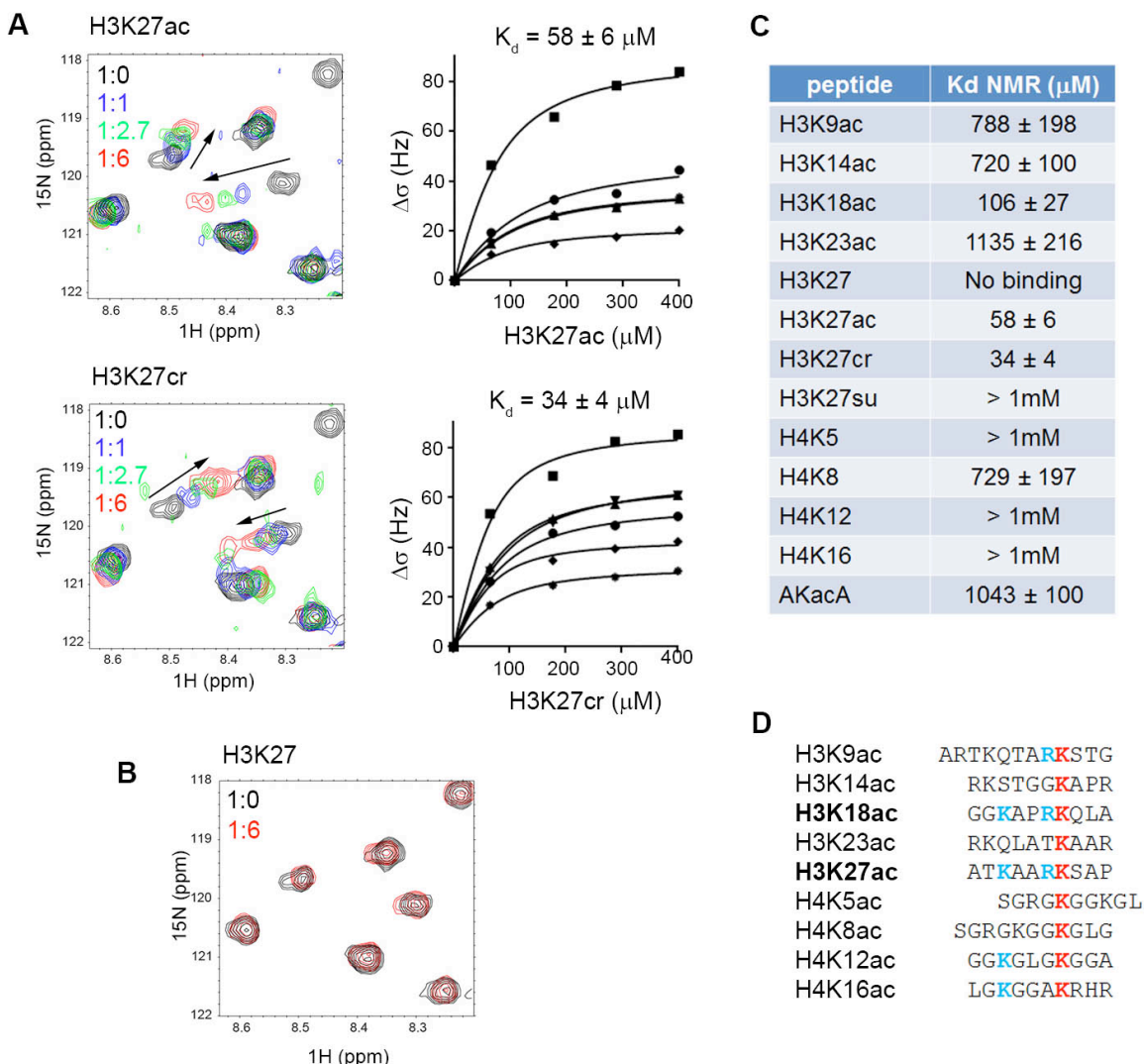
Figure 2.1. Characterization of the binding affinity and selectivity of the GAS41 YEATS domain. A. ^1H-^{15}N HSQC spectra of the 65 μM ^{15}N-labeled YEATS domain (black) titrated with H3K27ac and H3K27cr peptides. The YEATS–peptide ratio is labeled. Graphs show the chemical shift titrations used to determine K_D values. B. ^1H-^{15}N HSQC spectra of the 65 μM

[15]N-labeled YEATS domain (black) with 400 μM unmodified H3K27 peptide (red). C. Binding affinities for a series of peptides toward the YEATS domain of GAS41 determined from NMR titrations. D. Sequence alignment of the peptides used in the binding experiments. Acetyl-lysine is shown in red, and positively charged residues upstream of acetyl-lysine are labeled in cyan.

C. Results and Discussion

C.1. The GAS41 YEATS domain preferentially binds H3K18ac and H3K27ac peptides

Previous studies revealed that the human YEATS domain-containing proteins AF9 and ENL bind histone H3 peptides with acetylated K9, K18, and K27, with low- to mid-micromolar affinities [13-14]. To determine the binding affinity and specificity of the GAS41 YEATS domain, we tested a series of peptides derived from histones H3 and H4 with mono-acetylated lysine residues encompassing major sites for lysine acetylation. The recombinant GAS41 YEATS domain yields a well-dispersed ^{1}H-^{15}N HSQC spectrum, which enabled us to perform nuclear magnetic resonance (NMR) titration experiments and determine the binding affinity of selected histone peptides (**Figure 2.2.**).
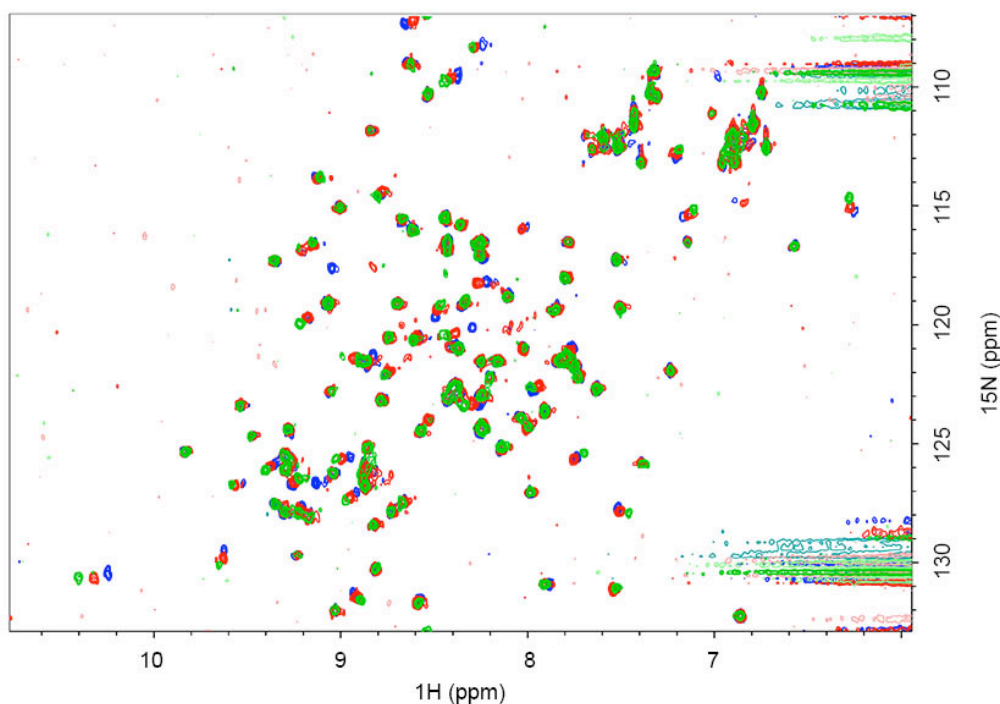
Figure 2.2. GAS41 YEATS domain binds acetylated-Histone H3 peptides H3K14ac and H3K27ac. ^{1}H-^{15}N HSQC of 60 μM GAS41 YEATS domain (blue) with 400 μM H3K27ac (green) and 400 μM H3K14ac (red).

We found that all acetylated peptides bind to the YEATS domain with relatively

modest affinities, and the most potent binding was observed for H3K27ac (K_D = 58 μM) and

H3K18ac (K_D = 106 μM) (**Figure 2.1.A.** and **2.1.C.**). All remaining peptides bind to the

GAS41 YEATS domain with at least 10-fold lower affinity (**Figure 2.1.C.**). Contradictory to

the recent report, we have observed a low affinity of GAS41 YEATS toward H3K14ac (K_D =

720 μM) (**Figure 2.3.**). No binding was detected for the non-acetylated H3K27, indicating

that acetylated lysine is essential for the interaction with the GAS41 YEATS domain (**Figure**

2.1.B.).
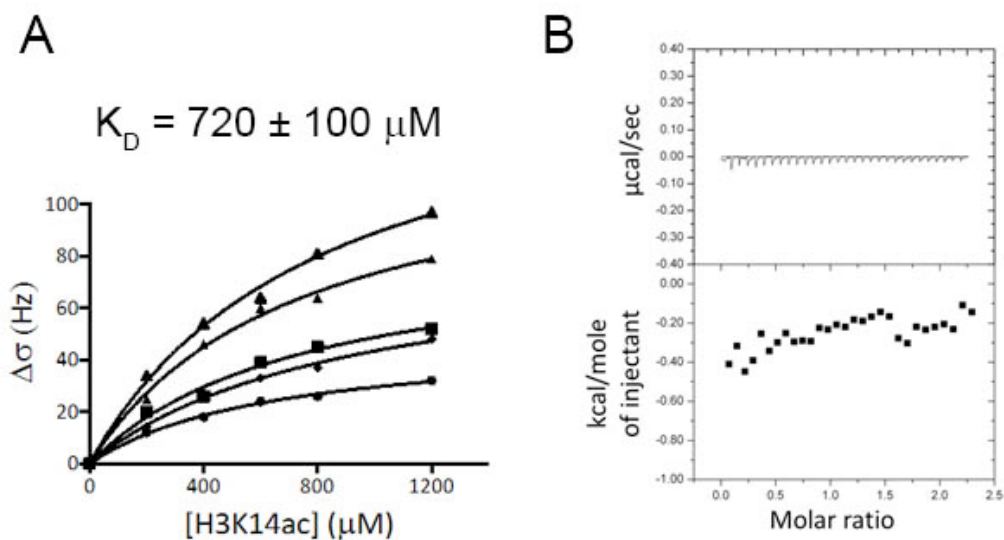
Figure 2.3. Characterization of H3K14ac binding to GAS41 YEATS domain. A. Chemical shift titrations used to determine K_D of H3K14ac for GAS41 YEATS domain by NMR. **B.** Results of ITC experiment to determine affinity of H3K14ac for GAS41 YEATS domain.

To assess the contribution of acetylated lysine, we tested the tripeptide AKacA and found that it binds to the GAS41 YEATS domain with a K_D of ~1 mM, making it comparable to weakly binding acetylated H3 and H4 peptides (**Figure 2.1.C.**). Next, we performed a sequence alignment of all tested peptides and found that two common features of the two peptides binding with the highest affinities, namely H3K18ac and H3K27ac, are the presence of an arginine residue directly preceding the acetylated lysine and an additional lysine located four residues upstream of the acetyl-lysine. This result suggests that these positively charged residues are likely important for the recognition of acetylated histone H3 by GAS41.
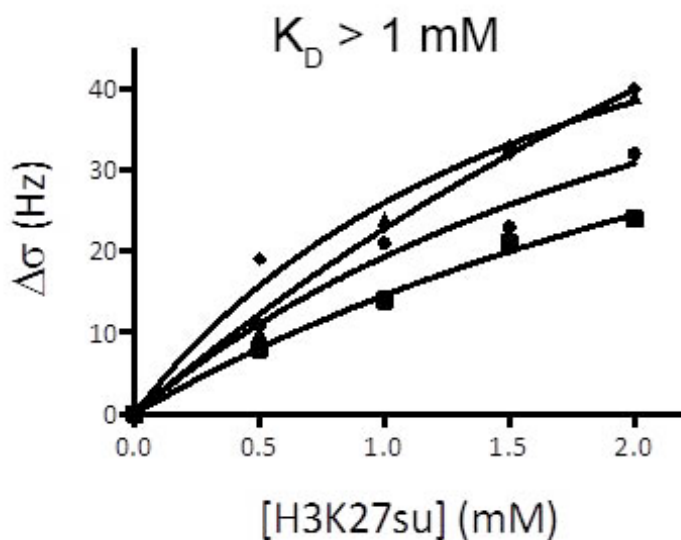
Figure 2.4. Characterization of H3K27su binding to GAS41 YEATS domain. Chemical shift titrations used to determine K_D of H3K27su for GAS41 YEATS domain by NMR-titration experiments.

Our data demonstrate that the binding affinities of GAS41 YEATS toward acetylated H3 peptides are relatively modest, comparable to those of ENL and YEATS2[14,16], and in agreement with recently reported data for GAS41[6]. Several recent reports revealed that YEATS domains recognize crotonylated lysine residues with higher affinity than that for acetylated lysines[16-19]. Among these proteins, YEATS2 was shown to have the largest selectivity (approximately seven-fold) toward crotonylated over acetylated lysine[16]. To assess whether GAS41 YEATS binds crotonylated lysine, we tested the binding of the H3K27cr peptide and found an approximate two-fold improvement in the binding affinity ($K_D = 34\ \mu$M) compared with H3K27ac (**Figure 2.1.A** and **2.1.C.**). According to the recent studies, the GAS41 YEATS domain binds H3 peptide succinylated on K112 at acidic pH (6.0)[15]. We have tested binding of an H3 peptide with succinylated K27 (H3K27su) at pH 7.5, and found only modest binding with K_D >1 mM (**Figures 2.1.C.** and **2.4.**) suggesting that GAS41 is not recognizing K27su at physiological pH.

C.2. Full-length GAS41 YEATS is dimeric in HEK293T cells

The sequence of GAS41 reveals the presence of a coiled-coil domain immediately following the YEATS domain (**Figure 2.5.**), and previous studies have shown that the GAS41 coiled-coil region is dimeric in solution[20]. We expressed the C-terminal fragment of GAS41 encompassing residues 149-227, which includes the coiled-coil domain. Using circular dichroism (CD) spectra, we demonstrate helical secondary structure of the C-terminal coiled-coil in solution (**Figure 2.6.**). Thermal denaturation indicates melting transition with $T_m =$ 29˚C (**Figure 2.5.B.**), which is consistent with the previously reported stability of GAS41 coiled-coil derived peptides[20].
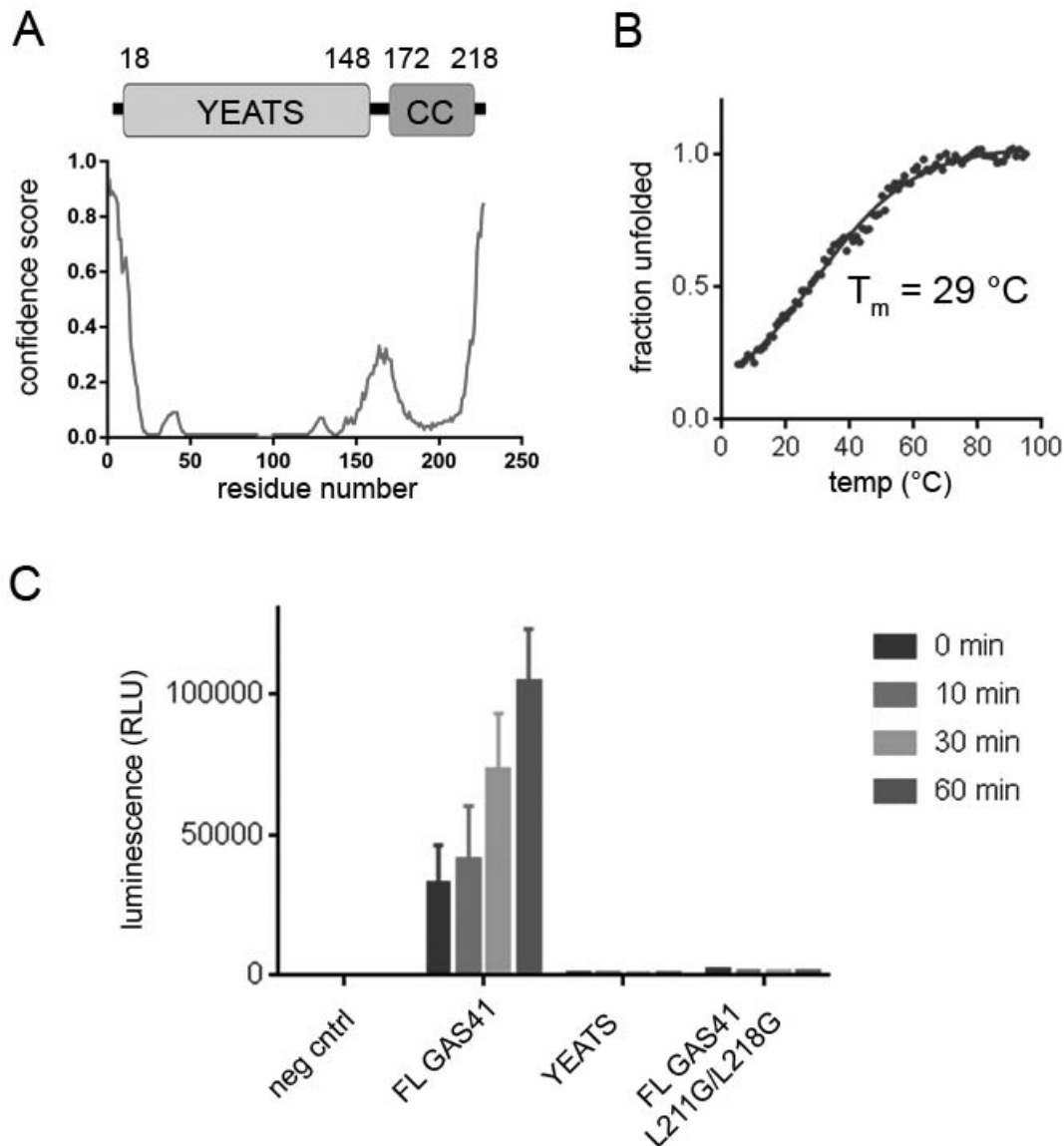
Figure 2.5. GAS41 is dimeric. A. Disorder prediction for GAS41 using PSIPRED server (http://bioinf.cs.ucl.ac.uk/psipred). B. Thermal denaturation of C-terminal fragment of GAS41 (residues 149-227) determined from CD spectra by monitoring signal at 208 nm. C. NanoBit assay showing luminescence signal in HEK293T cells transfected with three GAS41 constructs cloned into LgBit and SmBit vectors. Luminescence was measured 48h following the transfection after adding Nano-Glo® Live Cell Reagent at four time points.

To test whether full-length GAS41 forms a dimer in cells, we employed the NanoBit protein-protein interaction assay[21]. We expressed GAS41 as fusions with LgBit and SmBit proteins in HEK293T cells, and observed a strong luminescence signal, indicating that full-

length GAS41 is indeed dimeric in eukaryotic cells (**Figure 2.5.C.**). On the contrary, we have

not observed dimerization of the YEATS domain alone (**Figure 2.5.C.**). Indeed, the

dimerization of GAS41 has been suggested in previous studies, and L211G and L218G

mutations disrupting the coiled-coil motif impaired GAS41-mediated activation of p53 tumor

suppressor pathway[22]. We tested GAS41 L211G/L218G double mutant in NanoBit assay and

found that these mutations disrupted GAS41 dimerization (**Figure 2.5.C.**).
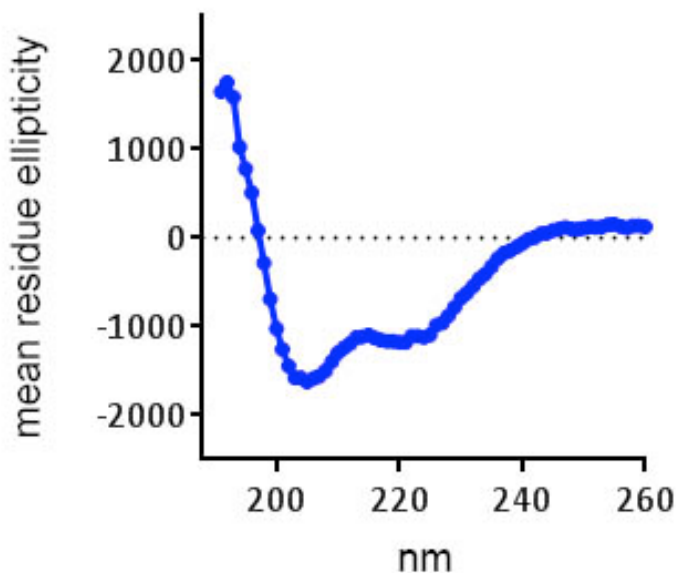
Figure 2.6. GAS41 C-terminus presents Coiled-coil character by Circular Dichroism (CD) spectroscopy. CD spectrum of 50 μM GAS41 C-terminal fragment (residues 149 – 227) in 20 mM sodium phosphate, pH 6.5, 250 mM NaF at 25 ˚C.

C.3. Full-length GAS41 YEATS domain binds with increased affinity to di-acetylated histone H3

We hypothesized that full-length, dimeric GAS41 may simultaneously recognize di-acetylated histone H3 through a bivalent interaction mode, which could result in enhanced

binding affinity. To characterize the interaction of full-length GAS41 with acetylated

peptides, we employed bio-layer interferometry using selectively biotinylated GAS41. First, we tested mono-acetylated H3K18ac and H3K27ac peptides, and found mid-micromolar K_D values (36 μM and 81 μM, respectively, [**Figure 2.7.A.** and **2.7.B.**]) that are consistent with the NMR results for the isolated YEATS domain (**Figure 2.1.**). Subsequently, we tested the binding of di-acetylated peptides and found low-micromolar affinities for H3K18acK27ac (K_D = 3.2 μM) and H3K23acK27ac (K_D = 5.0 μM) (**Figures 2.7.C.** and **2.7.D.**). This represents a >10-fold enhancement in the binding affinity of the full-length GAS41 for di-acetylated H3 peptides.
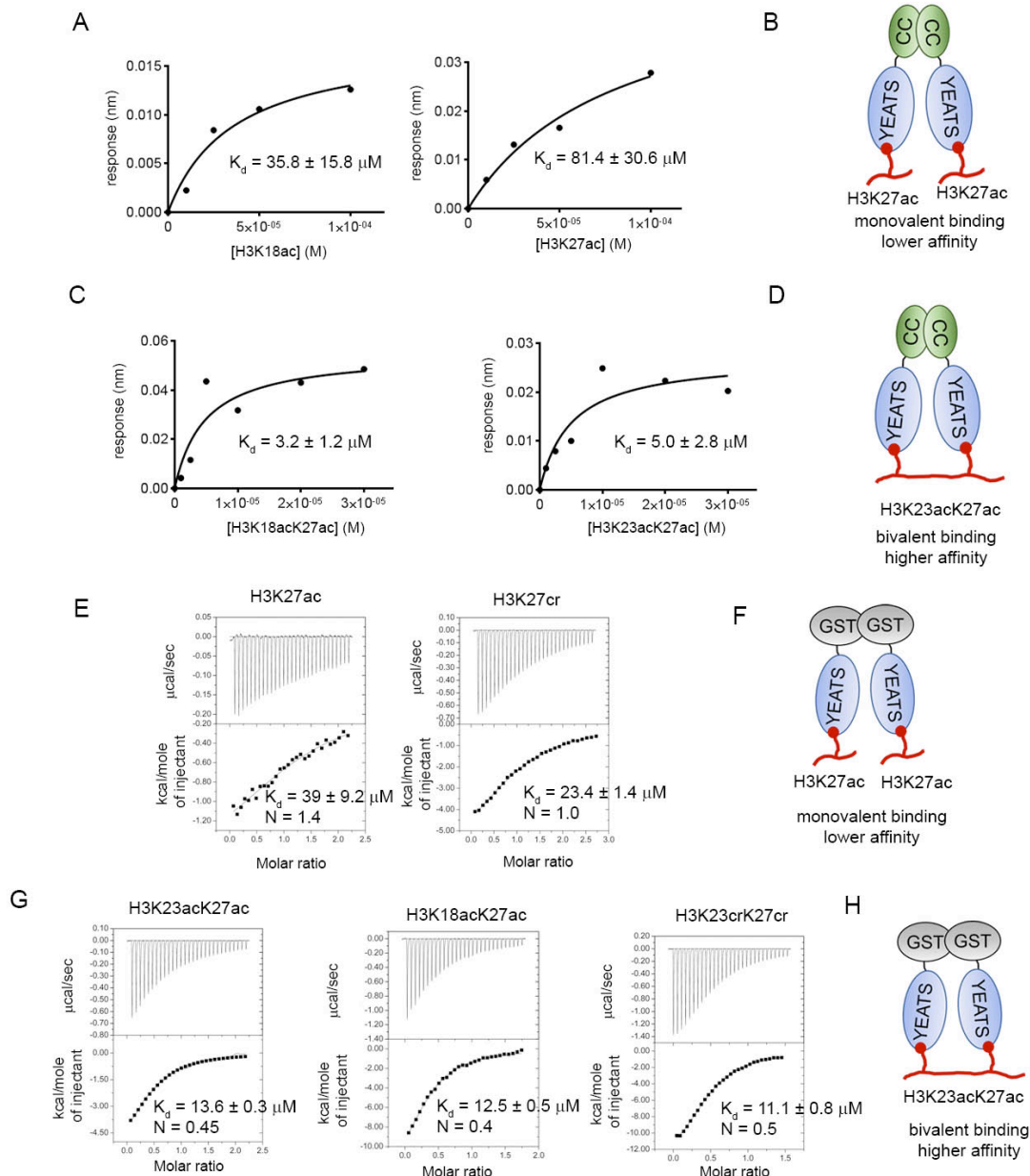
Figure 2.7. Affinity of acylated-H3 peptides towards full-length GAS41 and GST-YEATS constructs. A. Binding of H3K18ac and H3K27ac to the full-length GAS41 determined using bio-layer interferometry experiments and corresponding model of studied protein-peptide interaction B. C. Binding of H3K18acK27ac and H3K23acK27ac to the full-length GAS41 determined using bio-layer interferometry experiments and corresponding model is shown in panel D. E. Determination of the affinity and stoichiometry of binding between GST-YEATS and H3K27ac and H3K27cr peptides using ITC and corresponding model is shown in panel F. G. Characterization of the binding of GST-YEATS and di-acylated H3 peptides using ITC and corresponding model in panel H.

C.4. Dimerization of YEATS domain increases the affinity toward the di-acetylated histone H3

To test whether dimerization of the YEATS domain increases the affinity for di-acetylated H3 peptides in a model system, we purified a dimeric GST-YEATS construct and performed isothermal titration calorimetry (ITC) experiments (**Table 2.1.**). We found that GST-YEATS binds mono-acetylated H3K27ac peptide with a 1:1 stoichiometry, and with a binding affinity similar to that of an isolated YEATS domain (K_D = 39 μM, **Figures 2.7.E. and 2.7.F.**). We also confirmed that crotonylated lysine slightly increases the binding affinity of H3K27cr for GST-YEATS (K_D = 23.4 μM, **Figure 2.7.E.**). Subsequently, we tested the di-acetylated H3K23acK27ac peptide, and found that it binds to GST-YEATS with K_D = 15.0 μM and 1:2 stoichiometry (a single molecule of H3K23acK27ac binds two molecules of YEATS domain [**Figure 2.7.G.**]). A very similar binding affinity and a very similar stoichiometry were observed for H3K18acK27ac (K_D = 12.8 μM and a 1:2 stoichiometry [**Figure 2.7.G.**]). Therefore, dimerization of the YEATS domain in a model system using GST fusion leads to two- to three-fold stronger binding of di-acetylated peptides than of mono-acetylated H3K27ac. The stoichiometry clearly indicates that the dimerized YEATS domain recognizes di-acetylated H3, and the enhanced affinity versus that of mono-acetylated H3 most likely results from a bivalent binding mode (**Figure 2.7.H.**). We also found a very similar effect for a double crotonylated H3K23crK27cr peptide (**Figure 2.7.G.**), which binds to GST-YEATS with a K_D = 11.1 μM and 1:2 stoichiometry, consistent with an approximately two- to three-fold enhanced affinity compared with that of H3K27cr.

Table 2.1. Binding affinity, stoichiometry and thermodynamic parameters for GAS41 GST-YEATS and H3-derived peptide obtained from ITC.

Peptide	K_D (µM)	N	ΔH (cal/mol)	ΔS (cal/mol/K)
H3K27ac	39.0±9.2	1.4	-9,190±550	-7.7
H3K27cr	23.4±1.4	1.0	-6,850±190	-1.9±3.4
H3K23acK27ac	15.0±1.3	0.51	-8,940±860	-7.9±2.7
H3K18acK27ac	12.8±0.1	0.4	-11,800±3400	-17.1±11.6
H3K23crK27cr	9.6±0.7	0.48	-9,200±10	-8.0±0.3

Average values from two independent experiments are reported.

C.5. Di-acetylated H3 peptide pulls-down GAS41 from eukaryotic cells

We also assessed whether di-acetylated H3K23acK27ac peptide can interact with full-length GAS41 expressed in eukaryotic cells. Transfection of HEK293T cells with the construct encoding wild-type GAS41 resulted in robust overexpression of the full-length protein (**Figure 2.10.**). We then used these cells to perform the pull-down, using biotinylated, di-acetylated H3K23acK27ac and mono-acetylated H3K27ac peptides. While both peptides pulled-down GAS41, the interaction with H3K23acK27ac is considerably stronger when compared than that with the H3K27ac peptide (**Figure 2.10.**). To confirm that this effect is specific, we introduced W93A mutation that is known to impair YEATS domain function[6,9] and found that it leads to the disruption of the binding (**Figure 2.10.**). Altogether, pull-down experiments validate the enhanced affinity of full-length GAS41 towards di-acetylated H3, and are consistent with K_D values determined in biochemical experiments.
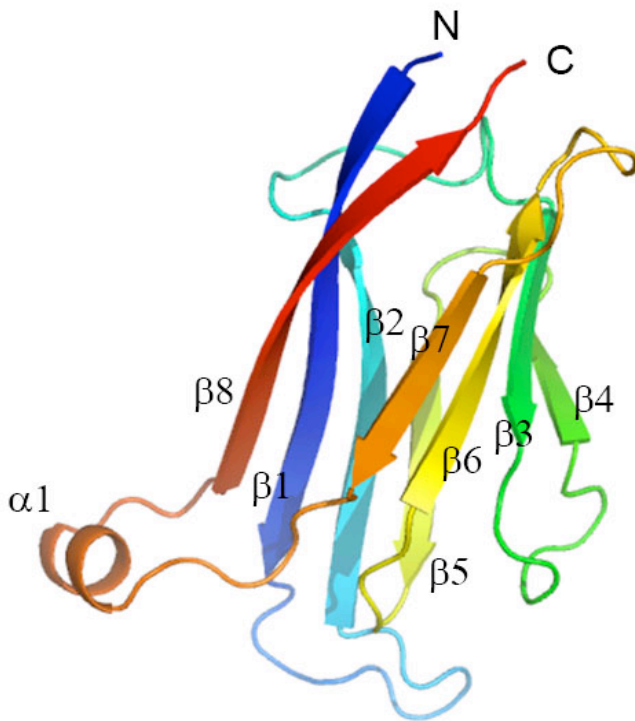
Figure 2.8. Monomeric structure of GAS41 YEATS. A ribbon diagram showing the overall structure of GAS41 YEATS. It is colored according to the sequence by a rainbow color from the N-terminus (blue) to the C-terminus (Red). Crystal structure of GAS41 YEATS domain encompassing residues 20-146. Secondary structure elements are labeled.

C.6. Structural basis for the recognition of di-acetylated histone H3 by the GAS41 YEATS domain

To determine how the GAS41 YEATS domain recognizes acetylated histone peptides, we used X-ray crystallography. First, we determined the structure of the YEATS domain at 2.1 Å resolution (**Figure 2.8.** and **Table 2.2.**). The GAS41 YEATS domain adopts an immunoglobulin fold with a two-layer β-sandwich consisting of eight antiparallel β-strands (**Figure 2.8.**), and is similar to the previously reported structures of the YEATS domains from AF9, ENL, and YEATS2 (**Figure 2.9.A.**). We found a short α-helical region encompassing

residues 123–129 as a new structural element not observed in other YEATS domains (**Figures 2.8. and 2.9.A.**).
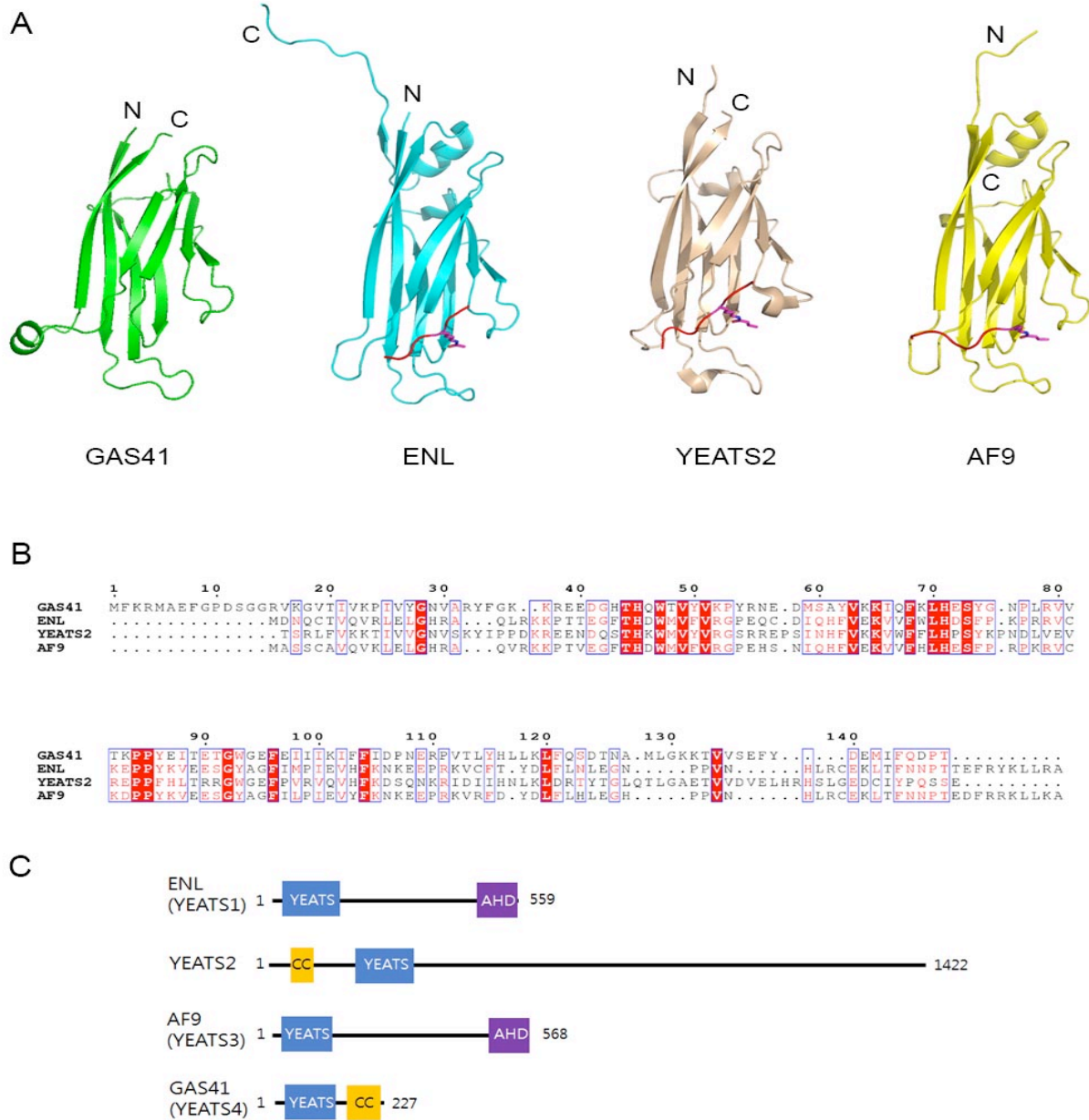
Figure 2.9. Comparison of the structures of human YEATS domains. **A**. Structures of YEATS domains from: GAS41 (green, PDB 5VNA), ENL-H3K27ac complex (cyan, PDB 5J9S), YEATS2-H3K27cr complex (wheat, PDB 5IQL), AF9-H3K9cr complex (yellow, PDB 5HJB). **B**. Alignment of human YEATS domain sequences. **C**. Cartoon representation of

domain structures of YEATS domain proteins. CC – coiled coil predicted using COILS server; AHD – ANC1 homology domain.

Analysis of the packing in the crystal structure revealed four YEATS domain molecules in the asymmetric unit, with two molecules showing accessible sites for the binding of acetylated peptides. To determine the crystal structure of a complex with the histone peptide, we performed soaking of the YEATS domain crystals with di-acetylated peptides and found the best diffraction (2.4 Å resolution) for H3K23acK27ac complex.
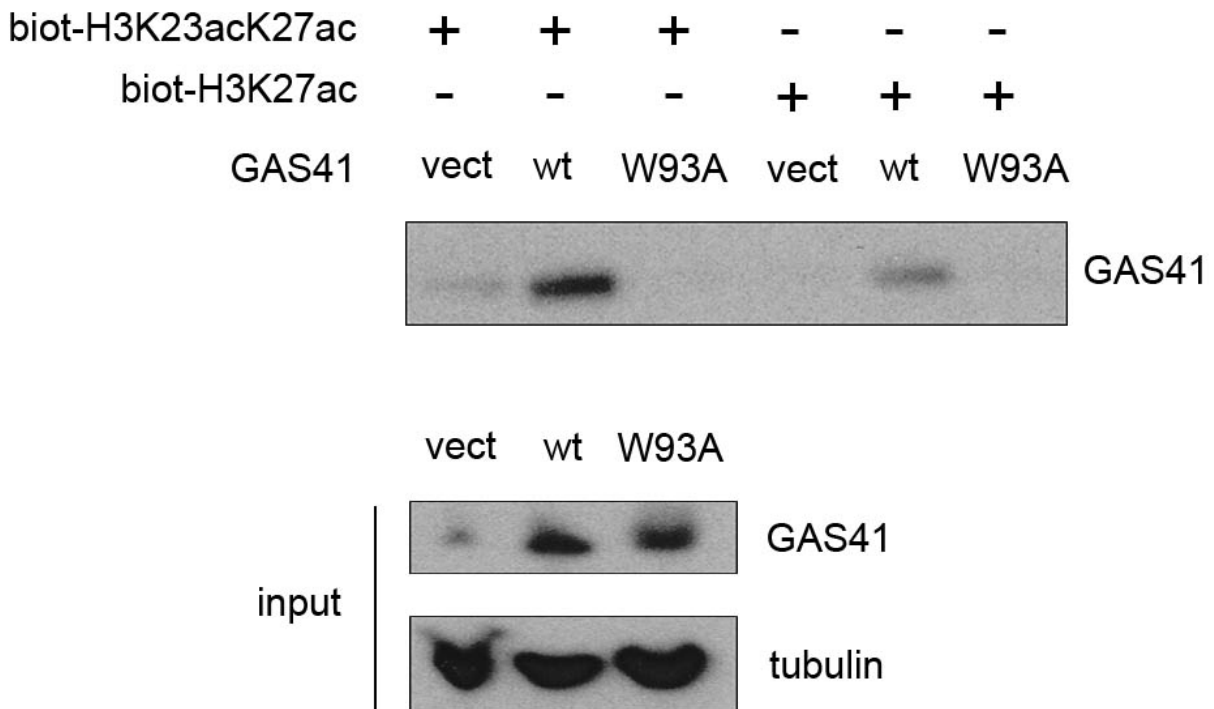
Figure 2.10. Peptide pull-down assay with biotinylated H3K23acK27ac and H3K27ac peptides using 293T cells expressing GAS41. Samples pulled-down using streptavidin beads were immunoblotted with anti-GAS41 antibody. GAS41 and tubulin were detected in input samples (whole cell lysates).

Interestingly, we found that di-acetylated H3K23acK27ac binds simultaneously to the two YEATS domains in the crystal structure and that acetylated K23 and K27 residues occupy the corresponding pockets in two different molecules of the YEATS domain (**Figure 2.12.A.**). Importantly, this binding mode is consistent with a 1:2 binding stoichiometry from the ITC experiment (**Figure 2.7.G.**).
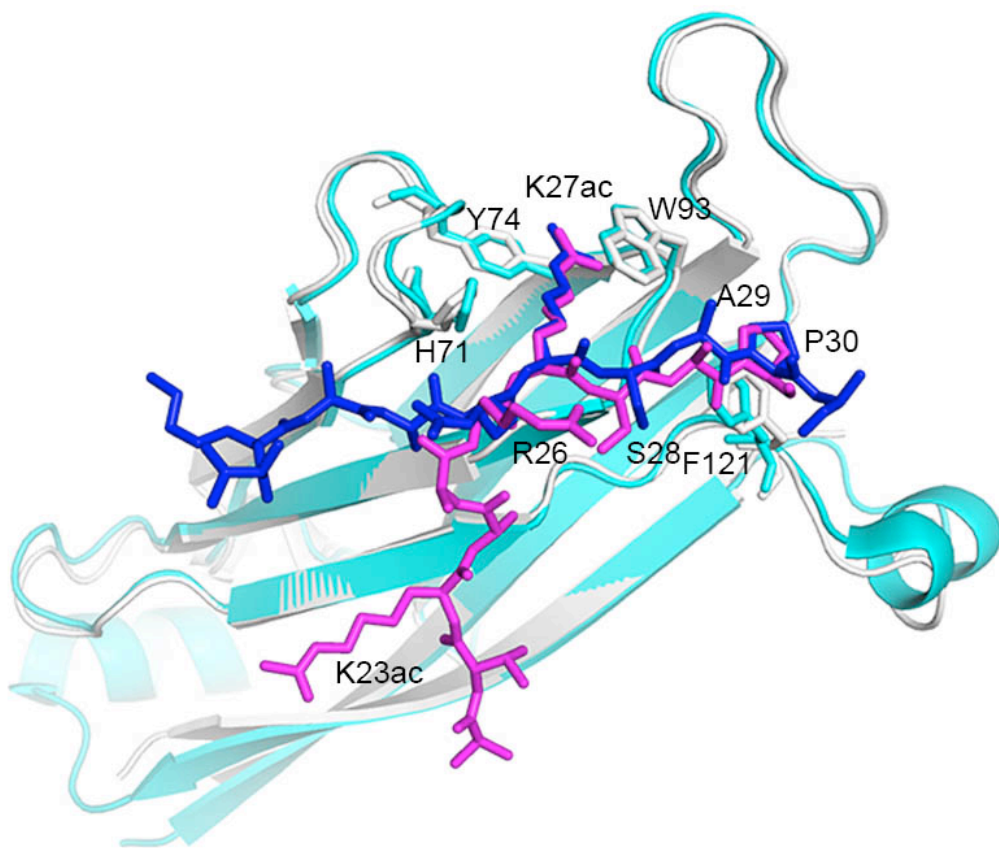
Figure 2.11. Superposition of GAS41 YEATS domain crystal structures with bound acetylated-histone peptide. YEATS domain determined in this study is white and H3K23acK27ac peptide is in magenta. YEATS domain from the complex with H3K27ac (PDB 5XTZ) is cyan and H3K27ac peptide is blue. Selected residues in histone peptides and YEATS domains are labeled.

Analysis of the crystal structure reveals that only very short segments of the H3 peptide are in contact with the YEATS domains, and the major contacts involve recognition of

43

the acetylated K23 and K27 side chains (**Figure 2.12.B.**). Acetylated lysine fits into the channel comprised of the aromatic side chains of H71, Y74, W93, and F96. The acetyl group is recognized by a network of hydrogen bonds, including interactions with the backbone amides of W93 and G94, as well as the side chain of S73 (**Figure 2.12.B.**). The binding mode of the H3 fragment surrounding K27ac is very similar to the recently reported structure of the GAS41 YEATS domain bound to H3K27ac peptide (**Figure 2.11.**). Both structures show the presence of a hydrophobic interaction between H3P30 and GAS41 F121, which may explain the enhanced specificity towards H3K27ac peptide (**Figure 2.11.**). Notably, the structure of the YEATS domain with bound H3K23acK27ac reveals a reversed direction of polypeptide around H3K23ac, suggesting some plasticity in recognition of acetylated lysine residues. Overall, the crystal structure of GAS41 YEATS with bound H3K23acK27ac represents a snapshot of the bivalent recognition mode of di-acetylated histone H3 by full-length, dimeric GAS41.
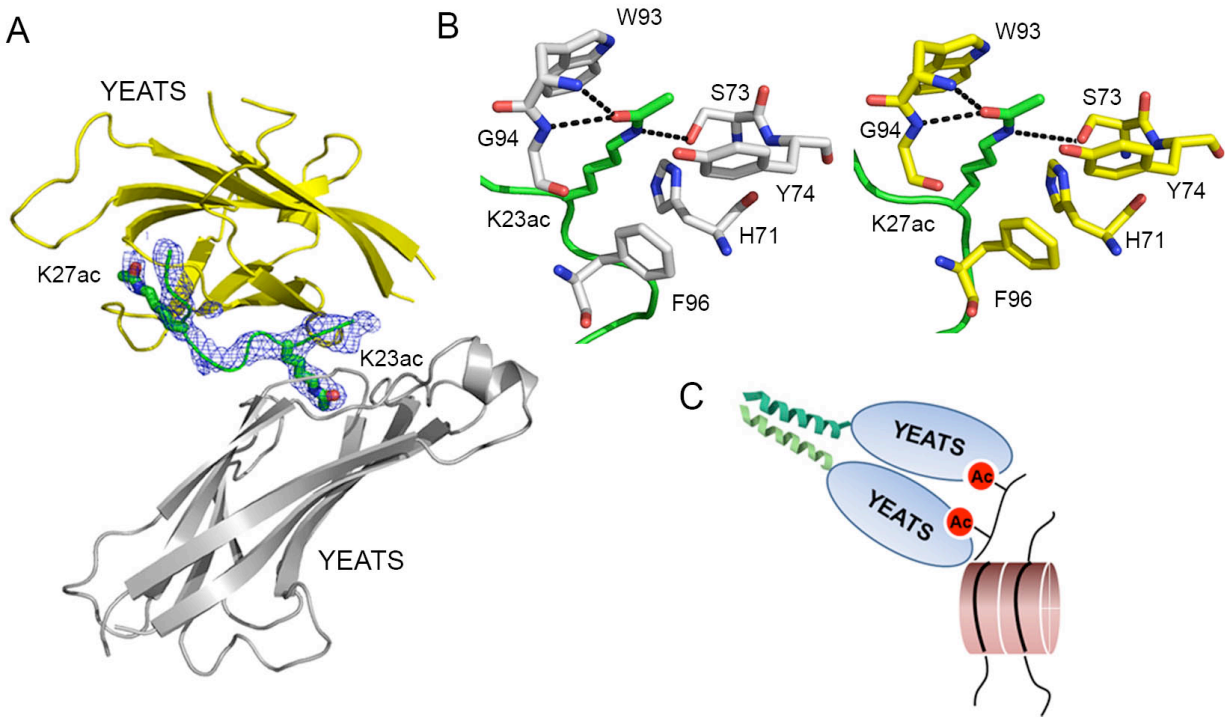
Figure 2.12. Crystal structure of GAS41 YEATS in complex with H3K23acK27ac. A. Structure of the GAS41 YEATS domain with bound H3K23acK27ac determined at 2.4 Å resolution showing Fo - Fc omit electron density maps contoured at 3.0 σ level for H3K23acK27ac (blue mesh). The YEATS domains are shown as yellow and white ribbons, and the H3K23acK27ac peptide is shown in green, with acetylated lysines shown as sticks. B. Details of the binding of K23ac and K27ac to the GAS41 YEATS domains. C. Model of the bivalent recognition of di-acetylated histone protein by full-length GAS41.

D. Conclusions

GAS41 belongs to a four-member family of human proteins, characterized by the presence of the conserved YEATS domain. GAS41 is the shortest protein in this family and is composed of two domains, namely, the N-terminal YEATS and C-terminal coiled-coil (**Figure 2.5.**). In this study, we investigated the binding of the GAS41 YEATS domain to a series of histone H3- and H4-derived peptides with acetylated lysine residues and found that it preferentially recognizes H3K18ac and H3K27ac with modest affinities (K_D values ranging from 35 to 106 μM). The presence of the C-terminal coiled-coil suggests GAS41

dimerization[20,22]. Indeed, we have validated that full-length GAS41 is dimeric in HEK293T cells. Disorder prediction for the GAS41 sequence indicates that the linker between the YEATS and the coiled-coil domain is relatively short and consists of approximately 20 amino acids (**Figure 2.5.**). This implicates the close proximity of the two YEATS domains in dimeric GAS41. In this study, we found that full-length GAS41 preferentially binds to di-acetylated H3 peptides with a ~10-fold enhanced affinity when compared with that for mono-acetylated H3 (**Figure 2.7.A.** and **2.7.C.**). Importantly, we observed a more efficient pull-down of GAS41 using di-acetylated H3K23acK27ac when compared to that of mono-acetylated H3K27ac (**Figure 2.10.**), further validating GAS41 as a reader of di-acetylated histone H3. Our results suggest a model in which full-length GAS41 recognizes di-acetylated histone in a unique, bivalent mode (**Figure 2.12.C.**). We further validated this model by testing the binding of dimeric GST-YEATS domain and found enhanced affinity towards di-acylated (including both di-acetylated and di-crotonylated) versus mono-acylated H3 peptides (**Figure 2.7.E**. and **2.7.G.**).

Histone acetylation is an epigenetic modification important for regulation of gene expression. Recent ChIP-seq experiments revealed co-localization of GAS41 to gene promoters enriched with H3K14ac and H3K27ac[6]. Acetylated lysines H3K14, H3K18 and H3K27 are enriched at promoters of actively transcribed genes[6,23], and enhanced affinity of GAS41 towards di-acetylated H3 may explain why >90% of GAS41 is bound to highly-acetylated gene promoter regions[6]. We also observed approximately two-fold enhanced binding of GAS41 YEATS to peptides with crotonylated versus acetylated lysine. Further studies are required to establish whether GAS41 is a di-acetyl-lysine or di-crotonyl-lysine reader under physiological conditions.

YEATS domain proteins are recently characterized readers of acylated histones, recognizing both acetylated and crotonylated lysine[13, 24]. Among human homologs, only AF9 has been reported to bind acetylated histone proteins with relatively high affinity ($K_D = 2.1 - 3.7$ μM)[13,24], while the K_D values for ENL and YEATS2 interactions are weaker (K_Ds >30 μM)[14,16], and are comparable to the affinity of the GAS41 YEATS domain for mono-acetylated H3[6]. We found that the affinity of GAS41 for di-acetylated H3 is significantly improved through a higher-order interaction involving dimerization of the YEATS domain. Whereas a bivalent binding mode has not been proposed for other members of the YEATS domain family, sequence analysis predicts the presence of the coiled-coil domain in YEATS2 (**Figure 2.9.C.**). This suggests that higher-order interactions might be present in other YEATS protein readers.

Acknowledgements

E. Materials and Methods

47

Cloning, Expression and Protein Purification

The synthetic gene encoding *Human* GAS41 YEATS (residues 13-158) was ordered from Life Technologies and subcloned using the *Bam*HI and *Hind*III restriction sites into the pQE-80L expression vector (Qiagen) with an N-terminal hexahistidine (His$_6$) tag. Recombinant plasmid pQE80L-GAS41 (residues 13-158) was transformed into *E. coli* strain BL21(DE3). Transformants were grown in ^{15}N-labeled M9 minimal medium containing ampicillin at 37 °C until reaching the OD$_{600}$ is between 0.6 and 0.8. After induction with 0.25 mM isopropyl 1-thio-β-D-galactopyranoside (IPTG), cultures were grown for an additional 16 h at 18 °C. Harvested cells were resuspended in lysis buffer (50 mM Tris, pH 7.5, 300 mM NaCl and 1 mM TCEP) and lysed using a cell disrupter. Clarified lysate was applied to Ni–NTA (Qiagen) affinity column. The column was extensively washed with lysis buffer containing 35 mM imidazole and eluted with lysis buffer containing 200 mM imidazole. The eluted pure fractions were pooled and dialyzed against 50 mM Tris, pH 7.5, and 200 mM NaCl buffer and then concentrated to ~60 μM. The final purified ^{15}N-labeled GAS41 (13-158) was used in NMR studies.

Codon-optimized cDNAs of full-length human GAS41 was synthesized by Life Technologies and amplified GAS41 YEATS (residues 1-148) by the polymerase chain reaction (PCR). The PCR product was subcloned using the *Bam*HI and *Eco*RI restriction sites into the pGST-parallel vector [25] with an N-terminal GST tag followed by a TEV cleavage site. Resulting plasmid pGST-GAS41(1-148) was transformed into *E. coli* strain BL21(DE3). Transformed cells were grown in Luria broth medium with ampicillin selection. After 18 h induction with 0.2 mM IPTG at 18°C, cells were harvested by centrifugation and resuspended in lysis buffer containing 50 mM Tris, pH 7.5, 300 mM NaCl and 1 mM TCEP and lysed using a cell

disruptor. The soluble fraction of the cell lysate was then loaded onto a glutathione-Sepharose 4B (GE Healthcare) affinity column. The column was thoroughly washed with buffer containing 50 mM Tris, pH 8.0, and 500 mM NaCl and eluted with 10 mM reduced glutathione. GST-tagged GAS41(1-148) protein was used in the ITC experiment. The eluted proteins were proteolytically cleaved with TEV protease, followed by S-75 size exclusion chromatography purification into buffer containing 20 mM Tris, pH 7.5, and 300 mM NaCl. The final purified GAS41(1-148) was used in the crystallization trial.

The GAS41 coiled coil region (residues 149-227) was cloned from the synthetic *human GAS41* gene by PCR. The PCR product was subcloned using the *Bam*HI and *Hind*III restriction sites into pMocr vector with an N-terminal hexahistidine (His$_6$) tag followed by a TEV cleavage site. GAS41(149-227) protein was expressed as inclusion bodies and solubilized in buffer with 6 M Guanidine hydrochloride. Re-folding was performed using dialysis to 50 mM Tris, pH 7.5, 150 mM NaCl, 2 mM DTT buffer following the cleavage with TEV protease. N-terminal Mocr-His$_6$ was extracted by re-application to Ni-NTA resin. Protein was dialyzed extensively against Storage Buffer (50 mM tris pH 7.5, 150 mM NaCl, 1 mM TCEP) and concentrated for storage at $-80\ ^{\circ}$C.

NMR Titration

All NMR spectra were collected at 30°C on a 600 MHz Bruker Advance III spectrometer equipped with a cryogenic probe, running Topspin version 2.1. Binding of histone peptides to GAS41 YEATS (residues 13-158) were characterized by measuring chemical shift perturbations of selected amide resonances on the ^{1}H-^{15}N HSQC spectra of 60 μM ^{15}N-labeled GAS41$^{13\text{-}158}$ titrated with peptides at molar ratios of 1.0, 2.7, 4.3 and 6.0. Dissociation

constants were determined from least-squares fitting of chemical shift perturbations as a

function of ligand concentration:

Equation 2.1.

$$\delta_i = \frac{\{b - \sqrt{(b^2 - 4 \times a \times c)}\}}{2a}$$

given $a = (K_A/\delta_b) \times [P_t]$, $b = 1 + K_A([L_{ti}] + [P_t])$, and $c = \delta_b \times K_A \times [L_{ti}]$, where δ_i is the

absolute change in chemical shift for each titration point, $[L_{ti}]$ is the total ligand concentration

at each titration point, $[P_t]$ is the total protein concentration, $K_A = 1/K_D$ is the binding

constant, and δ_b is the chemical shift of the resonance in question in the complex[26]. K_D and δ_b

were used as fitting parameters in analysis[26].

Isothermal Titration Calorimetry

The measurements were performed using a VP-ITC titration calorimetric system (MicroCal)

at 25°C. GST-tagged GAS41 YEATS (residues 1-148) was dialyzed extensively against ITC

buffer consisting of 50 mM phosphate, pH 7.5, 150 mM NaCl and 2 mM β-mercaptoethanol.

Histone-derived peptides were directly dissolved in ITC buffer at 500 μM concentration. The

titration curve was obtained by injecting 10 μL aliquots of histone-derived peptides into the

cell containing 50 μM GST-GAS41(1-148), at a time interval of 200 s. All samples were

degassed by vacuum aspiration for 20 min prior to measurements. ITC titration data were

analyzed with a single-site fitting model using Origin 7.0 software.

Determination of binding affinity using bio-layer interferometry

To determine the affinity of full-length GAS41 towards H3 peptides, we employed Bio-Layer

Interferometry experiments using an Octet Red 96 instrument (ForteBio, Inc.). Mono-

biotinylated GAS41 was obtained by co-expression of Avi-tagged GAS41 with BirA enzyme

in *E. coli* BL21(DE3) cells as previously described [27]. The purification procedure for

biotinylated GAS41 was identical to that for the His-tagged GAS41 YEATS (residues 13-158) protein. Prior to protein immobilization, streptavidin biosensor tips were incubated in 50 mM Tris, pH 7.5, and 200 mM NaCl buffer for 600 s. Subsequently, protein was loaded onto tips for 600 s, followed by 1200 s equilibration step. Peptide binding experiments were performed in following order: 600 s equilibration, 300 s association, and 600 s dissociation. Experimental set-up was performed using Octet Data Acquisition Software, and data were analyzed by Octet Data Analysis Software (Pall ForteBio, LLC.). The signal was plotted as a function of ligand concentration to determine K_D, using Prism software (GraphPad Software, Inc.). To correct for drift during association, the slope during the drift period was fit by linear regression, and the product of slope and time (in seconds) was subtracted from the signal.

Circular Dichroism Spectroscopy

Spectra and thermal denaturation experiments were recorded at a GAS41(149-227) concentration of 50 μM in 20 mM sodium phosphate, pH 6.5, and 250 mM sodium fluoride using a Jasco CD-810 spectropolarimeter with constant N_2 flushing. Rectangular cells with 1 mm path length were used, and a circular water bath was used to control temperature of the optic cell chamber. Protein spectra were averaged from three wavelength-scans collected at 0.1 nm intervals from 178 to 260 nm. Baseline spectrum of buffer was recorded prior to the protein spectra being recorded. Mean residue ellipticity was determined by the equation

Equation 2.2.

$$\Theta = (\Theta_{obs} \times \text{mrw})/(10 \times l \times c)$$

where Θ_{obs} is the observed ellipticity (in millidegrees), mrw is mean residue molecular weight, l is optical path length of the cell (in centimeters), and c is peptide concentration. For

thermal denaturation experiments, we recorded CD spectra over temperature gradient from 5 – 95˚C at 1°C degree increments.

X-Ray Crystallography

Protein crystals of GAS41 YEATS(1-148) were obtained by sitting drop method using 6-8.5 mg mL^{-1} protein mixed in a 1:1 volume ratio with a precipitation solution containing 100 mM CHES, pH 9.4, and 1.24 – 1.36 M ammonium sulfate. The crystals were cryoprotected in a reservoir solution supplemented with 25% ethylene glycol. X-ray diffraction data were collected on LS-CAT beam line 21-ID-F at the Advanced Photon Source. The data were then indexed, integrated, and scaled using the HKL2000 suite[28]. The structure was determined by molecular replacement method with the CCP4 version of MOLREP[29] using the structure of yeast Taf14 YEATS domain (PDB code: 3QRL) as a search model. To obtain a histone peptide-bound complex structure, native GAS41 crystals were soaked in a 1:2 mixture of protein buffer and reservoir solution containing saturating amounts of peptide. Model building and structure refinement were carried out using WinCOOT[30] and Phenix.refine[31]. The data statistics are summarized in Table 2.2.

GAS41 pull-down assay

Full-length GAS41 and GAS41 W93A mutant was cloned into the pCMV vector and used to transfect HEK293T cells using FuGENE6 (Promega). Cell lysates were prepared in lysis buffer (PBS with 1 % Triton X-100 and protease inhibitor cocktail) and sonicated. Biotinylated H3K23acK27ac and H3K27ac peptides were incubated with streptavidin magnetic beads for 6 h at 4 °C. Cell lysates were precleared with streptavidin magnetic beads for 2 h at 4 °C and incubated with the biotinylated peptides immobilized on streptavidin magnetic beads overnight at 4 °C. The beads were washed 10 times with wash buffer (PBS

with 0.2% Triton X-100). Pulled-down proteins were analyzed using Western blot with the

GAS41 antibody (sc-393708 from Santa Cruz).

NanoBiT protein-protein interaction assay

The NanoBiT protein-protein interaction assays (Promega) were conducted according to the

manufacturer's instructions. Briefly, full-length GAS41, YEATS(1-158), and the GAS41

L211G/L218G mutant were cloned into pBiT1.1-C[TK/LgBiT] and pBiT2.1-C[TK/SmBiT]

vectors for NanoBiT system. All plasmids were verified by sequencing. HEK293T cells

were plated into 96-well plates at 1×10^4 cells per well. Each Lg-BiT and Sm-BiT plasmids

were co-transfected using FuGENE HD (Promega) the next day and incubated for 48 h. After

we added the Nano-Glo Live Cell Reagent to each well, the luminescence was measured on 0,

10, 30, and 60 min using the PHERAstar FS instrument (BMG Labtech). The positive control

(Lg-PRKAR2A and Sm-PRKACA) and the negative control vectors were provided by the

manufacturer.

Table 2.2. Data collection and refinement statistics of crystal structures of GAS41 YEATS, and of GAS41 YEATS in complex with H3K23acK27ac peptide.

Crystal Form	GAS41 YEATS	GAS41 YEATS – H3K23acK27ac complex
Diffraction Data		
X-Ray source	APS 21ID-F	APS 21ID-F
Wavelength (Å)	0.97872	0.97872
Space group	$P2_12_12_1$	$P2_12_12_1$
Cell dimensions		
a, b, c (Å)	70.16, 79.93, 121.64	70.79, 80.31, 121.66
α, β, γ (°)	90.00, 90.00, 90.00	90.00, 90.00, 90.00
Resolution (Å)	50 – 2.1 (2.18 – 2.10)	50 – 2.4 (2.44 – 2.40)
R_{sym}	0.097 (0.65)	0.078 (0.492)
$I / \sigma I$	27.12 (3.0)	28.0 (3.0)
Completeness (%)	99.6 (99.9)	97.5 (99.9)
Redundancy	7.2 (6.6)	7.0 (6.8)
Refinement		
Resolutions (Å)	48.40 – 2.10	48.67 – 2.40
No. reflections	38,396	25,753
R_{work} /R_{free}	17.38/22.23	18.62/24.77
No. atoms		
Protein	4,210	4,198
Ligand/ion	67	127
Water	394	132
B-factors		
Protein	39.37	58.96
Ligand/ion	81.87	81.84
Water	49.02	50.85
RMSD		
Bond length (Å)	0.02	0.015
Bond angle (°)	1.957	1.643
Ramachandran plot (%)		
Favored	98.59	97.42
Allowed	1.41	2.39
Outliers	0.0	0.2
PDB ID	5VNA	5VNB
Values in parentheses indicate the highest-resolution shell.		

F. References

1.   Fischer, U.; Heckel, D.; Michel, A.; Janka, M.; Hulsebos, T.; Meese, E., Cloning of a novel transcription factor-like gene amplified in human glioma including astrocytoma grade I. *Hum Mol Genet* **1997,** *6* (11), 1817-22.
2.   Schulze, J. M.; Wang, A. Y.; Kobor, M. S., YEATS domain proteins: a diverse family with many links to chromatin modification and transcription. *Biochem Cell Biol* **2009,** *87* (1), 65-75.
3.   Schulze, J. M.; Wang, A. Y.; Kobor, M. S., Reading chromatin: insights from yeast into YEATS domain structure and function. *Epigenetics* **2010,** *5* (7), 573-7.
4.   Doyon, Y.; Selleck, W.; Lane, W. S.; Tan, S.; Cote, J., Structural and functional conservation of the NuA4 histone acetyltransferase complex from yeast to humans. *Mol Cell Biol* **2004,** *24* (5), 1884-96.
5.   Cai, Y.; Jin, J.; Florens, L.; Swanson, S. K.; Kusch, T.; Li, B.; Workman, J. L.; Washburn, M. P.; Conaway, R. C.; Conaway, J. W., The mammalian YL1 protein is a shared subunit of the TRRAP/TIP60 histone acetyltransferase and SRCAP complexes. *J Biol Chem* **2005,** *280* (14), 13665-70.
6.   Hsu, C. C.; Shi, J.; Yuan, C.; Zhao, D.; Jiang, S.; Lyu, J.; Wang, X.; Li, H.; Wen, H.; Li, W.; Shi, X., Recognition of histone acetylation by the GAS41 YEATS domain promotes H2A.Z deposition in non-small cell lung cancer. *Genes Dev* **2018,** *32* (1), 58-69.
7.   Harborth, J.; Elbashir, S. M.; Bechert, K.; Tuschl, T.; Weber, K., Identification of essential genes in cultured mammalian cells using small interfering RNAs. *J Cell Sci* **2001,** *114* (Pt 24), 4557-65.
8.   Zimmermann, K.; Ahrens, K.; Matthes, S.; Buerstedde, J. M.; Stratling, W. H.; Phivan, L., Targeted disruption of the GAS41 gene encoding a putative transcription factor indicates that GAS41 is essential for cell viability. *J Biol Chem* **2002,** *277* (21), 18626-31.
9.   Hsu, C. C.; Zhao, D.; Shi, J.; Peng, D.; Guan, H.; Li, Y.; Huang, Y.; Wen, H.; Li, W.; Li, H.; Shi, X., Gas41 links histone acetylation to H2A.Z deposition and maintenance of embryonic stem cell identity. *Cell Discov* **2018,** *4*, 28.
10.   Bannister, A. J.; Kouzarides, T., Regulation of chromatin by histone modifications. *Cell Res* **2011,** *21* (3), 381-95.
11.   Filippakopoulos, P.; Picaud, S.; Mangos, M.; Keates, T.; Lambert, J. P.; Barsyte-Lovejoy, D.; Felletar, I.; Volkmer, R.; Muller, S.; Pawson, T.; Gingras, A. C.; Arrowsmith, C. H.; Knapp, S., Histone recognition and large-scale structural analysis of the human bromodomain family. *Cell* **2012,** *149* (1), 214-31.
12.   Zhao, D.; Li, Y.; Xiong, X.; Chen, Z.; Li, H., YEATS Domain-A Histone Acylation Reader in Health and Disease. *J Mol Biol* **2017,** *429* (13), 1994-2002.
13.   Li, Y.; Wen, H.; Xi, Y.; Tanaka, K.; Wang, H.; Peng, D.; Ren, Y.; Jin, Q.; Dent, S. Y.; Li, W.; Li, H.; Shi, X., AF9 YEATS domain links histone acetylation to DOT1L-mediated H3K79 methylation. *Cell* **2014,** *159* (3), 558-71.
14.   Wan, L.; Wen, H.; Li, Y.; Lyu, J.; Xi, Y.; Hoshii, T.; Joseph, J. K.; Wang, X.; Loh, Y. E.; Erb, M. A.; Souza, A. L.; Bradner, J. E.; Shen, L.; Li, W.; Li, H.; Allis, C. D.; Armstrong, S. A.; Shi, X., ENL links histone acetylation to oncogenic gene expression in acute myeloid leukaemia. *Nature* **2017,** *543* (7644), 265-269.

15.   Wang, Y.; Jin, J.; Chung, M. W. H.; Feng, L.; Sun, H.; Hao, Q., Identification of the YEATS domain of GAS41 as a pH-dependent reader of histone succinylation. *Proc Natl Acad Sci U S A* **2018,** *115* (10), 2365-2370.

16.   Zhao, D.; Guan, H.; Zhao, S.; Mi, W.; Wen, H.; Li, Y.; Zhao, Y.; Allis, C. D.; Shi, X.; Li, H., YEATS2 is a selective histone crotonylation reader. *Cell Res* **2016,** *26* (5), 629-32.

17.   Li, Y.; Zhao, D.; Chen, Z.; Li, H., YEATS domain: Linking histone crotonylation to gene regulation. *Transcription* **2016**, 1-6.

18.   Zhang, Q.; Zeng, L.; Zhao, C.; Ju, Y.; Konuma, T.; Zhou, M. M., Structural Insights into Histone Crotonyl-Lysine Recognition by the AF9 YEATS Domain. *Structure* **2016,** *24* (9), 1606-12.

19.   Andrews, F. H.; Shinsky, S. A.; Shanle, E. K.; Bridgers, J. B.; Gest, A.; Tsun, I. K.; Krajewski, K.; Shi, X.; Strahl, B. D.; Kutateladze, T. G., The Taf14 YEATS domain is a reader of histone crotonylation. *Nat Chem Biol* **2016,** *12* (6), 396-8.

20.   Hagen, S.; Mattay, D.; Rauber, C.; Muller, K. M.; Arndt, K. M., Characterization and inhibition of AF10-mediated interaction. *J Pept Sci* **2014,** *20* (6), 385-97.

21.   Oh-Hashi, K.; Hirata, Y.; Kiuchi, K., SOD1 dimerization monitoring using a novel split NanoLuc, NanoBit. *Cell Biochem Funct* **2016,** *34* (7), 497-504.

22.   Park, J. H.; Roeder, R. G., GAS41 is required for repression of the p53 tumor suppressor pathway during normal cellular proliferation. *Mol Cell Biol* **2006,** *26* (11), 4006-16.

23.   Wang, Z.; Zang, C.; Rosenfeld, J. A.; Schones, D. E.; Barski, A.; Cuddapah, S.; Cui, K.; Roh, T. Y.; Peng, W.; Zhang, M. Q.; Zhao, K., Combinatorial patterns of histone acetylations and methylations in the human genome. *Nat Genet* **2008,** *40* (7), 897-903.

24.   Li, Y.; Sabari, B. R.; Panchenko, T.; Wen, H.; Zhao, D.; Guan, H.; Wan, L.; Huang, H.; Tang, Z.; Zhao, Y.; Roeder, R. G.; Shi, X.; Allis, C. D.; Li, H., Molecular Coupling of Histone Crotonylation and Active Transcription by AF9 YEATS Domain. *Mol Cell* **2016,** *62* (2), 181-93.

25.   Sheffield, P.; Garrard, S.; Derewenda, Z., Overcoming expression and purification problems of RhoGDI using a family of "parallel" expression vectors. *Protein Expr Purif* **1999,** *15* (1), 34-9.

26.   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* **2003,** *327* (5), 1121-33.

27.   Gray, F.; Cho, H. J.; Shukla, S.; He, S.; Harris, A.; Boytsov, B.; Jaremko, L.; Jaremko, M.; Demeler, B.; Lawlor, E. R.; Grembecka, J.; Cierpicki, T., BMI1 regulates PRC1 architecture and activity through homo- and hetero-oligomerization. *Nat Commun* **2016,** *7*, 13343.

28.   Otwinowski, Z.; Minor, W., Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol* **1997,** *276*, 307-26.

29.   Vagin, A.; Teplyakov, A., Molecular replacement with MOLREP. *Acta Crystallogr D Biol Crystallogr* **2010,** *66* (Pt 1), 22-5.

30.   Emsley, P.; Lohkamp, B.; Scott, W. G.; Cowtan, K., Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* **2010,** *66* (Pt 4), 486-501.

31.   Adams, P. D.; Afonine, P. V.; Bunkoczi, G.; Chen, V. B.; Davis, I. W.; Echols, N.; Headd, J. J.; Hung, L. W.; Kapral, G. J.; Grosse-Kunstleve, R. W.; McCoy, A. J.; Moriarty, N. W.; Oeffner, R.; Read, R. J.; Richardson, D. C.; Richardson, J. S.; Terwilliger, T. C.; Zwart,

P. H., PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* **2010,** *66* (Pt 2), 213-21.

Chapter 3

Development of small-molecule inhibitors targeting Histone H3 acetyl-lysine recognition by GAS41 YEATS domain

A. Abstract

Classified by its N-terminal YEATS domain that confers histone reader function, GAS41 is responsible for recognition of acyl-modified lysine residues of Histone H3. GAS41 is an epigenetic reader protein responsible for recognition of modified histone lysine residues. Within the context of gene regulation, reader proteins act with writer and eraser proteins to govern chromatin remodeling and transcriptional programming. Recent work has suggested that GAS41 is an oncogenic driver of a major subtype of Non-Small Cell Lung Cancer, Lung Adenocarcinoma, and further that the YEATS domain in particular is indispensable for this oncogenic activity of GAS41. Such reports identify GAS41 as an attractive target for pharmacological inhibition, yet there are no published small-molecule inhibitors. Here we present the identification and optimization of small-molecule inhibitors targeting site-specific Histone H3 acetyl-lysine recognition by GAS41. We performed a fragment screen against GAS41 YEATS by protein-observed, solution-NMR spectroscopy, and identified a fragment hit, **6EE9**. **6EE9** shows modest, high-micromolar activity in blocking GAS41 YEATS domain-Histone H3 peptide H3K23crK27cr complex formation *in vitro*. Guided by a combination of biochemical assays, biophysical studies and structural biology we optimized small-molecule inhibitors **DLG-157** and **DLG-189** with low-micromolar activities, an over

hundred-fold improvement in potency over fragment hit. Subsequently, we developed dimeric inhibitors to bind two GAS41 YEATS domains simultaneously; this mimicked the bivalent mode of Histone H3 binding. Dimeric inhibitors **DLG-231** and **DLG-233** presented low-nanomolar activities in inhibiting GAS41 YEATS-H3K23crK27cr *in vitro*, and studies to demonstrate on-target activity and anti-proliferative effects in GAS41-dependent NSCLC cell lines are ongoing in our lab. Our efforts represent the first in-class inhibitors targeting GAS41 YEATS. This series of compounds may hold potential as chemical probes to characterize the roles of GAS41 in tumorigenesis, and further may aid in rational design of inhibitors as a possible means of therapeutic intervention against GAS41-dependent NSCLCs.

B. Introduction

B.1. GAS41 is a pharmacological target for Non-Small Cell Lung Cancers

In otherwise normal cell physiology, GAS41 functions as a histone reader protein via its YEATS domain. GAS41 YEATS recognizes site-specific acetylation at Histone H3 lysine residues K14, K18, K23, and K27[1,2]. *In cell* studies demonstrate that GAS41 associates with mono-acetylated H3K14ac and H3K27ac[1] and di-acetylated H3K23acK27ac[2] and subsequently recruits SRCAP chromatin-remodeling complexes, which govern H2A.Z deposition and further maintain poised chromatin for PRC2 complex recruitment[1].

An emerging body of literature has implicated GAS41 in oncogeneses of various subtypes of Non-Small Cell Lung Cancers (NSCLCs). GAS41 is overexpressed in three main NSCLC subtypes: lung andenocarcomina (LAC), lung squamous cell carcinoma (LSCC), and large cell carcinoma (LCC)[3]. In LAC cell lines, GAS41 overexpression was associated with Histone 2A.Z deposition[4], and shRNA-mediated knockdown of GAS41 attenuated cell proliferation[4].

The GAS41 YEATS domain is directly implicated in oncogenic transformation of GAS41-dependent NSCLCs. In LAC cell lines, following knockdown of GAS41 by shRNA, subsequent transfection of *wild-type* GAS41 was able to rescue the oncogenic phenotype associated with GAS41-dependent NSCLCs, as indicated by cell proliferation and tumor size[4]. However, GAS41-mutants Y74A and W93A were neither able to restore cell proliferation nor anchorage-independent cell growth[4]. With regards to mutant residues, several studies have revealed the mechanistic and structural bases of Histone H3 acetyl- and crotonyl-lysine recognition by GAS41 YEATS. Side chains of residues Y74 and W93 play a crucial role in GAS41 YEATS histone reader function, via formation of the acyl-lysine binding pocket[2]. Moreover, previously published *in cell* studies from our lab showed W93A mutation disrupted GAS41 binding to both di-acetylated Histone H3 peptides H3K23acK27ac and mono-acetylated Histone H3K27ac[2].

These findings beget the mechanism by which GAS41 promotes tumorigenesis in LAC. In one report, in H1993 cell lines, knockdown of GAS41 was shown to increase expression of p21 protein, notably of the *p53* tumor suppressor pathway that regulates cell cycle, DNA repair, and apoptosis[5]. In addition, in a different GAS41-dependent LAC cell line, H226, GAS41 knockdown was shown to increase levels of both p21 and p53[5]. In this same study, knockdown of GAS41 was shown to attenuate cell growth and induce senescence[5]. Several studies have implicated GAS41 in governing the *p53* pathway, via both phosphorylation of p53 and transcriptional activation of *p53* pathway genes *p21* and *p14^ARF*[6,7]. As the *p53* pathway is inactivated in approximately 50% of all lung cancers[3], targeting GAS41 may be viable in treatment of GAS41-dependent NSCLCs with mutations in the *p53* tumor suppressor pathway. Yet, further study to characterize the mechanism of on-target

activity of GAS41 small-molecule inhibitors would be imperative. Altogether, these studies offer evidence in favor of GAS41 as an oncogenic driver of one of the primary subtypes of NSCLCs, and posits a rationale in targeting GAS41 YEATS as a mean of therapeutic intervention against GAS41-dependent NSCLCs.

B.2. Blocking recognition of Histone H3 acyl-lysine by human YEATS domains with small molecules

There are four YEATS domain-containing proteins in humans (ENL, Af9, GAS41, YEATS2), classified by their respective N-terminal YEATS domains, which are responsible for recognition of acyl-lysine residues of histones[8]. The last half-decade has yielded numerous studies that identify roles for YEATS domains in various physiological processes, including regulation of histone lysine modifications, chromatin remodeling, and transcriptional regulation, among others[1,2,9,10,11,12]. In parallel with such studies, YEATS domains have also emerged as pharmacological targets, on the bases of reports implicating deregulation of YEATS domains' functions in various diseases[4,13,14,15]. Yet, small-molecule inhibitors targeting YEATS domains have largely remained elusive. Peptidomimetic, HTS, and fragment-based approaches have yielded inhibitors of ENL and AF9 YEATS domains (referred to herein as ENL YEATS and AF9 YEATS, respectively) with sub-micromolar activity *in vitro*, but poor selectivity for either ENL or AF9 YEATS and limited on-target activity in cells[16,17,18]. No GAS41 or YEATS2 YEATS inhibitors have been reported in the literature.

B.3. Fragment-based approach to develop inhibitors of Histone H3 acyl-lysine recognition by GAS41

GAS41 histone reader function comprises a challenging target for inhibitor development. Biophysical, structural and physiochemical characteristics of the acetyl-lysine recognition site[2] are not consistent with those of a "druggable pocket": a deep, well-defined binding site with polar surface area that additionally demonstrates tight binding affinity for its endogenous ligand[19]. Rather, GAS41 YEATS displays an open-ended, narrow, channel-like binding site for inhibitor development that demonstrates shape complementarity for acetyl- and crotonyl-lysine[2,4]. This histone reader pocket is structurally conserved with those of other human YEATS domains ENL, AF9, and YEATS2, characterized by its "aromatic cage" that is formed by the aromatic side chains of phenylalanine, tyrosine, and tryptophan[8]. Previous studies in our lab showed that GAS41 YEATS binds mono-acetylated and mono-crotonylated Histone H3 peptides with mid- to high-micromolar affinities (mono-acylated peptides with strongest affinities demonstrate $\sim$34 – 106 μM K_Ds), whereas di-acetylated and di-crotonylated Histone H3 peptides present improvement in affinities by at least an order of magnitude, in the low-micromolar range ($\sim$3 – 5 μM K_Ds)[2].

To develop small-molecule inhibitors of Histone H3 acyl-lysine recognition by GAS41 YEATS, we undertook a fragment-based approach. Here we present the identification and optimization of a novel class of small molecules targeting GAS41. We identified a ligand for GAS41 YEATS by NMR-based fragment screening that shows modest activity in blocking GAS41 histone reader function. Subsequently, we employed a structure-guided approach to lead optimization of inhibitors to low-nanomolar potency *in vitro*. Our efforts represent the first in-class inhibitors of GAS41.
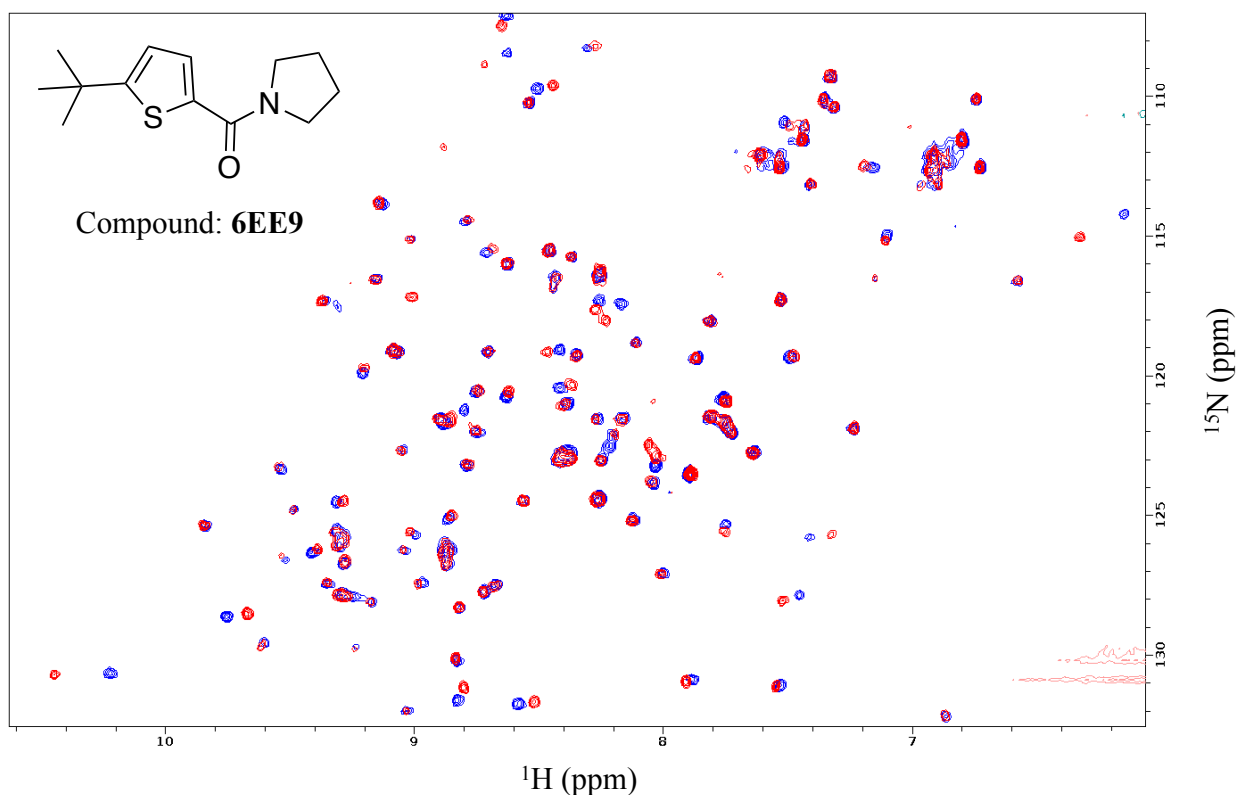
Figure 3.1. 6EE9 binds GAS41 YEATS domain. ^{1}H-^{15}N HSQC spectrum of 60 μM ^{15}N-GAS41$^{13-158}$ (blue), and spectrum in presence of 1 mM 6EE9 (red). Chemical structure of fragment hit **6EE9** shown in inset.

C. Results and Discussion

C.1. Identification of fragment hit 6EE9 by NMR-based screening

To identify compounds binding the GAS41 YEATS domain, we screened a library of approximately 600 fragment-like small molecules against ^{15}N-GAS41 YEATS domain (GAS41 residues 13 – 158) by NMR experiment ^{1}H-^{15}N HSQC. We screened fragments in mixtures of 10 compounds, at 500 μM per compound, and used chemical shift perturbations on ^{1}H-^{15}N HSQC spectra in presence of compound mixtures to identify binders. Subsequent

to the initial screening, we performed deconvolution experiments to partition individual fragment hits.

The screen yielded fragment hit **6EE9** (5-(tert-butyl)thiophene-2-yl)(pyrrolidin-1-yl)methanone), identified by global chemical shift perturbations of amide proton resonances on the ^{1}H-^{15}N HSQC spectrum of ^{15}N-GAS41 YEATS in the presence of 1 mM **6EE9** (**Figure 3.1.**). However, we were unable to map the binding site of **6EE9** by NMR experiment, as we have yet to assign the ^{1}H-^{15}N HSQC spectrum of GAS41 YEATS. Efforts to perform resonance assignment are ongoing in our lab.

A

GST-YEATS—FAM-di-cr-Histone H3 peptide

$K_D = 0.93 \pm 0.38$ μM

B
- 6EE9: $IC_{50} = 260 \pm 10$ μM
- H3K23acK27ac: $IC_{50} = 126 \pm 21$ μM
- H3K18acK27ac: $IC_{50} = 56 \pm 2$ μM
- H3K14acK27ac: $IC_{50} = 60 \pm 6$ μM
- H3K27ac: $IC_{50} = 260 \pm 20$ μM
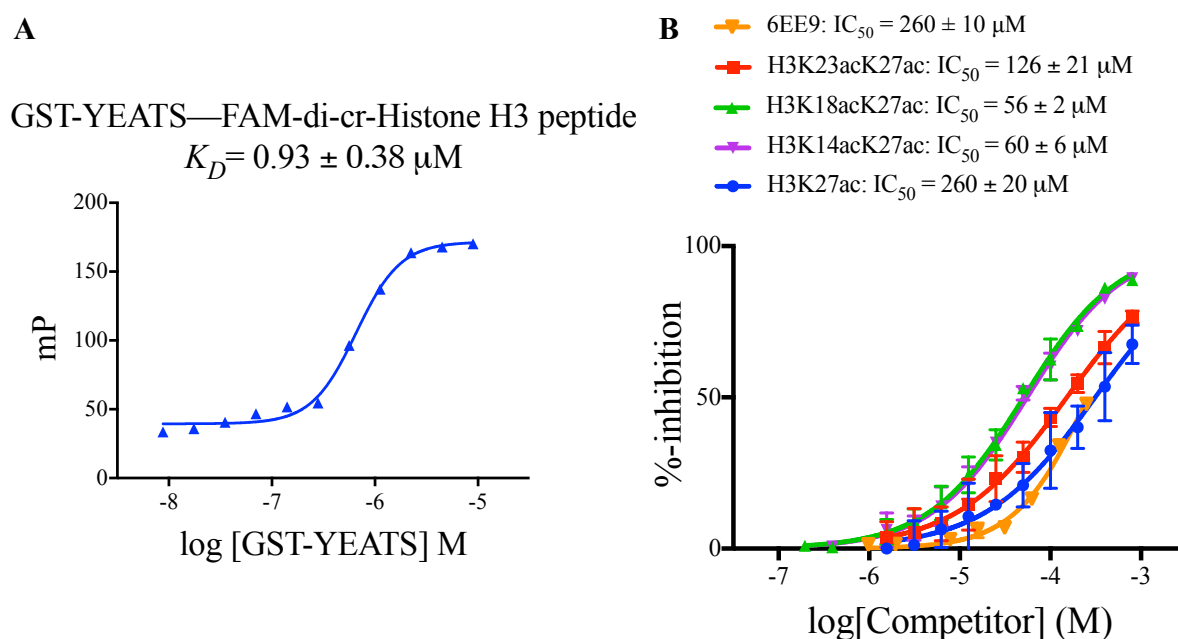
Figure 3.2. Development of FP assay to characterize IC_{50} value of fragment hit 6EE9.
A. Determination of affinity of FAM-H3K23crK27cr for GST-YEATS. **B.** IC_{50} values for **6EE9** and for Histone H3 mono-acetylated and di-acetylated peptides in blocking GAS41 YEATS binding to H3K23crK27cr probe. Mean values ± SD and curves reported from n ≥ 2 independent experiments. Data courtesy of Alyssa Winkler (Cierpicki Lab).

C.2. Development of FP and AlphaScreen Competition assays to characterize inhibitory activity of DLG-series of compounds

To characterize inhibitory activity of hit **6EE9**, we established an FP assay to quantify its IC_{50} value. We designed fluorphore-conjugated Histone H3 peptide with di-crotonylated lysine residues, FAM-H3K23crK27cr. To recapitulate dimeric GAS41 YEATS in a model system *in vitro*, we expressed and purified glutathione S-transferase (GST)-fused GAS41 YEATS domain (GST-GAS41$^{1-148}$). FAM-H3K23crK27cr probe binds GST-YEATS with sub-micromolar affinity ($K_D = 0.93 \ \mu$M) by FP assay (**Figure 3.2.A.**). This is consistent with previously published work from our lab that demonstrates H3K18acK27ac and H3K23acK27ac bind both GST-YEATS and full-length GAS41 with low-micromolar affinity, as assessed by ITC and BLI, respectively[2].

6EE9 presents high-micromolar inhibitory activity in blocking H3K23crK27cr, $IC_{50} = 260 \ \mu$M (**Figure 3.2.B.**). This activity is comparable to that of mono-acetylated Histone H3 peptide H3K27ac, and two- to four-fold weaker than that of di-acetylated Histone H3 peptides H3K14acK27ac, H3K18acK27ac, and H3K23acK27ac (IC_{50} values of 60 μM, 56 μM, and 126 μM, respectively [**Figure 3.2.B.**]). Modest potency in the high-micromolar range, which is witnessed among many fragment hits, does not exclude compounds from being suitable candidates for inhibitor optimization[20].
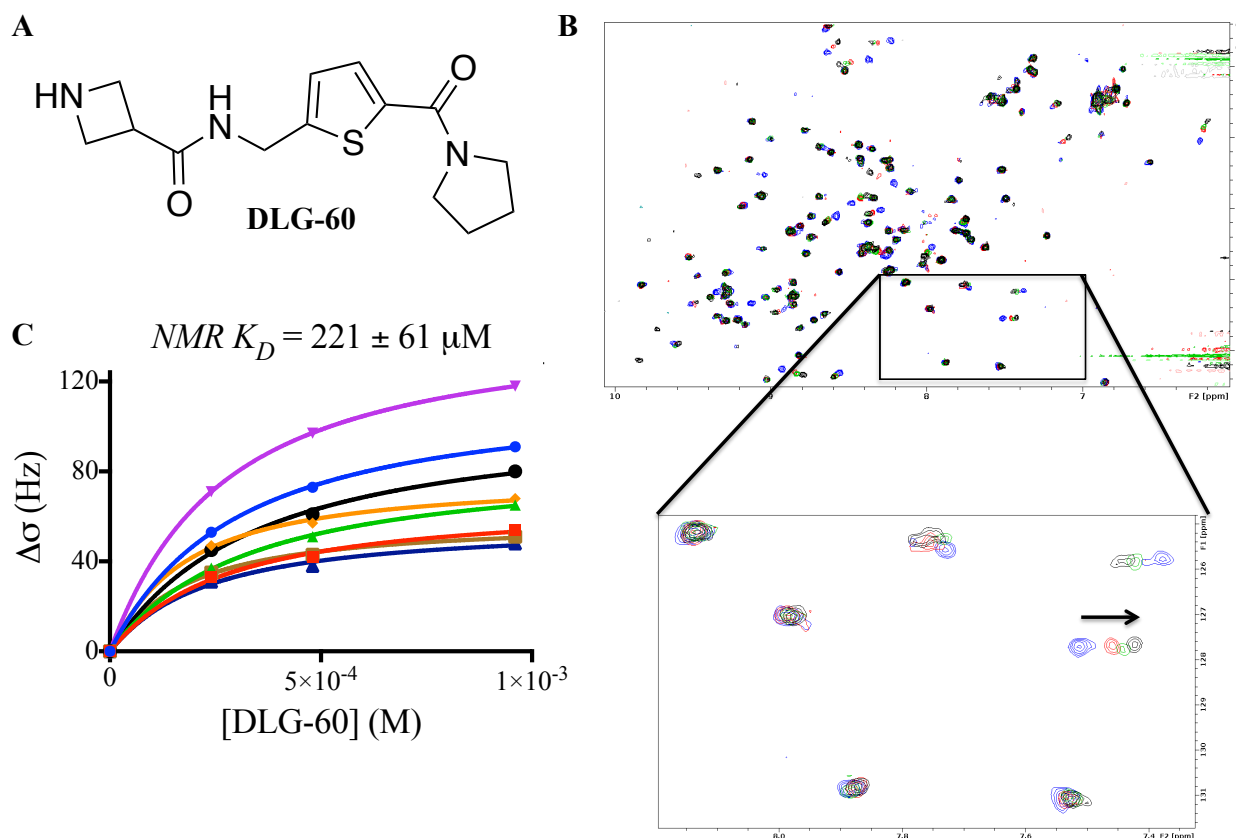
Figure 3.3. Biophysical characterization of DLG-60 binding GAS41 YEATS. A. Chemical structure of **DLG-60**. **B.** ^{1}H-^{15}N HSQC spectrum of 60 μM GAS41 YEATS (blue), and in presence of 250, 500, and 1000 μM **DLG-60**, colored red, green, and black, respectively. **C.** NMR-titration experiment to determine the K_D of **DLG-60** for GAS41 YEATS. K_D value $\pm$ SD averaged from eight amide proton resonances.

To characterize potencies of small-molecule inhibitors in an orthogonal assay, we developed AlphaScreen assay using hexa-histidine-tagged GAS41 YEATS domain (residues 13-158, referred to herein as his$_6$-YEATS) and biotinylated-H3K23crK27cr. Fragment hit

6EE9 presents high-μM IC$_{50}$, 340 μM, consistent with data obtained from FP Assay (**Table 3.1.**).

Compound	R$_1$	R$_2$	IC$_{50}$ (FP) (μM)	IC$_{50}$ (Alpha) (μM)
6EE9			260 ± 10	340 ± 80
6EE9-12			190*	N.d.
DLG-36			37%-inhibition @ 50 μM	N.d.
DLG-60			~500	290 ± 100
DLG-113			240 ± 40	55 ± 11

Table 3.1. SAR of commercially available and in-house synthesized 6EE9 analogues. Development of GAS41 inhibitors. IC$_{50}$ values determined by Fluorescent Polarization (FP) and AlphaScreen (Alpha) Competition Assays, as indicated. Mean values ± SD reported from $n \geq 2$ independent experiments (unless otherwise indicated). *One experiment. Data courtesy Alyssa Winkler (Cierpicki Lab).

C.3. Exploration and characterization of 6EE9 analogues

We first explored ~20 commercially available analogues of fragment hit **6EE9**, and identified compound **6EE9-12** with trifluoroethoxy-modified azetidine substituted for pyrrolidine, and with chloro-substitution for tert-butyl on thiophene (**Table 3.1.**). This compound showed a roughly ~1.5-fold improvement in potency by FP assay, $IC_{50} = 190\ \mu M$. To control for effects of trifluoroethoxy-modified azetidine, we synthesized an in-house analogue of **6EE9-12**, **DLG-36**, with tert-butyl-substituted thiophene (conserved with fragment hit **6EE9**), and this compound showed 37%-inhibition at 50 μM by FP. We were not able to perform full titration due to compound solubility, yet such percent-inhibition at 50 μM compound, by comparison, indicates an improvement in potency. We thus incorporated azetidine moiety substituted for pyrrolidine moving forward in compound design.

In parallel, we explored modifications on the thiophene moiety. We found that replacing tert-butyl with an amide group was well tolerated, as assessed by protein-NMR experiment (data not shown). Further, substitution of azetidine at the amide group improved binding. This yielded compound **DLG-60** (**Figure 3.3.A.**). NMR-titration experiments were performed to determine the affinity of **DLG-60** for GAS41 YEATS (**Figure 3.3.B.**). **DLG-60** binds GAS41 YEATS with a K_D of 221 μM, fit using a previously established model (**Figure 3.3.C.**) [21]. We then characterized **DLG-60** inhibition by AlphaScreen assay and found **DLG-60** $IC_{50} = 290\ \mu M$ (**Table 3.1.**). In total, our efforts to explore modifications at both thiophene and pyrrolidine modifications guided our design of more potent small molecules.
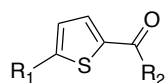
Compound (DLG-)	R$_1$	R$_2$	IC$_{50}$ (FP) (μM)	IC$_{50}$ (Alpha) (μM)
149			18 ± 4	5.2 ± 0.3
150			163 ± 16	43 ± 11
157			10.7 ± 2.1	3.0 ± 0.3
189			4.6 ± 1.4	2.2 ± 0.4

Table 3.2. SAR of pyrrolidine-based GAS41 inhibitors. Development of GAS41 inhibitors. IC$_{50}$ values determined by Fluorescent Polarization (FP) and AlphaScreen (Alpha) Competition Assays, as indicated. Mean values ± SD reported from $n \geq 2$ independent experiments. Data courtesy of Alyssa Winkler (Cierpicki Lab).

C.4. Molecular details of DLG-60 interaction with GAS41

To establish the binding mode of **DLG-60** to the GAS41 YEATS domain, we solved the crystal structure of the GAS41 YEATS domain (GAS41[1-148]) in complex with **DLG-60**, to 2.7 Å (**Table 3.6.**). The structure reveals a narrow, channel-like binding pocket comprised of H43, H71, S73, Y74, W93, G94, E95, and Y96, for which **DLG-60** presents shape complementarity (**Figure 3.4.A.**). The carbonyl group of **DLG-60** forms two hydrogen

bonds, one with each of the backbone amides of W93 and G94 (**Figure 3.4.B.**), which is

consistent with interactions made by H3K23acK27ac[2]. In addition, the thiophene of **DLG-60**

makes edge-to-face pi-pi stacking interactions with the side chains of W93, H71, and F96

(**Figure 3.4.B.**). Notably, **DLG-60** occupies the same binding site as acetyl-lysine of

H3K23acK27ac.

Figure 3.4. Crystal structure of DLG-60 in complex with GAS41 YEATS. A. *mFo-DFc*

map, contoured to *1σ* and colored black showing compound occupancy at binding site.

GAS41 YEATS represented as cartoon, colored gray. Selected residues shown as sticks,

salmon; compound, green. B. Molecular details of **DLG-60** interaction with GAS41.

Coloring scheme and representation same as in A. Hydrogen bonds indicated by dashed lines.

C.5. Development of pyrrolidine-based GAS41 inhibitors

Using the crystal structure of **DLG-60** in complex with GAS41 YEATS and SAR

derived from studies of both commercially available and in-house 6EE9 analogues, we

designed compounds with improved potency for GAS41 YEATS. Compound **DLG-113**, with one CH_2 removed from the linker to the amide group, yielded an approximate two-fold improvement in activity by FP over **DLG-60**, $IC_{50} = 240\ \mu M$ (**Table 3.1.**). By AlphaScreen, this improvement is over five-fold, $IC_{50} = 55\ \mu M$ (**Table 3.1.**). In addition, NMR-titration experiments were performed to determine the affinity of **DLG-113** for GAS41 YEATS, and we found $K_D = 121\ \mu M$ (**Figure 3.5.**).
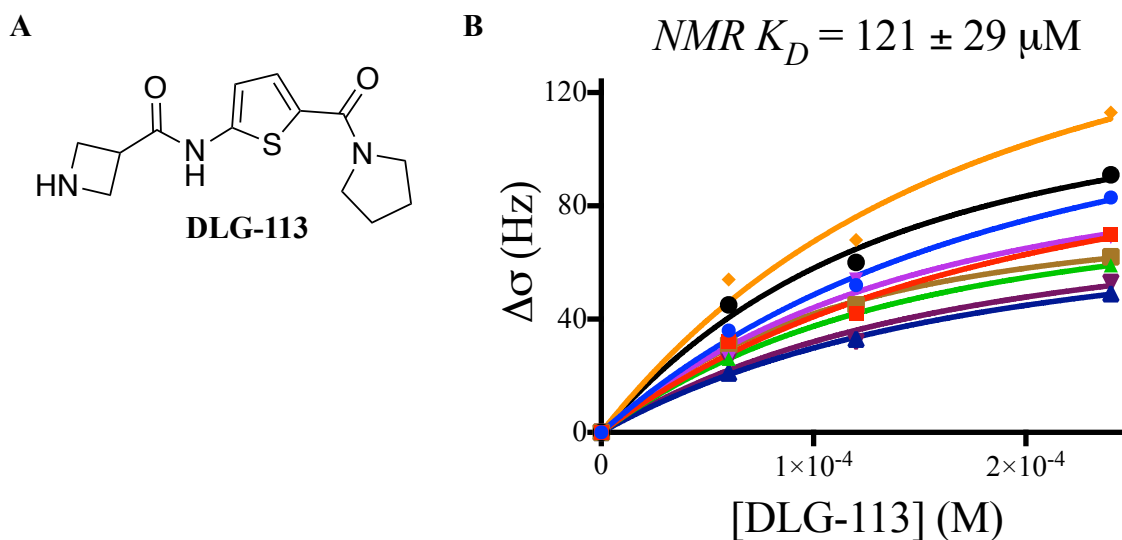
Figure 3.5. Determination of affinity of DLG-113 for GAS41 YEATS. A. Structure of **DLG-113**. B. NMR-titration experiment to extract K_D of **DLG-113** for GAS41 YEATS. K_D value ± SD averaged from ten amide proton resonances.

Further modification of the azetidine group on the amide, via substitution with pyrrolidine, conferred an over-10-fold improvement in activity: compound **DLG-149** presents an IC_{50} value of $18\ \mu M$ by FP Assay. Notably, the S-enantiomer, **DLG-150**, is approximately nine-fold weaker than R-enantiomer **DLG-149** (**DLG-150** $IC_{50} = 163\ \mu M$) (**Table 3.2.**). This

preference for the R-enantiomer was further validated in the AlphaScreen Competition Assay, as **DLG-149** is over eight-fold more potent than **DLG-150** in blocking complex formation, 5.2 μM versus 43 μM.

C.6. Development of analogues with substituted azetidine ring

Commercially available analogue **6EE9-12** showed that replacing pyrrolidine with azetidine was associated with an improvement in binding (**Table 3.1.**), and thus we explored substitutions on the azetidine moiety at this site. Azetidine substituted with thiazole, **DLG-157**, conferred nearly two-fold improvement from **DLG-149**, approaching low-micromolar inhibitory activity with IC$_{50}$ = 10.7 μM (**Table 3.2.**). Further modification with a benzothiazole substituent yielded an approximately 2.5-fold boost in activity, as **DLG-189** presented 4.6 μM IC$_{50}$ (**Table 3.2.**). In parallel with characterization by FP Competition Assay, **DLG-157** and **DLG-189**, demonstrated low-micromolar IC$_{50}$ values in AlphaScreen Competition Assay, 3 μM and 2.2 μM, respectively.

To determine the affinity of lead compounds **DLG-157** and **DLG-189** for GAS41 YEATS, ITC experiments were performed. **DLG-157** binds GAS41 YEATS with K_D = 8.2 μM, and **DLG-189**, with K_D = 5.4 μM (**Figure 3.6.**). By order of magnitude and by relative improvements in potencies, K_D values are consistent with IC$_{50}$ values from FP and AlphaScreen competition assays for **DLG-157** and **DLG-189**.

A DLG-157

C DLG-189

B

Time (min)

D

Time (min)

E
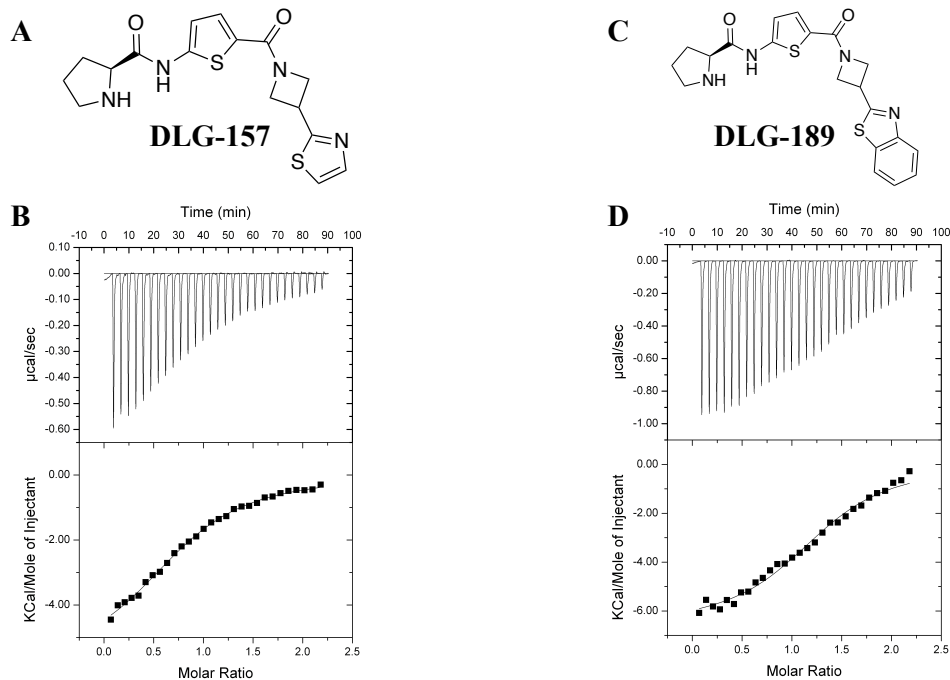
Compound	K_D (μM)	N	ΔS (cal/mol/K)	ΔH (cal/mol)
DLG-157	8.2 ± 3.2	1.1 ± 0.3	5.2 ± 2.5	-5420 ± 510
DLG-189	5.4 ± 0.1	1.2 ± 0.1	1.8 ± 0.4	-6646 ± 122

Figure 3.6. Determination of affinity of DLG-157 and of DLG-189 for GAS41 YEATS.
A. Structure of **DLG-157**. B. Isotherm showing binding of **DLG-157** to GAS41 YEATS.
Binding affinity and stoichiometry shown at inset. Values reported from $n \geq 2$ independent
experiments. C. Structure of **DLG-189**. D. Isotherm showing binding of **DLG-189** to
GAS41 YEATS. Binding affinity and stoichiometry shown at inset. Mean values ± SD
reported from $n \geq 2$ independent experiments. E. Binding affinity, stoichiometry, and
thermodynamic parameters. Mean values ± SD averaged from two independent experiments
are reported.

C.7. Structural studies reveal molecular basis of DLG-157 interaction with GAS41

To determine the molecular basis of **DLG-157** binding GAS41 YEATS, we co-crystallized GAS41 YEATS (GAS41 residues 1-148)—**DLG-157** complex and solved the high-resolution crystal structure of the complex at 2.10 Å (**Table 3.6.**). The structure reveals the binding mode of **DLG-157** to GAS41 YEATS, at the pocket comprised of H43, H71, S73, Y74, W93, G95, E95, and Y96 (**Figure 3.7.A.**) The carbonyl of **DLG-157** forms two hydrogen bonds, one with each of the backbone amide nitrogen atoms of W93 and G94 (**Figure 3.7.B.**). In addition, the thiophene moiety of **DLG-157** engages in edge-to-face pi-pi stacking interactions with side chains of W93, H71, and Y96.

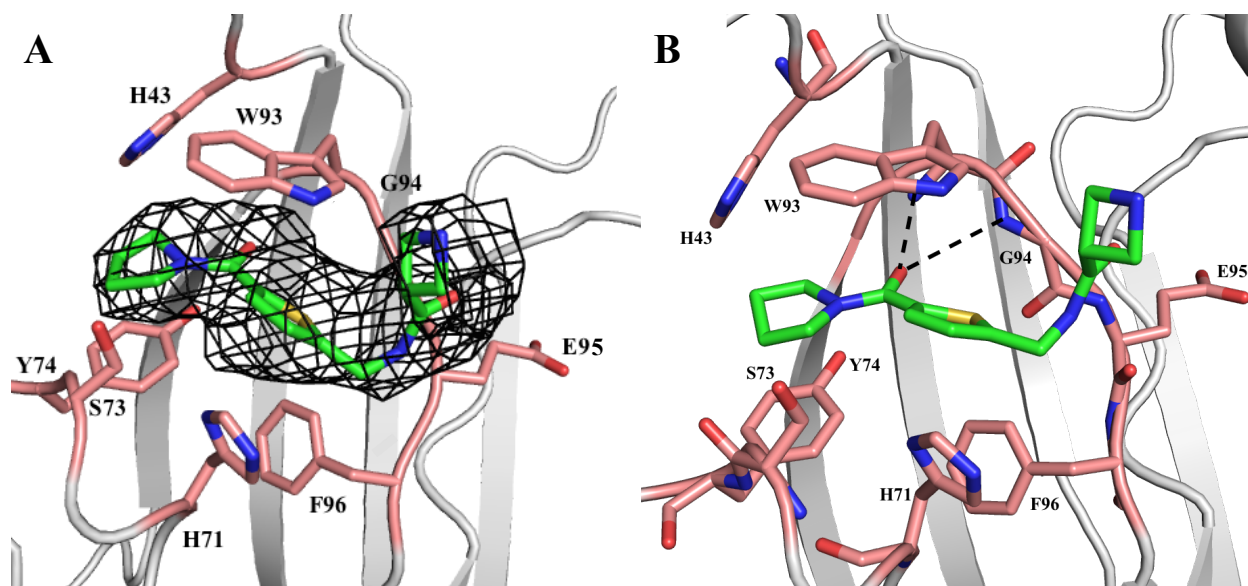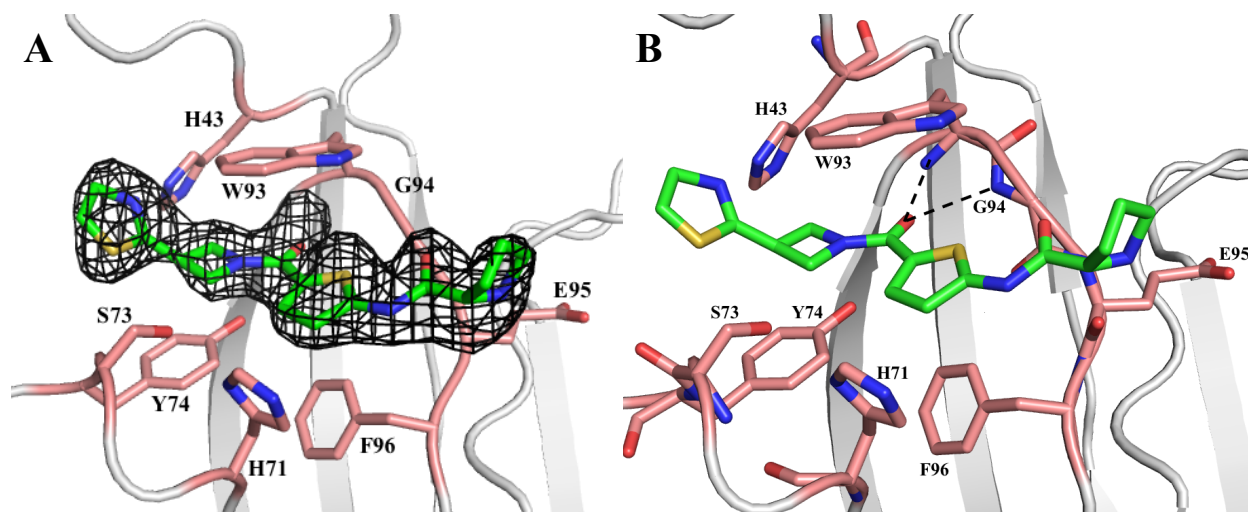**Figure 3.7. Crystal structure of DLG-157 in complex with GAS41 YEATS.** A. *mFo-DFc* map, contoured to *1σ* and colored black showing compound occupancy at binding site. GAS41 YEATS represented as cartoon, colored gray. Selected residues shown as sticks, salmon. Compound, green.. B. Molecular details of DLG-157 interaction with GAS41. Coloring scheme and representation same as in A. Hydrogen bonds indicated by dashed lines.

Many molecular details of the binding mode of **DLG-157** are conserved with **DLG-60**. That said, as **DLG-157** demonstrates an approximate 100-fold improvement in potency over **DLG-60**, the crystal structure of **DLG-157** in complex with GAS41 YEATS offers explanations for this compound's stronger activity. The structure of the **DLG-157** presents striking shape complementarity for the narrow, channel-like shape of the site of acetyl-lysine Histone H3 recognition by GAS41 YEATS (**Figure 3.7.B.**), and furthermore its thiazole-substituted azetidine moiety better occupies the acyl-lysine binding site than does the pyrrolidine moiety of **DLG-60**.
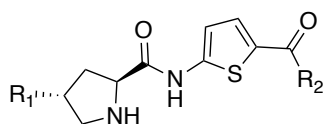
Compound (DLG-)	R_1	R_2	IC$_{50}$ (FP) (μM)	IC$_{50}$ (Alpha) (μM)
149	H		18 ± 4	5.2 ± 0.3
180			48 ± 0.5	28 ± 1.0

Table 3.3. SAR of monomeric GAS41 inhibitors with synthetic handles for dimerization.
IC$_{50}$ values determined by Fluorescent Polarization (FP) and AlphaScreen (Alpha) Competition Assays, as indicated. Mean values ± SD reported from $n \geq 2$ independent experiments. Data courtesy of Alyssa Winkler (Cierpicki Lab).

.
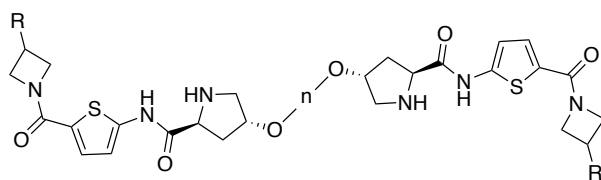
Compound (DLG-)	R	n ($-CH_2-$)	IC_{50} (FP) (nM)	IC_{50} (Alpha) (nM)
222		6	660 ± 240	61 ± 7
231		6	240 ± 40	4.5 ± 0.6
232		8	210 ± 2	10.1 ± 4.2
233		4	69 ± 27	10.6 ± 0.1

Table 3.4. SAR of dimeric GAS41 inhibitors. IC_{50} values determined by Fluorescent Polarization (FP) and AlphaScreen (Alpha) Competition Assays, as indicated. Mean values ± SD reported from $n \geq 2$ independent experiments. Data courtesy of Alyssa Winkler (Cierpicki Lab).

C.8. Development of dimeric inhibitors of GAS41

Our design incorporated dimerization of monomeric inhibitors (herein referred to as "dimeric inhibitors"), connected by a linker. To begin development of dimeric inhibitors, we synthetized analogues of **DLG-149** with ethoxy (**DLG-180**) modification substituted on the amide pyrrolidine. Ethoxy modifications served as synthetic handles, which would comprise linkers of dimeric inhibitors. Ethoxy substituted on pyrrolidine yielded a nearly three-fold

loss in activity for **DLG-180** as compared to **DLG-149** (48 μM to 18 μM) (**Table 3.3.**). By

AlphaScreen Competition Assay, as **DLG-180** presented an over five-fold loss in activity

compared to **DLG-149**, 28 μM to 5.2 μM (**Table 3.3.**). Nonetheless, despite the modest

reduction in compound potency associated with linker substitution, this did not discourage

development of dimeric inhibitors.

Figure legend (in plot):

- **DLG-157**: IC_{50} = 3,000 ± 300 nM
- **DLG-189**: IC_{50} = 2,200 ± 400 nM
- **DLG-222**: IC_{50} = 61 ± 7 nM
- **DLG-231**: IC_{50} = 4.5 ± 0.6 nM
- **DLG-232**: IC_{50} = 10 ± 4 nM
- **DLG-233**: IC_{50} = 10 ± 1 nM
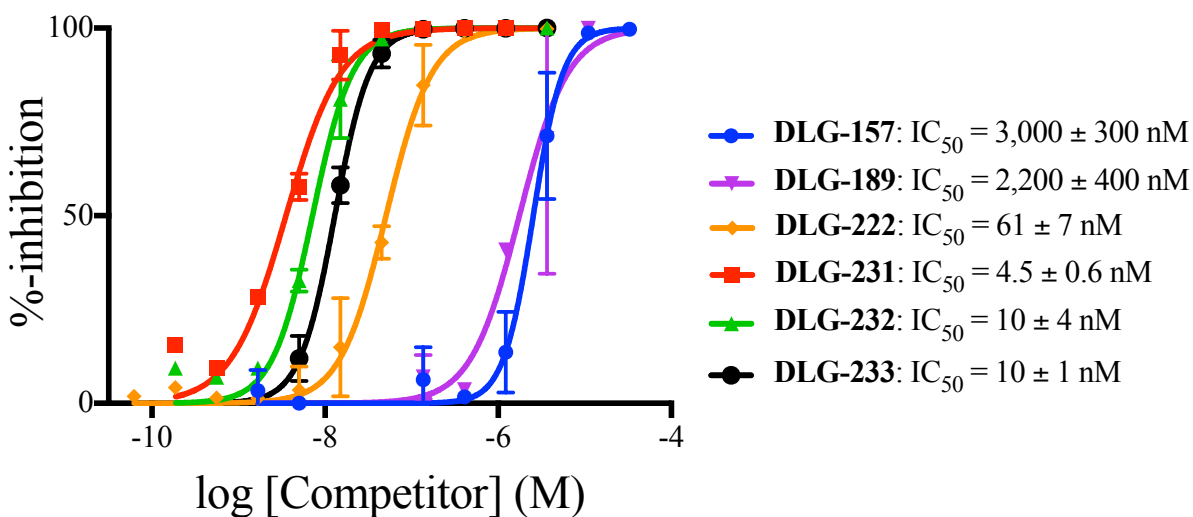
Figure 3.8. Dimeric inhibitors show ~300- to ~600-fold improvement over monomeric

inhibitors in IC$_{50}$ by AlphaScreen Competition Assay. Mean values ± SD reported from *n*

≥ 2 independent experiments. Data courtesy Alyssa Winkler (Cierpicki Lab).

On the bases of our most potent monomeric compounds, **DLG-157** and **DLG-189**, we

generated dimeric inhibitors **DLG-231** and **DLG-222**, respectively. **DLG-222** and **DLG-231**

demonstrate sub-micromolar IC_{50} values, 660 and 240 nM, respectively (**Table 3.4.**).

Notably, the linker composition (alkane) and linker length (*n* = 6, where *n* is CH_2) of both

compounds was identical. We developed **DLG-222** analogue **DLG-233** with saturated linker

(length *n* = 4), and **DLG-231** analogue **DLG-232** with saturated linker as well (length *n* = 8).

By FP assay, these compounds present several hundred-fold stronger IC_{50} values than monomeric analogues. **DLG-233** blocks complex formation with IC_{50} = 69 nM, and **DLG-232** with IC_{50} = 210 nM (**Table 3.4.**). However, limited FP assay sensitivity likely underestimates IC_{50} values.

Quantification of dimeric inhibitors' IC_{50} values by AlphaScreen Competition Assay revealed potent activities relative to those of unlabeled-acyl-lysine Histone H3 peptides. **DLG-222** (**DLG-189** analogue, $n = 6$) presents 61 nM IC_{50} (**Table 3.4.**). Interestingly, shortening linker length to $n = 4$ in **DLG-233** confers a nearly six-fold improvement in potency, IC_{50} = 10.7 nM. **DLG-231** (**DLG-157** analogue, $n = 6$) demonstrates 4.5 nM IC_{50}, yet increasing linker length to $n = 8$ yields an over two-fold loss in activity, IC_{50} = 10.1 nM. Relative to our strongest monomeric inhibitors **DLG-157** and **DLG-189**, dimeric inhibitors **DLG-231** and **DLG-233** represent approximately ~300- to ~600-fold improvements in potency (**Figure 3.8.**).

C.9. Dimeric inhibitors induce dimerization of GAS41 YEATS domain

To explain the improvement in compounds' inhibitory activities, and offer evidence for the bivalent mode of inhibitor binding, we validated that dimeric inhibitors induce dimerization of GAS41 YEATS domain. We developed *in vitro* biochemical assay by AlphaScreen to test dimerization of GAS41 YEATS domain. In brief, his$_6$-GAS41 YEATS (residues 13 – 158) and biotin-labeled GAS41 YEATS (1 – 148) were incubated, and titrated with DLG-231 (**Figure 3.9.**). Alpha Signal was plotted as a function of ligand concentration, and presented dose-dependence increases to mid-micromolar concentrations. The increase in signal is evidence of dimerization. At higher concentrations, approaching 100 μM, Alpha

78

Signal decreased, consistent with the Hook effect. This is likely attributed to saturation of individual YEATS domains with one molecule of dimeric inhibitors.
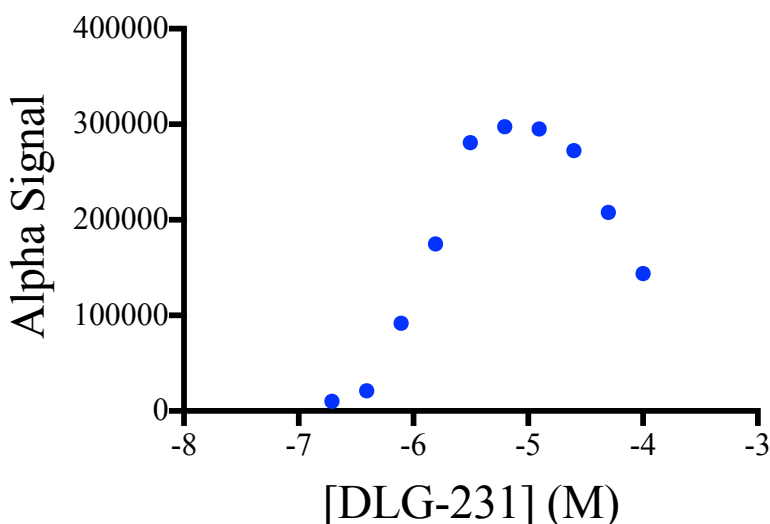
Figure 3.9. GAS41 YEATS domain *in vitro* dimerization assay by AlphaScreen. His-tagged GAS41 YEATS and biotinylated-GAS41 YEATS were incubated, and titrated with DLG-231. Alpha Signal plotted as function of DLG-231 concentration.

D. Conclusion

Here, we describe the design and development of the first in-class small-molecule inhibitors of GAS41. We identified hit compound **6EE9** that binds GAS41 YEATS, via protein-observed solution-NMR-based fragment screening of an in-house library of fragment-like small molecules. By *in vitro* biochemical assays FP and AlphaScreen, **6EE9** presents high-micromolar activity in inhibiting H3K23crK27cr recognition by GAS41 YEATS. Structure-guided design yielded lead monomeric inhibitors **DLG-157** and **DLG-189**. Both compounds presented low-micromolar IC_{50} values by FP and AlphaScreen Competition

Assays. In addition, we solved the crystal structure of **DLG-157** in complex with GAS41 YEATS to high-resolution, which revealed the molecular details of GAS41-inhibitor complex formation.

We subsequently developed lead dimeric inhibitors **DLG-231** and **DLG-233**. Dimeric inhibitor **DLG-231**, presents 4.5 nM IC_{50} in blocking GAS41 YEATS-H3K23crK27cr complex, and inhibitor **DLG-233**, 10.6 nM IC_{50}. Dimeric inhibitors represents an approximate ~300- to ~600-fold improvement in potency over lead monomeric inhibitors **DLG-157** and **DLG-189**. In addition, *in vitro* dimerization assay demonstrates that **DLG-231** and **DLG-233** compounds induce dimerization of GAS41 YEATS domain.

In total, our campaign has yielded a novel class of inhibitors targeting GAS41. As no small-molecule inhibitors of GAS41 have been reported as of this writing, our efforts provide insight into the structural basis of inhibition of site-specific Histone H3 acetyl- and crotonyl-lysine recognition by GAS41. With the marked increase in interest in biology of the YEATS family of histone reader domains in the past half-decade, as well as the emergence of GAS41 YEATS as a pharmacological target, blocking GAS41 YEATS- Histone H3 acyl-lysine complex may offer dual benefits, in basic biology and in the clinic: the former, via characterization of the mechanism of on-target activity of GAS41 YEATS small-molecule inhibitors that may help in elucidating the roles of GAS41 in both transcriptional regulation and in chromatin remodeling, among other physiological processes; the latter, via aiding in rational design of inhibitors targeting GAS41 in NSCLCs.

Acknowledgements

YEATS-Histone H3 acetyl-lysine protein-protein interaction. Additionally, I must acknowledge my co-authors on the *ACS Chem Bio* paper, Cho, *et al*., (adapted in Chapter 2), Hao Li, Hongzhi Miao, and EunGi Kim, for their work to perform structure-function studies and *in cell* studies of GAS41.

On the GAS41 small-molecule inhibitor development program, I must thank Dymytrii Listunov for both his efforts in synthetic medicinal chemistry on the series of compounds that bears his name, and his insight and discussions in the matter of inhibitor design. Also, I must thank Alyssa Winkler for her expertise in protein expression and purification and in chemical biology, via *in vitro* and *in cell* assay development. Hongzhi Miao, Trupta Purohit, Xiaoqing Zheng, and EunGi Kim have all contributed to the battery of functional assays in cells to quantify compounds' growth-inhibition and on-target activity, which has been exciting and rewarding to witness. Moreover, collaborator Venkat Keshamouni and his lab have made tremendous contributions, and I am looking forward to what the future holds for our labs.

E. Materials and Methods

Expression and Purification of Recombinant Proteins

^{15}N-GAS41 YEATS for NMR spectroscopy

For NMR experiments, the synthetic gene encoding *Human* GAS41 YEATS (residues 13 – 158) was ordered from Life Technologies and sub-cloned using BamHI and HindIII restriction sites into pQE-80L expression vector with an N-terminal hexa-histidine (His$_6$) tag. Recombinant plasmid pQE-80L-GAS41 YEATS (residues 13 – 158) was transformed into One Shot BL21(DE3) Chemically Competent *E. Coli* cells (Thermo Fisher Scientific). Transformants were grown in ^{15}N-labeled M9 minimal medium (Table 4.6) with 100 μg mL^{-1} Ampicillin at 37 $^{\circ}$C to OD$_{600}$ between 0.6 and 0.8. Protein expression was induced with 250

μM isopropyl 1-thio-β-D-galactopyranoside (IPTG), and cultures were incubated at 18 $^{\circ}$C for

16 h. Cells were harvested, and re-suspended in 50 mM Tris, pH 7.5, 300 mM NaCl, 30 mM

imidazole, 1 mM TCEP, and 500 μM PMSF for lysis by cell disrupter. Lysate was cleared at

14,500 rpm for 1 h at 4 $^{\circ}$C, and clarified lysate was loaded on Ni-NTA (Qiagen) affinity

column. The column was washed with Lysis Buffer, and eluted with Lysis Buffer containing

300 mM imidazole. Fractions were resolved by SDS-PAGE, and pure fractions were pooled

and dialyzed against 50 mM Tris, pH 7.5, and 500 mM NaCl at 4 $^{\circ}$C. Protein was

concentrated to ~30 μM, and then dialyzed against 50 mM Tris, pH 7.5, and 500 mM NaCl at

4 $^{\circ}$C overnight, followed by final dialysis against 50 mM Tris, pH 7.5, and 200 mM NaCl at 4

$^{\circ}$C for 4 h. Protein was then concentrated to 60 μM for NMR experiments.

GAS41 YEATS for AlphaScreen Competition Assay

For AlphaScreen Competition Assay experiments, GAS41 YEATS (residues 13 – 158)

was expressed and purified in the same manner as above, with the following exceptions:

protein was expressed in Luria Broth (LB) media; and, protein was dialyzed extensively

against 50 mM Tris, pH 7.5, and 500 mM NaCl at 4 $^{\circ}$C. Protein was then concentrated to ~25

μM for storage at -80 $^{\circ}$C.

Table 3.5. Recipe for 1x M9 Minimal Media.

Reagent	Final Concentration
Na_2HPO_4	4 g L^{-1}
KH_2PO_4	2 g L^{-1}
NaCl	0.5 g L^{-1}
Na_2SO_4	0.3 mM
Biotin	1 mg L^{-1}
Thiamin	1 mg L^{-1}
$MgSO_4$	1 mM
$CaCl_2$	1 mM
$(^{15}NH_4)_2SO_4$	1 g L^{-1}
Glucose	10 g L^{-1}
U-15N Bioexpress	5 mL L^{-1}

For 13C-, 15N- minimal media, substitute Glucose and Bioexpress with following:	
^{13}C-glucose	3 g L^{-1}
U-13C, U-15N Bioexpress	5 mL L^{-1}

GAS41 YEATS for ITC

For ITC experiments, codon-optimized GAS41 YEATS domain (residues 9 – 151) was synthesized by Life Technologies and sub-cloned into pet32a vector with N-terminal Thioredoxin (Trx)-His$_6$ tag and N-terminal PreScission cleavage site. Recombinant pet32a-GAS41 YEATS (residues 9 – 151) was transformed into One Shot BL21(DE3) *E. Coli* Chemically Competent cells (Thermo Fisher Scientific). Transformants were grown LB medium with 100 μg mL^{-1} Ampicillin at 37 ^{o}C to OD$_{600}$ between 0.6 and 0.8. Protein expression was induced with 250 μM IPTG, and cultures were incubated at 18 ^{o}C for 16 h. Cells were harvested, and re-suspended in 50 mM Tris, pH 7.5, 300 mM NaCl, 30 mM imidazole, 1 mM TCEP, and 500 μM PMSF for lysis by cell disrupter. Lysate was cleared at 14,500 rpm for 1 h at 4 ^{o}C, and clarified lysate was loaded on Ni-NTA (Qiagen) affinity column. The column was washed with Lysis Buffer, and eluted with Lysis Buffer containing 300 mM imidazole. Eluate was dialyzed against 50 mM Tris, pH 7.5, 300 mM NaCl, and 1 mM TCEP extensively at 4 ^{o}C, and proteolytically cleaved with PreScission enzyme. Following a final dialysis step against 50 mM Tris, pH 8.5, 50 mM NaCl, and 1 mM TCEP (referred to herein as Low Salt Buffer), sample was then loaded on Q-Sepharose (GE Healthcare Life Sciences) ion-exchange column. The column was washed with Low Salt Buffer, and eluted with 50 mM Tris, pH 8.5, 1 M NaCl, and 1 mM TCEP. Pure fractions were pooled, and dialyzed extensively against 50 mM Tris buffer, pH 7.5, and 150 mM NaCl at 4 ^{o}C for storage at -80 ^{o}C.

GST-GAS41 YEATS for FP Assay

For Fluorescence Polarization assay, Glutathione S-transferase (GST)-tagged GAS41 YEATS domain (residues 1 − 148, referred to herein as GST-YEATS) was cloned into pGST-parallel vector as previously described [2]. Transformants were grown LB medium with 100 μg mL^{-1} Ampicillin at 37 $^{\circ}$C to OD$_{600}$ between 0.6 and 0.8. Protein expression was induced with 250 μM IPTG, and cultures were incubated at 18 $^{\circ}$C for 16 h. Cells were harvested, and re-suspended in 50 mM Tris, pH 7.5, 300 mM NaCl, 1 mM TCEP, and 500 μM PMSF (herein referred to as Lysis Buffer) for lysis by cell disrupter. Lysate was cleared at 14,500 rpm for 1 h at 4 $^{\circ}$C, and clarified lysate was loaded on glutathione-Sepharose 4B (GE Healthcare Life Sciences) affinity column for 1 h at 4 $^{\circ}$C. The column was washed extensively with buffer containing 50 mM Tris, pH 7.5, 500 mM NaCl, and 1 mM TCEP, and eluted with buffer containing 50 mM Tris, pH 8.5, 500 mM NaCl, 10 mM reduced-glutathione, and 1 mM TCEP. Pure fractions were pooled and dialyzed extensively against 50 mM Tris, pH 7.5, 150 mM NaCl, and 1 mM TCEP at 4 $^{\circ}$C for storage at -80 $^{\circ}$C.

GAS41 YEATS for x-ray crystallography

For x-ray crystallography experiments, GST-GAS41 YEATS (residues 1 − 148) was expressed and purified in the same manner as above. Following affinity purification, protein was pooled and dialyzed against 50 mM Tris, pH 7.5, 500 mM NaCl, and 1 mM TCEP at 4 $^{\circ}$C, and proteolytically cleaved with Tobacco Etch Virus (TEV) protease enzyme. Sample was applied to glutathione-Sepharose 4B (GE Healthcare Life Sciences) column to extract GST-tag, and loaded on a Superdex 75 (GE Healthcare Life Sciences) size-exclusion column, pre-equilibrated in buffer containing 20 mM Tris, pH 7.5, and 300 mM NaCl). Protein was concentrated to ~350 μM for storage at -80 $^{\circ}$C.

Fragment screening by Protein-Observed Solution-NMR Spectroscopy

Fragment-screening samples were prepared with 60 μM ^{15}N-labeled GAS41 YEATS (residues 13 – 158) in buffer containing 50 mM Tris, pH 7.5, 200 mM NaCl, and 7% D$_2$O. The library of fragments used for screening was a combination of both commercially available compounds and compounds synthesized in-house. Fragments were screened in mixtures of ten compounds per sample at 500 μM final concentration per compound. Deconvolution experiments were performed with mixtures of three compounds per sample at 330 μM concentration per compound, and fragment hit confirmation was performed with 1000 μM compound. ^{1}H-^{15}N HSQC spectra were acquired at 30 °C on a 600 MHz Bruker Avance III spectrometer equipped with cryoprobe running Topspin version 2.1. Processing and spectral visualization were performed using NMRPipe and Sparky.

K_D Determination by Protein-Observed Solution-NMR Spectroscopy

^{1}H-^{15}N HSQC experiments were run at 30 °C with samples containing 60 μM ^{15}N-labeled GAS41 YEATS (residues 13 – 158) in buffer containing 50 mM Tris, pH 7.5, 200 mM NaCl, and 7% D$_2$O. For NMR-titration experiment with **DLG-60**, compound concentrations were 250, 500, and 1000 μM. For NMR-titration experiment with **DLG-113**, compound concentrations were 62.5, 125, and 250 μM. Dissociation constants were determined from least-squares fitting of chemical shift perturbations as a function of ligand concentration[21]:

Equation 3.1.

$$\delta_i = \frac{b - \sqrt{(b^2 - 4 \times a \times c)}}{2a}$$

with $a = (K_A/\delta_b) \times [P_t]$, $b = 1 + K_A([L_{ti}] + [P_t])$, and $c = \delta_b \times K_A \times [L_{ti}]$, where δ_i is the absolute change in chemical shift for each titration point, $[L_{ti}]$ is the total ligand concentration

at each titration point, $[P_t]$ is the total protein concentration, $K_A = 1/K_D$ is the binding

constant, and δ_b is the chemical shift of the resonance in question in the complex[21]. K_D and δ_b

were used as fitting parameters in analysis[21].

Fluorescence Polarization Assay

FP Assay to determine GST-YEATS—FAM-H3K23crK27cr affinity

5' 6-Fluorescein (FAM)-labeled di-crotonylated Histone H3 peptide probe

H3K23crK27cr (FAM-QLAT[K-cr]AAR[K-cr]SAPA-NH$_2$) was synthesized by Genscript.

Probe was dissolved in FP Storage Buffer (50 mM Tris, pH 7.5, 150 mM NaCl, and 1 mM

TCEP) at 1 mM concentration for storage at -20 $^\circ$C. To determine binding affinity of

H3K23crK27cr probe for GST-YEATS, probe was diluted to 2x 25 nM in FP Assay Buffer

(50 mM Tris, pH 7.5, 150 mM NaCl, 1 mM TCEP, 0.01% bovine serum albumin [BSA],

0.01% Tween-20) with 1% DMSO. GST-YEATS was prepared at 2x 10 μM in FP Assay

Buffer with 1% DMSO, and 10 two-fold serial dilutions of GST-YEATS in FP Assay Buffer

with 1% DMSO were prepared. Equal volumes of 2x H3K23crK27cr stock and 2x GST-

YEATS stock were incubated for 1 h at room temperature in dark. 15 μL per sample was

transferred to assay plate (Corning, product no. 4514) for reading. Using PHERAstar

microplate reader (BMG), fluorescence polarization was measured at 525 nm subsequent to

excitation at 495 nm. Data was fit in Prism 7.0 (GraphPad) program, with equation

log(agonist) vs. response – Variable slope (four parameters).

FP Competition Assay with FAM-H3K23crK27cr peptide

Compounds were first dissolved in DMSO, from which 100x (of starting concentration) mid-

stocks in DMSO were prepared. From 100x mid-stock, 10 two-fold dilutions of compound in

DMSO were performed. 1 μM GST-YEATS was prepared in FP Assay Buffer, and 100x

compound mid-stocks were added to 100-fold volumetric excess 1 μM GST-YEATS. Protein

was incubated with compound for 1 h at room temperature. In parallel with the first

incubation step, 100x 25 nM FAM-H3K23crK27cr was prepared in FP Assay Buffer. After 1

h incubation, 100x FAM-H3K23crK27cr was mixed with 100-fold volumetric excess sample.

Samples were prepared in 96-well assay plate (Corning, product no. 3356), and incubated for

1 h at room temperature in dark. Fluorescence polarization data was measured as described in

the aforementioned section. Data was fit in Prism 7.0 (GraphPad), and IC$_{50}$ values were

derived using the equation *log(inhibitor) vs. response – Variable slope (four parameters)*.

AlphaScreen Competition Assay

Compounds were first dissolved in DMSO, from which 100x (of starting

concentration) mid-stocks in DMSO were prepared. From 100x mid-stock, 10 three-fold

dilutions of compound in DMSO were performed.   100x compound mid-stocks were mixed

with 100-fold volumetric excess 1.25x 100 nM His$_6$-GAS41 YEATS (residues 13 – 158),

prepared in 50 mM HEPES, pH 7.5, 100 mM NaCl, 0.01% Tween-20, 0.05% BSA (referred

to herein as AlphaScreen Assay Buffer). Protein-compound solution was incubated for 1 h at

room temperature. 10x 25 nM biotinylated-, di-crotonylated-Histone H3 peptide biotin-

H3K23crK27cr was prepared in AlphaScreen Assay Buffer, and added to 10-fold excess

protein-compound solution. Biotinylated-peptide was incubated with protein-compound

solution for 1 h at room temperature. 20x Ni-Chelate Acceptor Bead solution (PerkinElmer)

was then prepared in AlphaScreen Assay Buffer, and added to 20-fold volumetric excess

sample. Sample was incubated for 1 h at room temperature in dark. Following acceptor bead

addition, 20x Streptavidin Donor Bead solution (PerkinElmer) was prepared in AlphaScreen

Assay Buffer, and added to 20-fold volumetric excess sample. Sample was incubated for 2 h

at room temperature in dark. AlphaScreen signal (Rubrene emission at 520 – 620 nM) was measured by PHERAstar microplate reader (BMG). Data was fit in Prism 7.0 (GraphPad), and IC$_{50}$ values were derived using the equation *log(inhibitor) vs. response – Variable slope (four parameters)*.

Isothermal Titration Calorimetry

GAS41 YEATS (residues 9 – 151) was dialyzed repeatedly against 50 mM phosphate buffer, pH 7.5, and 150 mM NaCl (referred to herein as ITC Buffer) at 4 $^{\circ}$C. Compounds **DLG-157** and **DLG-189** were dissolved in DMSO, and diluted to 400 μM in ITC Buffer with 1% DMSO. Protein and reference solutions were prepared with the addition of 1% DMSO. Titrations were performed using a VP-ITC titration calorimetric system (MicroCal) at 25 $^{\circ}$C. Calorimetric cell containing 40 μM GAS41 YEATS was titrated with compound (**DLG-157** or **DLG-189**) at 400 μM, in 10 μL injections at 180 s intervals. Data were analyzed by Origin 7.0 (OriginLab) to derive K_D and stoichiometry.

Crystallization of GAS41-inihibitor complexes

Soaking Experiment to Determine Crystal Structure of DLG-60 in complex with GAS41

Native-crystals of GAS41 YEATS (residues 1 – 148) were obtained using the sitting-drop technique with equal volumes 7.3 mg mL^{-1} protein in 20 mM Tris, pH 7.5, and 300 mM NaCl, and precipitant solution containing 0.1 M CHES, pH 9.5, and 1.26 M ammonium sulfate. Crystals were soaked in 2 mM **DLG-60** dissolved in reservoir for 15 minutes. Subsequently, the solution was exchanged and crystals were incubated in cryoprotectant solution, comprised of precipitant solution with 2 mM **DLG-60** and 30% glycerol, prior to freezing in liquid nitrogen.

Co-crystallization Experiment to Determine Crystal Structure of DLG-157 in complex with GAS41

Initial co-crystals of GAS41 YEATS (residues 1 – 148) in complex with **DLG-157** were obtained through screening using the hanging-drop vapor diffusion technique over one week at 4 °C. GAS41 YEATS at 7.3 mg mL^{-1} (400 μM) was incubated with 450 μM **DLG-157** in 20 mM Tris, pH 7.5, and 300 mM NaCl for six hours at room temperature prior to screening. To optimize crystals, protein-inhibitor complex assembly was repeated as described above. Subsequently, equal volumes of protein-inhibtor complex solution were mixed with precipitant solution containing 0.1 M Tris, pH 7.3, 0.2 M MgCl$_2$, and 2.2 M NaCl. Crystals grew over one week at 4 °C. Prior to data collection, crystals were exchanged into cryoprotectant solution, comprised of precipitant solution with 750 μM **DLG-157**, 1% DMSO, and 25% ethylene glycol, followed by freezing in liquid nitrogen.

Crystallographic Data Collection and Structure Determination

Diffraction data for GAS41 YEATS-inhibitor complexes were collected at the 21-ID-F and 21-ID-G beam lines at the Life Sciences Collaborative Access Team at the Advanced Photon Source. Data were integrated and scaled using HKL-2000 [22], and structures were solved by molecular replacement with MOLREP using known native-GAS41 YEATS structure as a search model. Refinement of structures was performed using REFMAC [23], COOT [24], CCP4 program suite [25], and PHENIX program suite [26]. Structures were validated using MOLPROBITY [27] and ADIT [28] servers. Data collection and structure refinement statistics are presented in Table 4.4.

Table 3.6. Data collection and refinement statistics for GAS41-inhibitor complexes

	GAS41^{YEATS}-DLG-60	GAS41^{YEATS}-DLG-157
PDB Code	N/A	N/A
Space group	P2$_1$2$_1$2$_1$	C121
Cell dimensions *a, b, c* [4]	70.5, 80.6 121.6	117.3, 112.5, 63.2
Cell dimensions α, β, γ (deg)	90, 90, 90	90, 116.7, 90
Resolution [4]	2.71 (2.76 - 2.71)	2.10 (2.17 - 2.10)
Unique reflections	19305 (937)	42637 (4184)
R_{merge}	0.217 (1.38)	0.091 (0.48)
$I/\sigma I$	20.5 (2.0)	11.7 (2.5)
Completeness (%)	99.7 (99.0)	98.1 (98.2)
Redundancy	12.4 (10.4)	3.9 (3.8)
	Refinement	
R_{work}/R_{free} (%)	21.5/25.3	20.6/23.0
No. atoms		
Protein	3957	4222
Water	0	640
Mean B-factors (Å^2)	60.93	36.65
RMS deviations		
Bond lengths [4]	0.005	0.013
Bond angles (deg)	0.79	1.48
	Ramachandran plot	
Most favored regions (%)	95.71	98.79
Additional allowed regions (%)	3.86	1.01

F. References

1   Hsu, C. C. *et al.* Gas41 links histone acetylation to H2A.Z deposition and maintenance of embryonic stem cell identity. *Cell Discov* **4**, 28, doi:10.1038/s41421-018-0027-0 (2018).

2   Cho, H. J. *et al.* GAS41 Recognizes Diacetylated Histone H3 through a Bivalent Binding Mode. *ACS Chem Biol* **13**, 2739-2746, doi:10.1021/acschembio.8b00674 (2018).

3   Herbst, R. S., Heymach, J. V. & Lippman, S. M. Lung cancer. *N Engl J Med* **359**, 1367-1380, doi:10.1056/NEJMra0802714 (2008).

4   Hsu, C. C. *et al.* Recognition of histone acetylation by the GAS41 YEATS domain promotes H2A.Z deposition in non-small cell lung cancer. *Genes Dev* **32**, 58-69, doi:10.1101/gad.303784.117 (2018).

5   Pikor, L. A. *et al.* YEATS4 is a novel oncogene amplified in non-small cell lung cancer that regulates the p53 pathway. *Cancer Res* **73**, 7301-7312, doi:10.1158/0008-5472.CAN-13-1897 (2013).

6   Park, J. H. & Roeder, R. G. GAS41 is required for repression of the p53 tumor suppressor pathway during normal cellular proliferation. *Mol Cell Biol* **26**, 4006-4016, doi:10.1128/MCB.02185-05 (2006).

7   Park, J. H., Smith, R. J., Shieh, S. Y. & Roeder, R. G. The GAS41-PP2Cbeta complex dephosphorylates p53 at serine 366 and regulates its stability. *J Biol Chem* **286**, 10911-10917, doi:10.1074/jbc.C110.210211 (2011).

8   Zhao, D., Li, Y., Xiong, X., Chen, Z. & Li, H. YEATS Domain-A Histone Acylation Reader in Health and Disease. *J Mol Biol* **429**, 1994-2002, doi:10.1016/j.jmb.2017.03.010 (2017).

9   Zhang, Q. *et al.* Structural Insights into Histone Crotonyl-Lysine Recognition by the AF9 YEATS Domain. *Structure* **24**, 1606-1612, doi:10.1016/j.str.2016.05.023 (2016).

10   Wang, Y. *et al.* Identification of the YEATS domain of GAS41 as a pH-dependent reader of histone succinylation. *Proc Natl Acad Sci U S A* **115**, 2365-2370, doi:10.1073/pnas.1717664115 (2018).

11   Zhao, D. *et al.* YEATS2 is a selective histone crotonylation reader. *Cell Res* **26**, 629-632, doi:10.1038/cr.2016.49 (2016).

12   Li, Y. *et al.* AF9 YEATS domain links histone acetylation to DOT1L-mediated H3K79 methylation. *Cell* **159**, 558-571, doi:10.1016/j.cell.2014.09.049 (2014).

13   Erb, M. A. *et al.* Transcription control by the ENL YEATS domain in acute leukaemia. *Nature* **543**, 270-274, doi:10.1038/nature21688 (2017).

14   Wan, L. *et al.* ENL links histone acetylation to oncogenic gene expression in acute myeloid leukaemia. *Nature* **543**, 265-269, doi:10.1038/nature21687 (2017).

15   Mi, W. *et al.* YEATS2 links histone acetylation to tumorigenesis of non-small cell lung cancer. *Nat Commun* **8**, 1088, doi:10.1038/s41467-017-01173-4 (2017).

16   Heidenreich, D. *et al.* Structure-Based Approach toward Identification of Inhibitory Fragments for Eleven-Nineteen-Leukemia Protein (ENL). *J Med Chem* **61**, 10929-10934, doi:10.1021/acs.jmedchem.8b01457 (2018).

17   Moustakim, M. *et al.* Discovery of an MLLT1/3 YEATS Domain Chemical Probe. *Angew Chem Int Ed Engl* **57**, 16302-16307, doi:10.1002/anie.201810617 (2018).

18   Li, X. *et al.* Structure-guided development of YEATS domain inhibitors by targeting pi-pi-pi stacking. *Nat Chem Biol* **14**, 1140-1149, doi:10.1038/s41589-018-0144-y (2018).

19   Cheng, A. C. *et al.* Structure-based maximal affinity model predicts small-molecule druggability. *Nat Biotechnol* **25**, 71-75, doi:10.1038/nbt1273 (2007).

20   Erlanson, D. A., Fesik, S. W., Hubbard, R. E., Jahnke, W. & Jhoti, H. Twenty years on: the impact of fragments on drug discovery. *Nat Rev Drug Discov* **15**, 605-619, doi:10.1038/nrd.2016.109 (2016).

21   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).

22   Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol* **276**, 307-326 (1997).

23   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, doi:10.1107/S0907444996012255 (1997).

24   Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* **60**, 2126-2132, doi:10.1107/S0907444904019158 (2004).

25   Collaborative Computational Project, N. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr D Biol Crystallogr* **50**, 760-763, doi:10.1107/S0907444994003112 (1994).

26   Adams, P. D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* **66**, 213-221, doi:10.1107/S0907444909052925 (2010).

27   Davis, I. W. *et al.* MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res* **35**, W375-383, doi:10.1093/nar/gkm216 (2007).

28   Yang, H. *et al.* Automated and accurate deposition of structures solved by X-ray diffraction to the Protein Data Bank. *Acta Crystallogr D Biol Crystallogr* **60**, 1833-1839, doi:10.1107/S0907444904019419 (2004).

Chapter 4

Identification of Thiourea-Based Inhibitors of the B-Cell Lymphoma 6 BTB Domain via NMR-Based Fragment Screening

A. Abstract

Protein-protein interactions (PPI) between the transcriptional repressor B-cell lymphoma 6 (BCL6) BTB domain (BCL6^{BTB}) and its co-repressors have emerged as a promising target for anti-cancer therapeutics. However, identification of potent, drug-like inhibitors of BCL6^{BTB} has remained challenging. Using NMR-based screening of a library of fragment-like small molecules, we have identified a thiourea compound (7CC5) that binds to

BCL6[BTB]. From this hit, the application protein-NMR spectroscopy, x-ray crystallography, and biochemical and biophysical studies, guided an investment in synthetic medicinal chemistry, which yielded an inhibitor **15f** that demonstrated over 100-fold improved potency for the BCL6[BTB]. This gain in potency was achieved by a unique binding mode that mimics the binding mode of the corepressor SMRT into the aromatic and the HDCH sites. The structure-activity relationship based on these new inhibitors will have a significant impact on the rational design of novel BCL6 inhibitors, facilitating the identification of therapeutics for the treatment of BCL6-dependent tumors.

B. Introduction

B-Cell lymphoma 6 (BCL6) is an oncoprotein associated with multiple cancers.[1] As the master regulator of the germinal center (GC) reaction, BCL6 is constitutively expressed in diffuse large B-cell lymphomas (DLBCL).[2,3] It is one of the most upregulated genes in BCR-ABL-positive B-cell precursor acute lymphoblastic leukemia (B-ALL) cells after exposure to BCR-ABL tyrosine kinase inhibitors (TKIs).[4] BCL6 is also upregulated by mixed lineage leukemia (MLL) fusion oncoproteins in B-ALLs with MLL translocation.[5] Moreover, patients with blast-phase chronic myeloid leukemia (BP-CML) cannot be cured by BCR-ABL TKIs because CML stem cells are not fully dependent on BCR-ABL.[6] However, CML stem cells express high levels of BCL6, and targeting BCL6 eliminated these cells.[6] Furthermore, the BCL6 locus is often amplified in breast cancers, and BCL6 maintains the survival of aggressive triple-negative breast cancer (TNBC) cells.[7] Therefore BCL6 has broad oncogenic roles in many cancer subtypes.

BCL6 mediates tumorigenesis through its repression of DNA-damage-sensing genes, such as *ATR*, and cell-death and cell-cycle-checkpoint genes, such as *CHEK1, TP53,*

CDKN1A, CDKN2A, P14ARF, CDKN1B, and *EP300*.[8-14] Blocking BCL6-repressor activity induces expression of these genes, resulting in rapid cell death.[10,15-18] Although different types of cancers are heterogeneous, a majority of them are BCL6-dependent.[19]

BCL6 has an N-terminal broad-complex, tramtrack, and brick-a-brac (BTB) domain (BCL6[BTB]) that mediates transcriptional repression, a C-terminal C_2H_2 zinc finger responsible for DNA binding and inflammatory reactions,[20] and a linker region containing a second repression domain.[21] BCL6[BTB] forms an obligate homodimer. The interface of the dimer forms a lateral groove (LG), which is employed in the recruitment of the corepressors silencing mediator of retinoid- and thyroid-hormone receptor (SMRT), nuclear-receptor co-repressor (NCoR), and BCL6 co-repressor (BCoR).[22,23] This unique BCL6[BTB]-LG-corepressor interaction is essential for the repression activity of BCL6.[24,25]

BCL6 knockout is lethal and most knockout mice die within 4-6 weeks of a severe inflammatory syndrome.[26] This phenotype raises concern regarding potential side effects of BCL6 inhibitors. However, exposure to BCL6[BTB]-LG peptide inhibitors[16] did not induce inflammation or other toxic effects.[16] Moreover, mice expressing BCL6 with point mutations that abrogate corepressor binding lived healthy lives without inflammatory syndrome.[27] Taken together, BCL6[BTB]-LG function is only essential in GC B-cells and in tumors, but it is dispensable for the other functions of BCL6. Indeed the effect of BCL6 in macrophages appears to be linked to its zinc fingers competing for DNA binding with STAT proteins.[27] Because BCL6[BTB]-LG-targeting inhibitors do not affect BCL6 DNA binding, its function in macrophages is preserved. Suppression of GC B-cells, especially in a transient manner, would likewise be non-toxic in humans. Therefore, BCL6[BTB]-LG inhibitors have favorable safety profiles and will likely be better tolerated than currently approved anticancer therapies.

Small molecule inhibitors of BCL6[BTB] are known (**Figure 4.1.**). For instance, Takeda developed inhibitors using fragment-based screening, which disrupted BCL6 interaction with a BCoR peptide in a micromolar range.[28,29] Astra Zeneca developed macrocycles that could bind the BCL6[BTB]-LG with nanomolar affinity but had no effect on DLBCL proliferation.[30] Note that no target-engagement studies were performed in either study, so it is not known whether these compounds could actually block BCL6-mediated formation of repression complexes or for how long this might occur. Recently, Boehringer developed two small molecules that bind to BCL6[BTB]-LG with nanomolar activity.[31] One of the compounds resulted in degradation of BCL6 in a DNA-binding dependent manner, while the non-degrading compounds had weak activity against BCL6 in cells.[31] Our laboratories reported the discovery of inhibitors 79-6 and FX1.[32] FX1 binds to the BCL6[BTB]-LG with improved potency ($K_D = 7 \ \mu M$) compared with 79-6 ($K_D = 129 \ \mu M$) and even with SMRT ($K_D = 30 \ \mu M$) in a microscale thermophoresis (MST) assay.[32] It maintains high specificity for BCL6[BTB] over other BTB-containing proteins,[32] and shows good specificity towards BCL6-dependent lymphoma cell lines over BCL6-independent cells.[32] Results from *in vivo* studies show that FX1 is nontoxic at high concentrations and demonstrated excellent efficacy in both GCB- and ABC-type DLBCL xenograft models.[32] Compounds 79-6 and FX1 exhibit promising biological profiles, yet these inhibitors also have potential limitations. First, they only bind to the top aromatic pocket and leave the rest of the LG unoccupied. We speculate that new inhibitors with extended structures that occupy a larger portion of the LG will have improved inhibitory activities. Moreover, the rhodanine-based scaffold of inhibitors 79-6 and FX1 can potentially be problematic for further development of as clinical candidates.[33,34]
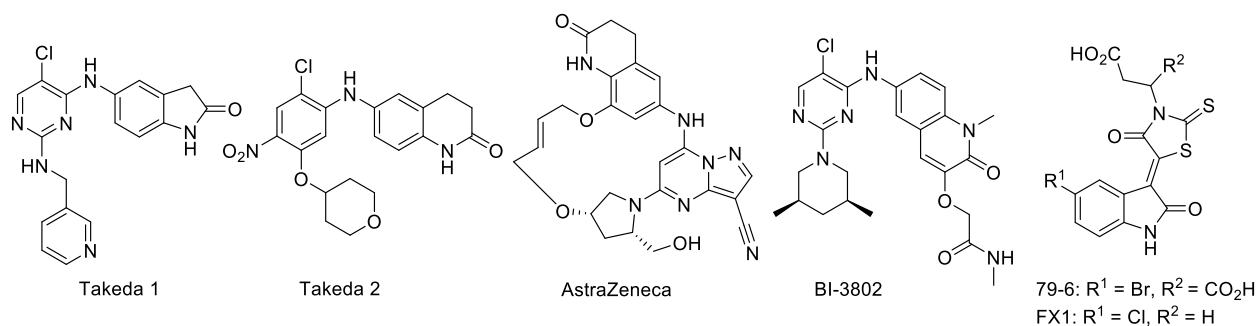
Figure 4.1. Chemical structures of previously reported inhibitors for the BCL6[BTB].

Herein we report the identification and optimization of a novel series of thiourea-based inhibitors of BCL6[BTB]. The scaffold was first identified by screening a fragment-like small molecule library, using protein-observed NMR spectroscopy, and subsequently the binding mode of the fragment hit in complex with BCL6[BTB] was determined using x-ray crystallography. We employed structure-guided design and protein-NMR-based methodology to lead development of a new class of inhibitors targeting BCL6-co-repressor complex formation. This study presents the design strategy, chemical synthesis, and experimental characterization of this series of compounds.

C. Results and Discussion

C.1. Identification and characterization of hit 7CC5

To identify small molecules binding to BCL6[BTB], we began by screening an in-house library of approximately 1500 chemically diverse, fragment-like small molecules by protein-NMR spectroscopy. Recombinant, isotopically labeled BCL6[BTB] was used for ^{1}H-^{15}N HSQC experiments. Identification of binders was based on ^{1}H- and ^{15}N-chemical shift perturbations on the ^{1}H-^{15}N HSQC NMR spectra. Screening was initially performed with 10 compounds per sample at 500 μM final concentrations of each compound, followed by deconvolution experiments on individual compounds to identify hits. Through this screen, we discovered

97

compound 7CC5 (1-phenethyl-3-(pyridin-3-yl) thiourea) (**Figure 4.2.A.**). To determine the affinity of this compound, we performed NMR-titration experiments and estimated the K_D of 7CC5 to be 3.2 mM (**Figure 4.2.B.**). Despite their relatively weak binding affinities, fragment-like compounds with mM affinities are usually suitable candidates for further optimization.[35]
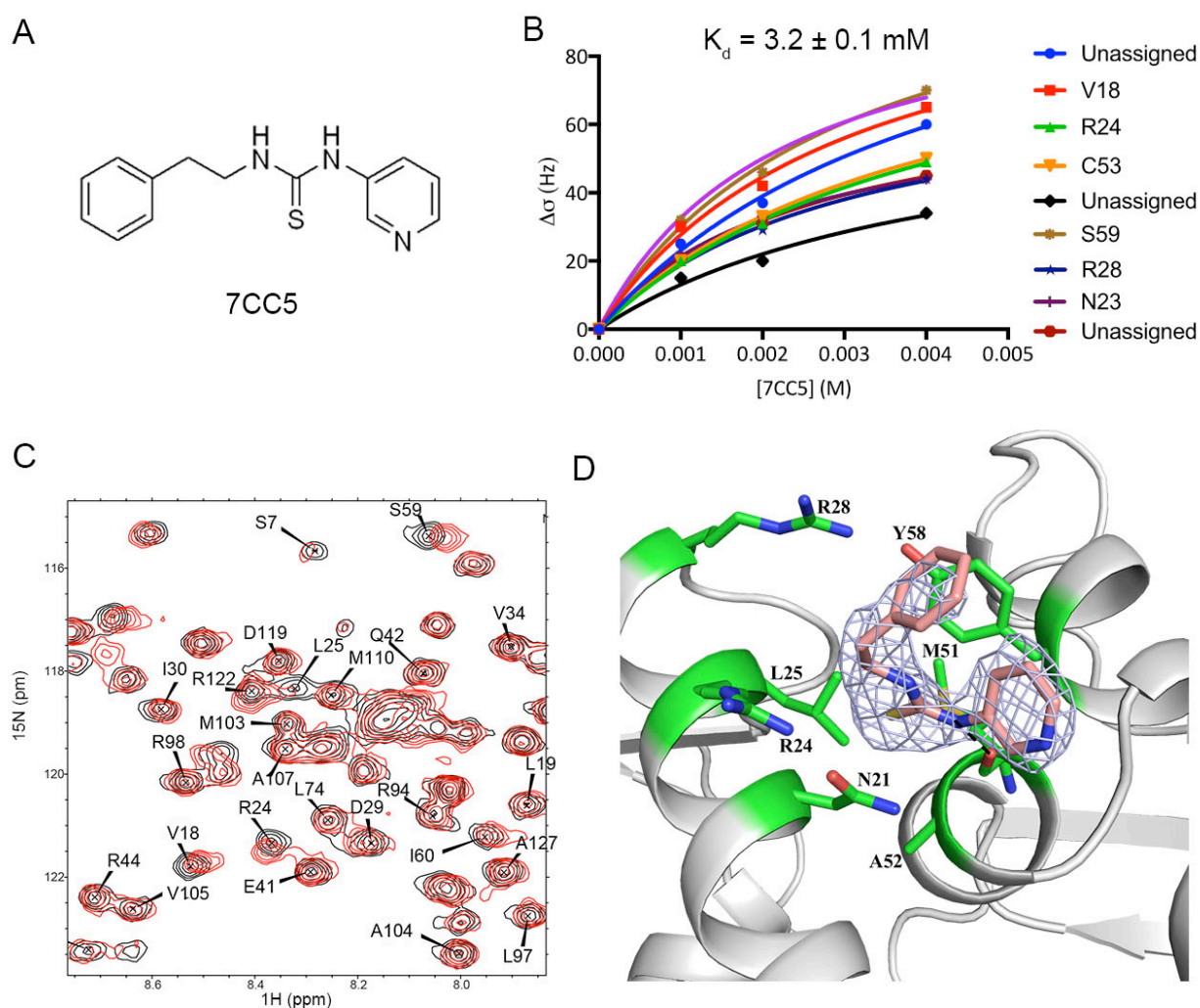
Figure 4.2. Identification and characterization of fragment hit 7CC5. A. Chemical structure of hit 7CC5. B. NMR titration experiment to determine K_D value for binding of 7CC5 and BCL6^{BTB}. C. Superposition of the ^{1}H-^{15}N HSQC spectra of 150 μM BCL6^{BTB} with 5% DMSO (black) and 500 μM 7CC5 (red). D. Crystal structure of BCL6^{BTB}-7CC5 complex

(PDB 6C3N). The *mFo-DFc* electron density map contoured at 3.0 σ is shown for 7CC5 and selected protein residues involved in the interactions with 7CC5 are shown as sticks.

To map the binding site of hit 7CC5 on BCL6[BTB], we analyzed chemical shift perturbations, and found that binding of 7CC5 results in several chemical shift perturbations for residues in the BCL6[BTB]-LG (**Figure 4.2.C.**). These chemical shift perturbations are consistent with those observed for FX1, suggesting the same binding site[1]. To accurately establish the binding mode of compound 7CC5, we determined a crystal structure of compound 7CC5 in complex with BCL6[BTB] at 2.5 Å resolution (**Figure 4.2.D.**). The structure reveals that compound 7CC5 binds in a well-defined pocket in the BCL6[BTB]-LG. The sulfur atom of the thiourea group occupies a hydrophobic site comprised of Asn21, Leu25, Met51, Ala52, and Tyr58. One of the thioamide hydrogens forms a hydrogen bond (2.59 Å) with backbone carbonyl of Met51. Similar interactions have been observed with other BCL6[BTB] inhibitors[17,28-32]. The pyridine ring binds in a site above the α-helix with one ring carbon in contact distance with the CG and CD1 atoms of Tyr58.

C.2. New inhibitor design

On the basis of the binding mode of fragment hit 7CC5 in complex with BCL6[BTB], we identified the pyridine moiety as growth vector by which we could extend inhibitors to the HDCH site (His14-Asp17-Cys53-His116) on BCL6[BTB]. This pocket is occupied by the side chain of Ile1425 in BCL6[BTB]-SMRT[BBD] complex, and moreover I1425A mutation abolishes competitive activity of SMRT[BBD] peptide by FP assay[24,25]. To guide our investment in synthetic medicinal chemistry, we first explored commercially available analogues of 7CC5, and subsequently characterized and ranked compounds' binding by protein-NMR experiments. We then incorporated modifications that improved binding in our in-house

synthetic medicinal chemistry campaign, as we designed next generation 7CC5 analogues

targeting the HDCH site.

C.3. Chemistry

Synthetic medicinal chemistry has been described extensively in the **Materials**

and Methods section (later in the chapter), and in Cheng and Linhares, *et al.*, *J. Med. Chem.*,

2018[36].
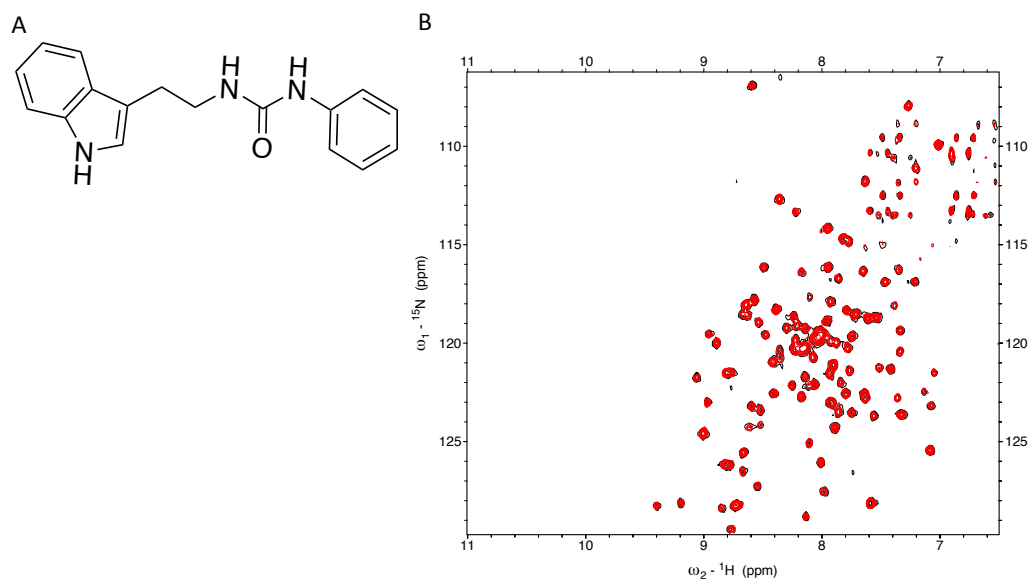
Figure 4.3. 7CC5 urea analogue 7CC5-15 does not bind BCL6$^{\text{BTB}}$. (A) Structure of 7CC5-15. (B) Superposition of ^{1}H-^{15}N HSQC spectrum of BCL6$^{\text{BTB}}$ in 5% DMSO (black), and with 500 μM 7CC5-15 (red).

Table 4.1. Characterization of the binding of BCL6 inhibitors using NMR.

a) Chemical shift perturbations for six selected amide resonances determined at 500 μM concentration of the compounds and 250 μM concentration of BCL6[BTB]. Values represent combined amide proton (HN) and nitrogen (N) chemical shift perturbations in Hz.

b) 6PA value (in Hz) represents a sum of chemical shift perturbations for six selected amides.

c) K_D values determined from NMR titration experiments based on at least five amide resonances. Limited solubility of compounds 7CC5-14 and **7a-h** precluded accurate determination of K_D values.

cmpd	T62[a] (Hz)	T48[a] (Hz)	F61[a] (Hz)	N23[a] (Hz)	R28[a] (Hz)	V18[a] (Hz)	6 PA[b] (Hz)	NMR K_D[c] (μM)
7CC5	5	5	6	12	10	11	49	3200 ± 100
7CC5-14	17	14	16	23	23	23	116	N.d.
7a	11	11	11	20	19	16	88	N.d.
7b	47	42	47	64	65	58	323	N.d.
7c	32	22	29	37	39	26	185	N.d.
7d	37	29	38	45	48	40	237	N.d.
7e	23	24	22	32	34	31	166	N.d.
7f	10	11	12	17	14	8	72	N.d.
7g	54	51	54	78	73	70	380	N.d.
7h	16	13	12	19	18	17	95	N.d.
7i	74	45	58	81	77	56	391	124 ± 6
13	58	39	51	71	74	70	363	191 ± 46
15a	35	42	33	64	64	50	288	219 ± 32
15b	38	31	41	52	56	59	277	283 ± 79
15c	48	36	41	69	67	62	323	226 ± 70
15d	56	54	51	83	86	60	390	245 ± 29
15e	41	46	31	71	77	50	316	405 ± 72
15f	57	63	61	108	102	82	473	44 ± 16
15g	36	55	33	101	92	68	385	153 ± 63
15h	28	49	42	74	72	55	320	221 ± 74

C.4. Biological evaluation.
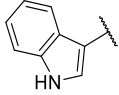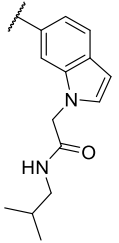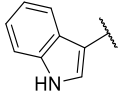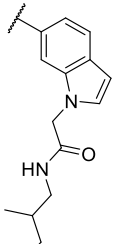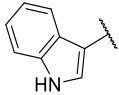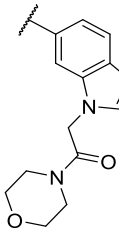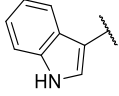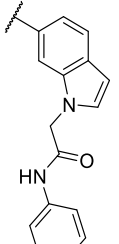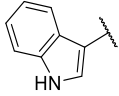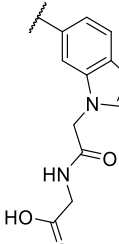
To initially characterize the binding and rank the potencies of new inhibitors, we used

NMR. Because of relatively low solubilities, we were unable to perform full titrations to

determine K_D values for many compounds. Instead, to rank compounds, we measured the sum

of chemical shift perturbations of six selected amide (Thr62, Thr48, Phe61, Asn23, Arg28, and Val18) resonances (6PA) observed on the ^{1}H-^{15}N HSQC spectra (**Tables 4.1.** and **4.2.**). The 6PA value of hit 7CC5 was measured to be 49 Hz. Compound 7CC5-14, a commercial analog of 7CC5, indicated an increased 6PA value of 116 Hz. Keeping R_1 as an indole group and replacing the Ph group of 7CC5-14 with a second indole, we synthesized compound **7a** with a 6PA value of 88 Hz. 7CC5 urea analog 7CC5-15 showed no obvious chemical shift perturbations (**Figure 4.3**), highlighting the importance of the thiourea group for this class. Next, using the nitrogen of the R_2 indole as a handle, we synthesized compounds **7b-7i** with various groups targeting the HDCH site (**Table 4.2**). Considering the distance between the aromatic and HDCH sites, we chose a methylene amide linker for the new compounds.. When *i*-Pr amine was used, compound **7b** showed a 3.7-fold improved 6PA value compared to that of compound **7a**. Interestingly however, when structurally similar *c*-Pr amine was used, the corresponding inhibitor, **7c**, indicated a much smaller 6PA value. The *n*-Pr analog, **7d**, had a 6PA value of 237 Hz while the branched *i*-Bu analog, **7e**, indicated a 6PA value of 166 Hz. These results indicated that the branching effect at the C_β of the amide bond might play an important role for more favorable binding affinity. This result was further confirmed by inhibitor **7f** employing a longer aliphatic tail. The decrease of the 6PA values from inhibitor **7d** to inhibitors **7e** and **7f** indicated that the large flexible groups might not interact efficiently with the HDCH site of BCL6^{BTB}. We were pleased to see that the morpholine analog, **7g**, indicated a large 6PA value. When a Ph group was included as a tail, the resulting compound, **7h**, indicated a significantly decreased 6PA value. Compound **7i**, which employed a carboxylic acid tail, presented a large 6PA value of 391 Hz.

Table 4.2. Chemical structures and 6PA values of inhibitors 7CC5, 7CC5-14, and 7a-i
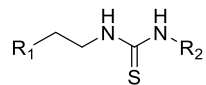
Cmpd	R$_1$	R$_2$	6PA (Hz)
7CC5			49
7CC5-14			116
7a			88
7b			323
7c			185
7d			237

7e			166
7f			72
7g			380
7h			95
7i			391
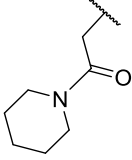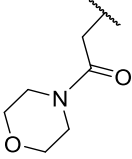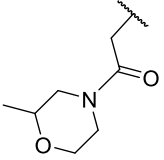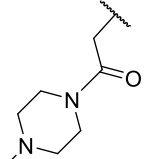
One common limitation of inhibitors **7a-i** was the relatively poor aqueous solubility (<200

μM in aqueous solution containing 1% DMSO). To address this, we synthesized new

compounds **13** and **15a-h** with a methyl piperazine fragment built off the C3 position of the right arm indole (**Table 4.3.**). Although the ionizable nitrogen atoms of the piperazine ring can help the aqueous solubility of inhibitors, the terminal tertiary amine group may also form additional interactions with residue Glu115, as previously shown in BCL6^{BTB} inhibitors.[28-31] The 6PA value of compound **13** was 363 Hz. Compared with that of inhibitor **13**, the *i*-Pr analog, **15a**, showed a decrease in 6PA value (**Table 4.3.**). Further extension of the hydrophobic tail of the inhibitor gave compound **15b**, which had an even lower 6PA value. Inhibitor **15c**, with a free carboxylic acid group to occupy the HDCH site, showed a 6PA value of 323 Hz. When secondary amines were used, inhibitors **15d-15h** demonstrated significant improvements in their 6PA values. Specifically, inhibitor **15f** showed the largest 6PA value of 473 Hz. Note that with the additional 1-methylpiperazine substitution, new inhibitors **13** and **15a-h** showed significantly improved solubility (>1000 μM in aqueous solution containing 1% DMSO).

Table 4.3. Chemical Structures, 6PA, and NMR K_D values of inhibitors 13 and 15a-h
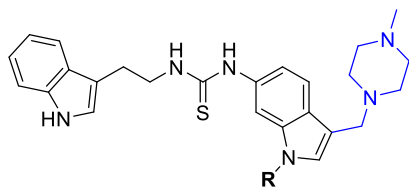
Cmpd	R	6PA (Hz)	NMR K_D (μM)
13		363	191 ± 46
15a		288	219 ± 32
15b		277	283 ± 79
15c		323	226 ± 70
15d		390	245 ± 29

15e		316	405 ± 32
15f		473	44 ± 16
15g		385	153 ± 63
15h		320	221 ± 74

Increased solubility of compounds **13** and **15a-h** permitted full NMR-titration experiments to determine binding affinities. K_D values were calculated by fitting chemical shift perturbations as functions of compound concentration using a previously established model (**Table 4.3.**)[37].

The majority of these compounds bind to BCL6 with K_D values ranging from 153 to 405 μM, and the most potent affinity was observed for **15f** (K_D = 44 μM, **Figure 4.5.B.**). The increased affinity of **15f** is consistent with the most extensive perturbations on the ^{1}H-^{15}N HSQC spectra (6PA value of 473 Hz, **Table 4.3** and **Figure 4.4.**). We have further validated the binding of compound **15f** using isothermal-titration calorimetry (ITC) and obtained very similar affinity, K_D = 36 ± 5 μM (**Figure 4.5.C.**). Moreover, compound **15f** represents an improvement in affinity for BCL6^{BTB} by two orders of magnitude relative to hit 7CC5.
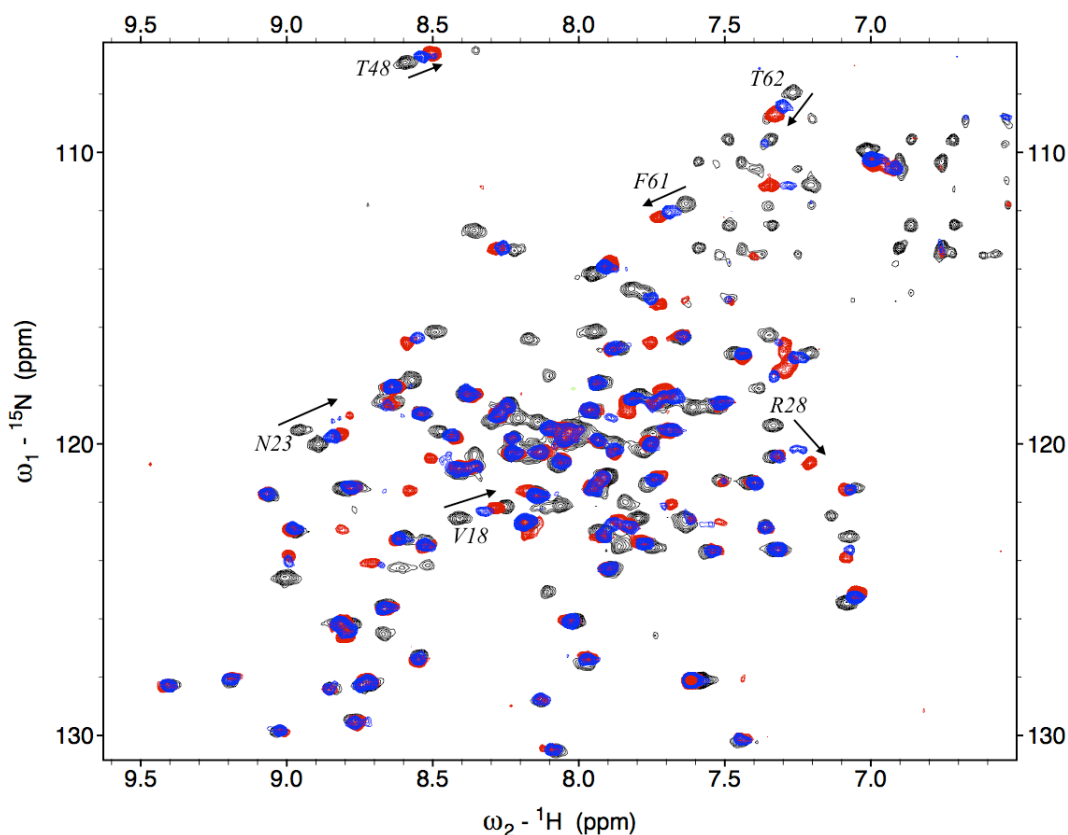
Figure 4.4. **NMR-titration experiments show dose-dependence in Chemical Shift Perturbations (CSPs) of amide resonances used in 6PA value calculation.** ^{1}H-^{15}N HSQC spectrum of BCL6^{BTB} in presence of 5% DMSO (black), overlaid with spectra of 150 μM (blue) and 500 μM (red) **15f**. Amide resonances used in 6PA calculation are highlighted. Arrows show directions of CSPs.

In order to establish the binding modes of optimized BCL6 inhibitors, we selected **15a** ($K_D = 219$ μM) and the most potent, **15f,** and determined high-resolution crystal structures of the complexes with BCL6^{BTB} (**Figure 4.5.D.** and **4.5.E.**). In both structures, **15a** and **15f** occupy a significant part of the LG, and the positions of the thiourea group are identical to fragment hit 7CC5 (**Figures 4.5.F.**). In addition, the piperazine moiety is in close proximity to Glu115 and forms a favorable electrostatic interaction. The R groups in **15a** and **15f** bind in the HDCH site, which is composed of the backbones of Ala52 and Cys53, and to the side

chains of His14, Asp17, Val18 and Asn21. The affinity of **15f** represents a 6-fold improvement over that of **15a**, and on the basis of the crystal structures, the morpholine group in **15f** extends more and therefore better occupies the HDCH site when compared with the isopropyl group in **15a** (**Figure 4.5.F.**).

We have subsequently plotted 6PA against K_D values and found a very good correlation (**Figure 4.6.**). This correlation demonstrates that for compounds binding with fast-to intermediate-exchange kinetics, it is possible to rank their affinities on the basis of quantification of chemical shift perturbations. Such approach might be particularly valuable in cases where K_Ds cannot be determined from NMR-titration experiments due to limited solubility of the ligands (as for the compounds in **Table 4.2**, for example). However, the limitation of using 6PA values renders its applicability only to ranking compounds with a similar scaffold.
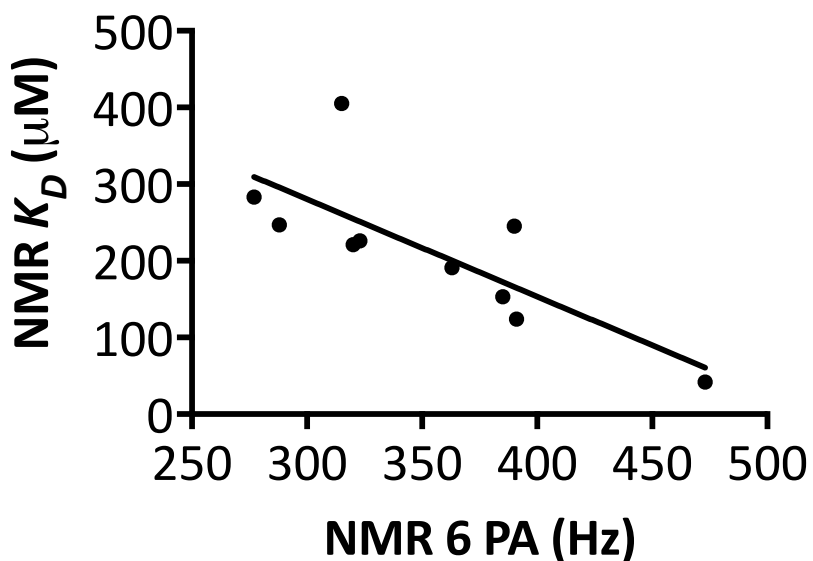
Figure 4.5. Correlation of 6PA and K_D values determined from NMR experiments.
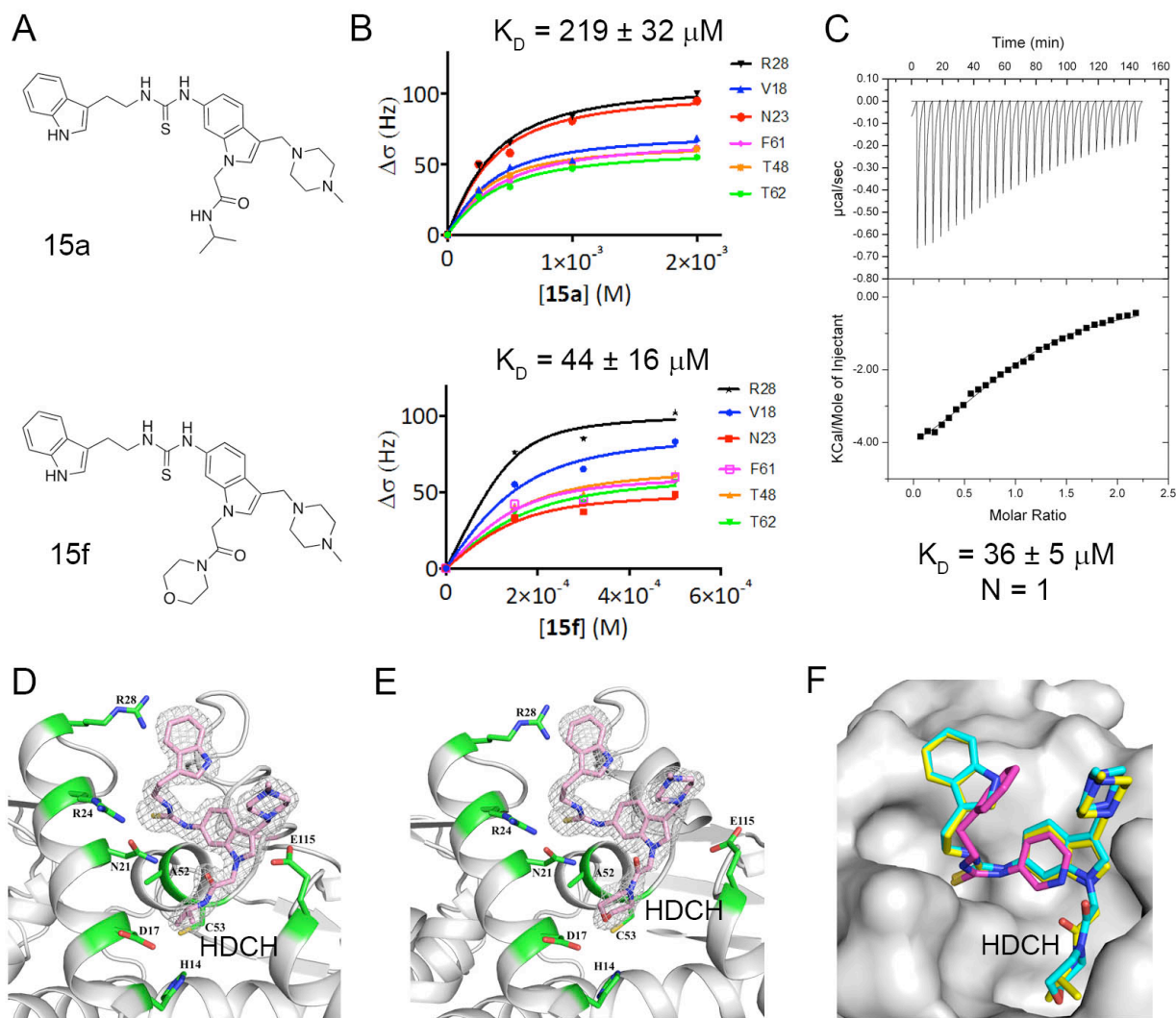
Figure 4.6. Characterization of the binding of 15a and 15f to BCL6^{BTB}. A. Chemical structures of **15a** and **15f**. B. NMR titration experiment to determine K_D values for binding of **15a** and **15f** to BCL6^{BTB}. C. Binding affinity of **15f** for BCL6^{BTB} determined using isothermal titration calorimetry. D. Crystal structure of BCL6^{BTB}-**15a** complex (PDB 6CQ1). The 2mFo-DFc electron density map contoured at 1.0 sigma is shown for **15a** and selected protein residues involved in the interactions with **15a** are shown as sticks. E. Crystal structure of BCL6^{BTB}-**15f** complex (PDB 6C3L). The 2mFo-DFc electron density map contoured at 1.0 sigma is shown for **15f** and selected protein residues involved in the interactions with **15f** are shown as sticks. F. Superposition of compounds 7CC5 (magenta), **15a** (yellow), and **15f** (cyan) from BCL6^{BTB}-complex crystal structures. HDCH site is labeled.

We further characterized the inhibitory activity of the top inhibitors, **7i**, **13**, **15a**, **15c**, **15e**, **15f**, and **15g**, in biochemical and cellular assays (**Table 4.4.**). We established an

AlphaLisa assay to test the inhibitory potencies of compounds in competing with the binding of the BCOR peptide to BCL6^{BTB}. The differential cell-killing assays were then performed for inhibitors using a panel of four BCL6-dependent lymphoma cells (OCI-Ly1, OCI-Ly7, SUDHL4, and SUDHL6) along with three BCL6-independent lymphoma cells (K422, Toledo, and Ly1B50). The IC$_{50}$ of inhibitor **7i** was measured to be 170 μM, but this compound showed no effect for all seven lymphoma cell lines at the concentration of 500 μM. With the additional methylpiperazine fragment, inhibitor **13** indicated an IC$_{50}$ value of 80 μM, although it had weak cellular activity only for the SUDHL6 cells. *i*-Pr-amide analog, **15a,** showed an IC$_{50}$ value of 142 μM, and the carboxylic acid compound **15c** showed an IC$_{50}$ value of 112 μM. Neither of these inhibitors showed any effects in the cell-killing assays. Although the NMR-based K_D value of inhibitor **15e** was determined to be over 400 μM, it showed an IC$_{50}$ value of 34 μM in the AlphaLisa assay. The GI$_{50}$ values (i.e., the concentration that causes 50% reduction of cancer-cell proliferation) of compound **15e** for BCL6-dependent lymphoma cells varies between 33 to 103 μM, while the same inhibitor indicated no effect for all tested BCL6-independent lymphoma cells at the concentration of 500 μM. The morpholine analog, **15f**, showed the best inhibitory potency for both the NMR-based K_D and the AlphaLisa assay. However, inhibitor **15f** only indicated weak potency against the OCI-Ly7 cells and no obvious activity against other tested BCL6-dependent lymphoma cells. By introducing a methyl group to the morpholine fragment, inhibitor **15g** indicated decreased NMR- and AlphaLisa-based potency compared to compound **15f**. Compound **15g** only showed activity in the OCI-Ly7 cell line. In general, the IC$_{50}$ values of the tested inhibitors indicated the same trend as that of the NMR K_D values (**Figure 4.7.**) except for compound **15e**. The current series showed relatively weak cellular activity. We reasoned that these thioureas may have limited cell

permeability because of the size, number of rotatable bonds, and multiple ionizable functional

groups of the inhibitors. The slightly improved cellular activity of the piperidine or

morpholine amide analogs (**15e-g**) might be due to their increased rigidity.

Table 4.4. Top inhibitors' activities using AlphaLisa and cellular viability assays

Cmpd	AlphaLisa IC$_{50}$ (μM)	GI$_{50}$ BCL6 dependent cell lines (μM)				GI$_{50}$ BCL6 independent cell (μM)		
		OCI-Ly1	OCI-Ly7	SUDHL4	SUDHL6	K422	Toledo	Ly1B50
7i	170 ± 4	>500	>500	>500	>500	>500	>500	>500
13	80 ± 3	>500	>500	>500	100 ± 2	>500	>500	>500
15a	142 ± 7	>500	>500	>500	>500	>500	>500	>500
15c	112 ± 5	>500	>500	>500	>500	>500	>500	>500
15e	34 ± 7	100 ± 1	33 ± 4	103 ± 18	60 ± 1	>500	>500	>500
15f	27 ± 4	>500	193 ± 12	>500	>500	>500	>500	>500
15g	62 ± 17	>500	116 ± 13	>500	>500	>500	>500	>500

D. Conclusion

  In conclusion, we have identified and optimized a series of thiourea inhibitors of the

BCL6^{BTB}. We used NMR-based fragment screening to identify the initial hit, and solved the

crystal structure of fragment hit in complex with BCL6^{BTB}. We employed protein-NMR, x-

ray crystallography, and biophysical characterization by ITC, and biochemical assay

AlphaLISA to design and develop more potent inhibitors. Our efforts led to the identification

of compounds **15e** and **15f**. Compound **15f** is one of the most potent BCL6 inhibitors in our

series and represents a two-order of magnitude improvement in activity when compared to the

original fragment hit, 7CC5. We have also tested binding of **15f** to the three related BTB domain proteins: KAISO, MIZ1 and LRF and found no interactions (**Figure 4.8.**), which suggests that the thiourea scaffold is selective for BCL6. We have not observed potent activity in cell-based assays for this class of BCL6 inhibitors, likely because of insufficient permeability of the compounds. Nevertheless, the availability of the crystal structure and validation of the key role of the HDCH pocket for inhibitor binding identify this scaffold as a valuable core structure for further modifications to develop potent BCL6 inhibitors. Investigations of new inhibitors, by replacing the thiourea fragment in the current series with high-affinity scaffolds, are ongoing in our laboratories.
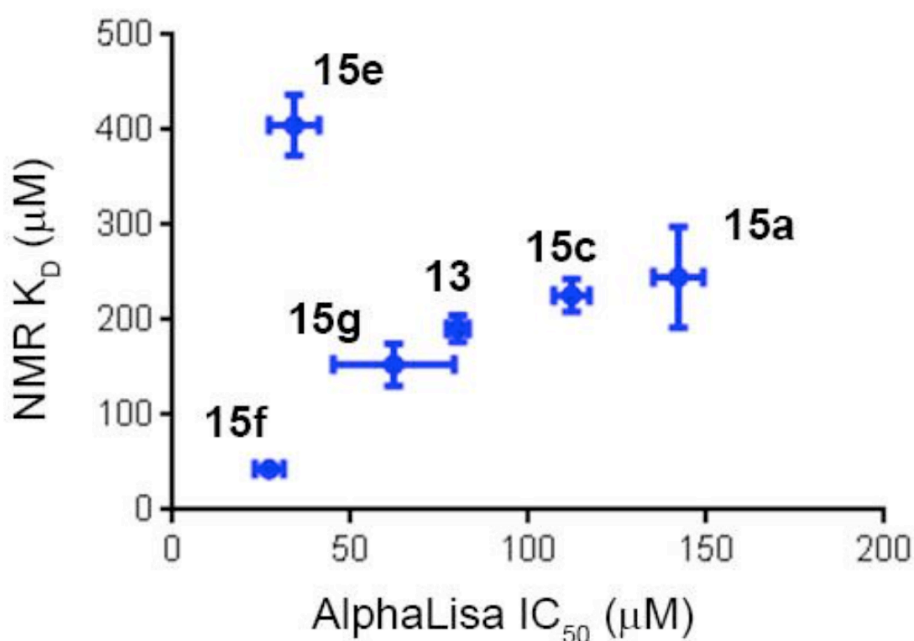
Figure 4.7. Correlation of IC$_{50}$ values from AlphaLISA experiment and K_D determined from NMR.

Acknowledgements

113

First, I must recognize Huimin Cheng, Ph.D., and Fengtian Xue, Ph.D., of the Univerisity of Maryland School of Pharmacy, for contributions to inhibitor design and synthesis of inhibitors **7a-i**, and **15a-i**. Additionally, of the University of Maryland School of Pharmacy, I thank Wenbo Yu, Ph.D., Yong Ai, Ph.D., and Alex MacKerell, Jr., Ph.D., for computational studies. I am grateful for the efforts of Mariano G. Cardenas, Ph.D., and Ari Melnick, M.D., of Weill Cornell Medical College, for biochemical assay development and *in cell* studies. Finally, I am especially appreciative of Alyssa Winkler of the Cierpicki Lab for assistance in biophysical studies NMR-titration experiments to determine K_Ds, and of Tomasz Cierpicki, Ph.D., for oversight in biophysical and structural studies, and in inhibitor design.

E. Materials and Methods

General Information

All reagents were purchased without further purification unless otherwise noted. Reactions were monitored using thin-layer chromatography (TLC) on commercial silica gel plates (GF254). Visualization of the developed plates was performed under UV light (254 nm). Flash column chromatography was performed on silica gel (200-300 mesh). ^{1}H and ^{13}C NMR spectra were recorded on a Varian INOVA 400 MHz NMR spectrometer at 25 °C. Chemical shifts (δ) are reported in ppm referenced to an internal tetramethylsilane standard, or the DMSO-d_6 residual peak (δ 2.50) for ^{1}H NMR. Chemical shifts of ^{13}C NMR are reported relative to CDCl$_3$ (δ 77.0) or DMSO-d_6 (δ 39.5). The following abbreviations were used to describe peak splitting patterns when appropriate: br s = broad singlet, s = singlet, d = doublet, t = triplet, q = quartet, m = multiplet. Coupling constants, J, were reported in Hertz unit (Hz). High-resolution mass spectra (HRMS) were obtained on a JEOL AccuTOF with ESI/APCI

ion sources coupled to an Agilent 1100 HPLC system. HPLC analysis was performed on a Agilent 1100 series (DAD detector) fitted with a C-18 reversed-phase column (XTerra RP 18, 5 μM, 4.6 × 250mm) with a flow rate of 0.8 mL/min using CH_3CN-H_2O mobile phase and detected under 254 nm UV light. The purity of final products is >95%. Compounds 7CC5 and 7CC5-14 were purchased from Maybridge and Vitas, respectively.

New Compound Synthesis and Characterization

The synthesis of inhibitors **7a-i** is detailed in Scheme 4.1. Alkylation of the indole nitrogen of the commercially available 6-nitroindole (**1**) using methyl bromoacetate in the presence of K_2CO_3 generated compound **2** in good yields. Catalytic hydrogenation of the nitro group in compound **2** using Pd/C provided amino compound **3** in modest yields. Amino compound **3** was converted into the isothiocyanate **4** using 1,1'-thiocarbonylbis(pyridine-2(1*H*)-one) in high yields. The resulting isothiocyanate group of compound **4** reacted with tryptamine to give thiourea compound **5** in modest yields. Saponification of methylester **5** using LiOH·H_2O generated carboxylic acid **6**, which was submitted to a 1-[bis(dimethylamino)methylene]-1*H*-1,2,3-triazolo[4,5-*b*]pyridinium 3-oxid hexafluorophosphate (HATU)-mediated amide bond formation reaction to give the final inhibitors **7a-i** in modest yields.

Scheme 1. Synthesis of compounds **7a-i**.*a*
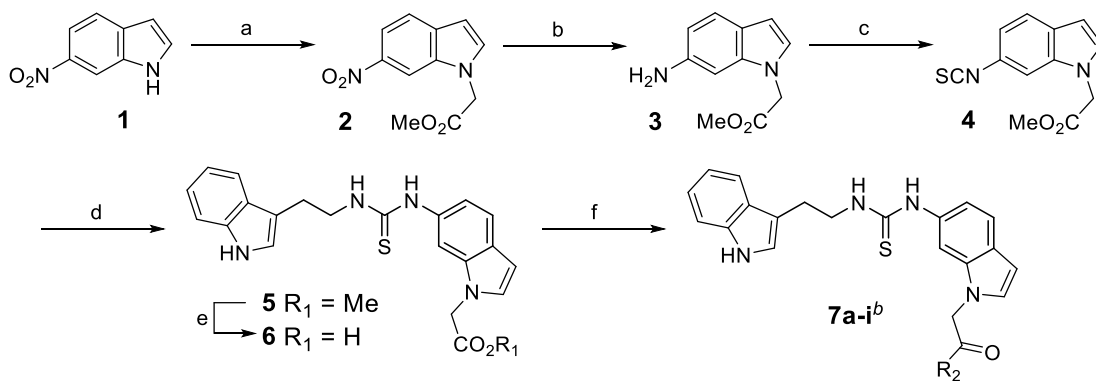
^aReagents and conditions: (a) methyl bromoacetate, K$_2$CO$_3$, DMF, 60 °C, overnight, 86%; (b) Pd/C, H$_2$, MeOH:THF = 1:1, r.t., overnight, 52%; (c) 1,1'-thiocarbonylbis(pyridine-2(1H)-one), CH$_2$Cl$_2$, r.t., 2 h, 90%; (d) tryptamine, CH$_2$Cl$_2$, r.t., 3 h, 76%; (e) LiOH·H$_2$O, MeOH:THF:H$_2$O = 3:1:1, 70 °C, 3 h, 82%; (f) amines, HATU, DIPEA, DMF, r.t., overnight, 42-86%.^b The synthesis of **7a** and **7i** were slightly different from compounds **7b-h**

The synthesis of inhibitors **15a-h** started with 6-nitro-1H-indole-3-carbaldehyde **8** (**Scheme 4.2.**). Alkylation of the indole nitrogen of compound **8** using methyl bromoacetate in the presence of K$_2$CO$_3$ generated compound **9** in good yields. Reductive amination of aldehyde **9** with 1-methylpiperazine using NaBH(OAc)$_3$ generated compound **10** in modest yields. Reduction of the nitro group of compound **10** using catalytic hydrogenation provided compound **11** in modest yields. The amino compound **11** was converted into isothiocyanate **12** using 1,1'-thiocarbonylbis (pyridine-2(1H)-one). The isothiocyanate group of compound **12** reacted with tryptamine to give thiourea **13** in modest yields. Saponification of methylester **13** using LiOH·H$_2$O generated carboxylic acid **14**, which was submitted to an HATU-mediated amide bond formation reaction to give the final inhibitors **15a-h** in modest yields. More extensive and characterization, including 1-D spectra analyses of compounds, has been described previously[36].
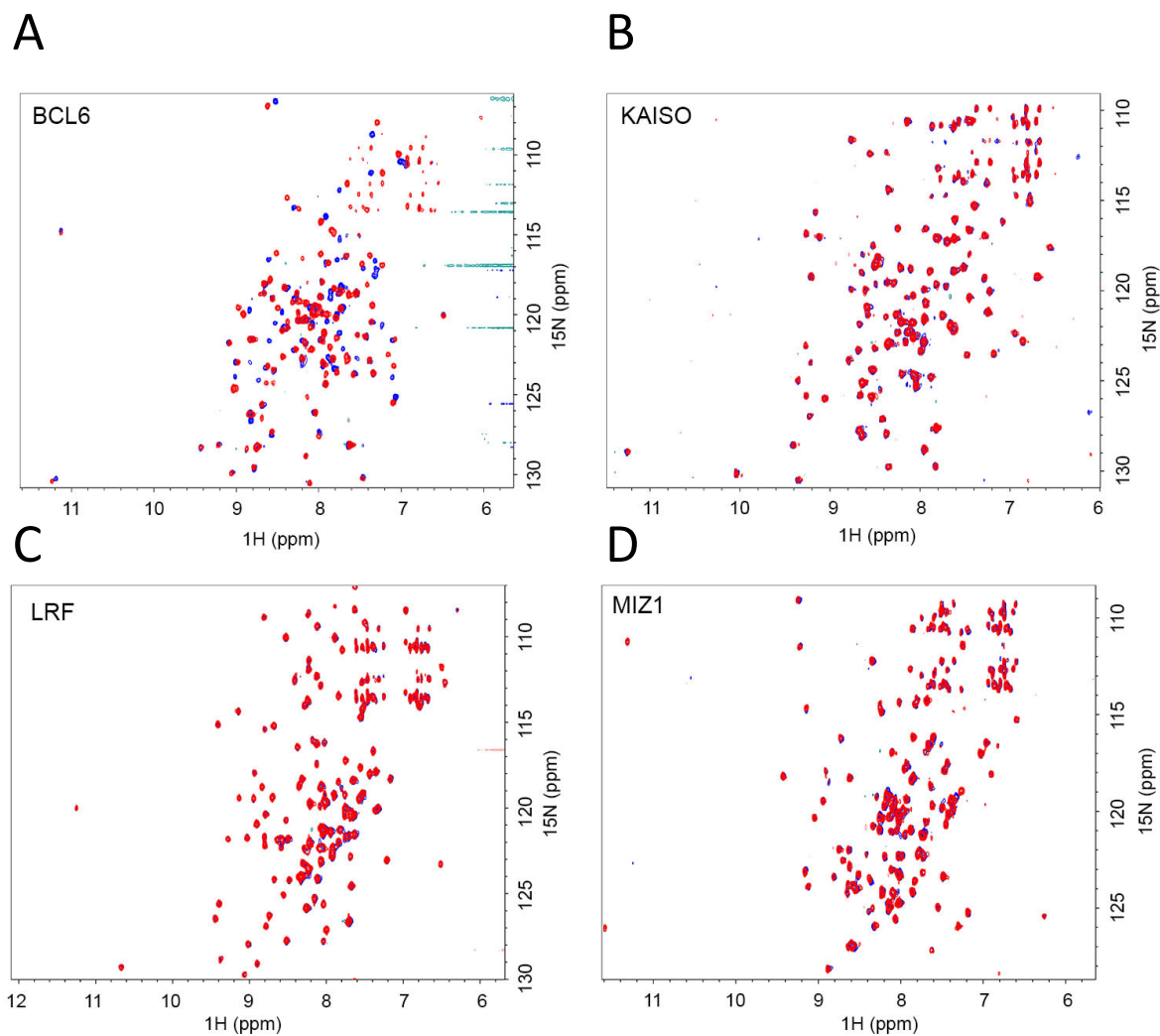
Figure 4.8. 15f is selective for BCL6^{BTB}. Superposition of ^{1}H-^{15}N HSQC spectra of A. 150 μM BCL6^{BTB} in 5% DMSO (red), and in presence of 500 μM **15f** (blue). B. 150 μM KAISO^{BTB} in 5% DMSO (red), and in presence of 500 μM **15f** (blue). C. 150 μM LRF^{BTB} in 5% DMSO (red), and in presence of 500 μM **15f** (blue). D. 150 μM Miz1^{BTB} in 5% DMSO (red), and in presence of 500 μM **15f** (blue).

Scheme 4.2. Synthesis of inhibitors **13**, **15a-h**.[a]
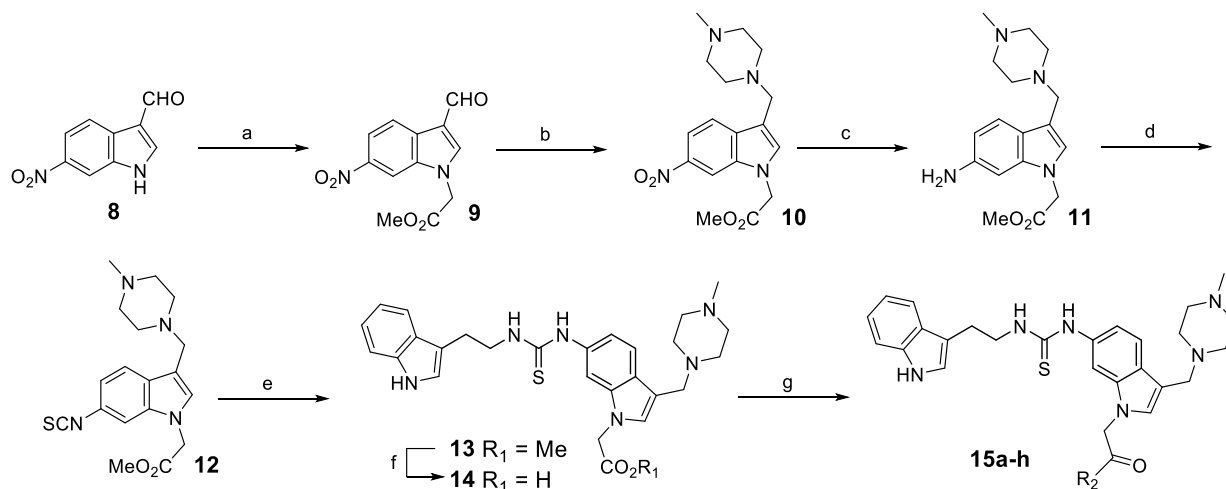
[a]Reagents and conditions: (a) methyl bromoacetate, K_2CO_3, DMF, 70 °C, overnight, 79%; (b) 1-methylpiperazine, NaBH(OAc)$_3$, CH$_2$Cl$_2$, r.t., overnight, 58%; (c) Pd/C, H$_2$, MeOH, r.t., 3 h, 72%; (d) 1,1'-thiocarbonylbis(pyridine-2(1H)-one), CH$_2$Cl$_2$, r.t., 4 h; (e) tryptamine, CH$_2$Cl$_2$, r.t., overnight, 33% for two steps; (f) LiOH·H$_2$O, MeOH:THF:H$_2$O = 3:1:1, r.t., 3 h, 80%; (g) amines, HATU, DIPEA, DMF, r.t., overnight, 32-65%.

BCL6^{BTB}-BCoR AlphaLisa Assay

We designed and optimized an Alphascreen assay using BCL6^{BTB}-His$_6$ protein, biotinylated BCOR peptide (Biotin-Ahx-RSEIISTAPSSWVVPGP-COOH, Biosynthesis) and Alphascreen acceptor and donor beads (AlphaLisa Histidine detection kit, PerkinElmer). In this optimized low volume assay, we incubated the recombinant protein diluted in AlphaLisa immunoassay buffer with different concentrations of the small molecules for 30 min. Then, we added the acceptor and donor beads, together with the BCOR peptide. After 1h of incubation in the dark and agitation, we measured the luminescence produced by the acceptor beads after excitation of the donor beads with a laser. This setup produced a signal/background ratio of 20 and a Z' Score of 0.8 that was calculated as follows:

Equation 4.1.

Z' score: 1 – $\dfrac{(3*(\text{SD negative control} + \text{SD positive control}))}{(\text{Av negative control} - \text{Av positive control})}$

Using 0.2% DMSO as a negative control and 80 µM of SMRT Peptide (H_2N-

LVATVKEAGRSIHEIPR-CO_2H, Biosynthesis) as a positive control.

Expression and Purification of Recombinant Proteins

cDNA encoding BCL6[BTB] (1-129; C8Q, C67R, C84N, and E99V) was cloned into a pet32a

vector containing an N-terminal thioredoxin-His$_6$ tag with an N-terminal thrombin-cleavage

site. Cells were grown in either LB or ^{15}N-labeled M9 medium with ampicillin selection, and

protein was expressed with 0.25 mM IPTG for 16 hours at 18 ^{o}C. *E. Coli* cells were lysed in a

buffer containing 50 mM Tris, pH 7.5, 300 mM NaCl, 30 mM imidazole, 1 mM TCEP, and

0.5 mM PMSF. Protein was purified using Ni-NTA resin, eluted with lysis buffer containing

300 mM imidazole. The eluate was proteolytically cleaved with Thrombin, and applied to Ni-

NTA to remove the thioredoxin-His$_6$ tag. In the final step, protein was dialyzed to 20 mM

Tris, pH 7.5, 150 mM NaCl, and 1 mM TCEP, and frozen at -80 ^{o}C. For protein-observed

NMR experiments, cDNA encoding BCL6[BTB] (1-129; C67R, C84N, E99V), LRF[BTB] (1-130),

KAISO[BTB] (1-122), and Miz1[BTB] (1-115) were cloned into a pet32a vector containing an *N*-

terminal thioredoxin-His$_6$ tag with an *N*-terminal PreScission cleavage site. Proteins were

expressed and purified in similar manner as described above. Following affinity purification,

proteins were cleaved with PreScission protease and thioredoxin-His$_6$ tag was removed using

Ni-NTA resin. Finally, BCL6[BTB], Miz1[BTB], and KAISO[BTB] proteins were dialyzed to 50 mM

Tris, pH 7.5, 150 mM NaCl, and 1 mM TCEP, and used in NMR experiments. LRF[BTB] was

dialyzed to 50 mM HEPES, pH 7.5, 150 mM NaCl, and 1 mM TCEP.

Fragment Screening by Protein-Observed NMR Spectroscopy

The fragment library used for screening was a combination of commercially available fragment-like small molecules and compounds synthesized in-house. Samples for fragment screening were prepared with 150 μM ^{15}N-labeled BCL6^{BTB} (1-129; C67R, C84N, E99V) in buffer containing 50 mM Tris, pH 7.5, 150 mM NaCl, 1 mM TCEP, and 7.5% D$_2$O. Fragments were screened in mixtures of 10 compounds per sample at 500 μM final concentrations in 5% DMSO. ^{1}H-^{15}N HSQC spectra were acquired at 30 ^{o}C on a 600 MHz Bruker Avance III spectrometer equipped with cryoprobe, running Topspin version 2.1. Processing and spectral visualization was performed using NMRPipe and Sparky. Ranking of compound affinity was performed by calculating respective 6 PA values measured at 500 μM compound concentration for the six most perturbed amide resonances on the ^{1}H-^{15}N HSQC spectra of BCL6^{BTB}.

Calculation of 6 PA Values

6 PA values represent sum of ^{1}H- and ^{15}N-chemical shift perturbations (Hz) calculated for the backbone amides of Thr62, Thr48, Phe61, Asn23, Arg28, and Val18 (Table 3.2, Figure 3.6). NMR samples for the 6 PA measurements contained 500 μM compound (in 5% DMSO) and 250 μM BCL6^{BTB} (1-129; C67R, C84N, E99V).

K_D Determination by Protein-Observed NMR Spectroscopy

^{1}H-^{15}N HSQC experiments were performed at 30 ^{o}C with samples containing 150 μM BCL6^{BTB} (residues 1-129; C67R, C84N, E99V) in 50 mM Tris, pH 7.5, 150 mM NaCl, 1 mM TCEP, 7.5% D$_2$O, and 5% DMSO, by adding increasing concentrations of compound 7CC5 (1 mM, 2 mM, and 4 mM). For more potent compounds, the concentrations were varied from 150 to 1000 μM and 250 μM BCL6^{BTB} (1-129; C67R, C84N, E99V). Dissociation constants

were determined from least-squares fitting of chemical shift perturbations as a function of ligand concentration:

Equation 4.2.

$$\delta_i = \frac{\{b - \sqrt{(b^2 - 4 \times a \times c)}\}}{2a}$$

with $a = (K_A/\delta_b) \times [P_t]$, $b = 1 + K_A([L_{ti}] + [P_t])$, and $c = \delta_b \times K_A \times [L_{ti}]$, where δ_i is the absolute change in chemical shift for each titration point, $[L_{ti}]$ is the total ligand concentration at each titration point, $[P_t]$ is the total protein concentration, $K_A = 1/K_D$ is the binding constant, and δ_b is the chemical shift of the resonance in question in the complex[37]. K_D and δ_b were used as fitting parameters in analysis[37].

Isothermal Titration Calorimetry

BCL6[BTB] was dialyzed repeatedly against 50 mM phosphate buffer, pH 7.5, 150 mM NaCl, and 1 mM TCEP (referred to herein as ITC Buffer) at 4 °C. **15f** was dissolved in DMSO and diluted to 1 mM in ITC Buffer containing 5% DMSO. Additionally, 5% DMSO was added to protein and buffer (reference) solutions. Titrations were performed using a VP-ITC titration calorimetric system (MicroCal) at 25 °C. The calorimetric cell, containing 100 μM BCL6[BTB], was titrated with **15f** at 1 mM, in 10 μL aliquots and at 300 s intervals. Data were analyzed by Origin 7.0 (OriginLab) to obtain the K_D and stoichiometry.

Crystallization of BCL[BTB]-Inhibitor Complexes

Initial crystals of native-BCL6[BTB] (1-129; C8Q, C67R, C84N, and E99V) were obtained by screening with the hanging-drop vapor-diffusion technique. Crystals were further optimized using the sitting-drop technique, with equal volumes (1.5 μL) of protein (9 mg/mL in 20 mM Tris, pH 7.5, 150 mM NaCl, and 1 mM TCEP) and precipitant solution (1.6 M ammonium formate and 0.1 M Tris, pH 7.5). Crystals formed overnight at 4 °C. For 7CC5, crystals were

transferred to the mother liquor solution containing molar excess of 7CC5 and 2% DMSO to soak for 2 h, and were then transferred to the mother liquor solution containing molar excess of 7CC5, 2% DMSO, and 30% glycerol for cryoprotection prior to freezing in liquid nitrogen. For compound **15a**, 3 mg/mL BCL6^{BTB} (1-129; C8Q, C67R, C84N, and E99V) was incubated with 2 mM compound (a 10-fold molar excess) for 6 h at room temperature. Crystals of BCL6^{BTB}-**15a** complex were obtained through screening using the hanging-drop vapor diffusion technique at 4 °C overnight in 0.2 M NaCl, 0.1 M sodium cacodylate, pH 6.0, and 8% (w/v) PEG-8,000. Before data collection, crystals were transferred into mother liquor containing 2 mM **15a**, 2% DMSO, and 30% glycerol for cryoprotection prior to freezing in liquid nitrogen. For compound **15f**, 6 mg/mL BCL6^{BTB} (1-129; C8Q, C67R, C84N, and E99V) was incubated with 2 mM compound for 6 h at room temperature. Initial crystals were obtained by screening with the hanging-drop vapor-diffusion technique. Crystals were further optimized using the sitting-drop technique, with equal volumes (1.5 μL) of protein-inhibitor complex and precipitant solution (0.1 M sodium acetate, pH 4.5, and 27% (w/v) PEG-3,350). Crystals grew within one week at 4 °C. Before data collection, crystals were transferred to the mother liquor solution containing 2 mM **15f**, 2% DMSO, and 30% glycerol for cryoprotection prior to freezing in liquid nitrogen.

Crystallographic Data Collection and Structure Determination

Diffraction data for the BCL6^{BTB}-inhibitor complexes were collected at the 21-ID-F and 21-ID-G beam lines at the Life Sciences Collaborative Access Team at the Advanced Photon Source. Data were integrated and scaled using HKL-2000[38], and structures were solved by molecular replacement with MOLREP using the known native BCL6^{BTB} structure for the search model. Refinement for structures was performed using REFMAC[39], COOT[40], the

CCP4[41] program suite, and the PHENIX[42] program suite. The structures were validated using the MOLPROBITY[43] and ADIT[44]. Data collection and structure refinement statistics are presented in **Table 4.5**.

Cell Lines

Sources: Human embryonic kidney 293T cells, and human DLBCL cells Toledo, SUDHL-4 and SUDHL-6, cells were purchased in ATCC. Human DLBCL cell lines OCI-Ly1 and OCI-Ly7 were from the DSMZ German collection of Microorganisms and cell cultures. OCI-Ly1-B50 are OCI-Ly1 cells that were grown in the presence of increasing concentrations of RI-BPI and grow independently of BCL6 inhibitors. DLBCL cell lines, TMD8 and HBL-1 were kindly provided by Dr. Louis Staudt, National Institute oh Health (NIH, MD). All human cell lines were identified and authenticated by DNA genotyping previous to use by Biosynthesis (Lewisville, TX). 293T cells were grown in Dulbecco's modified Eagle's medium supplemented with 10% fetal bovine serum (FBS) (Life Technologies Corp., Grand Island, NY) and 1% penicillin–streptomycin (Life Technologies Corp., Grand Island, NY). DLBCL cell lines OCI-Ly1, OCI-Ly1-B50 and OCI-Ly7 were grown in Iscove's medium (Life Technologies Corp., Grand Island, NY), 10% FBS and 1% penicillin–streptomycin. DLBCL cell lines Toledo, SUDHL-4, HBL-1, and SUDHL-6 were cultured in 90% RPMI medium (Life Technologies Corp., Grand Island, NY), 10% FBS, 2 mM glutamine, 10 mM Hepes (Life Technologies Corp., Grand Island, NY), and 1% penicillin–streptomycin. All cell lines were cultured at 37°C in a humidified atmosphere of 5% CO2.

Growth-Inhibition Determination

Exponentially growing DLBCL cell lines were treated in three replicates with small molecules for 48 hours. Cell viability was then determined with the fluorescent redox dye CellTiter-Blue

(Promega, Madison, WI). Fluorescence was determined with the Synergy NEO microplate

reader (BioTek, North Brunswick Township, NJ) and the percentage of cell viability of the

drug-treated cells was normalized to their vehicle. The drug effect was calculated as 100 –

(percent viability). Dose-effect curves were used to determine the drug concentration that

inhibits the growth of cell lines by 50% compared with vehicle (GI_{50}) with the CompuSyn

software (Biosoft, Cambridge, UK). Experiments were performed in triplicate.

Table 4.5. Crystallographic Data Collection and Refinement Statistics

	$BCL6^{BTB}$-7CC5	$BCL6^{BTB}$-**15f**	$BCL6^{BTB}$-**15a**
PDB Code	6C3N	6C3L	6CQ1
Space group	$P2_12_12_1$	P1	P1
Cell dimensions *a, b, c* (Å)	34.1 85.1 117.7	30.9 39.6 55.2	31.0, 39.6, 55.2
Cell dimensions α β γ (deg)	90 90 90	83.7 73.8 67.0	83.7, 73.6, 67.0
Resolution (Å)	2.53 (2.62 - 2.53)	1.46 (1.51 - 1.46)	1.70 (1.76 – 1.70)
Unique reflections	11,920 (1054)	38,467 (3801)	22,269 (2398)
R_{merge}	0.116 (0.438)	0.066 (0.316)	.062 (.347)
$I/\sigma I$	16.6 (4.4)	11.4 (2.1)	11.5 (2.5)
Completeness (%)	84.4 (87.5)	95.9 (93.6)	86.6 (93.6)
Redundancy	6.3 (6.8)	2.0 (2.0)	2.0 (1.9)
	Refinement		
R_{work}/R_{free} (%)	20.7/26.7	19.9/22.4	18.1/21.7
No. atoms			
Protein	1937	1974	1991
Water	86	286	242
Mean B-factors ($Å^2$)	45.33	29.62	32.58
RMS deviations			
Bond lengths (Å)	0.003	0.002	.012
Bond angles (deg)	0.52	0.48	1.26
	Ramachandran plot		
Most favored regions (%)	97.10	98.36	98.35
Additional allowed regions (%)	2.90	1.64	1.65

F. References

(1)   Cardenas, M. G.; Oswald, E.; Yu, W.; Xue, F.; MacKerell, A. D.; Melnick, A. M. The expanding role of the BCL6 oncoprotein as a cancer therapeutic target. *Clin. Cancer Res.* **2017**, *23*, 885-893.

(2)   Ye, B. H.; Cattoretti, G.; Shen, Q.; Zhang, J.; Hawe, N.; de Waard, R.; Leung, C.; Nouri-Shirazi, M.; Orazi, A.; Chaganti, R. S.; Rothman, P.; Stall, A. M.; Pandolfi, P. P.; Dalla-Favera, R. The BCL-6 proto-oncogene controls germinal-centre formation and Th2-type inflammation. *Nature genet.* **1997**, *16*, 161-170.

(3)   Fukuda, T.; Yoshida, T.; Okada, S.; Hatano, M.; Miki, T.; Ishibashi, K.; Okabe, S.; Koseki, H.; Hirosawa, S.; Taniguchi, M.; Miyasaka, N.; Tokuhisa, T. Disruption of the Bcl6 gene results in an impaired germinal center formation. *J. Exp. Med.* **1997**, *186*, 439-448.

(4)   Duy, C.; Hurtz, C.; Shojaee, S.; Cerchietti, L.; Geng, H.; Swaminathan, S.; Klemm, L.; Kweon, S. M.; Nahar, R.; Braig, M.; Park, E.; Kim, Y. M.; Hofmann, W. K.; Herzog, S.; Jumaa, H.; Koeffler, H. P.; Yu, J. J.; Heisterkamp, N.; Graeber, T. G.; Wu, H.; Ye, B. H.; Melnick, A.; Muschen, M. BCL6 enables Ph+ acute lymphoblastic leukaemia cells to survive BCR-ABL1 kinase inhibition. *Nature* **2011**, *473*, 384-388.

(5)   Geng, H.; Brennan, S.; Milne, T. A.; Chen, W. Y.; Li, Y.; Hurtz, C.; Kweon, S. M.; Zickl, L.; Shojaee, S.; Neuberg, D.; Huang, C.; Biswas, D.; Xin, Y.; Racevskis, J.; Ketterling, R. P.; Luger, S. M.; Lazarus, H.; Tallman, M. S.; Rowe, J. M.; Litzow, M. R.; Guzman, M. L.; Allis, C. D.; Roeder, R. G.; Muschen, M.; Paietta, E.; Elemento, O.; Melnick, A. M. Integrative epigenomic analysis identifies biomarkers and therapeutic targets in adult B-acute lymphoblastic leukemia. *Cancer discov.* **2012**, *2*, 1004-1023.

(6)   Hurtz, C.; Hatzi, K.; Cerchietti, L.; Braig, M.; Park, E.; Kim, Y. M.; Herzog, S.; Ramezani-Rad, P.; Jumaa, H.; Muller, M. C.; Hofmann, W. K.; Hochhaus, A.; Ye, B. H.; Agarwal, A.; Druker, B. J.; Shah, N. P.; Melnick, A. M.; Muschen, M. BCL6-mediated repression of p53 is critical for leukemia stem cell survival in chronic myeloid leukemia. *J. Exp. Med.* **2011**, *208*, 2163-2174.

(7)   Walker, S. R.; Liu, S.; Xiang, M.; Nicolais, M.; Hatzi, K.; Giannopoulou, E.; Elemento, O.; Cerchietti, L.; Melnick, A.; Frank, D. A. The transcriptional modulator BCL6 as a molecular target for breast cancer therapy. *Oncogene* **2015**, *34*, 1073-1082.

(8)   Ranuncolo, S. M.; Polo, J. M.; Melnick, A. BCL6 represses CHEK1 and suppresses DNA damage pathways in normal and malignant B-cells. *Blood Cells Mol. Dis.* **2008**, *41*, 95-99.

(9)   Ci, W.; Polo, J. M.; Cerchietti, L.; Shaknovich, R.; Wang, L.; Yang, S. N.; Ye, K.; Farinha, P.; Horsman, D. E.; Gascoyne, R. D.; Elemento, O.; Melnick, A. The BCL6 transcriptional program features repression of multiple oncogenes in primary B cells and is deregulated in DLBCL. *Blood* **2009**, *113*, 5536-5548.

(10)   Ranuncolo, S. M.; Polo, J. M.; Dierov, J.; Singer, M.; Kuo, T.; Greally, J.; Green, R.; Carroll, M.; Melnick, A. Bcl-6 mediates the germinal center B cell phenotype and lymphomagenesis through transcriptional repression of the DNA-damage sensor ATR. *Nature immunol.* **2007**, *8*, 705-714.

(11)   Phan, R. T.; Saito, M.; Basso, K.; Niu, H.; Dalla-Favera, R. BCL6 interacts with the transcription factor Miz-1 to suppress the cyclin-dependent kinase inhibitor p21 and cell cycle arrest in germinal center B cells. *Nature immunol.* **2005**, *6*, 1054-1060.

(12)   Phan, R. T.; Dalla-Favera, R. The BCL6 proto-oncogene suppresses p53 expression in germinal-centre B cells. *Nature* **2004**, *432*, 635-639.

(13)   Cerchietti, L. C.; Hatzi, K.; Caldas-Lopes, E.; Yang, S. N.; Figueroa, M. E.; Morin, R. D.; Hirst, M.; Mendez, L.; Shaknovich, R.; Cole, P. A.; Bhalla, K.; Gascoyne, R. D.; Marra, M.; Chiosis, G.; Melnick, A. BCL6 repression of EP300 in human diffuse large B cell lymphoma cells provides a basis for rational combinatorial therapy. *J. Clin. Invest.* **2010**, *120*, 4569-4582.

(14)   Parekh, S.; Polo, J. M.; Shaknovich, R.; Juszczynski, P.; Lev, P.; Ranuncolo, S. M.; Yin, Y.; Klein, U.; Cattoretti, G.; Dalla Favera, R.; Shipp, M. A.; Melnick, A. BCL6 programs lymphoma cells for survival and differentiation through distinct biochemical mechanisms. *Blood* **2007**, *110*, 2067-2074.

(15)   Polo, J. M.; Dell'Oso, T.; Ranuncolo, S. M.; Cerchietti, L.; Beck, D.; Da Silva, G. F.; Prive, G. G.; Licht, J. D.; Melnick, A. Specific peptide interference reveals BCL6 transcriptional and oncogenic mechanisms in B-cell lymphoma cells. *Nat. Med.* **2004**, *10*, 1329-1335.

(16)   Cerchietti, L. C.; Yang, S. N.; Shaknovich, R.; Hatzi, K.; Polo, J. M.; Chadburn, A.; Dowdy, S. F.; Melnick, A. A peptomimetic inhibitor of BCL6 with potent antilymphoma effects in vitro and in vivo. *Blood* **2009**, *113*, 3397-3405.

(17)   Cerchietti, L. C.; Ghetu, A. F.; Zhu, X.; Da Silva, G. F.; Zhong, S.; Matthews, M.; Bunting, K. L.; Polo, J. M.; Fares, C.; Arrowsmith, C. H.; Yang, S. N.; Garcia, M.; Coop, A.; Mackerell, A. D., Jr.; Prive, G. G.; Melnick, A. A small-molecule inhibitor of BCL6 kills DLBCL cells in vitro and in vivo. *Cancer Cell* **2010**, *17*, 400-411.

(18)   Cerchietti, L. C.; Polo, J. M.; Da Silva, G. F.; Farinha, P.; Shaknovich, R.; Gascoyne, R. D.; Dowdy, S. F.; Melnick, A. Sequential transcription factor targeting for diffuse large B-cell lymphomas. *Cancer Res.* **2008**, *68*, 3361-3369.

(19)   Ci, W.; Polo, J. M.; Melnick, A. B-cell lymphoma 6 and the molecular pathogenesis of diffuse large B-cell lymphoma. *Curr. Opin. Hematol.* **2008**, *15*, 381-390.

(20)   Ye, B. H. BCL-6 in the pathogenesis of non-Hodgkin's lymphoma. *Cancer Invest.* **2000**, *18*, 356-365.

(21)   Zhang, H.; Okada, S.; Hatano, M.; Okabe, S.; Tokuhisa, T. A new functional domain of Bcl6 family that recruits histone deacetylases. *Biochim. Biophys. Acta* **2001**, *1540*, 188-200.

(22)   Huynh, K. D.; Bardwell, V. J. The BCL-6 POZ domain and other POZ domains interact with the co-repressors N-CoR and SMRT. *Oncogene* **1998**, *17*, 2473-2484.

(23)   Huynh, K. D.; Fischle, W.; Verdin, E.; Bardwell, V. J. BCoR, a novel corepressor involved in BCL-6 repression. *Genes. Dev.* **2000**, *14*, 1810-1823.

(24)   Ahmad, K. F.; Melnick, A.; Lax, S.; Bouchard, D.; Liu, J.; Kiang, C. L.; Mayer, S.; Takahashi, S.; Licht, J. D.; Prive, G. G. Mechanism of SMRT corepressor recruitment by the BCL6 BTB domain. *Mol. Cell* **2003**, *12*, 1551-1564.

(25)   Ghetu, A. F.; Corcoran, C. M.; Cerchietti, L.; Bardwell, V. J.; Melnick, A.; Prive, G. G. Structure of a BCOR corepressor peptide in complex with the BCL6 BTB domain dimer. *Mol. Cell* **2008**, *29*, 384-391.

(26)   Bunting, K. L.; Melnick, A. M. New effector functions and regulatory mechanisms of BCL6 in normal and malignant lymphocytes. *Curr. Opin. Immunol.* **2013**, *25*, 339-346.

(27)   Huang, C.; Hatzi, K.; Melnick, A. Lineage-specific functions of Bcl-6 in immunity and inflammation are mediated by distinct biochemical mechanisms. *Nature Immunol.* **2013**, *14*, 380-388.

(28)   Yasui, T.; Yamamoto, T.; Sakai, N.; Asano, K.; Takai, T.; Yoshitomi, Y.; Davis, M.; Takagi, T.; Sakamoto, K.; Sogabe, S.; Kamada, Y.; Lane, W.; Snell, G.; Iwata, M.; Goto, M.; Inooka, H.; Sakamoto, J. I.; Nakada, Y.; Imaeda, Y. Discovery of a novel B-cell lymphoma 6 (BCL6)-corepressor interaction inhibitor by utilizing structure-based drug design. *Bioorg. Med. Chem.* **2017**, *25*, 4876-4886.

(29)   Kamada, Y.; Sakai, N.; Sogabe, S.; Ida, K.; Oki, H.; Sakamoto, K.; Lane, W.; Snell, G.; Iida, M.; Imaeda, Y.; Sakamoto, J.; Matsui, J. Discovery of a B-cell lymphoma 6 protein-protein interaction inhibitor by a biophysics-driven fragment-based approach. *J. Med. Chem.* **2017**, *60*, 4358-4368.

(30)   McCoull, W.; Abrams, R. D.; Anderson, E.; Blades, K.; Barton, P.; Box, M.; Burgess, J.; Byth, K.; Cao, Q.; Chuaqui, C.; Carbajo, R. J.; Cheung, T.; Code, E.; Ferguson, A. D.; Fillery, S.; Fuller, N. O.; Gangl, E.; Gao, N.; Grist, M.; Hargreaves, D.; Howard, M. R.; Hu, J.; Kemmitt, P. D.; Nelson, J. E.; O'Connell, N.; Prince, D. B.; Raubo, P.; Rawlins, P. B.; Robb, G. R.; Shi, J.; Waring, M. J.; Whittaker, D.; Wylot, M.; Zhu, X. Discovery of pyrazolo[1,5-a]pyrimidine B-cell lymphoma 6 (BCL6) binders and optimization to high affinity macrocyclic inhibitors. *J. Med. Chem.* **2017**, *60*, 4386-4402.

(31)   Kerres, N.; Steurer, S.; Schlager, S.; Bader, G.; Berger, H.; Caligiuri, M.; Dank, C.; Engen, J. R.; Ettmayer, P.; Fischerauer, B.; Flotzinger, G.; Gerlach, D.; Gerstberger, T.; Gmaschitz, T.; Greb, P.; Han, B.; Heyes, E.; Iacob, R. E.; Kessler, D.; Kolle, H.; Lamarre, L.; Lancia, D. R.; Lucas, S.; Mayer, M.; Mayr, K.; Mischerikow, N.; Muck, K.; Peinsipp, C.; Petermann, O.; Reiser, U.; Rudolph, D.; Rumpel, K.; Salomon, C.; Scharn, D.; Schnitzer, R.; Schrenk, A.; Schweifer, N.; Thompson, D.; Traxler, E.; Varecka, R.; Voss, T.; Weiss-Puxbaum, A.; Winkler, S.; Zheng, X.; Zoephel, A.; Kraut, N.; McConnell, D.; Pearson, M.; Koegl, M. Chemically induced degradation of the oncogenic transcription factor BCL6. *Cell Rep.* **2017**, *20*, 2860-2875.

(32)   Cardenas, M. G.; Yu, W.; Beguelin, W.; Teater, M. R.; Geng, H.; Goldstein, R. L.; Oswald, E.; Hatzi, K.; Yang, S. N.; Cohen, J.; Shaknovich, R.; Vanommeslaeghe, K.; Cheng, H.; Liang, D.; Cho, H. J.; Abbott, J.; Tam, W.; Du, W.; Leonard, J. P.; Elemento, O.; Cerchietti, L.; Cierpicki, T.; Xue, F.; MacKerell, A. D., Jr.; Melnick, A. M. Rationally designed BCL6 inhibitors target activated B cell diffuse large B cell lymphoma. *J. Clin. Invest.* **2016**, *126*, 3351-3362.

(33)   Baell, J. B.; Holloway, G. A. New substructure filters for removal of pan assay interference compounds (PAINS) from screening libraries and for their exclusion in bioassays. *J. Med. Chem.* **2010**, *53*, 2719-2740.

(34)   Mendgen, T.; Steuer, C.; Klein, C. D. Privileged scaffolds or promiscuous binders: a comparative study on rhodanines and related heterocycles in medicinal chemistry. *J. Med. Chem.* **2012**, *55*, 743-753.

(35)   Erlanson, D. A.; Fesik, S. W.; Hubbard, R. E.; Jahnke, W.; Jhoti, H. Twenty years on: the impact of fragments on drug discovery. *Nature Res. Drug Discov.* **2016**, *15*, 605-619.

(36)   Cheng, H.,; Linhares, B. M.; Yu, W.; Cardenas, M. G.; Ai, Y.; Jiang, W.; Winkler, A.; Cohen, S.; Melnick, A.; MacKerell, Jr., A.; Cierpicki, T.; Xue, F. Identification of Thiourea-Based Inhibitors of B-Cell Lymphoma 6 BTB Domain via NMR-Based Fragment Screening and Computer-Aided Drug Design. J. Med. Chem. **2018**, *61*, 7573-7588.

(37)   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.* **2003**, *327*, 1121-1133.

(38)   Otwinowski, Z.; Minor, W. [20] Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **1997**, *276*, 307-326.

(39)   Murshudov, G. N.; Vagin, A. A.; Dodson, E. J. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D Biol. Crystallogr.* **1997**, *53*, 240-255.

(40)   Emsley, P.; Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* **2004**, *60*, 2126-2132.

(41)   The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D Biol. Crystallogr.* **1994**, *50*, 760-763.

(42)   Adams, P. D.; Afonine, P. V.; Bunkoczi, G.; Chen, V. B.; Davis, I. W.; Echols, N.; Headd, J. J.; Hung, L. W.; Kapral, G. J.; Grosse-Kunstleve, R. W.; McCoy, A. J.; Moriarty, N. W.; Oeffner, R.; Read, R. J.; Richardson, D. C.; Richardson, J. S.; Terwilliger, T. C.; Zwart, P. H. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* **2010**, *66*, 213-221.

(43)   Davis, I. W.; Leaver-Fay, A.; Chen, V. B.; Block, J. N.; Kapral, G. J.; Wang, X.; Murray, L. W.; Arendall, W. B., 3rd; Snoeyink, J.; Richardson, J. S.; Richardson, D. C. MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res.* **2007**, *35*, W375-383.

(44)   Yang, H.; Guranovic, V.; Dutta, S.; Feng, Z.; Berman, H. M.; Westbrook, J. D. Automated and accurate deposition of structures solved by X-ray diffraction to the Protein Data Bank. *Acta Crystallogr. D Biol. Crystallogr.* **2004**, *60*, 1833-1839.

Chapter 5

Conclusions and Future Directions

A. Conclusions

A.1. Motivation

Numerous biological systems and biochemical pathways rely on protein-protein interactions (PPIs) and multi-protein complexes to govern signal transduction, post-translational modifications, cell metabolism, protein synthesis and turnover, and transcriptional regulation, among other fundamental biological functions[1,2]. In particular, chromatin remodeling and transcriptional control are governed in part by epigenetic mechanisms, notably histone lysine modifications, including acetylation, methylation, and crotonylation[3,4,5,6]. These modifications are regulated by epigenetic proteins, classified as either reader, writer, or eraser proteins[7]. Altogether, epigenetic mechanisms comprise an integral component of transcriptional regulation mechanisms[8], and moreover are deregulated in a number of human cancers, particularly hematologic malignancies[9,10]. As epigenetic systems are often regulated via PPIs and multi-protein complexes, PPIs may therefore figure prominently as pharmacological targets. Yet, these binding interfaces tend to be challenging to block with small molecules, due to a number of biophysical, physiochemical, and topological properties at PPI interfaces[11,12].

A.2. Dissertation summary and discussion

Here, we present biophysics- and structural biology-based approaches to target protein-protein interactions of epigenetic reader complexes. First, we characterize the biophysical and structural basis of Histone H3 acetyl-lysine recognition by GAS41 via its YEATS domain (discussed in Chapter 2, adapted from Cho, *et al*., 2018[13]). In brief, we present a model of bivalent recognition of di-acetylated Histone H3 by GAS41. Remarkably, we demonstrate that bivalent recognition of di-acetylated Histone H3 by two GAS41 molecules confers an improvement in potency by an order of magnitude over mono-acetylated Histone H3 recognition. In addition, we reveal the molecular details of acetyl-lysine binding the GAS41 YEATS domain.

We subsequently incorporated this information into design and development of GAS41 inhibitors, specifically targeting its YEATS domain; as of this writing, there are no reported inhibitors of GAS41. We used a fragment-based approach to identify a ligand for GAS41 YEATS. To grow our fragment hit into potent small-molecule inhibitors, we undertook an extensive investment in synthetic medicinal chemistry. Design of compounds was guided by the application of protein-NMR, x-ray crystallography, biophysical studies, and biochemical assay development.
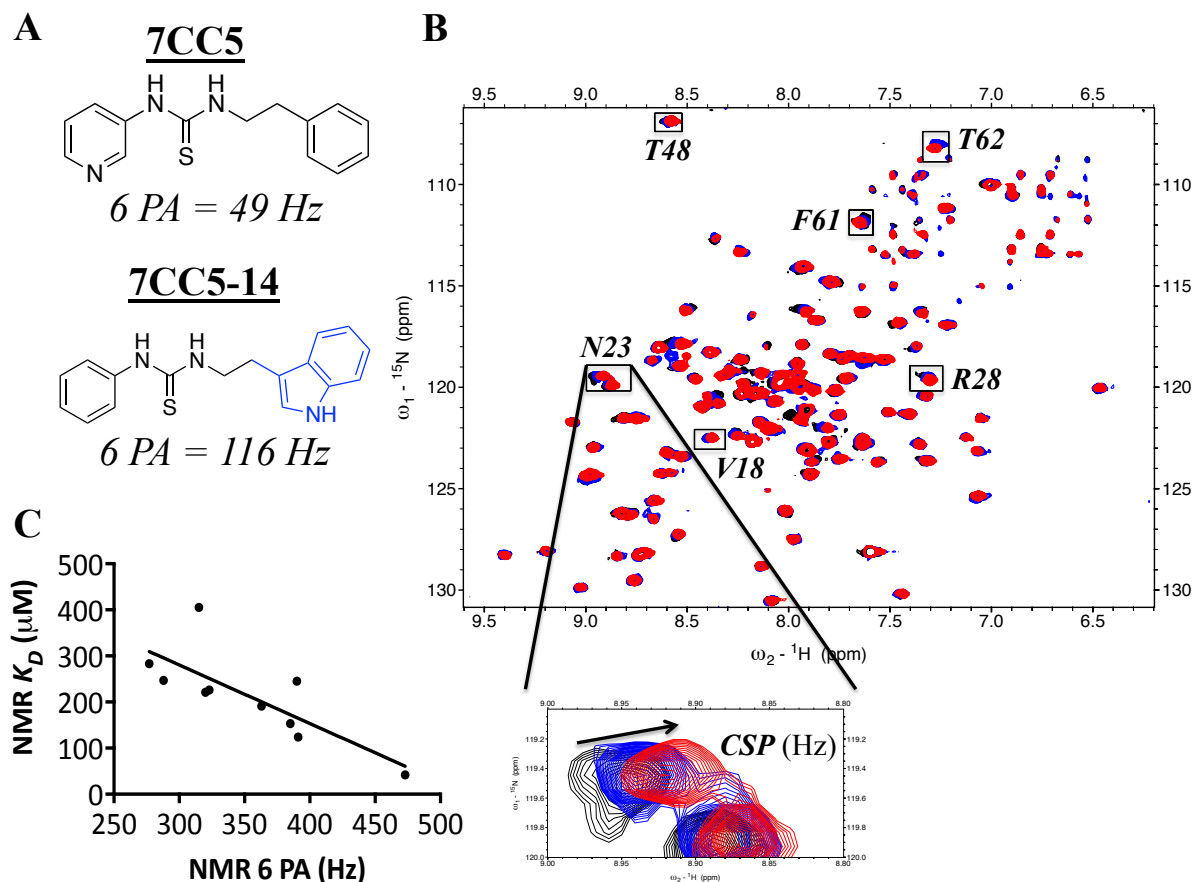
Figure 5.1. Method to rank compound binding by summation of Chemical Shift Perturbations (CSPs) of resonances on ^{1}H-^{15}N HSQC spectrum in presence of compound. (A) Chemical structures of fragment hit **7CC5** and analogue **7CC5-14**. *6 PA* value is sum of Chemical Shift Perturbations (CSPs) of six most perturbed amide resonances on ^{1}H-^{15}N HSQC spectra of 150 μM BCL6^{BTB} in presence of 500 μM compound. (B) ^{1}H-^{15}N HSQC spectra of 150 μM BCL6^{BTB} (black), overlaid with spectra in presence of 500 μM **7CC5** (blue) and 500 μM **7CC5-14** (red). Assigned resonances indicated. Inset shows calculation of CSP of resonance assigned to Asn23 of BCL6^{BTB} as representative perturbation for *6 PA* value. (C) Correlation of 6PA and K_D values determined from NMR experiments. Figure panels adapted from Cheng and Linhares, *et al.*, 2018[14].

Altogether, our campaign to block acetylated-Histone H3 recognition by GAS41

highlights the tremendous value of studies in basic biology to characterize the molecular

details of PPIs prior to undertaking inhibitor discovery programs. Biophysics- and structural

biology-guided approaches suggested interactions to target in GAS41-inhibitor complex

formation, which would prove tremendously valuable to improving our compounds'

potencies. This informed rational design of GAS41 inhibitors, in which a fragment hit with high-micromolar activity was improved to a class of small-molecule inhibitors with low-nanomolar potency *in vitro*.

Finally, we report a fragment-based approach to develop inhibitors blocking BCL6-co-repressor complex formation. We identified a small-molecule hit by protein-NMR-based fragment screening, and resolved the binding mode to atomic resolution by determination of the crystal structure of BCL6[BTB]-hit complex (discussed in Chapter 4, adapted from Cheng and Linhares, *et al.*, 2018[14]). Structure determination led the design of inhibitors to target a hydrophobic pocket that is occupied in BCL6-co-repressor complex. To improve binding of fragment analogues in early stages of the project, we developed protein-NMR-based methodology to rank compounds' binding by summation of chemical shift perturbations of selected amide resonances on ^{1}H-^{15}N HSQC spectra in presence of compounds. We improved potency of compounds by two orders of magnitude, and high-resolution crystal structures of most potent inhibitors in complex with BCL6[BTB] provide insight into the structural basis of inhibition of BCL6-co-repressor interaction.

First, this study emphasizes the importance of protein-fragment hit structure determination to high-resolution in FBLD campaigns. Fragments tend to bind with weak potencies, yet nonetheless may be suitable candidates for further development into potent inhibitors. This is attributed to fragment binding mode, synthetic tractability, and identification of growth vectors[15]. The crystal structure of protein target-fragment hit complex would therefore be advantageous in assessing the fragment hit as a starting point for an investment in synthetic medicinal chemistry, and in design of inhibitors. Second, this study provides a method to rank weakly binding fragment hits in early stages of FBLD

projects (**Figure 5.1.**, discussed more extensively later in this chapter). Given compounds'

weak affinities (high-micromolar to low-millimolar K_Ds), often limited solubility, and the

possibility of intrinsic fluorescence at high concentrations, comprehensive biochemical and

biophysical characterization may not be feasible. In the context of early-stage FBLD

projects, in parallel with x-ray crystallography, this strategy can be utilized to identify

modifications that either improve or abolish binding, to generate more potent compounds.

B. Future Directions

**B.1. *In cell* studies to characterize GAS41 reader function of di-acetylated
Histone H3 under physiological conditions**

Our model promoting bivalent mode of recognition of di-acetylated Histone H3 by

GAS41 begets the question of the physiological significance of GAS41 as a reader of di-

acetylated Histone H3, versus that of GAS41 as a reader of mono-acetylated Histone H3. A

recent report by Hsu, *et al*., 2018, demonstrated GAS41 is co-localized to promoters enriched

with H3K14ac and H3K27ac[16]. This is consistent with prior work that has established an

association between transcriptional activation and enrichment of H3K14ac, H3K18ac, and

H3K27ac[17]. Further studies to determine if GAS41 co-localizes at promoters simultaneously

enriched with H3K18ac and H3K27ac, to recapitulate *in vivo* the model of bivalent

recognition of di-acetylated Histone H3 recognition by GAS41, may thereby suggest the role

of GAS41 in the regulation of transcription as a reader of di-acetylated Histone H3.

B.2. Structure-based design to optimize monomeric inhibitors of GAS41

For small-molecule inhibitors that can compete off di-acetylated Histone H3 *in vivo*,

we aim to develop more potent monomeric inhibitors of GAS41. The co-crystal structure of

GAS41 YEATS-**DLG-157** complex shows **DLG-157** occupies the acetyl-lysine binding

pocket of GAS41 YEATS. Furthermore, an intricate network of hydrogen bonds, hydrophobic contacts (in the form of edge-to-face pi-pi stacking interactions), and shape complementarity drive complex formation. Structure-guided design efforts will be pursued to improve potency of DLG-series inhibitors, to better occupy this site and make novel contacts with GAS41.

B.3. Demonstrate on-target activity by DLG-series inhibitors in functional assays in cells

To move forward with inhibitors as lead candidates, we need to demonstrate on-target activity and anti-proliferative effects in cells. GAS41 is a pharmacological target for therapeutic intervention against NSCLCs, and moreover the YEATS domain specifically is required for oncogenic transformation[18]. We are currently pursuing *in cell* studies to assess on-target efficacy and anti-proliferative effects, and to elucidate mechanism-of-action of small-molecule inhibitors in GAS41-dependent and in GAS41-knockdown NSCLC cell lines. This work is ongoing both in our lab, and in collaboration in the lab of Venkat Keshamouni, in the Department of Internal Medicine. To facilitate *in cell* studies, we have undertaken structure-guided design approach to optimize compounds with improved metabolic stability and cell permeability, among other properties, while maintaining compound potency.

B.4. NMR-based methodology to rank fragment binding in early stages of FBLD campaigns

Here we present a technique to characterize and subsequently rank binding of compounds with weak affinities, in the high-micromolar to low-millimolar K_D range, by summation of chemical shift perturbations of selected amide resonances on ^{1}H-^{15}N HSQC spectra in the presences of ligand (**Figure 5.1.** and reported previously[14]). We contend that

this is a viable method to guide an investment in synthetic medicinal chemistry in early stages of fragment-based lead discovery projects, against challenging targets that may be intractable to other binding assays, or in cases in which fragments' scaffolds present synthetic tractability, demonstrable growth vectors, and/or interesting binding modes that warrant development, yet may not be pursued due to weak binding that would hamper compound optimization. As the notion of "druggability" is being questioned, and more challenging targets are being pursued in academic and industrial research labs, we would kindly recommend research groups to utilize this method.

B.5. Validation of pharmacological inhibition of BCL6-co-repressor interactions as tractable means of therapeutic intervention against BCL6-dependent DLBCLs

As discussed in Chapter 3 (adapted from Cheng and Linhares, *et al*., 2018[14]), fragment-based approach guided by biophysics and structural biology methodologies yielded insights into the structural and molecular bases of inhibition of BCL6-co-repressor complex formation. Lead compound presented mid-micromolar potency that, in tandem with likely poor compound permeability, limited *in cell* studies of on-target activity and anti-proliferative effects in BCL6-dependent DLBCL cell lines. Nonetheless, as our compounds demonstrate a novel scaffold to block BCL6-co-repressor interaction, our study begets questions as to where this finding stands within the context of the BCL6-inhibitor development field.

A peptidomimetic approach to inhibit BCL6 demonstrated on-target activity *in vivo* and anti-proliferative effects in BCL6-dependent DLBCL cell lines[19]. Early programs to develop small molecule inhibitors did not prove fruitful: the first small-molecule inhibitor, 79-6, demonstrated mid- to high-micromolar affinity for BCL6^{BTB} and possessed a rhodanine moiety, renown for its promiscuity[20]. (For this reason, rhodanine moieties tend to be avoided

in drug discovery[21].) On the basis of this previous generation of inhibitors, small-molecule FX-1 was identified, with low-micromolar potency for BCL6^{BTB} *in vitro,* yet modest *in cell* and *in vivo* activities, as well as the conserving the rhodanine scaffold, precluded further development[22]. In more recent years, industrial groups at Takeda and Astra-Zeneca identified novel inhibitor scaffolds by HTS approaches that were improved to sub-micromolar potencies *in vitro*, however, compounds have yet to display strong on-target activity or efficacy by *in cell* and *in vivo* studies[23,24]. Beyond design of reversible small-molecule inhibitors, covalent inhibitors[25] and small-molecule degraders[26] of BCL6 have demonstrated superior efficacy in cells as compared to non-covalent inhibitors, yet do not present strong on-target activity and anti-proliferative effects, and additionally possess poor bioavailability that has limited *in vivo* study. As of this writing, no compounds targeting BCL6 have entered the clinic. At this juncture, inhibitors of any classification, non-covalent, covalent, or degraders, have repeatedly failed to show the efficacy by *in cell* and *in vivo* studies that would be expected of pre-clinical lead compounds. This may as well be a function of BCL6's vaunted status as a target for therapeutic intervention in treatment of DLBCLs, as well as may be attributed to the compounds themselves. Further study is warranted to ascertain if pharmacological inhibition of BCL6 constitutes a viable approach to target BCL6-dependent DLBCLs.

C. References

1   Gavin, A. C. *et al.* Functional organization of the yeast proteome by systematic analysis of protein complexes. *Nature* **415**, 141-147, doi:10.1038/415141a (2002).
2   Stelzl, U. *et al.* A human protein-protein interaction network: a resource for annotating the proteome. *Cell* **122**, 957-968, doi:10.1016/j.cell.2005.08.029 (2005).
3   Bannister, A. J. & Kouzarides, T. Regulation of chromatin by histone modifications. *Cell Res* **21**, 381-395, doi:10.1038/cr.2011.22 (2011).
4   Javaid, N. & Choi, S. Acetylation- and Methylation-Related Epigenetic Proteins in the Context of Their Targets. *Genes (Basel)* **8**, doi:10.3390/genes8080196 (2017).
5   Zhang, Q. *et al.* Structural Insights into Histone Crotonyl-Lysine Recognition by the AF9 YEATS Domain. *Structure* **24**, 1606-1612, doi:10.1016/j.str.2016.05.023 (2016).

6   Zhao, D. *et al.* YEATS2 is a selective histone crotonylation reader. *Cell Res* **26**, 629-632, doi:10.1038/cr.2016.49 (2016).

7   Allis, C. D. & Jenuwein, T. The molecular hallmarks of epigenetic control. *Nat Rev Genet* **17**, 487-500, doi:10.1038/nrg.2016.59 (2016).

8   Diaz-Eufracio, B. I., Naveja, J. J. & Medina-Franco, J. L. Protein-Protein Interaction Modulators for Epigenetic Therapies. *Adv Protein Chem Struct Biol* **110**, 65-84, doi:10.1016/bs.apcsb.2017.06.002 (2018).

9   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, doi:10.1111/imr.12244 (2015).

10   Kanwal, R. & Gupta, S. Epigenetic modifications in cancer. *Clin Genet* **81**, 303-311, doi:10.1111/j.1399-0004.2011.01809.x (2012).

11   Petta, I., Lievens, S., Libert, C., Tavernier, J. & De Bosscher, K. Modulation of Protein-Protein Interactions for the Development of Novel Therapeutics. *Mol Ther* **24**, 707-718, doi:10.1038/mt.2015.214 (2016).

12   Thompson, A. D., Dugan, A., Gestwicki, J. E. & Mapp, A. K. Fine-tuning multiprotein complexes using small molecules. *ACS Chem Biol* **7**, 1311-1320, doi:10.1021/cb300255p (2012).

13   Cho, H. J. *et al.* GAS41 Recognizes Diacetylated Histone H3 through a Bivalent Binding Mode. *ACS Chem Biol* **13**, 2739-2746, doi:10.1021/acschembio.8b00674 (2018).

14   Cheng, H. *et al.* Identification of Thiourea-Based Inhibitors of the B-Cell Lymphoma 6 BTB Domain via NMR-Based Fragment Screening and Computer-Aided Drug Design. *J Med Chem* **61**, 7573-7588, doi:10.1021/acs.jmedchem.8b00040 (2018).

15   Erlanson, D. A., Fesik, S. W., Hubbard, R. E., Jahnke, W. & Jhoti, H. Twenty years on: the impact of fragments on drug discovery. *Nat Rev Drug Discov* **15**, 605-619, doi:10.1038/nrd.2016.109 (2016).

16   Hsu, C. C. *et al.* Recognition of histone acetylation by the GAS41 YEATS domain promotes H2A.Z deposition in non-small cell lung cancer. *Genes Dev* **32**, 58-69, doi:10.1101/gad.303784.117 (2018).

17   Wang, Z. *et al.* Combinatorial patterns of histone acetylations and methylations in the human genome. *Nat Genet* **40**, 897-903, doi:10.1038/ng.154 (2008).

18   Hsu, C. C. *et al.* Gas41 links histone acetylation to H2A.Z deposition and maintenance of embryonic stem cell identity. *Cell Discov* **4**, 28, doi:10.1038/s41421-018-0027-0 (2018).

19   Polo, J. M. *et al.* Specific peptide interference reveals BCL6 transcriptional and oncogenic mechanisms in B-cell lymphoma cells. *Nat Med* **10**, 1329-1335, doi:10.1038/nm1134 (2004).

20   Cerchietti, L. C. *et al.* A small-molecule inhibitor of BCL6 kills DLBCL cells in vitro and in vivo. *Cancer Cell* **17**, 400-411, doi:10.1016/j.ccr.2009.12.050 (2010).

21   Mendgen, T., Steuer, C. & Klein, C. D. Privileged scaffolds or promiscuous binders: a comparative study on rhodanines and related heterocycles in medicinal chemistry. *J Med Chem* **55**, 743-753, doi:10.1021/jm201243p (2012).

22   Cardenas, M. G. *et al.* Rationally designed BCL6 inhibitors target activated B cell diffuse large B cell lymphoma. *J Clin Invest* **126**, 3351-3362, doi:10.1172/JCI85795 (2016).

23 McCoull, W. *et al.* Discovery of Pyrazolo[1,5-a]pyrimidine B-Cell Lymphoma 6 (BCL6) Binders and Optimization to High Affinity Macrocyclic Inhibitors. *J Med Chem* **60**, 4386-4402, doi:10.1021/acs.jmedchem.7b00359 (2017).

24 Yasui, T. *et al.* Discovery of a novel B-cell lymphoma 6 (BCL6)-corepressor interaction inhibitor by utilizing structure-based drug design. *Bioorg Med Chem* **25**, 4876-4886, doi:10.1016/j.bmc.2017.07.037 (2017).

25 Sameshima, T. *et al.* Discovery of an Irreversible and Cell-Active BCL6 Inhibitor Selectively Targeting Cys53 Located at the Protein-Protein Interaction Interface. *Biochemistry* **57**, 1369-1379, doi:10.1021/acs.biochem.7b00732 (2018).

26 Kerres, N. *et al.* Chemically Induced Degradation of the Oncogenic Transcription Factor BCL6. *Cell Rep* **20**, 2860-2875, doi:10.1016/j.celrep.2017.08.081 (2017).

Structure-Guided Optimization of Potent Small-Molecule and Peptidomimetic

Inhibitors of Menin-Mixed Lineage Leukemia 1 Protein-Protein Interaction

*The text and data presented here are adapted from the following sources:

Borkin, D., Klossowski, S., Pollock, J., Miao, H., Linhares, B. M., Kempinska, K., Jin, Z., Purohit, T., Wen, B., He, M., Sun, D., Cierpicki, T., & Grembecka, J. "Complexity of Blocking Bivalent Protein-Protein Interactions: Development of a Highly Potent Inhibitor of the Menin-Mixed-Lineage Leukemia Interaction". *J Med Chem*. **61**, 4832-50 (2018).

Fortuna, P., Linhares, B. M., Purohit, T., Pollock, J., Cierpicki, T., Grembecka, J., & Berlicki, L. "Covalent and noncovalent constraints yield a figure eight-like conformation of a peptide inhibiting the menin-MLL interaction". *Manuscript under review*.

A. Abstract

Chromosomal rearrangement at the *MLL* gene has been implicated in a number of human leukemias. Menin is an established oncogenic co-factor of MLL, and moreover targeting the Menin-MLL protein-protein interaction has demonstrated on-target activity in *in cell* functional assays, and anti-cancer effects *in vivo*. Progress towards potent, selective, and

metabolically stable small-molecule inhibitors with sufficient pharmacokinetic profiles has revealed tremendous insight into the molecular details of the structural basis of inhibition of the Menin-MLL interaction. Here we present a two-pronged approach, to design and develop both peptidomimetic and small-molecule inhibitors targeting Menin. On the bases of extensive biophysical and structural studies to characterize the molecular recognition of MLL by Menin, we report a peptidomimetic inhibitor of Menin, **25**, that mimics the binding modes of several crucial residues of MLL in complex with Menin. Furthermore, we explored modifications on **25** that make novel contacts with Menin, distinct from those made in Menin-MLL complex formation, which provide insight into the development of potent inhibitors of the Menin-MLL interaction. By *in vitro* biochemical assay, **25** presents low-nanomolar activity in inhibiting Menin-MLL complex formation.

In parallel, using the high-resolution crystal structure of small-molecule inhibitor, **MI-503**, in complex with Menin, we optimized the thienopyrimidine class of small-molecule inhibitors to generate compound **MI-1481** with improved activity by *in vitro* biochemical assays, *in cell* functional assays, and *in vivo* efficacy. **MI-1481** enantiomer **MI-1482**, however, did not demonstrate the same effects *in vitro* or in cells. To explain such differences at the molecular level, we solved the crystal structures of Menin in complex with **MI-1481** and **MI-1482**, respectively, to high-resolution, from which we were able to identify both polar interactions and hydrophobic contacts with Menin unique to **MI-1481** complex. As seen in high-resolution structure determination of **25** in complex with Menin, the molecular details of recognition of **MI-1481** and **MI-1482** by Menin demonstrated contacts that we had not witnessed previously in Menin complex with MLL, or with previous generations of inhibitors. This may elucidate a rationale for targeting specific sites on Menin previously unoccupied

with small molecules in inhibitor development campaigns. Moreover, SAR studies of peptidomimietic **25** and of **MI-1481** and **MI-1482** reveal insight into previously unexplored chemical modifications that may be advantageous in development of potent inhibitors. In summation, we highlight biophysics- and structural biology-based approaches to design and optimize potent peptidomimetic and small-molecule inhibitors of the Menin-MLL interaction, which facilitate a greater understanding of the structural and mechanistic bases of inhibition of Menin.

B. Introduction

B.1. Menin-MLL protein-protein interaction is a pharmacological target

The Menin-Mixed Lineage Leukemia 1 (referred to herein as MLL) protein-protein interaction constitutes a major pharmacological target for therapeutic intervention against human acute myeloid leukemias (AML) and lymphoid leukemias (ALL)[1,2]. Notably, the *MLL* gene is translocated in approximately 5 – 10% of acute leukemias in adults[3], and 70 – 80% of acute leukemias in infants[4,5]. *MLL* translocations yield MLL fusion proteins[6], which interact with Menin to drive MLL-associated leukemogenesis[7]. *MLL* rearrangements have been demonstrated to affect transcriptional programming and epigenetic regulation, particularly expression of *HOX* genes and HOX cofactor *MEIS1*, with profound implications in oncogenic transformation[8,9,10]. Clinical outcomes for infant ALL with *MLL* rearrangement are poor: the five-year event-free survival (EFS) rates historically have varied from 20 – 40%, in contrast to those with wild-type *MLL*, which have been over 60%[11]. Regarding clinical outcomes in adults, one study of 85 patients with t(4,11) rearrangements (which yield MLL-AF4 fusion protein) reported an EFS rate of 34%, which is lower than the EFS rates of 40 – 45% overall in adults presenting ALL of any and all sub-types[11].

B.2. Structural and mechanistic bases of Menin-MLL interaction guide inhibitor development

The Menin-MLL interaction comprises an essential part of the larger MLL SET1-like histone methyltransferase complex, which governs expression of *HOX* family and *MEIS1* genes[7,10,12,13]. Mechanistic and biophysical characterization demonstrated the N-terminal MLL fragment (MLL[1-46]) possesses low-nanomolar affinity (K_D = 6.8 nM) for Menin, and mapped MLL Menin Binding Motif 1 (MBM1) and Menin Binding Motif 2 (MBM2) to MLL[4-15] and MLL[23-40], respectively[14]. Menin-MLL[4-43] competition experiments by Fluorescence Polarization (FP) assay demonstrated MBM1 is more potent in competing off MLL, thereby promoting the MBM1 as a site to target in small-molecule inhibitor development[14].

Table A.1. Binding affinities and inhibitory activities of various MLL fragments for Menin and for inhibition of Menin-MLL complex formation, respectively. K_D values present affinity for Menin, by Isothermal Titration Calorimetry experiments or by Fluorescence Polarization assay[14]. Fluorescence Polarization assay, by fluorescein-labeled MLL[4-15] or MLL[4-43] peptides, was used to quantify IC$_{50}$ values[14]. For both K_Ds and IC$_{50}$s, average values from two or three independent experiments ± SD are presented.

Peptide	K_D (nM) (Method)	IC$_{50}$ (MLL[4-15]) (nM)	IC$_{50}$ (MLL[4-43]) (nM)
MLL[160]	9.8 ± 2.5 (ITC)	N.d.	N.d.
MLL[46]	6.8 ± 1.7 (ITC)	N.d.	N.d.
MLL[4-15] (MBM1)	72 ± 22 (ITC) 53 ± 4.2 (FP)	230 ± 16	490 ± 70
MLL[23-40] (MBM2)	1,400 ± 424 (FP)	5,800 ± 200	37,000 ± 8500
MLL[4-15] F9A	~80,000 (FP)	~500,000	N.d.
MLL[4-15] P10A	N.d.	6,700 ± 1,600	N.d.
MLL[4-15] P13A	N.d.	11,400 ± 1,200	N.d.

Alanine scan of MLL MBM1 by FP competition assay revealed that Phe9, Pro10, and Pro13 are necessary to the high-affinity Menin-MLL interaction, as MLL[4-15] mutants F9A, P10A, and P13A present approximate 2,300-, 30- and 50-fold losses in IC_{50}s for blocking Menin-MLL[4-15] complex[14]. Affinity determination showed MLL[4-15] F9A binds Menin with ~80 μM K_D, over 1,000-fold weaker than wild-type MLL[4-15] (53 nM, both values determined by FP assay)[14]. Structural studies revealed that MLL MBM1 binds a negatively-charged cavity of Menin that is centrally located, and moreover that Phe9, Pro10, and Pro13 occupy hydrophobic pockets at this site (**Figure A.2.** and **A.3.**)[15,16]. In total, the high-affinity of MBM1, as well as the aforementioned crucial residues in Menin-MLL complex formation, suggest that mimicry of Phe9, Pro10, or Pro13 binding modes by peptidomimetics or small molecules may be requisite in the development of potent inhibitors[17].

B.3. Previous campaigns to target Menin-MLL interaction

A 2012 study by Grembecka, *et al*., reported the first small-molecule inhibitors of the Menin-MLL interaction[18]. Discovered by high-throughput screening (HTS), hit compound **MI-1** was comprised of a thienopyrimidine ring (thereby conferring the name to the thienopyrimidine class), with dimethyl-substituted thiazole connected by piperazine linker (**Figure A.1.**)[18]. Substitution of the saturated cyclohexane moiety at the thienopyrimidine with a propyl group yielded **MI-2**, which demonstrated ~4.5-fold improvement in IC_{50} in inhibiting Menin-MLL[MBM1] complex formation *in vitro* (446 nM versus 1.9 μM)[18]. **MI-2** inhibited oncogenic transformation, affected expression of target genes and induced hematopoietic differentiation in MLL-fusion (MLL-AF9) transformed bone marrow cells, and inhibited proliferation and induced apoptosis in leukemia cells with *MLL* translocations[18].

Subsequently, structure-guided inhibitor design was employed. The crystal structure of **MI-2** in complex with Menin to high-resolution revealed an intricate hydrogen bond network made by the thienopyrimidine moiety, and occupation of the P13 pocket by dimethyl-substituted thiazole[15]. In addition, the structure presented the binding mode of the propyl moiety at the Phe9 pocket, which suggested that modifications could be explored at this site[15]. Trifluoroethyl group was substituted for *n*-propyl to optimize shape complementarity and engage in dipolar interactions between trifluoromethyl and the backbone carbonyl of His181, which generated **MI-2-2** possessing 22 nM K_D and 46 nM IC_{50} values[15]. Consistent with stronger affinity and *in vitro* activity, **MI-2-2** demonstrated superior efficacy in cells over a panel of functional assays to assess on-target activity[15].

MI-1
$IC_{50} = 1.9\ \mu M$

MI-2
$K_D = 158\ nM$
$IC_{50} = 446 \pm 28\ nM$

MI-2-2
$K_D = 22\ nM$
$IC_{50} = 46 \pm 16\ nM$
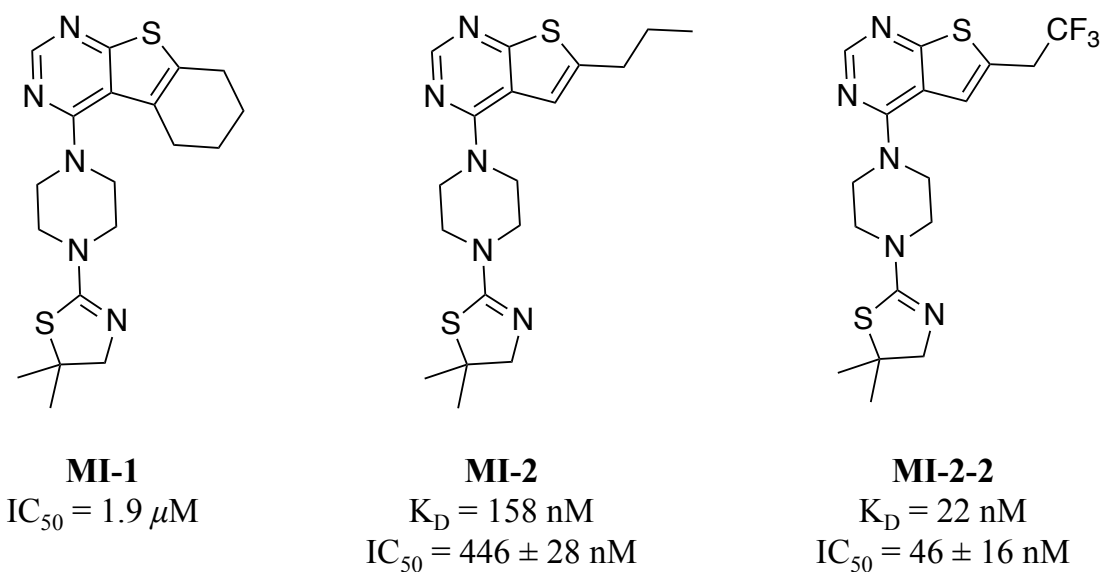
Figure A.1. Chemical structures of selected compounds from previous generation of thienopyrimidine class of Menin inhibitors. MI-1 identified by HTS[18]. K_D values derived from ITC experiment, adapted from[15,18]. IC_{50} values measured by Fluorescence Polarization (FP) Competition Assay using fluorescein-labeled MLL[MBM1] (MLL[4-15]) interaction, adapted from[15,18]. For both K_D and IC_{50}s, average values from two or three independent experiments ± SD are presented.

MI-2-2 showed neither sufficient metabolic stability nor potent cellular activity, and thus compounds appropriate for *in vivo* study would require further development[19]. Extensive investment in synthetic medicinal chemistry yielded **MI-503**, which presented 9.3 nM K_D for Menin, and 14.7 nM and 33 nM IC_{50} values in inhibiting Menin-MLL[4-15] and Menin-MLL[4-43] interactions, respectively (**Table A.4.**)[19]. **MI-503** showed on-target activity in MLL-rearranged leukemia cells, and blocked hematological tumors and progression of MLL-rearranged leukemia *in vivo*[19]. Yet, as Menin binds MLL with low-nanomolar affinities (Menin-MLL[46] K_D = 6.8 nM, and Menin-MLL[160] K_D = 9.8 nM), more potent inhibitors with low-nanomolar (<10 nM) IC_{50} values would need to be developed to compete off MLL *in vivo*[14,20].

C. Results and Discussion

C.1. Cyclization of peptidomimetic to optimize three-dimensional conformation of inhibitor scaffold in binding Menin

As previous biophysical studies had identified MLL residues Phe9, Pro10, and Pro13 as essential to the high affinity interaction with Menin[14], we therefore sought to mimic contacts in development of peptidomimetic inhibitors (**Table A.1.** and **Figure A.2.**). Using MBM1 fragment MLL[1-12], cyclization of the peptide spearheaded our design strategy (**Figure A.4.**). The introduction of covalent linkages would serve to constrain the conformation of peptidomimetic inhibitors, thereby facilitating the development of peptides with well-defined three-dimensional structures (Fortuna, *et al*., *manuscript under review*). The crystal structure of Menin-MLL[7-13] revealed an intramolecular hydrogen bond between the backbone carbonyl of Phe9 and the backbone amide of Arg12, that effectively cyclizes the MLL MBM1 binding motif (**Figure A.3.**); this interaction may stabilize the conformation of the MLL for molecular recognition by Menin[14]. We thus cyclized MLL[4-15] by substitution of an unsaturated four-

carbon linker at the gamma-carbons of the side chains of Arg8 and Arg12, to generate

peptidomimetic inhibitor **6** (**Figure A.4.**). Notably, the removal of Arg8 and Arg12 side

chains yielded an approximate 500-fold loss in activity, as peptidomimetic inhibitor **6** presents

IC_{50} of 41 μM towards Menin-MLL[4-15], compared to 84 nM for MLL[4-15]. The crystal

structure of MLL[4-15] in complex with Menin reveals that the side chain of Arg12 forms three

hydrogen bonds in total with Glu359 and Glu363 of Menin, and moreover presents shape

complementarity to occupy a negatively-charged pocket at the MLL-binding site of Menin

(**Figure A.2.** and **A.3.**). The loss in IC_{50} due to peptide cyclization is noteworthy, albeit, yet

not entirely unexpected in the early stages of peptidomimetic inhibitor development and

peptide cyclization. At such juncture in inhibitor development, optimization of linker length

and saturation would be premature; linker length and saturation can have profound affects on

cyclic peptide conformation, and therefore may abolish molecular recognition. We thus

sought to improve inhibitor potency by substitutions focused on MLL[7-13] at the Menin-MLL

interface, at which point we would return to optimization of linker length and saturation to

explore conformations of peptidomimetic inhibitors optimal for Menin recognition.
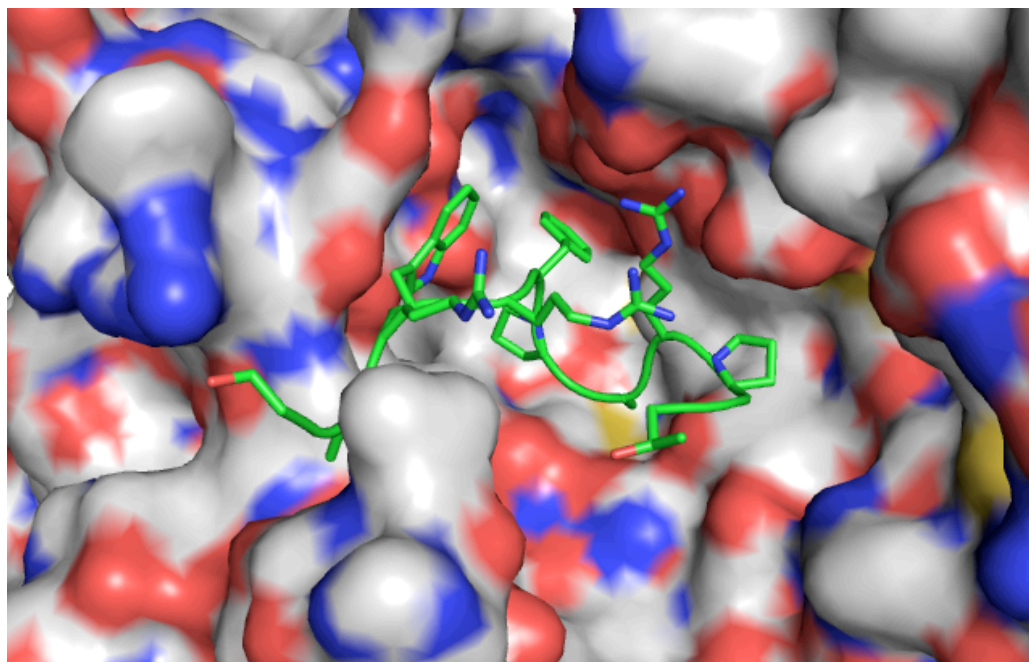
Figure A.2. Crystal structure of MLL1$^{4\text{-}15}$ in complex with Menin. Menin-MLL$^{4\text{-}15}$ (PDB code 4GQ6) backbone shown in cartoon representation, side chains shown as sticks, colored green. Menin shown as surface, in gray.

C.2. Structure-based optimization of peptidomimetic inhibitors guided by molecular details of Menin-MLL$^{4\text{-}15}$ interaction

In parallel with peptide cyclization, we aimed to increase inhibitory activity of the peptide by modifications to side chains of MLL residues 7 – 13. We first replaced C-terminal Gly-Thr-NH$_2$ with C-terminal dipeptide with Arg-NH$_2$, to generate peptidomimetic inhibitor **7**. This improved IC$_{50}$ over 50-fold, from 41 μM to 770 nM (**Table A.2.**).

Alanine-scanning mutagenesis had revealed MLL$^{4\text{-}15}$ F9A mutant conferred an over 1,000-fold loss in affinity for Menin (53 nM versus ~80 μM K_Ds for MLL$^{4\text{-}15}$ and MLL$^{4\text{-}15}$ F9A, respectively, by FP assay), and 2,300-fold loss in activity in inhibiting Menin-MLL$^{4\text{-}15}$ complex formation[14]. As Phe9 buries deep within a well-defined hydrophobic pocket of Menin, we installed a fluoro-modification at the *meta*-position on the phenylalanine side chain in **19**. This improved activity by 2.75-fold, from 770 to 280 nM (**Table A.2.**).

In addition to Phe9, Pro10 and Pro13 significantly contribute to the high affinity Menin-MLL interaction: the same study that characterized F9A mutant activity additionally showed P10A and P13A mutations yielded approximately 30- and 50-fold decreases in inhibitory activity in blocking Menin-MLL[4-15] interaction, by FP assay[14]. Pro10 presents shape complementarity for a hydrophobic pocket comprised of Phe238, A242, and C241, as does Pro13 for a pocket comprised of Tyr319, Met322, and Tyr323. Both synthetic tractability and steric considerations precluded further optimization at Pro10 and Pro13 positions, and with the significant contributions that Pro10 and Pro13 make to MLL recognition, we did not see modifications at these sites as warranted.
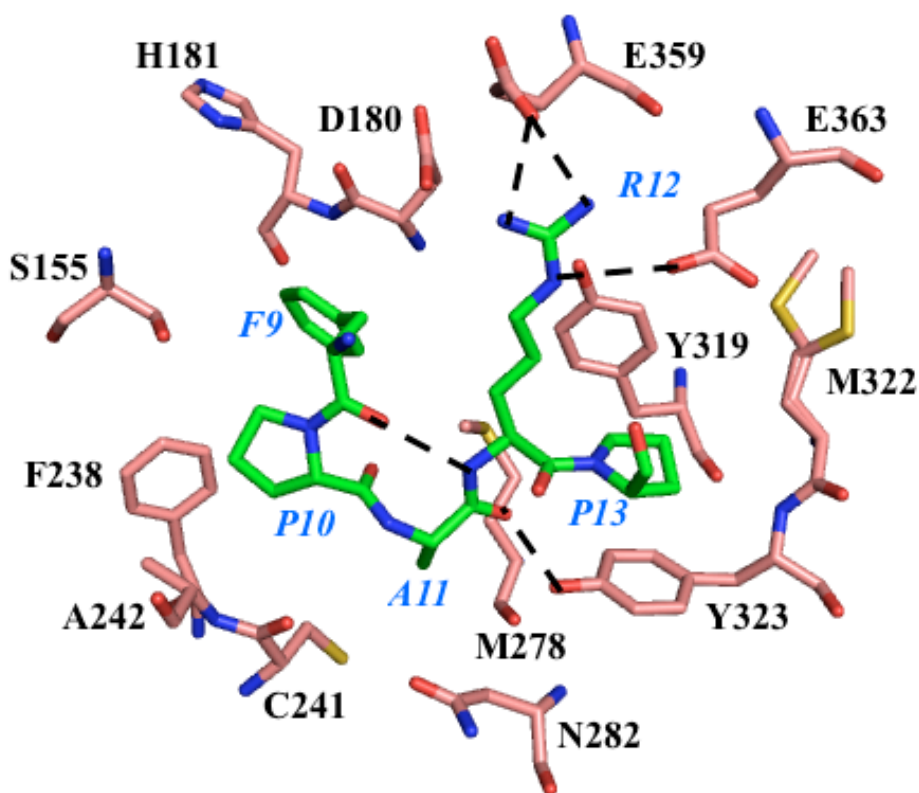

Figure A.3. Molecular details of MLL recognition by Menin. Selected interactions between Menin and MLL[9-13] shown. MLL[9-13] in stick representation, colored green. Menin

residues, stick representation, in salmon. Hydrogen bonds represented as dashed lines. PDB code 4GQ6.

In a report by Huang, *et al.*, mutagenesis studies suggested that the methyl side chain of Ala11 presents shape complementarity for a hydrophobic pocket in Menin-MLL binding, as MLL[6-25] mutant A11R binds with 5.4 μM K_D, 65-fold weaker than the 82 nM K_D of wild-type MLL[6-25] [21]. We explored modifications at Ala11 binding pocket. Inhibitor **20**, with cyclobutane substituted for the methyl side chain of Ala11, demonstrated a nine-fold improvement in IC$_{50}$: 83 over 770 nM (**Table A.2.**). In combination, fluoro-phenylalanine modification at Phe9 and cyclobutane substitution at the side chain of Ala11 yielded an improvement in IC$_{50}$ value to 64 nM for **21**, an order of magnitude improvement over previous generation inhibitor **7** (IC$_{50}$ = 770 nM) (**Table A.2.**).

MLL[4-15]
IC$_{50}$ MLL[4-15] (nM) =
84 ± 11

Peptidomimetic inhibitor **6**
IC$_{50}$ MLL[4-15] (nM) =
41,000 ± 11,300
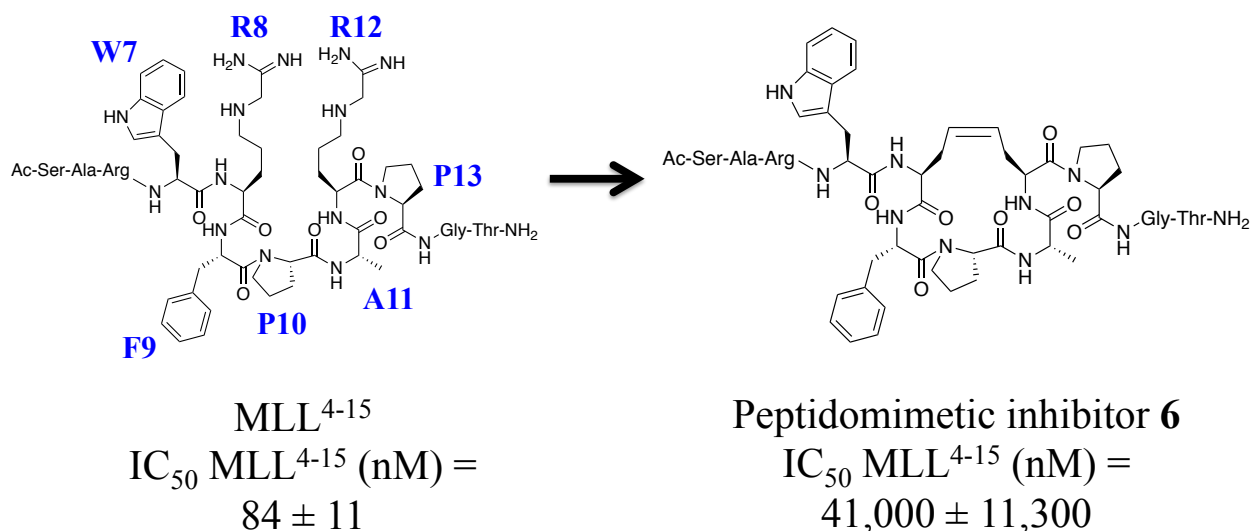
Figure A.4. MLL MBM1 fragment (MLL[4-15]) to serve as basis for design of cyclic peptidiomimetic inhibitors. Residues labeled in blue represent MLL[7-13] peptide, notably Phe9, Pro10, and Pro13, that contributes significantly to binding Menin[14]. IC$_{50}$ values determined by Fluorescence Polarization assay using fluorescein-labeled MLL MBM1 peptide (MLL[4-15]), adapted from Fortuna, *et al.*, *manuscript under review*. Average values from two or three independent experiments ± SD is presented.

C.3. Optimization of linker moiety, and of N- and C-termini, yields peptidomimetic inhibitor 25 with nanomolar potency for Menin

Subsequent to modifications at MLL[7-13], we explored substitutions at the N- and C-termini, and in the linker region for cyclization. We shortened the N-terminus by removal of Serine residue in **23**, and witnessed a modest decrease in activity (**23** IC_{50} MLL[4-43] = 111 nM, compared to **22** IC_{50} MLL[4-43] = 43 nM) (**Table A.3.**). At the C-terminus, we found that substitution of positively-charged residues improved activity, and first added non-natural amino acid Ornithine to increase activity ~5.5-fold from 770 to 142 nM (SAR not shown; this generation was prior to addition fluoro-group on Phe9 and cyclobutyl substitution at Ala11). Subsequently, the addition of an Arginine tri-peptide (Arg_3) to **23** yielded **25**, with improved activity towards Menin-MLL[4-43] nearly six-fold to 19 nM (**Table A.3.**).

Finally, we returned to the cyclization moiety, to explore the affects of saturation on binding. We found that peptidomimetic inhibitor **25** analogue **28**, with saturated carbon-carbon, showed slightly weaker IC_{50} for Menin-MLL[4-43], 27 nM (compared to **25** IC_{50} MLL[4-43] = 19 nM) (**Table A.3.**). This would suggest that the more rigid linker stabilizes peptidomimetic inhibitor **25** in a three-dimensional conformation that is advantageous to molecular recognition by Menin, compared to that of **28**. Furthermore, **25** demonstrates the most potent IC_{50} in blocking Menin-MLL[4-43]. Interestingly, the IC_{50}s of **25** and **28** towards Menin-MLL[4-15] are comparable, 5.6 and 5.5 nM, respectively (**Table A.3.**). Yet, the stronger inhibitor activity of **25** towards Menin-MLL[4-43], encompassing MLL motifs MBM1 and MBM2, led us to pursue structural studies of **25** in complex with Menin to determine the structural basis of inhibition.
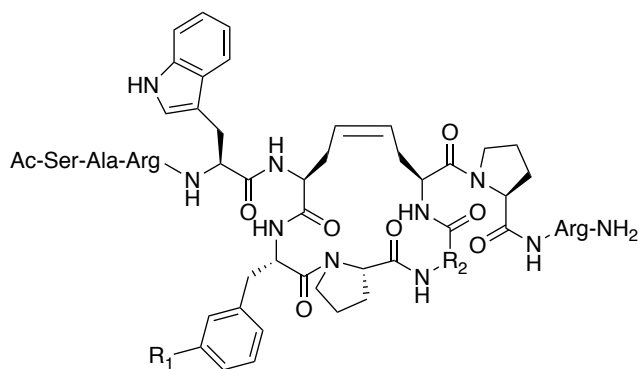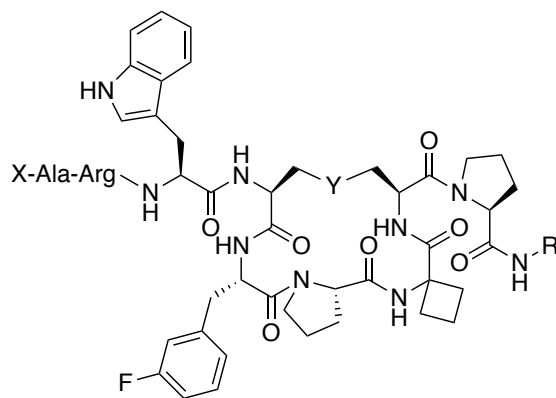
Inhibitor	R_1	R_2	IC_{50} MLL[4-15] (nM)
7	(H)	(isopropyl)	770 ± 80
19	(F)	(isopropyl)	281 ± 20
20	(H)	(cyclobutyl)	83 ± 15
21	(F)	(cyclobutyl)	64 ± 20

Table A.2. SAR for peptidomimietic inhibitor 7 derivatives with modifications at Phe9 and Ala11. IC_{50} values determined by Fluorescence Polarization assay using fluorescein-labeled MLL MBM1 peptide (MLL[4-15]), adapted from Fortuna, *et al.*, *manuscript under review*. Average values from two or three independent experiments ± SD is presented.

C.4. High-resolution structure determination reveals molecular basis of inhibition of Menin-MLL interaction by peptidomimietic inhibitor 25

To determine the molecular basis of Menin-MLL inhibition by **25**, we determined the crystal structure of Menin-**25** complex to 1.50 Å (**Table A.5.**). First, the crystal structure confirms that **25** binds at the same site as MLL[4-15]. Moreover, the cyclization strategy was demonstrated to be effective in constraining the three-dimensional conformation of the

peptide: the binding mode of the backbone of peptide **25** is largely conserved with that of

MBM1 fragment MLL$^{4\text{-}15}$ (**Figure A.5.**).

Inhibitor	X	Y	R	IC$_{50}$ MLL$^{4\text{-}15}$ (nM)	IC$_{50}$ MLL$^{4\text{-}43}$ (nM)
22	Ac-Ser-	-CH=CH-	-Orn-NH$_2$	43 ± 10	9.4 ± 0.1
23	Ac-	-CH=CH-	-Orn-NH$_2$	111 ± 1	37 ± 2
25	Ac-	-CH=CH-	-Orn-Arg$_3$-NH$_2$	19 ± 3	5.6 ± 1.0
28	Ac-	-CH$_2$-CH$_2$-	-Orn-Arg$_3$-NH$_2$	27 ± 1	5.5 ± 2.1

Table A.3. SAR for final generation of peptidomimietic inhibitors. IC$_{50}$ values determined by Fluorescence Polarization assay using fluorescein-labeled MLL MBM1 peptide (MLL$^{4\text{-}15}$), and MLL MBM1 and MBM2 (MLL$^{4\text{-}43}$), adapted from Fortuna, *et al*., *manuscript under review*. Average values from two or three independent experiments $\pm$ SD is presented.

Superposition of the crystal structure of Menin-MLL$^{4\text{-}15}$ with the crystal structure of

Menin-**25** additionally reveals that **25** closely mimics side chains Trp7, Phe9, Pro10, and

Pro13 of MLL$^{4\text{-}15}$. In particular, fluoro-phenylalanine of **25** buries deep within a hydrophobic

pocket formed by the backbone of residues Ser155, His180, and Asp181, and the side chain of

Ala182 (**Figure A.7.**). Moreover, the binding modes of proline moieties of **25** are conserved

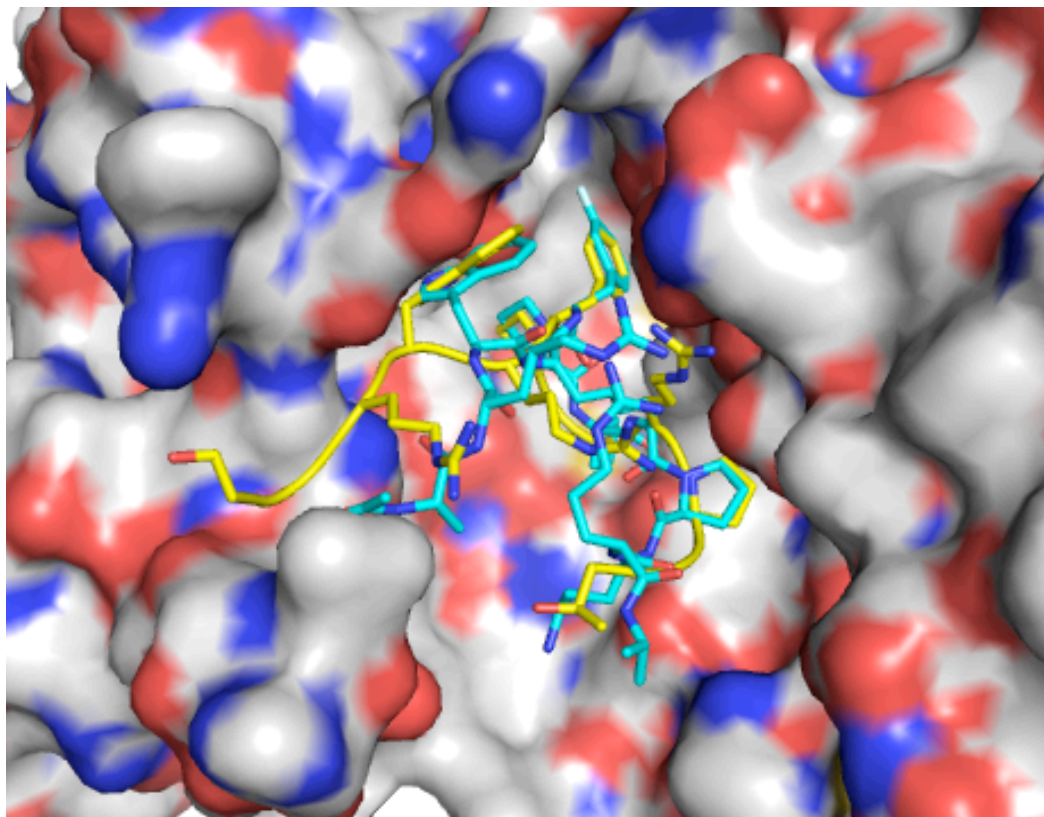with those of Pro10 and Pro13 of MLL$^{4-15}$ in complex with Menin.

Figure A.5. Crystal structure of peptidomimetic inhibitor 25 in complex with Menin.
Menin-**25** (PDB code 6OPJ) crystal structure, overlaid with Menin-MLL$^{4-15}$ (PDB code
4GQ6) crystal structure. Menin colored gray, shown in surface representation. MLL$^{4-15}$
colored yellow; backbone shown in cartoon representation, side chains shown as sticks. **25**
colored cyan; shown in stick representation.

In addition to hydrophobic contacts, the crystal structure reveals that **25** forms a

comparable hydrogen bonding network to MLL (**Figure A.6.**). The side chain of the first

arginine residue of the C-terminal arginine tri-peptide forms two hydrogen bonds with the side

chain of Glu359 of Menin. This is conserved with Arg12 of MLL, which makes two

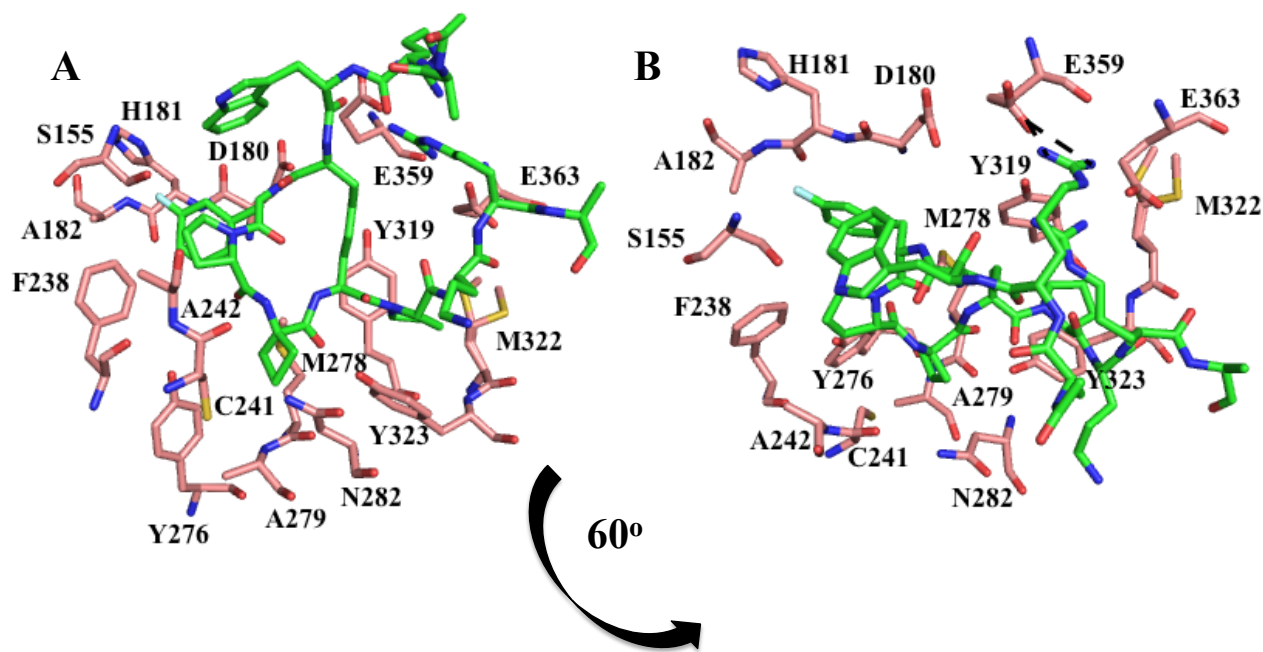hydrogen bonds with the side chain of Glu359, as well.

Figure A.6. Crystal structure reveals binding mode of peptidomimetic inhibitor 25 in complex with Menin. A. Structure of **25** in complex with Menin. Selected residues of Menin shown in stick representation, colored salmon. **25** shown in stick representation, green. B. Same as A., rotated 60°.

Notably, the structure reveals novel contacts made by modifications to **25** that are distinct from contacts made by MLL. In particular, compared to the methyl side chain of Ala11 of MLL, the cyclobutane substitution more fully occupies a hydrophobic pocket comprised of Cys241, Met278, Ala279, Tyr279, and the side chain of Asn282. We were unable to resolve two arginine residues of the C-terminal arginine tri-peptide, and the C-terminal amide cap. Nonetheless, our model provides insight into the structural basis of inhibition of Menin-MLL complex by **25**.
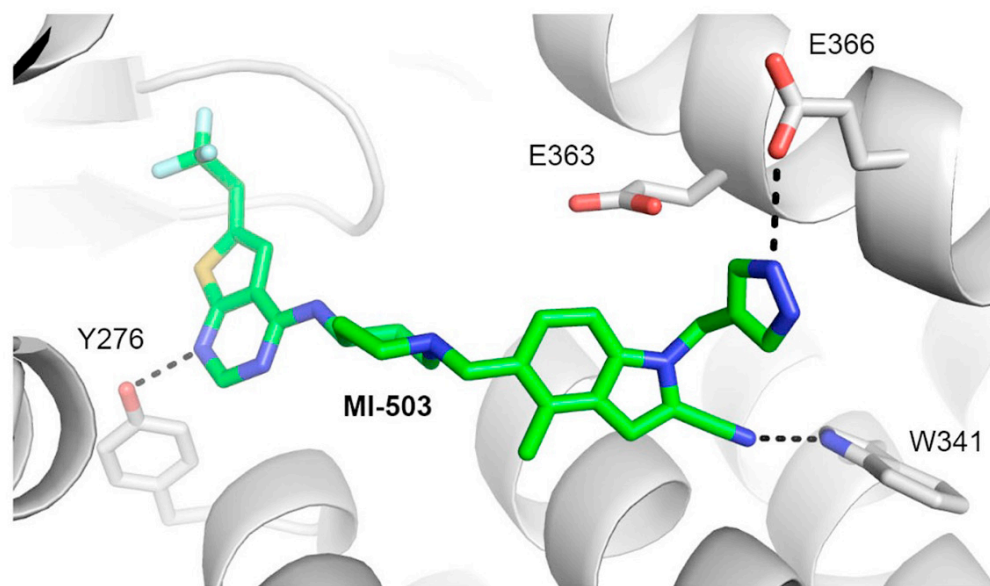
Figure A.7. Crystal structure of MI-503 in complex with Menin. Menin shown in cartoon representation, colored gray. Side chains of selected residues shown as sticks. **MI-503** shown in stick representation, colored green. Hydrogen bonds indicated by dashed lines. Figure adapted from Borkin, *et al*. (2018)[20].

C.5. Structure-based design of potent small-molecule inhibitors of the Menin-MLL interaction on the basis of MI-503

From the starting point of the previous generation of small-molecule inhibitors of Menin, we aimed to use structural biology-based approaches to design more potent small molecules targeting the Menin-MLL interaction. **MI-503** presents low- to mid-nanomolar IC_{50} values in blocking Menin-MLL[4-15] and Menin-MLL[4-43] by FP competition assay, 15 nM and 33 nM, respectively[19,20].
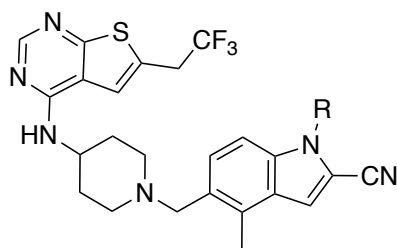
Compound	R	K_D (nM)	IC$_{50}$ MLL$^{4\text{-}43}$ (nM)
MI-503		24.2 ± 0.8	33 ± 8.5
MI-568 (RS)		N.d.	7.5 ± 2.1
MI-1481 (S)		9.0 ± 1.9	3.6 ± 0.9
MI-1482 (R)		N.d.	123 ± 21

Table A.4. SAR for MI-503 and derivatives with substitutions at indole nitrogen. K_D values measured by BLI, adapted from[20]. IC$_{50}$ values measured by Fluorescence Polarization assay using fluorescein-labeled MLL$^{4\text{-}43}$, adapted from[20]. For both K_D and IC$_{50}$s, average values from two or three independent experiments ± SD are presented.

The crystal structure of **MI-503** in complex with Menin reveals the molecular details of the Menin-**MI-503** interaction (**Figure A.7.**). The side chain of Tyr276 forms a hydrogen bond with the thienopyrimidine moiety. In addition, the cyano substituent acts a hydrogen bond acceptor with the side chain of Trp341. Finally, the side chain of Glu366 makes a hydrogen bond with the diazole moiety. Otherwise, **MI-503** makes several crucial hydrophobic contacts with Menin in occupying the F9 and P13 pockets, and presents shape complementarity for the binding site, as previously reported[19].
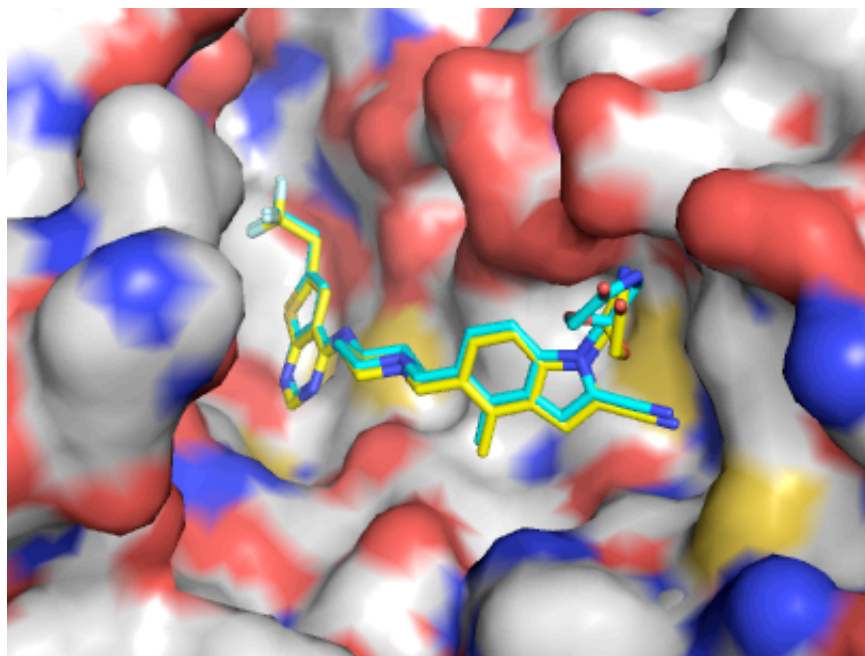
Figure A.8. Crystal structures of MI-1481 and MI-1482 in complex with Menin. Crystal structure of Menin-**MI-1481** (PDB code 6BXY), overlaid with crystal structure of Menin-**MI-1482** (PDB code 6BY8). Menin colored gray, shown in surface representation. **MI-1481** and **MI-1482** colored yellow and cyan, respectively, shown in stick representation.

C.6. Development of MI-503 analogues with substituted morpholine ring

We explored various saturated heterocycles for the diazole moiety, and identified **MI-568** as a potent inhibitor of Menin-MLL[4-43]. This compound possesses a morpholine ring with *para*-substituted carbonyl, introduced at the indole nitrogen via methyl linker. The racemic mixture presented low-nanomolar activity by *in vitro* biochemical assay FP, IC_{50} = 7.5 nM (**Table A.4.**). To quantify activities of enantiomers, we synthesized both *S*- and *R*- enantiomers, **MI-1481** and **MI-1482**, respectively. **MI-1481** presented greater than two-fold improvement in potency over racemic mixture by FP assay, IC_{50} = 3.6 nM, whereas **MI-1482** demonstrated approximately 35-fold weaker activity than **MI-1481**, IC_{50} = 123 nM (**Table**

A.4.). We further characterized **MI-1481** in functional assays and *in vivo*, and this compound

showed *in cell* and *in vivo* efficacy, as reported previously[20].

C.7. Affinity determination of MI-1481 and MI-503 by Bio-layer Interferometry

To ascertain the improvement in affinity of **MI-1481** for Menin, compared to **MI-503**,

we performed Bio-layer Interferometry (BLI) experiments using Octet Red to determine

compounds' respective K_D values. Consistent with IC$_{50}$s, **MI-1481** demonstrates an

approximate three-fold improvement in affinity for Menin over **MI-503** (MI-1481 K_D = 9 nM,

versus **MI-503** K_D = 24 nM) (**Table A.4.**). Moreover, characterization of binding kinetics

reveals comparable association rates, yet **MI-1481** displays a markedly slower dissociation

rate than **MI-503**, which is manifested in its tighter affinity for Menin.
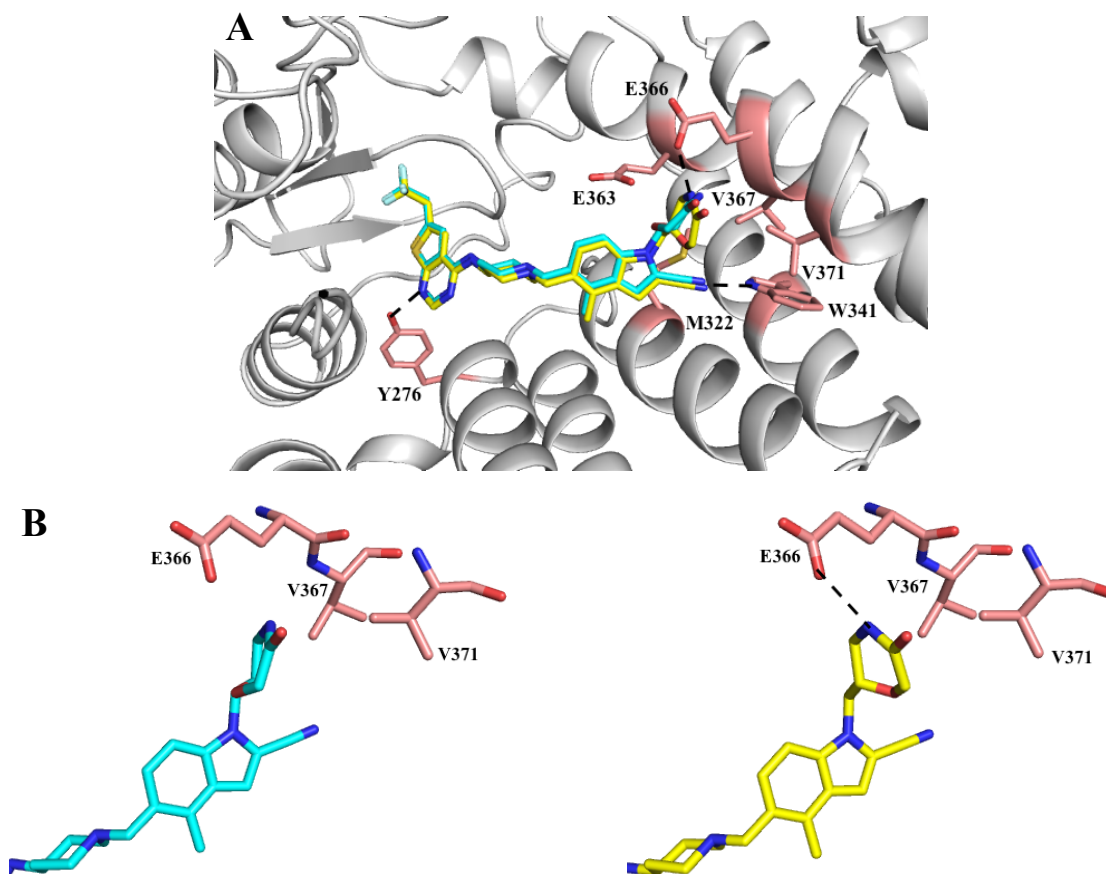
Figure A.9. Molecular details of MI-1481 and MI-1482 interaction with Menin. A. Crystal structure of Menin-**MI-1481 (**PDB code 6BXY**)**, overlaid with crystal structure of Menin-**MI-1482 (**PDB code 6BY8**)**. Menin colored gray, shown in cartoon representation. **MI-1481** colored yellow, shown as sticks; **MI-1481** colored cyan, shown as sticks. Side chains of selected residues shown as sticks, labeled. Hydrogen bonds indicated by dashed lines. B. Comparison of the binding modes of **MI-1481** and **MI-1482**. Selected residues of menin shown as sticks; representation otherwise the same as in A.

C.8. Crystal structures of MI-1481 and MI-1482 in complex with Menin, respectively, reveal structural basis of enantiomer selectivity

To explain the differences in inhibitory activities and binding affinities of **MI-1481** and **MI-1482**, we determined the crystal structure of Menin in complex with **MI-1481**, and in complex with **MI-1482**, respectively, to high-resolution (**Figures A.8.** and **A.9.,** and **Table A.5.**). The binding modes of **MI-1481** and **MI-1482** are conserved with the binding mode of **MI-503**. As in **MI-503**, in **MI-1481** and **MI-1482** the thienopyrimidine moiety makes a hydrogen bond with the side chain of Tyr276. In addition, the cyano group acts as a hydrogen bond acceptor with the side chain of Trp341. Also, **MI-1481** and **MI-1482** occupy Phe9 and Pro13 pockets of Menin, consistent with **MI-503**.

A

28

28 (200 nM)
k_{off} = 1.9*10^{-3} 1/s
k_{on} = 2.4*10^{5} 1/Ms
K_D = 7.7 nM

28 (150 nM)
k_{off} = 2.9*10^{-3} 1/s
k_{on} = 2.8*10^{5} 1/Ms
K_D = 10.4 nM

K_D (**28**)= 9.0 ± 1.9 nM

B

MI-503

MI-503 (200 nM)
k_{off} = 5.0*10^{-3} 1/s
k_{on} = 2.1*10^{5} 1/Ms
K_D = 23.6 nM

MI-503 (150 nM)
k_{off} = 6.1*10^{-3} 1/s
k_{on} = 2.5*10^{5} 1/Ms
K_D = 24.8 nM
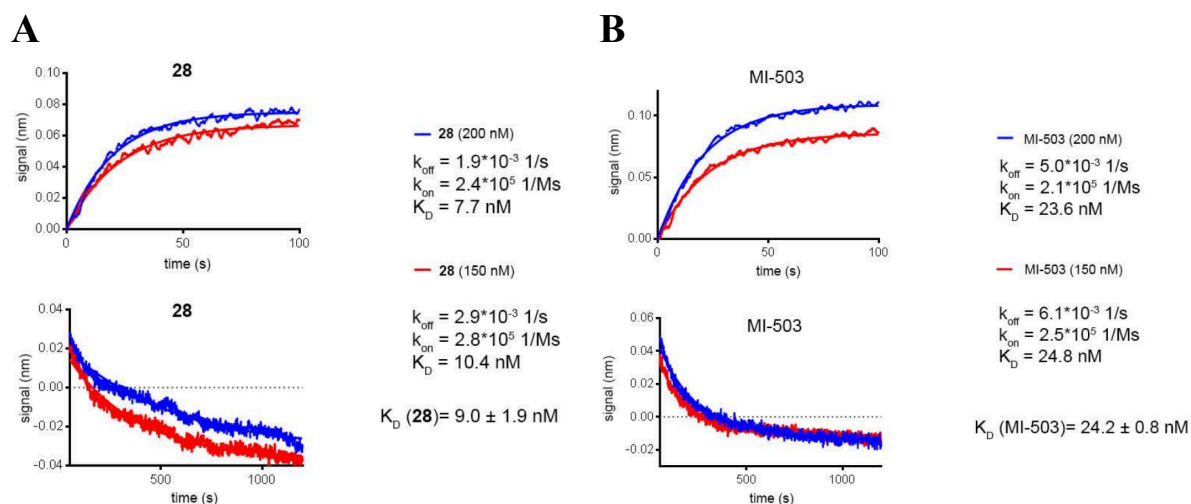
K_D (MI-503)= 24.2 ± 0.8 nM

Figure A.10. Bio-layer interferometry (BLI) studies by Octet Red to determine affinities of MI-503 and MI-1481 for Menin. (A, B). Association (top graph) and dissociation (bottom graph) kinetic assays for **MI-1481** (compound **28**) (**A**) and **MI-503** (**B**) are shown. K_D values determined from average values from two independent experiments ± SD are presented: 150 nM and 200 nM. Figure adapted from Borkin, *et al*. (2018)[20].

Crucially, the *S*-enantiomer **MI-1481** and *R*-enantiomer **MI-1482** present differences in the binding modes of the carbonyl-substituted morpholine moieties (**Figure A.9.**). In **MI-1481**, the morpholine forms a short hydrogen bond (~2.6 A) with the side chain of Glu366. In addition, the morpholine ring is orientated to make hydrophobic contacts with the side chains of Glu366, Val367, and Val371. In contrast, in **MI-1482** the morpholine does not form any hydrogen bond or other polar interactions, nor does it make any hydrophobic contacts, with Menin.

D. Conclusion

Here we present biophysics- and structural biology-based approaches to develop potent peptidomimetic and small-molecule inhibitors of the Menin-MLL interaction. As previous biophysical studies had characterized the structural and mechanistic bases of MLL recognition by Menin, we designed peptidomimetic inhibitors that targeted fundamentally critical interactions made by MLL in complex formation, namely occupying Phe9, Pro10, and Pro13 pockets. To maintain the three-dimensional conformation of peptide inhibitors, we designed a cyclization strategy with a four-carbon, unsaturated linker region. In parallel, structure-guided optimization led to substituting a fluoro-group at the *meta*-position on the phenyl moiety, and substituting a cyclobutane at a hydrophobic pocket occupied by the methyl side chain of Ala11 in Menin-MLL complex. In addition to hydrophobic contacts, we maintained polar interactions in our peptidomimetic by installing a C-terminal arginine moiety that formed a hydrogen bond network with Glu359 of Menin; this is conserved with

interactions made by Arg12 of MLL with Menin. Structural biology- and biophysics-guided development of peptidomimetic inhibitors yielded **25**, which demonstrated low-nanomolar potency *in vitro*. Finally, we determined the crystal structure of **25** in complex with Menin, which provided insight into the structural basis of blocking the Menin-MLL interaction. Further, such insight may be incorporated in the design and development of more potent small-molecule inhibitors.

Subsequently, we describe the structure-based inhibitor design to improve a previous generation of Menin inhibitors, on the basis of the crystal structure of Menin-**MI-503**, which yielded potent small-molecule inhibitors **MI-1481** and **MI-1482** that additionally possess on-target activity in cells, and *in vivo* efficacy. The binding mode of **MI-503** presented a diazole moiety that makes a hydrogen bond with Glu366 of Menin. Exploration of saturated heterocycles at this site identified a morpholine ring with a para-substituted carbonyl group as conferring an approximately six-fold improvement in activity over **MI-503**, 7.5 nM. This value, yet, was obtained from a racemic mixture. Synthesis of individual *R* and *S* enantiomers, **MI-1482** and **MI-1481**, respectively, presented ~35-fold differences in potency: **MI-1481** IC_{50} = 3.6 nM, versus **MI-1482** IC_{50} = 123 nM. To explain the differences in potencies, we solved crystal structures of Menin-**MI-1481** and Menin-**MI-1482** to high-resolution. The structures show that the morpholine moiety of *S* enantiomer **MI-1481** forms a hydrogen bond, and additionally engages in hydrophobic contacts, with Menin. These interactions are not present in complex formation of the *R* enantiomer **MI-1482** with Menin.

In total, our small-molecule inhibitor campaign reported here presents biophysics- and structural biology-based approaches to optimize small-molecule inhibitors **MI-1481** and **MI-1482** in complex with Menin. Given the prominence of the Menin as a pharmacological

target for therapeutic intervention against numerous different types of leukemias with substandard clinical outcomes, and the nanomolar affinity of the Menin-MLL protein-protein interaction, the need for potent, selective, and non-toxic Menin inhibitors with sufficient pharmacokinetic properties is pressing. Our efforts represent a new generation of the thienopyrimidine class of Menin inhibitors, and may facilitate future development of small molecules targeting the Menin-MLL interaction.

Acknowledgements

mentorship, financial support, and guidance on the Menin-inhibitor program in the lab over

the years.

Table A.5. Data collection and refinement statistics for menin-inhibitor complexes.

	MI-1481	MI-1482	25
Data collection			
Space group	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$
Cell dimensions			
a, b, c (Å)	47.32, 80.54, 122.34	48.99, 80.22, 124.72	47.87, 78.15, 123.70
α, β, γ (°)	90.0, 90.0, 90.0	90.0, 90.0, 90.0	90.0, 90.0, 90.0
Resolution (Å)	50-1.82 (1.86-1.82)	50-1.90 (1.93-1.90)	50-1.50 (1.56-1.50)
R_{sym} or R_{merge}	0.083 (0.677)	0.082 (0.508)	0.083 (0.574)
$I / \sigma I$	20.80 (2.00)	12.48 (2.49)	18.15 (2.80)
Completeness (%)	98.6 (95.3)	100.0 (100.0)	98.9 (98.1)
Redundancy	6.5 (6.1)	3.2 (2.9)	7.2 (7.3)
Refinement			
Resolution (Å)	40.27-1.82	38.53-1.90	25.98-1.50
No. reflections	42,152	39,551	53,932
R_{work} / R_{free}	17.38 / 21.27	15.06/18.59	16.84/19.76
No. atoms			
Protein	3675	3727	3726
Ligand/ion	99	98	78
Water	288	495	708
B-factors (Å^2)			
Protein	26.2	16.7	22.61
Ligand/ion	44.9	39.1	39.15
Water	37.0	30.6	38.24
R.m.s. deviations			
Bond lengths (Å)	0.015	0.009	0.008
Bond angles (°)	1.34	0.93	0.950

All diffraction data were obtained from a single crystal.
*Values in parentheses are for highest-resolution shell.

E. Experimental Methods

Expression and Purification of Menin

The expression and purification of Menin for x-ray crystallography and for biochemical experiments have been described previously[18,15]. To express and purify biotinylated-Menin, C-terminal Avi-tagged (AAALEGLNDIFEAQKIEWHE) Menin was co-expressed in *E. Coli* BL21 (DE3) cells with BirA enzyme, and purified as described previously[22].

Biochemical Characterization of Menin-MLL Inhibitors

Inhibition of Menin-MLL interaction by peptidiomimetics and small molecules was measured by Fluorescence Polarization (FP) Assay, using methods described previously[18,20].

Affinity Determination of MI-503 and MI-1481 for Menin by Bio-layer interferometry (BLI) using Octet Red

BLI assay was performed in 96-well microplates at room temperatures, with continuous shaking at 1000 rpm. Using Octet Red 96 system (ForteBio, Menlo Park, CA, USA), biotinylated-Menin was loaded on Super Streptavidin (SSA) biosensors (ForteBio, Inc.) by incubation of sensors in solution containing 200 nM biotinylated-Menin in assay buffer, comprised of 50 mM Tris, pH 7.5, 50 mM NaCl, 1 mM TCEP, 0.05% BSA, 0.01% Tween-20, and 2% DMSO. This incubation step was performed for 600 s, followed by 1200 s in assay buffer for equilibration. Next, loaded tips were incubated with compounds (**MI-1481** and **MI-503**) at 200 and 150 nM concentrations for 600 s association step, which was followed by 1200 s dissociation step. Data were generated automatically using Octet User Software (ForteBio, Inc.). Binding kinetics were analyzed in Prism (GraphPad Software, Inc.), using *one phase association* equation to ascertain K_{obs} and *one phase decay* equation for K_{off}. K_{on} was extracted using the equation

Equation 5.1.

$$K_{obs} = K_{on} \times [\text{Ligand}] + K_{off}$$

where [Ligand] is the free ligand concentration in solution. K_D was determined by K_{off}/K_{on}.

Crystallization of Menin Complexes with Inhibitors

For co-crystallization of Menin in complex with small-molecule inhibitors, crystals were obtained as described previously[15]. For Menin in complex with peptidomimetic inhibitor **25**, 2.8 mg mL^{-1} (50 μM) Menin was incubated with 200 μM **25**. Crystals were obtained in 0.2 M lithium sulfate, 0.1 M HEPES, pH 7.5, and 25% (w/v) PEG-3,350, using the sitting drop technique over several days incubation at 4 $^{\circ}$C. Before data collection, crystals were transferred into cryoprotectant solution comprised of crystallization solution with 20% PEG-550 MME, and subsequently flash-frozen in liquid nitrogen.

Crystallographic Data Collection and Structure Determination

Diffraction data of Menin complexes were collected at 21-ID-D, 21-ID-F, and 21-ID-G beam lines at the Life Sciences Collaborative Access Team at the Advanced Photon Source. Data were indexed and scaled in HKL-2000[23]. Structures were solved by molecular replacement using MOLREP with the apo-structure of human Menin (PDB code 4GPQ) as the search model. The model was refined using REFMAC[24], COOT[25], CCP4 program suite[26], and PHENIX program suite[27]. The structure was validated using the MOLPROBITY[28] and ADIT[29] servers. Data collection and refinement statistics are summarized in Table A.1. Coordinates and structure factors have been deposited in the Protein Data Bank: 6OPJ (Menin-peptidomimetic **25**), 6BXY (Menin-compound **MI-1481**), and 6BY8 (Menin-compound **MI-1482**).

F. References

1   Cox, M. C. *et al.* Chromosomal aberration of the 11q23 locus in acute leukemia and frequency of MLL gene translocation: results in 378 adult patients. *Am J Clin Pathol* **122**, 298-306, doi:10.1309/RX27-R8GJ-QM33-0C22 (2004).

2   Sorensen, P. H. *et al.* Molecular rearrangements of the MLL gene are present in most cases of infant acute myeloid leukemia and are strongly correlated with monocytic or myelomonocytic phenotypes. *J Clin Invest* **93**, 429-437, doi:10.1172/JCI116978 (1994).

3   Marschalek, R. Mechanisms of leukemogenesis by MLL fusion proteins. *Br J Haematol* **152**, 141-154, doi:10.1111/j.1365-2141.2010.08459.x (2011).

4   Tomizawa, D. *et al.* Outcome of risk-based therapy for infant acute lymphoblastic leukemia with or without an MLL gene rearrangement, with emphasis on late effects: a final report of two consecutive studies, MLL96 and MLL98, of the Japan Infant Leukemia Study Group. *Leukemia* **21**, 2258-2263, doi:10.1038/sj.leu.2404903 (2007).

5   Mann, G. *et al.* Improved outcome with hematopoietic stem cell transplantation in a poor prognostic subgroup of infants with mixed-lineage-leukemia (MLL)-rearranged acute lymphoblastic leukemia: results from the Interfant-99 Study. *Blood* **116**, 2644-2650, doi:10.1182/blood-2010-03-273532 (2010).

6   Liu, H., Cheng, E. H. & Hsieh, J. J. MLL fusions: pathways to leukemia. *Cancer Biol Ther* **8**, 1204-1211, doi:10.4161/cbt.8.13.8924 (2009).

7   Yokoyama, A. *et al.* The menin tumor suppressor protein is an essential oncogenic cofactor for MLL-associated leukemogenesis. *Cell* **123**, 207-218, doi:10.1016/j.cell.2005.09.025 (2005).

8   Zhang, Y., Chen, A., Yan, X. M. & Huang, G. Disordered epigenetic regulation in MLL-related leukemia. *Int J Hematol* **96**, 428-437, doi:10.1007/s12185-012-1180-0 (2012).

9   Armstrong, S. A. *et al.* MLL translocations specify a distinct gene expression profile that distinguishes a unique leukemia. *Nat Genet* **30**, 41-47, doi:10.1038/ng765 (2002).

10   Caslini, C. *et al.* Interaction of MLL amino terminal sequences with menin is required for transformation. *Cancer Res* **67**, 7275-7283, doi:10.1158/0008-5472.CAN-06-2369 (2007).

11   Winters, A. C. & Bernt, K. M. MLL-Rearranged Leukemias-An Update on Science and Clinical Approaches. *Front Pediatr* **5**, 4, doi:10.3389/fped.2017.00004 (2017).

12   Hughes, C. M. *et al.* Menin associates with a trithorax family histone methyltransferase complex and with the hoxc8 locus. *Mol Cell* **13**, 587-597 (2004).

13   Chen, Y. X. *et al.* The tumor suppressor menin regulates hematopoiesis and myeloid transformation by influencing Hox gene expression. *Proc Natl Acad Sci U S A* **103**, 1018-1023, doi:10.1073/pnas.0510347103 (2006).

14   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, doi:10.1074/jbc.M110.172783 (2010).

15   Shi, A. *et al.* Structural insights into inhibition of the bivalent menin-MLL interaction by small molecules in leukemia. *Blood* **120**, 4461-4469, doi:10.1182/blood-2012-05-429274 (2012).

16   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, doi:10.1074/jbc.M111.258186 (2011).

17   He, S. *et al.* 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, doi:10.1021/jm401868d (2014).

18   Grembecka, J. *et al.* Menin-MLL inhibitors reverse oncogenic activity of MLL fusion proteins in leukemia. *Nat Chem Biol* **8**, 277-284, doi:10.1038/nchembio.773 (2012).

19   Borkin, D. *et al.* Pharmacologic inhibition of the Menin-MLL interaction blocks progression of MLL leukemia in vivo. *Cancer Cell* **27**, 589-602, doi:10.1016/j.ccell.2015.02.016 (2015).

20   Borkin, D. *et al.* Complexity of Blocking Bivalent Protein-Protein Interactions: Development of a Highly Potent Inhibitor of the Menin-Mixed-Lineage Leukemia Interaction. *J Med Chem* **61**, 4832-4850, doi:10.1021/acs.jmedchem.8b00071 (2018).

21   Huang, J. *et al.* The same pocket in menin binds both MLL and JUND but has opposite effects on transcription. *Nature* **482**, 542-546, doi:10.1038/nature10806 (2012).

22   Fairhead, M. & Howarth, M. Site-specific biotinylation of purified proteins using BirA. *Methods Mol Biol* **1266**, 171-184, doi:10.1007/978-1-4939-2272-7_12 (2015).

23   Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol* **276**, 307-326 (1997).

24   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, doi:10.1107/S0907444996012255 (1997).

25   Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* **60**, 2126-2132, doi:10.1107/S0907444904019158 (2004).

26   Collaborative Computational Project, N. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr D Biol Crystallogr* **50**, 760-763, doi:10.1107/S0907444994003112 (1994).

27   Adams, P. D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* **66**, 213-221, doi:10.1107/S0907444909052925 (2010).

28   Davis, I. W. *et al.* MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res* **35**, W375-383, doi:10.1093/nar/gkm216 (2007).

29   Yang, H. *et al.* Automated and accurate deposition of structures solved by X-ray diffraction to the Protein Data Bank. *Acta Crystallogr D Biol Crystallogr* **60**, 1833-1839, doi:10.1107/S0907444904019419 (2004).