

**Discovery of Small Molecules to dissect the individual Roles
of GACKIX-Activator Complexes**

by

Jean M. Lodge

A dissertation submitted in partial fulfillment
of the requirements of the degree of
Doctor of Philosophy
(Chemistry)
in the University of Michigan
2017

Doctoral Committee:

Professor Anna K. Mapp, Chair

Professor Carol A. Fierke

Professor Robert T. Kennedy

Associate Professor Brandon T. Ruotolo

Assistant Professor Matthew B. Soellner

Jean M. Lodge

jmlodge@umich.edu

ORCID iD: orcid.org/0000-0001-5101-5470

© Jean M. Lodge 2017

Acknowledgements

I enjoyed living near the Huron River that forms an inviting landscape in Ann Arbor, MI, but more impressive is the flowing support from the people at the University of Michigan. I would like to first thank Prof. Anna Mapp who provided me with a fantastic project, and the freedom to play with it as I *scientifically* matured. Expert guidance and intellectual foresight also came from my committee members; Professors Fierke, Kennedy, Ruotolo, and Soellner. Our meetings increased my motivation to get back into the laboratory – not because I wanted to escape, but they had spotted interesting flaws or important missing experiments.

To friends affiliated with the Mapp Lab and Chemistry department, including, but not limited to Omari, Mel, Melanie, Matt, and Steve; thank you for the dynamic discussions and unwavering support. I was fortunate to be “buffered” into Chemical Biology research with outstanding instructors; Drs. Chinmay Majmudar, and William C. Pomerantz. Dr. Ningkun Wang was an exemplary role model, and a joy to work with - she changed the atmosphere of a lab, typically by promoting laughter during even the most tiresome experiments. Whenever I needed a brief escape, caffeine, and a confidant, there was Dr. Jessica Rabuck-Gibbons, who became an obliging friend and colleague. Small molecules were not my limiting reactant due to the hard-work of Drs. James Clayton and Meg Breen. Their organic friendship was their most valuable product along with their mechanism of support that helped yield me with this Ph.D. Positive bonds were formed and catalyzed by the kindness of Dr. Rachel Pricer and her family who also exposed me to many, new elements.

The extra-lab cursory encampments to Casey's – where the special conviviality and camaraderie were expertly curated by Norm, Chris, and Steve – were invaluable for the refreshment of my intellectual curiosity during my time in Ann Arbor.

During graduate school, I formed the strongest, irreversible bonds with my dogs. First with my elderly, but spirited golden retriever mix, Penny whose happiness was infectious – chasing after the row boats on the Huron River in Argo Park or the ducks in Gallup Park. I am forever in debt to my parents who cared for her during her last few months. My Golden Doodle, Teddy Bear was responsible for getting me through the latter half of graduate school. He started work as my service dog at 1.5 years old – my manners were likely comparable to his at that age if I could have run the marathons that he could as a toddler. While growls may be heard at research facilities new to service animals, only positive reinforcement came from Cathy Andrews at the Life Sciences Institute who treated Teddy like a worker and prevented “petting” harassments.

The enhancement of equanimity, necessary for lab work, was provided by my regular equestrian excursions. I crack the whip of thanks to Jenny and her horse, Vixen; to Mr. Head and his charges JoJo and Johnny. Mr. Head went an extra furlong to rein in my riding and social skills. He had a special knack for dismounting people from their “high horses” – this spurred on my mentally “firm-footed” return to lab. It was a brief abandonment in a hay field that encouraged me to “lose the draw reins, keep my head set,” and finally complete this Ph.D.

Of course - the best for last - thank you to my typically chill Minnesota-based family; Keith, Stephanie, and Fiona because life would be nonexistent, lonely, and dull without them.

Table of Contents

Acknowledgements	ii
List of Figures	viii
List of Tables	xi
List of Appendices	xii
List of Abbreviations	xiii
Abstract.....	xix
Chapter 1 Transcription is regulated by elaborate choreography between activators and coactivators.....	1
1.1 Introduction	1
1.2 What is transcription?.....	1
1.3 Coactivator play pivotal roles in mediating transcription.....	4
1.4 What constitutes a KIX domain?.....	6
1.5 How do different TADs interact with the KIX domain of CBP?.....	9
1.6 KIX masters intermolecular and intramolecular communication.	16
1.7 These additional binding partners like to hug KIX.....	19
1.8 Activators and coactivators are involved in diseases	20
1.9 Can small molecules effect and probe KIX-activator complexes?.....	24
1.10 What does this thesis offer?	29
1.11 References.....	30

Chapter 2 FP Tethering: a screening technique to rapidly identify compounds that disrupt protein-protein interactions.....	46
2.1 Abstract.....	46
2.2 Introduction to FP-based Tethering screens	46
2.3 Results.....	49
2.4 Conclusion.....	60
2.5 Materials and Methods.....	62
2.5.1 <i>Expression and purification of the KIX domain</i>	62
2.5.2 <i>Synthesis of fluorescent peptides</i>	62
2.5.3 <i>Fluorescence polarization binding assay</i>	62
2.5.4 <i>Fluorescence polarization Tethering screens</i>	63
2.5.5 <i>LC-MS validation of fragment hits</i>	64
2.5.6 <i>Determination of the dose response (DR₅₀)</i>	65
2.5.7 <i>Fragment Tethering</i>	65
2.5.8 <i>CD wavelength and variable temperature scans</i>	65
2.5.9 <i>Alkylation experiments</i>	66
2.6 References.....	67
Chapter 3 Dissecting Allosteric Effects of Activator-Coactivator Complexes Using a Covalent Small Molecule Ligand.....	72
3.1 Abstract.....	72
3.2 Introduction to the cooperativity between activators and coactivators...	73
3.3 Results.....	76
3.3.1 <i>Equilibrium Binding Measurements Define Extent of Cooperativity for Each</i>	

<i>Complex.</i>	77
3.3.2 <i>Testing for stabilizing effects with KIX mutants.</i>	82
3.4 Discussion on cooperativity	84
3.5 Conclusion	87
3.6 Methods and supplemental experimental information.	87
3.6.1 <i>Preparation of protein and peptides</i>	87
3.6.2 <i>Fluorescence anisotropy assays</i>	88
3.6.3 <i>Fluorescence stopped-flow kinetic assays</i>	89
3.6.4 <i>CD analysis of KIX mutants</i>	91
3.7 References.	92
Chapter 4 Covalent chemical co-chaperones for the CBP KIX domain	98
4.1 Abstract.	98
4.2 Introduction	99
4.3 Results and Discussion	101
4.3.1 <i>Converting 1-10 into irreversible analogues</i>	101
4.3.2 <i>Assessing target engagement of two 1-10 irreversible probes</i>	103
4.4 Conclusion.	106
4.5 Methods	108
4.5.1 <i>Protein Expression and Purification</i>	108
4.5.2 <i>Peptide Synthesis and Purification</i>	112
4.5.3 <i>Determination of the dose response (DR₅₀)</i>	113
4.5.4 <i>Alkylation reaction of KIX</i>	113
4.5.5 <i>Fluorescent anisotropy assays</i>	113
4.5.6 <i>Labeling KIX cysteine mutants</i>	115
4.5.7 <i>Alkylation reaction in growing E. coli</i>	116
4.5.8 <i>Cell viability assay</i>	116
4.5.9 <i>Neutravidin pull-down assay with the biotinylated 1-10 probe</i>	116

4.6 References.....	117
Chapter 5 Conclusions and Future Directions	122
5.1 Abstract.....	122
5.2 Conclusions.....	123
5.3 Future directions.....	125
5.3.1 <i>Expanding on the Tethering screen</i>	<i>125</i>
5.3.2 <i>Developing covalent probes.....</i>	<i>129</i>
5.3.3 <i>Alternative method to identify probes for homologous domains.....</i>	<i>131</i>
5.3.4 <i>Venturing to probe the cooperativity imparted by KIX</i>	<i>132</i>
5.3.5 <i>Examining coactivators and its complexes</i>	<i>135</i>
5.3.6 <i>Expanding into cellular systems</i>	<i>136</i>
5.4 Final remarks	138
5.5 References.....	139
Appendices.....	145

List of Figures

Figure 1-1 The activation of transcription is either simple or	3
Figure 1-2 The architecture of CBP/p300	5
Figure 1-3 Comparison of GACKIX motifs	7
Figure 1-4 Distinct residue differences between GACKIX motifs	8
Figure 1-5 Alignment of TADs that interact with the KIX domain of CBP/p300 ...	9
Figure 1-6 Schematic for the interactions of pKID and c-Myb with KIX	11
Figure 1-7 Sequences for different c-Myb constructs.....	14
Figure 1-8 Cartoon representations of the KIX-MLL complex.....	15
Figure 1-9 KIX contains an allosteric network.....	18
Figure 1-10 KIX has multiple binding partners.....	21
Figure 1-11 Small molecules that bind to KIX	26
Figure 1-12 Disulfide Tethering scheme	28
Figure 2-1 The Tethering methodology for site-directed ligand discovery	48
Figure 2-2 FP- and MS-based Tethering identified overlapping ligand sets for the MLL-binding site.....	51
Figure 2-3 The impact of the cysteine mutants within the pKID-site of KIX on the overall KIX structure was assessed by CD.....	52
Figure 2-4 The FP Tethering screen requires a compatible, cysteine-containing KIX target.	53
Figure 2-5 FP-based Tethering screen to identify new ligands that disrupt the interaction between KIX and pKID.	54
Figure 2-6 Labeling of KIX H602C with 3D4 and 6D11 decreases the affinity for	

pKID, c-Myb, and MLL.	57
Figure 2-7. 3D4 and 6D11 contain chemical features that supported their discovery in the FP-Tethering screen in comparison to other fragments.	58
Figure 2-8 The initial 6D11 and 3D4 irreversible derivatives suffer from poor labeling efficiency with the KIX H602C target.	60
Figure 3-1 Activator-KIX domain interactions	73
Figure 3-2 Comparing the conformations of KIX.....	75
Figure 3-3 Bar graph of relative K_D values of pKID•KIX complexes normalized to the K_D of pKID and free KIX constructs	77
Figure 3-4 Anisotropy binding curves of complexes of KIX wt, KIX I660V, KIX N627C, and KIX L664C binding to pKID	78
Figure 3-5 Stopped-flow binding kinetics for the observed association rates	80
Figure 3-6 Comparing the K_{on} and K_{off} between pKID and KIX complexes.....	81
Figure 3-7 Mutants aimed to stabilize KIX and induce positive cooperativity	83
Figure 3-8 Analyzing the helicity of KIX mutants.....	83
Figure 3-9 Equilibrium binding results for KIX S670N and S670K against c-Myb	84
Figure 3-10 Summary for the k_{on} and k_{off}	85
Figure 3-11 Stopped-flow traces of control experiments	90
Figure 4-1 Irreversible Tethering scheme	100
Figure 4-2 Effect of the alkylating agent, IAM on KIX	103
Figure 4-3 Other TADs and their interactions with alkylated KIX N627C.....	104
Figure 4-4 Selectivity of 1-10 irreversible alkylators for KIX N627C.....	105
Figure 4-5 Cell viability for the 1-10 alkylators	106
Figure 4-6 Visualizing specificity with the biotinylated 1-10d probe	107
Figure 4-7 Summary of DR_{50} and FP experiments to compare the effect of the	

irreversible probes with KIX N627C	115
Figure 5-1 The Tethering-based approach to develop selective probes for CBP KIX	124
Figure 5-2 Site-directed Tethering scheme	126
Figure 5-3 Development of irreversible probes for KIX	129
Figure 5-4 Examples of reactive warheads.....	130
Figure 5-5 Bumb-and-hole approach	132
Figure 5-6 Structural characterization of the coactivator, CBP or p300	135
Figure 5-6 The irreversible probe 1-10 vinyl sulfonamide was developed to selectively target the coactivator containing the KIX N627C mutant in cells.	139
Figure I-1 In vitro endonuclease cleavage assay to target the KIX gene.....	146
Figure I-2 Comparing the cas9 constructs using the <i>in vitro</i> endonuclease cleavage assay	147
Figure II-1 Phage life cycle is generally described in six steps for phage that are temperate and virulent.....	162
Figure II-2 Preventing phage infections	165
Figure II-3 Indicators of a phage infection	167
Figure II-4. Bacteria are infected by phage, but they adapt strategies to fight off phage infection	171

List of Tables

Table 1-1 Comparing KDs for the TADs at the MLL-binding site.....	16
Table 2-1 FP results for wild-type KIX and two cysteine mutants	53
Table 3-1 Association and dissociation rate constants of pKID·KIX complexes...	80
Table 4-1 Evaluating the 1-10 irreversible probes	102

List of Appendices

Appendix I Unfinished business: Addressing the cellular requirements for the irreversible Tethering strategy	145
I.1 Abstract	145
I.2 Introduction.....	145
I.3 Results	146
I.4 Discussion and Future Directions	148
I.5 Materials and Methods	150
<i>I.5.1 Protocol for Cas9 purification.....</i>	<i>150</i>
<i>I.5.2 Preparation and purification of RNA.....</i>	<i>153</i>
<i>I.5.3 In vitro cleavage assay</i>	<i>155</i>
I.6 References	156
Appendix II Where bacteria grow, phage strive to be.	160
II.1 Abstract.....	160
II.2 The physical features of bacteriophages.....	160
II.3 The life cycle(s) of bacteriophages	162
II.4 How to identifying a phage infection?.....	165
II.5 Cleaning-up after phage	167
II.6 The good qualities of bacteriophage	171
II.7 Conclusion.....	176
II.8 References.....	177

List of Abbreviations

Å	angstrom, unit of length, equal to one tenth of a millimicron
Ala, A	alanine
AML	acute myeloid leukemia
APS	ammonium persulfate
ARC105	activator-recruited cofactor 105 KDa component
Arg, R	arginine
Asn, N	asparagine
Asp, D	aspartic acid
ATL	adult T-cell leukemia
Bis-Tris	bis(2-hydroxyethyl) aminotris (hydroxymethyl) methane
β-ME	β-mercaptoethanol or 2-Sulfanylethan-1-ol
CARM1	coactivator-associated arginine methyltransferase
Cas	CRISPR associate protein
CBP	CREB-binding protein
CD	circular dichroism
<i>C. elegans</i>	<i>Caenorhabditis elegans</i>
CMV	cytomegalovirus
c-Myb	myeloblastosis proto-oncogene
CR	conserved region
CREB	cyclic-adenosine-monophosphate response-element-binding
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
CV	coefficient of variation

Cys, C	cysteine
cysKIX	KIX containing a cysteine mutation
bp	base pair
bZIP	basic leucine zipper
Da	Dalton
DBD	DNA binding domain
DMEM	Dublecco's Modified Eagle Medium
DMSO	dimethylsulfoxide
DNA	deoxyribonucleic acid
DR ₅₀	dose response 50, half maximal
DSB	double-stranded DNA breaks
DTT	dithiothreitol
E1A	adenovirus-encoded early region 1A
<i>E. coli</i>	<i>Escherichia coli</i>
EDA	ethylenediamine
EM	electron microscopy
EMSA	electrophoretic mobility shift assay
Eq.	equation
FBS	fetal bovine serum
FITC	fluorescein isothiocyanate
Fmoc	Fluorenylmethoxycarbonyl
FOXO	forkhead box class O
FP	fluorescence polarization
FPLC	fast protein liquid chromatography
FRET	fluorescence resonance energy transfer
G ⁰	Gibbs free energy, units in kilocalories per mole
GACKIX	Gal11, ARC105, CREB binding protein KIX domain

Gln, Q	glutamine
Glu, E	glutamic acid
Gly, G	glycine
H3K27	histone 3 lysine residue 3
HAT	histone acetyltransferase
HBZ	HTVL-1 basic leucine zipper virus type 1
HDR	homology-directed repair
HEK	human embryonic kidney cells
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
His, H	histidine
HIV-1	human immunodeficiency virus type 1
HPLC	high-throughput liquid chromatography
HSQC	heteronuclear single quantum coherence
HTVL-1	human T-cell leukemia virus type 1
IAM	iodoacetamide
IBid	interferon-binding domain
IC ₅₀	half maximal inhibitory concentration
IDR	intrinsically disordered region
Ile, I	isoleucine
IM-MS	ion mobility mass spectrometry
indels	insertions or deletions
IPTG	isopropyl β -D-1-thiogalactopyranoside
ITC	isothermal titration calorimetry
k _d	dissociation constant
k _i	inhibitor constant
k _{obs}	observed rate constant
k _{off}	dissociation rate constant

k_{on}	association rate constant
kbp	kilo base pair
KID	kinase inducible domain
KIX	kinase-inducible domain interacting
L	liter
LB	lysogene broth
LC	liquid chromatography
Leu, L	leucine
Lys, K	lysine
MED15	mediator complex subunit 15
MES	2-(N-morpholino)ethanesulfonic acid
Met, M	methionine
ml	milliliter
MLL	mixed-lineage leukemia
mP	polarization signal x1,000
MS	mass spectrometry
MW	molecular weight
m/z	mass-to-charge ratio
NaCl	sodium chloride
NCBD	nuclear coactivator binding domain
NHEJ	non-homologous end joining
Ni-NTA	nickel nitrilotriacetic acid
NLS	nuclear localization sequence
NMR	nuclear magnetic resonance spectroscopy
NP-40	nonidet P-40, octylphenoxypolyethoxyethanol
OD	optical density
p300	adenovirus E1A-associated 300 kDa protein

PBS	phosphate buffered saline
PCR	polymerase chain reaction
PDB ID	Protein Data Bank identifier
Pdr1	pleiotropic drug resistance transcription factor 1 orthologue
Phe, F	phenylalanine
PIC	preinitiation complex
PPI	protein-protein interaction
Pro, P	proline
PrOF NMR	protein-observed fluorine NMR spectroscopy
PTM	post-translational modification
PVDF	polyvinylidene difluoride
Q-TOF	quadrupole time-of-flight
RecQL5	RecQ Protein-Like 5 or ATP-Dependent DNA Helicase 5
RNA	ribonucleic acid
rpm	revolutions per minute
RSC	Royal Society of Chemistry
RT	room temperature, approximately 20-25 °C
Phe	phenylalanine
PIC	preinitiation complex
PolII	RNA polymerase II
S:B	signal-to-background ratio
SD	standard error
SDS-PAGE	sodium dodecyl sulfate-polyacrylamide gel electrophoresis
SE	standard error
Ser, S	serine
SOC	super optimal broth with 20 mM glucose
SREBP	sterol response element binding protein

TAD	transcriptional activation domain
TAMARA	tetramethylrhodamine
TAZ	transcriptional adapter zinc finger 2
TEMED	tetramethylethylenediamine
TF	transcription factor
TFE	1,1,1-tri-fluoro-ethanol
TFI	total fluorescence intensity
Thr, T	threonine
TR-FRET	time-resolved FRET
Trp, W	tryptophan
Tyr, Y	tyrosine
Val, V	valine

Abstract

Discovery of Small Molecules to dissect the individual Roles of GACKIX-Activator Complexes

by

Jean Marie Lodge

In gene expression, the GACKIX motif interacts with multiple activators to regulate diverse physiological processes including cellular growth, hematopoiesis, synaptic plasticity, and lipid homeostasis; thus, dysregulation of these complexes is also linked to many diseases from leukemia to Alzheimer's disease. Chemical tools are sought to understand the individual roles of activator-GACKIX complexes in transcription. However, the confounding problem has been finding small molecules that are selective in their interactions towards GACKIX because this small motif is found in multiple eukaryotic coactivators including CBP, its homolog p300 and the mediator subunit, MED15. In this dissertation, the site-directed strategy known as Tethering is applied to develop selective small molecule probes that target the KIX domain of one coactivator such as CBP versus p300.

Tethering readily detects fragments with an intrinsic affinity for the target by amplifying the strength of these interactions via the formation of a covalent disulfide bond between a disulfide fragment and a cysteine-containing KIX target. Traditional liquid-chromatography mass spectrometry applied to Tethering

screens ranked fragments by their ability to bind to the cysteine-containing KIX target. We developed a novel fluorescence polarization screen that streamlined the identification of small molecules inhibitors directed at the c-Myb and CREB-binding site of CBP KIX. More importantly, the FP Tethering screen has the potential to increase the applicability of Tethering to other important biological targets that would benefit from this approach.

Tethering identified fragments that impacted the assembly or disassembly of activator binding partners with KIX. For example, KIX N627C tethered to **1-10** disrupted MLL binding and enhanced the binding affinity of KIX for pKID by 2-fold. By using mutants and tethered KIX complexes in transient state stopped-flow experiments, we showed that the stabilization of the ternary complex contributed to the positive cooperativity. In cells, **1-10** could be a useful probe to study the KIX interactions on CREB(pKID)-responsive genes because the **1-10**-KIX N627C complex allosterically prolongs this interaction. The **1-10** disulfide was converted into a series of irreversible analogs that are compatible with the reducing environment in cells. This irreversible Tethering strategy demonstrated how fragments such as **1-10** selectivity have the potential to selectively target an engineered cysteine-containing KIX mutant of CBP *in cellulo*.

In conclusion, I built upon the Tethering strategy to selectively target dynamic activator-GACKIX complexes with small molecules; these fragments along with other biochemical techniques demonstrated how KIX regulates its interactions with different partners.

Chapter 1 Transcription is regulated by elaborate choreography between activators and coactivators.

1.1 Introduction

Transient, low affinity protein-protein interactions (PPIs) are primed for regulating biological processes by rapidly switching partners and engaging different pathways in response to environmental cues. Transcription is a process that depends heavily upon dynamic PPIs between activator and coactivators for proper gene expression. For instance, the KIX domain of the coactivators CBP and its homolog p300 recognizes over a dozen activators to regulate diverse physiological processes from development to cognitive function. Activators induce conformational changes within KIX that impact the functional binding at two intrinsic binding surfaces. This introductory chapter describes how activator-coactivator KIX domain complexes regulate transcription and discusses approaches and challenges to understand their interactions and functions.

1.2 What is transcription?

Transcription is the first step to produce a protein and so it is essential for life. The definition of transcription is quite simple – the process by which DNA is converted into RNA.¹ A transcription reaction in a test tube proceeds by incubating a DNA template, the appropriate nucleotide triphosphates, and a RNA polymerase enzyme if a simple, single polypeptide such as the bacteriophage T7 RNA polymerase is the chosen enzyme because it contains high intrinsic affinity

and specificity for its own 23-base-pair (bp) promoter to initiate transcription (Figure 1-1A).²⁻⁴

In cells transcription becomes far more complex. Cells have diverse sets of genes encoding proteins for opposing processes such as cellular growth and apoptosis, so specialized transcriptional networks restrict RNA polymerase to given genes in a spatial and temporal manner.⁵ Even simple, unicellular prokaryotes such as bacteria regulate transcription through subunits known as sigma factors (σ).^{6,7} These factors recognize distinct sets of promoter sequences to direct the formation of the multi-subunit RNA polymerase.¹ Transcription proceeds once the four catalytic subunits of the RNA polymerase and the single regulatory σ factor assemble at a given promoter (Figure 1-1B).⁸

In higher eukaryotes, such as humans, transcription is highly regulated through the organization of the genome and intricate protein networks. The human genome contains approximately three billion base pairs of DNA so stuffing almost two meters of DNA into the nucleus requires finesse.^{9,10} Therefore, 146 base pairs of DNA wrap around a core of histone octamers to form chromatin and the repeating chromatin units form nucleosomes.^{11,12} They condense further into euchromatin to produce transcriptionally inactive regions.¹³ Histone tails contain multiple post translational modification site (PTM) for phosphorylation, methylation, acetylation, ubiquitination and other chemical changes.^{14,15} These modifications act as a "histone code" to regulate the accessibility of DNA and consequently transcriptional activity.¹⁶⁻¹⁹ Histone acetyltransferases (HAT)

transfer an acetyl group to lysine residues on histone tails which neutralize the positively-charged ϵ -amino group of lysine and weakens the histone-DNA interaction.^{20,21} These regions are indicative of transcriptional activity and vary across cell types.¹⁸

At accessible core promoters, transcription initiation requires the preinitiation complex (PIC) including RNA polymerase II (PolII), and the general transcription factors (TFIIA, TFIIB, TFIID, TFIIE, TFIIIF, and TFIIFH).²²⁻²⁴ Initiation is the highly controlled, rate-limiting step for gene expression because PolII

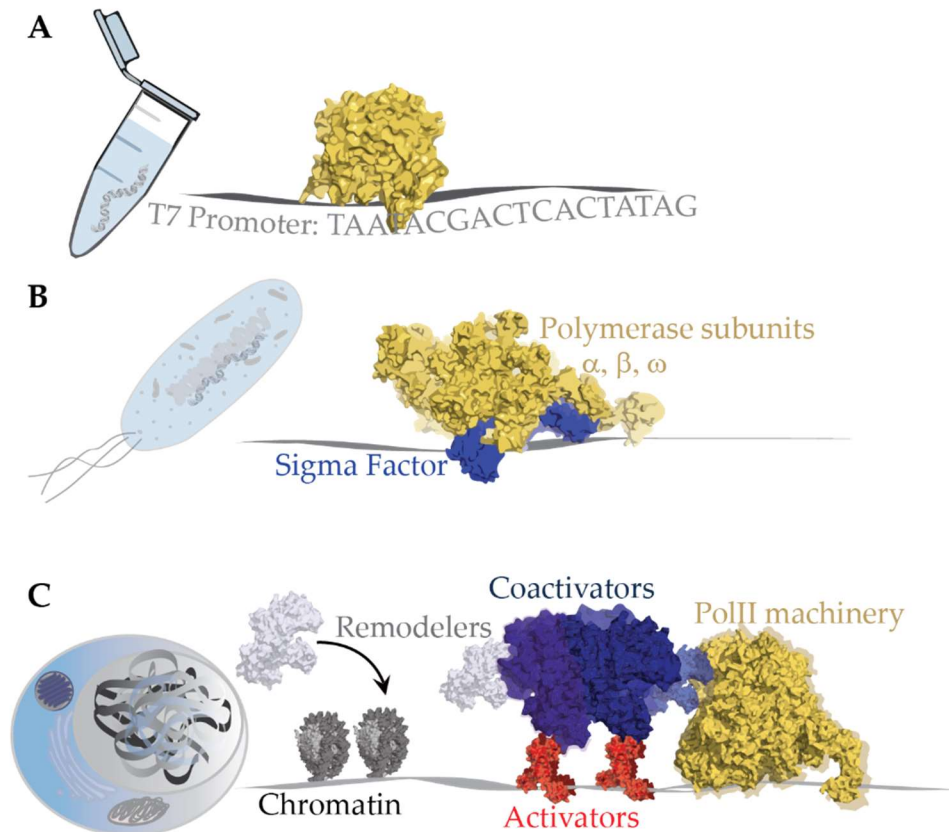


Figure 1-1 The activation of transcription is either simple or complex (A) In a test tube purified bacteriophage T7 RNA polymerase recognizes its promoter on a DNA template and catalyzes the synthesis of complementary RNA. PDB ID: 1CEZ (B) A model of how bacteria initiate transcription is shown with the assembly of sigma factor (blue), and RNA polymerase subunits (yellow). PDB ID: 4XLP (C) In multicellular organisms such as humans, DNA is compacted into nucleosomes (grey) and chromatin remodeling complexes modify these regions. To activate transcription TF (red) bind to DNA and recruit coactivators (blue). The coactivators recruit the PolII machinery (yellow) to initiate transcription. PDB IDs: 5FUR, 3AFA, 3K1F, 3RJ1, 4BHW, 3COQ

diffuses on and off the promoter resulting in pulses of mRNA production.²⁵ Site-specific transcription factors (TF) or activators promote active transcription by stabilizing the PIC at specific promoters.^{26,27} They bind to consensus sequences using a DNA-binding domain (DBD).²⁸ However, TF don't coordinate the assembly and stabilization of the PIC alone.²⁶ They lure in coactivator(s) through a flexible or intrinsically disordered recognition motif known as the transcriptional activation domain (TAD) (Figure 1-1C).²⁹⁻³¹

1.3 Coactivator play pivotal roles in mediating transcription

Coactivators link the requests of activators to the transcriptional machinery and they function as large dynamic and modular complexes.^{32,33} Some coactivators complexes associate as exchangeable subunits.^{34,35} Others connect multiple domains through flexible, intrinsically disordered linkers which permit them to independently interact with diverse TF.^{36,37} The individual interactions contribute to the overall function of the larger coactivator complex. Coactivators such as CBP and its homolog p300 coordinate interactions with over four hundred TFs.^{38,39} These interactions are mediated through the five domains; nuclear coactivator binding domain (NCBD) or interferon-binding domain (IBid),^{40,41} CH1 (TAZ1),^{42,43} CH2 (Bromo, PHD),⁴⁴⁻⁴⁶ CH3 (ZZ and TAZ2),⁴⁷ and KIX (Figure 1-2).⁴⁸ These master coactivators also contain a HAT domain which modifies both nucleosomes and other transcriptional regulators including itself.⁴⁹⁻⁵¹ The CBP and p300 HAT domain autoregulatory loop competes with the substrate-binding site which requires hyperacetylation and structural changes to fully activate the HAT domain.⁵¹ Structural changes imparted by the loop contribute to active transcription because it causes CBP or p300 to dissociate from the Mediator complex. Then TFIID containing the TATA-binding protein has access to the

promoter.^{52,53} The HAT domain regulates TF activity by altering protein or DNA interactions.⁵⁴ With the KIX-related TFs, c-Myb and CREB acetylation within their TADs enhances their transactivation potential by increasing their affinity for CBP.^{55,56} In the case of TFs such as p53 and SREBP1a, acetylation prevents ubiquitination and subsequent degradation.

CBP and its homolog p300 are pivotal for many cellular processes and consistent with their importance, the knock-out of CBP or p300 in mice is lethal.⁵⁷⁻⁵⁹ They share 61% overall sequence identity and 80-90% homology across conserved domains.⁶⁰ Due to their similarities CBP and p300 functions overlap acting as a buffer to make genetic networks more robust to cellular changes.^{61,62} They also maintain distinct tissue-specific and cell-specific processes.⁶³ From transfection-based cellular assays, most DNA-bound TFs stimulate transcription in the presence of either CBP or p300.⁶³ Therefore, their divergent functions depend on specific cellular contexts such as temporal-spatial regulators, or

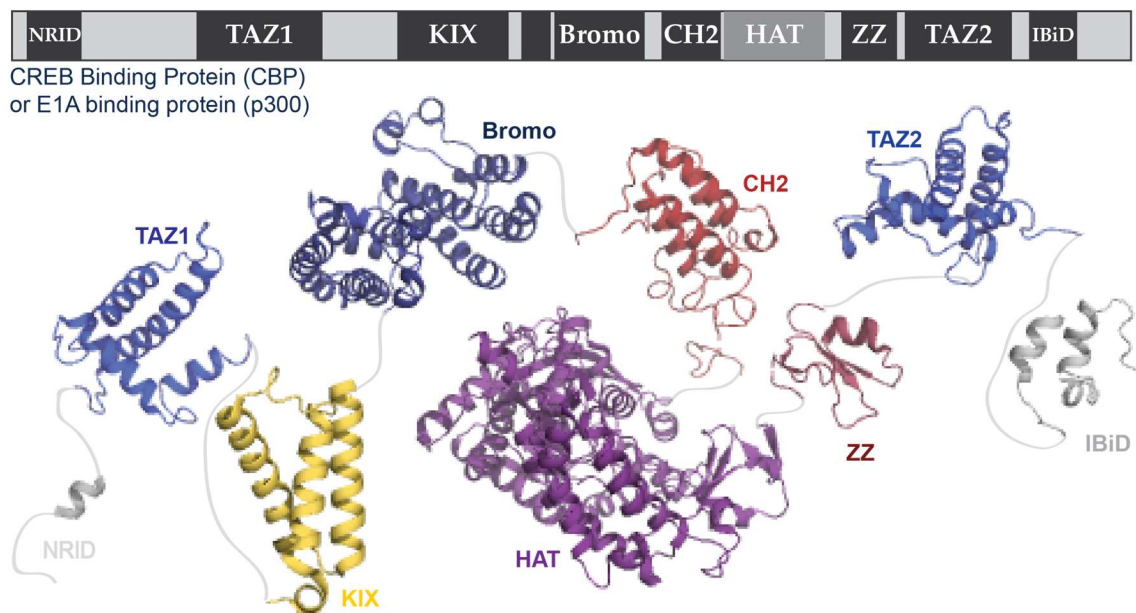


Figure 1-2 The architecture of CBP/p300 The schematic shows the arrangements of the CBP and p300 domains. PDB IDs: 1RDT, 1U2N, 4I9O, 4BHW, 3BIY, 2N1A, 5HP0, 1KBH

physiological cues.⁶³⁻⁶⁵

The core transcriptional players such as PolII and coactivator subunits are highly conserved across metazoans.^{66,67} This shows how accurate control over gene expression is vital for physiological processes.⁶⁸ To control these processes a highly versatile motif diverges into different co-regulators and contributes to other transcriptional events. This can be seen with the GACKIX motif.

1.4 What constitutes a KIX domain?

The GACKIX motif is found in mammals, yeast, plants, arthropods, and nematodes and it can tightly and efficiently regulate PPIs.⁶⁹ In the plant *Arabidopsis thaliana*, KIX domain proteins repress development and cell division through interactions with PEAPOD proteins and the RNA polymerase subunit B2.⁷⁰ Reduced expression of *KIX8* and *KIX9* genes results in higher crop yield due to increased leaf growth.^{71,72} In the fungi *Saccharomyces cerevisiae* and *Candida* species, the MED15/Gal11 subunit in the tail module of the Mediator complex contains a GACKIX motif that interacts with the pleiotropic drug resistance TF (Pdr1) to express genes in response to xenobiotics.⁷³ The genes encode membrane-spanning efflux pumps to expel toxins. The pathogenic fungi *Candida glabrata* resists antibiotics such as azoles through constitutive Pdr1 activity.⁷³ MED15/Gal11 KIX also participates in fatty acid metabolism. The TF Oaf1 recognizes fatty acids such as oleic acid, myristic acid, and lauric acid, which promotes its interaction with the MED15/Gal11 KIX domain to turn on fatty acid β -oxidation and peroxisomal genes.⁷⁴

In humans, the GACKIX motif is found in at least four transcriptional regulators (Figure 1-3A).⁷⁵ The KIX domain of CBP (residues 586-672) and p300 (residues 566-652) differ by nine amino acids and interact with the largest variety

of TF through two intrinsic binding sites. These interactions regulate diverse processes which deserve a more in-depth discussion in section 1.6. The ARC105/MED15 KIX domain is similar to the yeast MED15/Gal11 domain and it also regulates metabolic responses.^{60,69} Finally, a domain within the human RecQ protein-like 5 (RecQL5) minimally resembles the other GACKIX motifs yet functions similarly to the *Arabidopsis* KIX8 and KIX9 domains - it interacts with PolII to repress transcription during DNA damage response.

In terms of structure the GACKIX motif is small consisting of about 75 to 90 amino acids, and it is modestly stable (the free energy of unfolding, ΔG^0 of 3.5 – 4.1 kcal·mol⁻¹ for CBP KIX) at physiological pH.⁷⁶ The KIX domain of CBP is composed of three α -helices (α_1 597-611; α_2 623-640; α_3 646-669) with an amino-terminal 3_{10} helix (G_1 591-594). The carboxyl-terminus of α_1 connects by a loop (L_{12}) to a second 3_{10} helix (G_2 617-621). Core structural residues are highly conserved including W591, Y649 and the π -cation interaction between Y640 and R600 (Figure

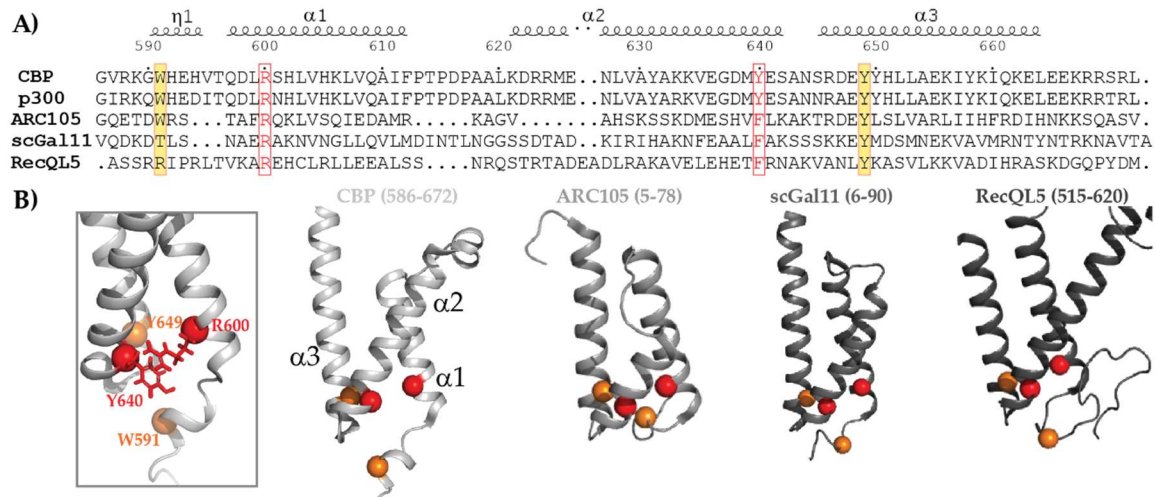


Figure 1-3 Comparison of GACKIX motifs A) The amino acid sequences are aligned using Crustal Omega (EMBL-EBI) and then compared to the structure of CBP KIX with ESPrIPT. B) In the box, the conserved core residues are shown on CBP KIX with the orange spheres as Y649 and W591 and the red spheres for Y640 and R600. To the right the full CBP KIX domain is shown as a grey-colored cartoon with its respective helices marked. Other available structures for GACKIX motifs share some of these core residues which are shown on the cartoon structures as spheres. PDB IDs: 2LXT for CBP KIX, 2GUT for ARC105/MED15, 2K0N for scGal11, 4BK0 for RecQL5

1-3B).⁶⁹ In the case of CBP, R600 is methylated by the coactivator-associated arginine methyltransferase 1 (CARM1) displacing activator-KIX interactions.^{77,78} R600 does not directly interact with activators so the methylation likely disrupts the structural integrity of the three-helix bundle.⁷⁹ KIX adopts a molten globule state *in vitro* when this arginine is mutated to methionine.

By altering certain residues, the GACKIX motif adjusts its binding partner preferences (Figure 1-4).⁸⁰ ARC105/MED15 KIX binds specifically to the sterol response element binding proteins, SREBP1a and SREBP2 which activates genes required for fatty acid, triglyceride, and cholesterol biosynthesis.⁸¹⁻⁸³ Within CBP KIX Y658 and K662 promote interactions between pKID and c-Myb, but the MED15 KIX domain lacks these residues rendering it unable to bind to them.⁸¹ The double mutant I64Y and D68K MED15 KIX can bind to c-Myb and pKID.⁸¹

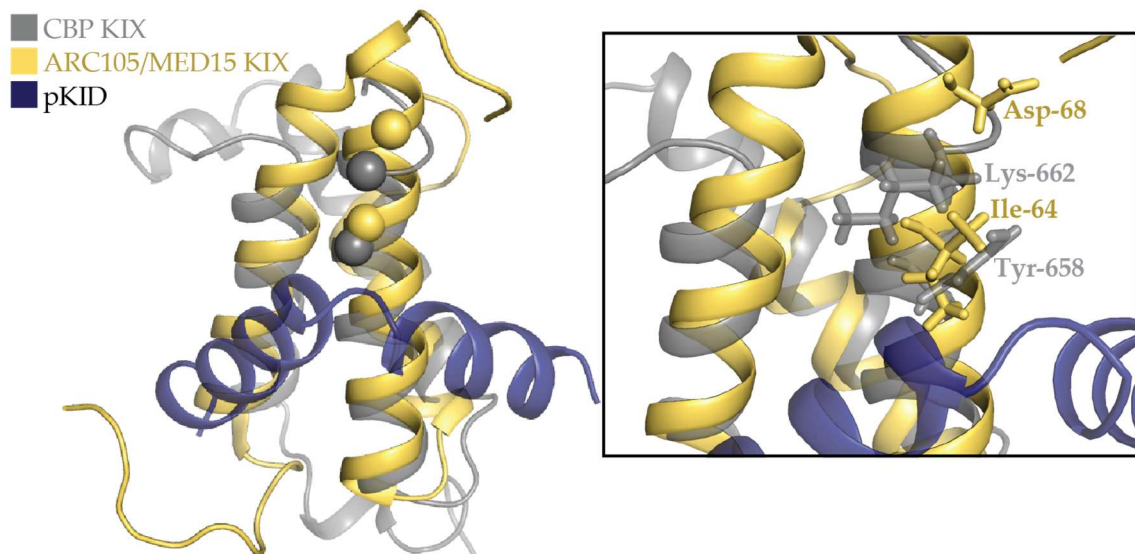


Figure 1-4 Distinct residue differences between GACKIX motifs The cartoon schematic shows the structural alignment for the NMR solution structure of CBP KIX in complex with pKID against the ARC105/MED15 KIX domain. Within the box the residues shown as sticks are important for the pKID interaction with KIX. By mutating these residues within ARC105/MED15 KIX it regains some ability to bind to pKID. PDB ID: 2GUT represents the NMR solution structure for human ARC105/MED15 KIX. PDB ID: 1KDX represent the NMR solution structure for mouse CBP KIX in complex with pKID.

Therefore, MED15 KIX with similar structural and functional characteristics as CBP or p300 KIX adopts limited, but specific binding partners. The binding profile of these coactivator domains may confer important transcriptional functions since CBP and p300 KIX also interact with SREBPs.^{84,85}

The challenge is to understand how CBP KIX recognizes a broad range of activators with disparate sequences to control gene expression. KIX consists of two defined binding sites often referred to as the “MLL site” and “pKID/c-Myb site” from the prominent activators that bind to the respective sites. Insights into their biological functions comes from first understanding their molecular and biophysical characteristics.

1.5 How do different TADs interact with the KIX domain of CBP?

The KIX binding partners share a distinct feature – their interacting regions contain various degrees of intrinsically disordered regions (IDRs) with minimal sequence similarities (Figure 1-5).

		ΦXXΦΦ
CREB 119-146	TDSQKRREILSRRPS	YRKILNDLSSDAP
c-Myb 284-315	GYNDEDPEKEKRIKE	LELLLMSTENELKGQAL
MLL 2826-2893	SDDCGNILPSD	IMDFVLKNTPSMQALGESPE
E2A-PCET 10-26		VGTDKELSDLLDFSMF
c-Jun 60-89	LLTSPDVGLLKLASPE	LERLI IQSSNGHIT
Tax 59-98	YIDGRVIGSALQFLIPR	LPSFPTQRTSKTLKVLTPPITHTT
HBZ 16-51		EDLLVEELVDGLLSLEEELKDKEEEKAVLDGLLSLE
p53 14-61		LSQETFSDLWKLLPENNVLSPLPSQAMDDLMLSPDDIEQWFTEDPGPD
FOXO3a CR2C		HYGNQTLQDLLTSDLS
FOXO3a CR3		GSLECDMESIRSELMDA

Figure 1-5 Alignment of TADs that interact with the KIX domain of CBP/p300 These KIX recognition sequences contain minimal similarities but conform to a generic motif of ϕ -X-X- ϕ - ϕ where ϕ is a bulky hydrophobic residue (orange text) and X is any residue (grey text).

These TADs lack hydrophobic cores that would drive them into independently, stable structures. They are enriched in disorder-promoting (A, G, S, Q, P) and charged residues (D, E, K, R).⁸⁶⁻⁸⁹ In other words, IDRs are minimalists as well as manipulators. They exploit the folding process to direct the function of

their binding partner(s) as they adopt structures against any partner that contains complementary chemical and steric properties.

Their recognition motifs make them highly selective in a complex cellular environment, yet they act as versatile regulators in a process known as “moonlighting”. This provides complexity to physiological processes.⁹⁰ They regulate diverse functions depending upon the combinations of partners directed within a hub such as CBP or p300.²⁸

The IDR of TF rely upon their highly versatile binding schemes; these raise the nagging question of what advantages disorder provides to biological function? Of course, IDRs have fewer steric and orientation restraints than globular proteins so they adapt to extended and complex binding surfaces. However, the biological advantage of IDRs is more complex and likely varies with each PPI and its concomitant binding affinity and specificity. The TADs of CREB (pKID) and c-Myb utilize a similar binding site around the α_1 and α_3 helices of KIX, yet their recognition of KIX occurs through distinct binding modes possibly reflecting their functions (Figure 1-6A). The phosphorylation of CREB (pKID) at the serine residue 133 promotes its interaction by 20-fold over the unmodified form. KIX also preferentially binds to pKID over c-Myb by approximately 7-fold.⁹¹ The magnitude of gene expression correlates to the strength of activator-coactivator complexes and this case it discriminates between inducible (CREB-dependent) and constitutive (Myb-dependent) gene expression.

The debate continues for what allows these IDRs to achieve functional binding affinity and specificity. Initially, they recognize partners by coupled folding and binding mechanisms. With an induced fit process, specific structural changes occur within the IDR that distinguish it from the bound versus the unbound state. The enthalpy of the complex compensates for the entropy loss from

the disorder-to-order transition. This requires an initial favorable interaction between the IDR and its partner before it commits to the complex. The pKID TAD follows this process to form two α -helices around KIX.⁹² The pKID-KIX complex initially forms weak hydrophobic contacts in a diffusional encounter complex followed by a folding intermediate where the α B of pKID remains partially unfolded (Figure 1-6B).^{93,94} Then the pKID· KIX complex stabilizes through complementary electrostatic and hydrophobic contacts in an enthalpy driven process.^{95,96}

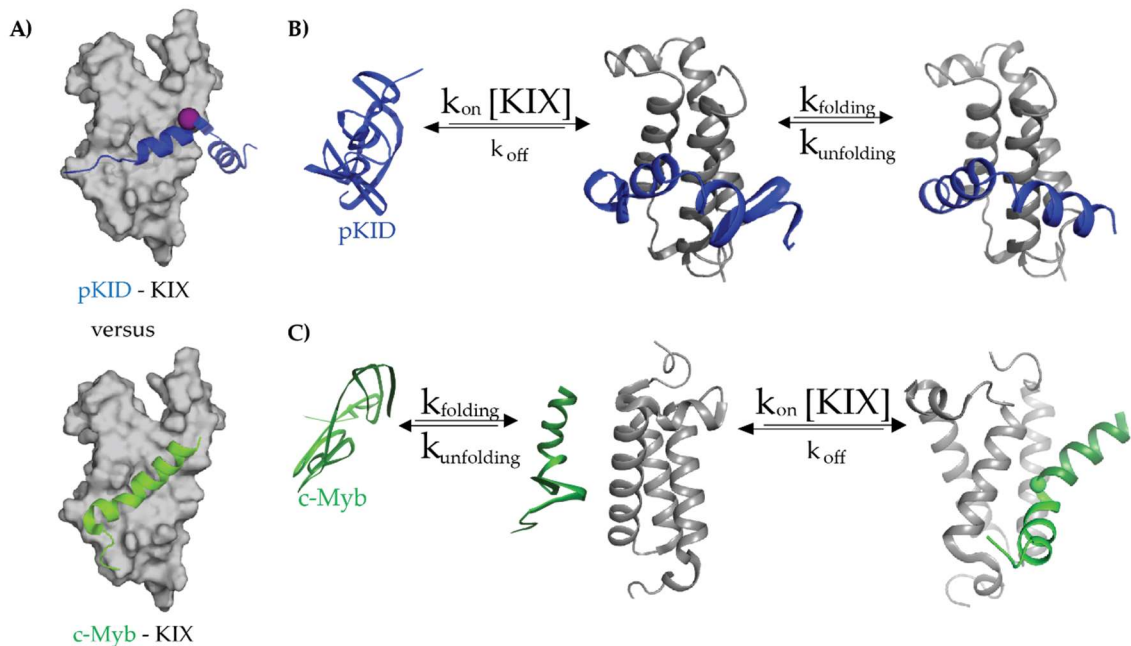


Figure 1-6 Schematic for the interactions of pKID and c-Myb with KIX (A) The TADs pKID (blue with purple sphere for the phosphorylated serine) and c-Myb (green) share an overlapping site within KIX (grey surface structure) (B) pKID is significantly disordered in solution and binds to KIX through an encounter complex intermediate before fully folding into two helices to form a stable complex with KIX. (C) c-Myb is largely disordered, but rapidly forms partial helical structures in solution. c-Myb binds to KIX through both a conformational selection and induced fit mechanism forming one stable helix. PDB IDs: 1KDX, 1SB0

Along with the coupled folding-after-binding process, the “fly casting” hypothesis purposes that the IDR flex out into a larger radius so they can rapidly initiate interactions by capturing their partners through complementary residues

on their flexible “fly lure”; the recognition motif within the IDR.⁹⁷ Coarse-grained molecular dynamics simulation using the pKID-KIX interaction shows that association rate constants (k_{on}) gradually increases with increasing disorder. The “fly casting” scheme is predicted to provide only 1.6-fold enhancement in the association rate⁹⁷ because the complex depends on its interaction strength at the binding site. As the disorder increases the intermolecular interactions must compensate for the entropy cost from the folding process. With too much disorder the intermolecular interactions cannot overcome the folding cost. Therefore, a good quality fly lure adds to a complex’s affinity by not only decreasing the association rate, but by enhancing the dissociation rate to an extent.⁹⁸

Electrostatic “steering” can contribute significantly to k_{on} and allows rates to exceed $10^7 \text{ M}^{-1}\cdot\text{s}^{-1}$.⁹⁹ These interactions are important for the formation of activators and coactivator domain complexes. The basal association rate constant measured at infinite salt concentration shows how an IDR produces a productive complex depending only upon diffusion and collisions between proteins. For KIX its partners reach basal association rate constants up to $10^5 - 10^6 \text{ M}^{-1}\cdot\text{s}^{-1}$ following diffusion-limited processes.^{93,100-102}

With conformational selection, an IDR freely acquires bound-like structures which stabilize it upon colliding into its partner. The residual structures reduce the entropic cost. The complex formation depends on the concentration of the high energy bound-like conformer and conformational selection should accelerate the association, k_{on} . However, an increase in k_{on} concretely correlates to a reduction in the free energy of the rate-limiting transition state, not necessarily due to a pre-formed conformer. Folding intermediates may increase k_{on} as the complex proceeds to stabilize helical, bound-like states. Additional structural evidence is required to show that the bound-like conformer contributes to the acceleration of

k_{on} through a conformational selection process.

An IDR may follow both mechanisms as seen with the c-Myb-KIX interaction which depends upon both favorable enthalpy and entropy.⁹⁶ The c-Myb32 (residues 284-315) tends to form an overall 80% helical ensemble freely in solution. However, c-Myb has disproportionate regions of helicity. The amino terminal from residues 290-301 populates an estimated 70% helicity while the propensity for spontaneous helix formation for the carboxyl-terminal region (302-310) is significantly smaller (~20%).¹⁰⁰ While c-Myb binds extremely fast proceeding through a two-step mechanism without observable intermediates, the disorder-to-ordered transitions for c-Myb23mer are faster along the estimated helix-coil transitions nanosecond time range.¹⁰³ Therefore, the helical ensembles transition more quickly than the rate at which c-Myb binds to KIX.¹⁰⁰ Due to the large helical propensity for the amino-terminal segment, c-Myb likely binds to KIX through a conformational selection model with fast pre-equilibrium while the carboxyl terminal proceeds through an induced fit process (Figure 1-6C). Isolating the amino-terminal region of c-Myb shows that increasing its helicity correlates directly to an increase in k_{on} .¹⁰⁰

Conflicting evidence argues for a purely induced-fit model.¹⁰⁴ For instance, chemically stabilizing the helicity of a shorter c-Myb construct with 1,1,1-trifluoro-ethanol (TFE) lowers k_{off} and has no significant effect on the overall k_{on} .¹⁰⁵ The additional electrostatic forces within the longer c-Myb construct may also contribute to the accelerated k_{on} rate because the basal association rate constants appear to form on the same time scale for the longer and shorter c-Myb forms

c-Myb TAD AAAAIQRHYNDEDPEKEKRIKELELLMSTENELKGQQVLPTQNHTCSYPGW
c-Myb25mer KEKRIKELELLMSTENELKGQQV
c-Myb32mer GYNDEDPEKEKRIKELELLMSTENELKGQQAL
oooooooooooooooooooooooooooo

Figure 1-7 Sequences for different c-Myb constructs Different KIX constructs have been used in studying its binding to KIX. One main difference in these constructs is that the amino terminal of KIX tends to form helical structures freely in solution. The leucine residue at 302 written in yellow text drives the interaction with KIX by positioning deep into a hydrophobic region.

(Figure 1-7).^{104,106} This represents how subtle changes within TF and other environmental conditions must be considered when comparing these complexes. General models remain consistent that c-Myb associates very quickly with KIX driven by the L302 that orients within a hydrophobic interface within KIX. Meanwhile the binding mechanism requires additional investigations to hash out this complex both *in vitro* and then *in cellulo*.

The MLL TAD binds to KIX at a different site formed by the G₂, α₁ and α₃ helices. MLL (residues 2844-2857) is only 13% helical in solution by circular dichroism (CD), and it forms a helical structure with 15 residues in a complex with KIX.¹⁰⁷ MLL induces a network of residues from the L₁₂-G₂- α₃ interface to reposition the G₂ helix and these structural changes promote energetically favorable hydrophobic contacts between KIX and MLL. Most notably F612 contacts MLL while other critical residues such as L628, Y631, L664 (to a minor extent I611, and I660) are also important for binding MLL (Figure 1-8).¹⁰⁸ As discussed with the c-Myb TADs, MLL constructs and changes in experimental design alter this complex. A shorter MLL construct lacking three negative charges outside of the helical region accelerates rather than decelerates with increasing salt concentration.¹⁰⁹ It's likely that the salt acts as a stabilizer for electrostatic charges near the MLL·KIX complex.¹⁰⁹ The longer MLL construct follows the general trend that the additional negative charges in MLL accelerate its association with

KIX.^{106,107}

Other activators utilize the same site as MLL with distinctive binding affinities and binding modes (Table 1-1). For instance, in differentiation and apoptosis a phosphorylated region of c-Jun (residues 47-89) is shown to contact KIX with moderate affinity.¹¹⁰ Another TF E2A principally recruits CBP or p300 through the KIX domain. The first activation domain (TAD1) of E2A (10-26) is 1 to 2% helical yet binds very quickly to KIX in an induced-fit process.¹⁰⁶ E2A contains another TAD2(294-407) that binds weakly to the same region within KIX. Another partner from a chromosomal translocation fuses the CREB-regulated transcriptional coactivator (CRCT1) to mastermind-like 2 (MAML2) coactivator and the product CRCT1-MAML2 TAD (150-190) directs the recruitment of CBP or p300 through KIX. CRCT1-MAML2 joins the activator club (c-Myb, pKID, MLL)

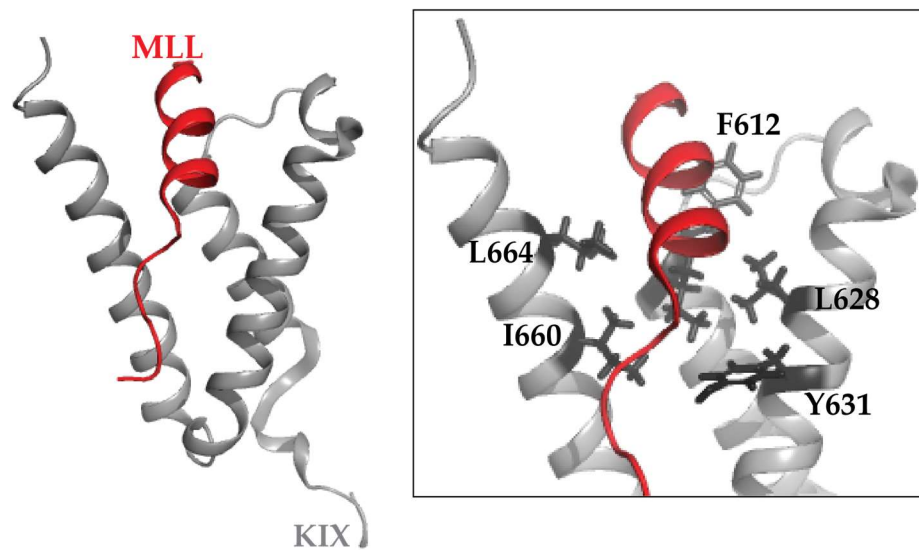


Figure 1-8 Cartoon representations of the KIX-MLL complex KIX is shown in grey and MLL is in red. The important residues that make-up the MLL-binding site within KIX are shown as black sticks. PDB ID: 2LXS

who contact the other site of KIX with low affinity (K_{DS} greater than 100 μM).¹¹¹

The human T-cell leukemia virus type 1 (HTVL-1) viral TF, Tax is intrinsically disordered with two possible helical regions (residues 60-71 and 83-89) that can contact the MLL-binding site of KIX.¹¹⁸ Another viral TF, HTVL-1 basic leucine zipper factor (HBZ) utilizes the VDGLL (residues 24-28) and LDGLL (residues 44-48) motifs to bind with high affinity to KIX.^{119, 27} These viral TFs may prefer the MLL-binding site, perhaps, because of its chemical and steric features that permit higher affinity interactions. This site also appears more “druggable” by its contacts with small molecule partners than the pKID-binding site as discussed in section 1.9. However, the endogenous TF·KIX relationship remains under investigation – especially for how they utilize this site.

Table 1-1 Comparing K_D s for the TADs at the MLL-binding site The reported dissociation constants (K_{DS}) are shown for different TADs that are known to bind within the MLL-binding site of KIX. The K_{DS} were measured as indicated with isothermal calorimetry (ITC), nuclear magnetic resonance spectroscopy (NMR), fluorescence polarization (FP) or calculated from stopped-flow transient kinetics experiments.

TAD	K_D (μM)	Method used
MLL19 (2844-2854)	2.8 ± 0.4	ITC ¹¹²
MLL (2842–2869)	2.1 ± 0.1	ITC ¹¹³
pJun (47-89)	30	NMR titration: ³¹ P chemical shifts ¹¹⁰
E2A-PCET (10-26)	12 ± 2	FP ¹¹⁴
E2A TAD1 (11-24)	5.2 ± 0.9	FP ¹¹⁵
E2A TAD2 (394-420)	21 ± 3	FP ¹¹⁵
CRCT1-MAML2 (150-190)	20 ± 5	ITC ¹¹⁶
HBZ (1-77)	2.3 ± 0.2	Calculated from stopped-flow, $K_D = k_{off}/k_{on}$ ¹⁰⁶
Tat (1-24)	11 ± 1	NMR chemical shift titrations ¹¹⁷

1.6 KIX masters intermolecular and intramolecular communication.

Although the flexibility of IDRs governs its interactions and functions - these also depend upon the dynamics of their binding partners. For instance,

conformational changes within KIX dramatically alter binding partner preferences at the second site leading to downstream effects in the transcriptional outcome.¹²⁰ This is shown by the binary KIX·MLL complex enhancing the affinity of pKID or c-Myb at the other site by 1.5 to 2-fold and vice versa. This process known as allostery occurs when a ligand at one site affects the function of another site through long-range, indirect changes. The cooperative effect offers another level of control to regulate PPIs.¹²¹

Allostery occurs through different mechanisms involving structural, thermodynamic and dynamic changes. In a purely dynamic process a ligand rigidifies its partner and limits its thermal fluctuations.^{122,123} Entropy drives the binding of the second ligand and a slight stiffening or limiting of motion within a protein provides favorable energy of a few kcal·mol⁻¹.¹²⁴ When the affinity increases for the second ligand then its binding mode changes by either how fast it associates or how long it resides with its partner. When the binary complex stabilizes the flexible and unfolded regions of the main partner or favorable intermolecular contacts become more available, the second ligand may associate more quickly increasing k_{on} .⁹⁹ The ternary complex can prolong the residence time of both ligands over the binary complex as shown by decrease in k_{off} . This allosteric mechanism appears to contribute to the ternary KIX complexes.

The allosteric effect with pKID depends on the stabilization of the MLL-KIX complex so entropy drives the allosteric effect.¹¹² The cooperativity between c-Myb and KIX·MLL occurs through an unfavorable decrease in entropy and a favorable increase in enthalpy. Hydrophobic contacts in the middle and C-terminal regions of the third helix contribute to binding while some favorable energy come from the KIX·MLL complex pushing residues E666 and E665 closer to the c-Myb residues K291 and K294 to promote favorable electrostatic side-chain

interactions.^{43,108} The c-Myb interaction with MLL·KIX lowers the k_{off} for the MLL·KIX·c-Myb complex by a factor of two.¹⁰⁹

KIX contains an allosteric network consisting of hydrophobic amino acids which MLL and c-Myb or pKID engage to stabilize flexible regions through the α_3 and the L₁₂-G₂ loop of KIX (Figure 1-9).^{108,125} From relaxation dispersion data for backbone amide ¹⁵N and ¹³C α nuclei the KIX·MLL complex is shown to exist as a disparate ensemble with 7% populating a higher energy conformation that displays a higher affinity for either c-Myb or pKID than the major lower energy state. The two states maintain an equilibrium. As the higher energy state is depleted through the formation of the ternary KIX complex additional higher

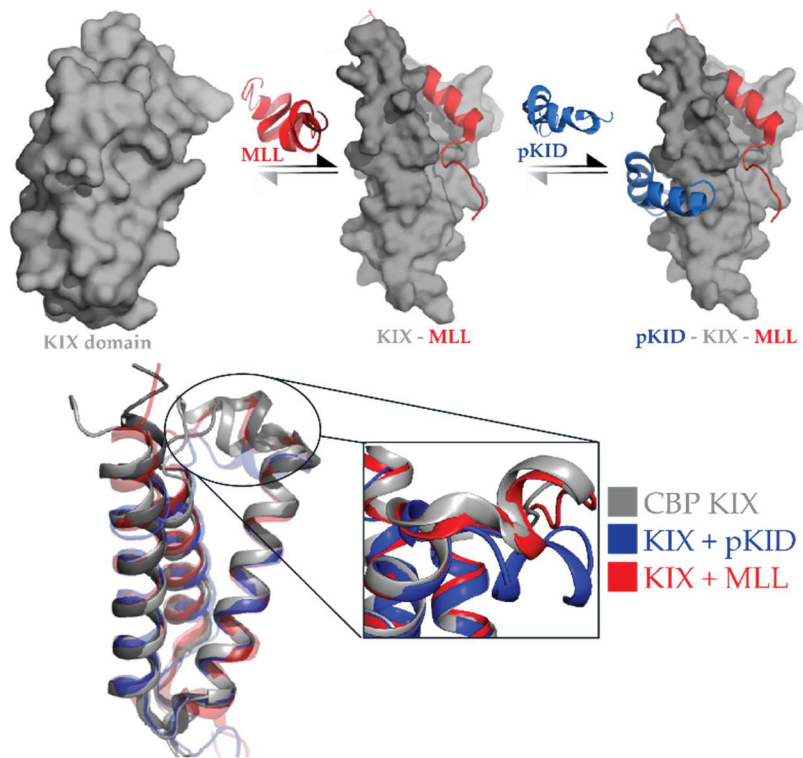


Figure 1-9 KIX contains an allosteric network. The first schematic shows the transition from apo KIX (grey) to binary KIX-MLL(red) and finally the ternary KIX-MLL-pKID(blue) complex. When MLL binds then pKID binds with a 2-fold higher affinity at the other site and vice versa. As shown in the cartoon KIX is flexible with the largest changes in the loop regions. These conformational changes contribute to the cooperativity between the two binding sites. PDB IDs: 2AGH, 1KDX, 2LXS, 4I9O

energy complexes form.¹⁰⁸ This suggests that conformational selection may contribute to the positive cooperativity. However, this contribution remains under debate because this favorable bound-like excited state should accelerate k_{on} yet its change is negligible for c-Myb and pKID with KIX·MLL using stopped-flow transient kinetic analysis.¹⁰⁶ This excited state may contribute to dynamic and structural allostery in which the MLL·KIX complex reduces the entropic cost for the second ligand by stiffening a favorable ternary conformation to decrease k_{off} for the ternary complex. However, this requires additional mechanistic and structural insight into the allosteric network.

1.7 These additional binding partners like to hug KIX

TFs regulate genes by their spatial and temporal presence. They make themselves or cellular signals more robust through factors such as competing with other TF for coactivators as well as utilizing alternative interactions to remain poised for active gene expression.¹²⁶ The latter may explain why FOXO3a and p53 contain tandem TADs that manipulate KIX through both binding sites. The FOXO3a construct consisting of the carboxyl-terminal segment of CR2 (CR2C) connected to CR3 with the (GGGS)₃ linker binds to KIX with two possible orientations that occur in equal populations.¹²⁷ However the binding modes are distinct. FOXO3a may appear promiscuous, but if one site within KIX is restricted then FOXO3a adopts another orientation to encourage the coactivators to remain in the vicinity.

The tumor suppressor p53 also interacts with KIX with tandem TADs in two different orientations to each of the binding sites within KIX.¹²⁸ These resemble “fuzzy” complexes because p53 retains disordered conformers in complex with

KIX. Both regions prefer the MLL-binding site although TAD2 outcompetes TAD1.

The moderate to low affinity ($86 \pm 18 \mu\text{M}$ for CR2C-CR3 FOXO3a and 14-27 μM for full-length p53) may support that these TF need to rapidly switch on and off. For active transcription these TF can further stabilize CBP and p300 because FOXO3a interacts with TAZ1(340-439) and TAZ2 (1764-1855) while p53 interacts with these domains it also associates with NCBD.^{126,127,129,130} These genes may require regulation from multiple transcriptional regulators such as a cooperative effort between themselves since promoters contain multiple FOXO3a recognition sites.¹²⁷ In both cases phosphorylation enhances their transcriptional activity. With FOXO3a phosphorylated S626 increases its affinity by 1.5-fold for KIX.¹³⁰ When the p53 TAD(13-57) is phosphorylated at Y18 and S20 then its affinity for KIX increases 15-fold relative to the unmodified form.¹²⁸ This demonstrates how KIX reacts to phosphorylation states to regulate its interactions.

1.8 Activators and coactivators are involved in diseases

Transcription regulates the variety and abundance of protein in cells. Differences between normal and disease states come from changes in transcription rates, messenger RNA degradation, translation and protein degradation, but transcription arguably makes the larger contribution to the biological outcomes.^{131,132} Since TFs drive gene expression, their actions govern downstream biological activity.^{133,134}

As highlighted in the previous two sections, KIX interacts with many different TF and these convey diverse functions (Figure 1-10). For instance, E2A regulates genes for cellular specification and maturation.¹³⁵ MLL regulates genes for hematopoiesis. Likewise, c-Myb regulates definitive hematopoiesis and it is down regulated as cells differentiate.¹³⁶⁻¹³⁹ CREB regulates hundreds of genes for

many processes. In response to various cellular cues CREB induces genes required for proliferation, survival and differentiation. Some of these depend on the interactions between CREB(pKID) and CBP or p300, but CREB utilizes other co-regulators.

Parsing these transcriptional pathways are important to understand physiological processes and diseases. Cancer represents a diseases-state where multiple processes are affected. Transcriptional networks are often hijacked which alters the concentrations and PPI networks.¹⁴⁰ For instance, the MLL chromosomal translocation fuses MLL to CBP to constitutively recruit CBP to MLL-regulated genes and constitutively express genes leading to overexpression of myeloid and lymphoid markers. E2A forms at least two chromosomal translocations associated with acute lymphoblastic leukemia. In one translocation, the E2A DBD is replaced with PBX-1.¹³⁵ E2A-PBX1 expression results in rapid proliferation of immortalized myeloid cells. This is also replicated by infecting primary bone marrow cells with a retrovirus expression E2A-PBX1 which leads to myeloid progenitors with sustained exponential growth. The CRTC-MAML2 translocation product activates

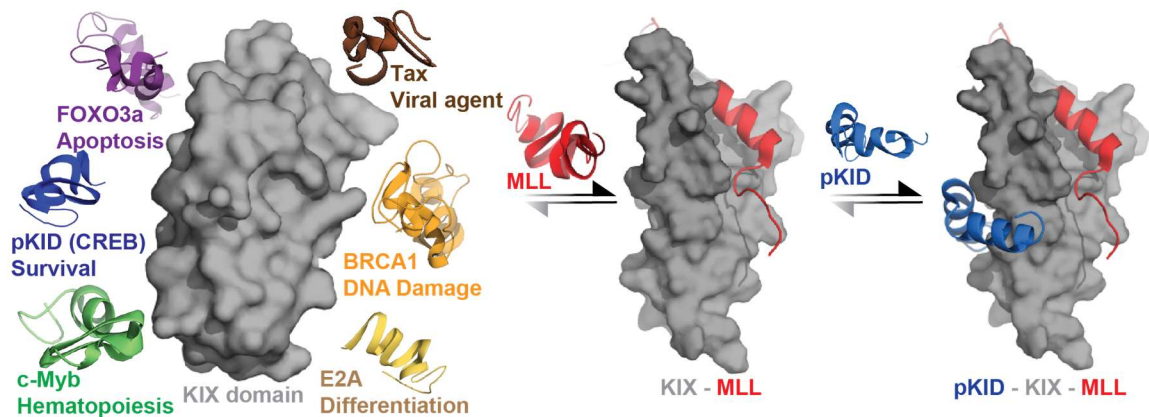


Figure 1-10 KIX has multiple binding partners. The activators that bind to KIX are associated with different biological processes as shown. These activators interact with KIX through at least of the two binding sites. PDB IDs: 2AGH, 2LXS, 2LXT

both myelocytomatosis oncogene (Myc) and CREB-responsive genes to drive tumor formations.¹¹⁶ The Myb proteins contribute to cancer through different schemes. In some cases, the *c-Myb* gene remains unaltered, but it is overexpressed. Cancer becomes “addicted” to it because it promotes undifferentiated growth. The *c-Myb* proto-oncogene likely contributes to proto-oncogenic activity because upon phosphorylation *c-Myb* may support normal cells through the G1/S phase in the cell cycle.¹⁴¹ The tumor suppressor p53 inhibits cell cycle progression, but *c-Myb* can compete with p53 to bind to KIX. This is one possible route for *c-Myb*’s involvement in leukemia.

Viral machinery hijack transcriptional regulators to replicate, and diseases result because of this viral activity. For instance, the virus HTVL-1 is present during the progression of adult T-cell leukemia (ATL). The viral TF Tax contacts the MLL-binding site of KIX to alter gene expression in cell cycle regulation and apoptosis early-on in HTVL-1 pathogenesis.¹¹⁹ Tax also directly stabilizes the DBD of CREB at sites within the HTVL-1 promoter. Although Tax hasn’t been shown to engage in the allosteric network Tax needs to directly recruit CBP or p300 in the absence of pKID.¹⁴² As Tax expression varies, HBZ can continue to disrupt gene expression. Due to its high affinity HBZ likely monopolizes the MLL-binding site of KIX and enhances the binding of *c-Myb* and pKID to KIX by 6-fold leading to proliferation of leukemia cells. Finally, from the virus HIV-1, Tat also competes with TF such as MLL, E2A and c-Jun for the MLL-binding site of KIX through its TAD (1-24) with moderate affinity.¹⁴² The HIV-1 genome is packaged into nucleosomes and requires HAT for its gene expression. Since these viruses utilize the MLL-binding site of KIX the other TFs such as MLL, E2A and c-Jun must compete for CBP and p300 so these TF-regulated genes are also disrupted

contributing to the leukemogenic states.¹¹⁷

KIX directly corresponds with adverse phenotypes. This has been demonstrated with the triple mutation (Y650A, A654Q, Y658A) that disrupts its interactions with c-Myb and pKID. Mice homozygous for these mutations in p300 have defects in hematopoiesis-related processes involving platelets counts, B cells, T cells and red blood cells indicative of altered c-Myb-dependent gene regulation.^{143,144} The analogous triple mutant within CBP produce minimal effects, possibly reflecting different expression patterns or preference of p300 over CBP in hematopoiesis.¹⁴³ Another mouse model containing the triple mutant within CBP aims to test KIX-dependent pKID and c-Myb interactions on long-term memory. From behavioral analyses, the mice show specific impairments in long-term memory occurring in the hippocampus, but not in the amygdala.¹⁴⁵ This demonstrates that regions of the brain have different CREB-dependent transcriptional requirements. In the hippocampi, the triple mutant reduces several genes (*Fosb*, *Dusp1*, *Icer*, *Bdnf4*) that depend on the KIX-CREB(pKID) interaction.¹⁴⁵ Rubinstein-Taybi syndrome (RTS) results from a genetic alterations in the *CBP* gene.¹⁴⁶ The development disorder is characterized by skeletal and cardiac abnormalities as well as mental retardation and occurs in 1 out of 125,000 births.¹⁴⁷ RTS develop from the monoallelic mutation of CBP and the loss of KIX-based activity likely contributes to the mental retardation as well as other defects.^{148,149}

Subtle cellular changes affect the fine-tuned control between TF and coactivators yet parsing out the individual players remains challenging. In certain transcriptional networks TF contribute to an observable phenotype only under certain conditions. Some gene networks are more robust so if one or two players are lost the system may function.¹⁵⁰ These subtle networks make it challenging to predict the phenotype from the genotype. As discussed here changes in these

transcriptional networks are important for normal functions and even when their concentrations are tightly controlled. If TF concentration increases in cellular models, then TF may sequester coactivators beyond the DNA activating region. This squelching effect depends on the concentration of the activating regions and the strength at which the TF engages its target. At the same time, imbalances in transcriptional regulators can be subtle or lacking because the transcriptional complex assembly provides some degree of buffering.¹⁵¹ There is a need to draw out the interactions that are important to specific transcriptional processes.

1.9 Can small molecules effect and probe KIX·activator complexes?

KIX·activator complexes elicit gene activation in response to various stimuli, but the specific transcriptional mechanisms remain elusive. Small molecules that target these interactions with high affinity can be useful to discriminate their phenotypic consequences from endogenous, lower-affinity activators.

Small molecule development for TF and coactivators requires addressing many hurdles including their transient and moderate affinity, promiscuity or selectivity, and binding surface structures. In terms of structure, a PPI surface spans 1,000 to 2,000 Å² while small molecules cover an average of 300 to 500 Å². Therefore, small molecules must efficiently engage a region within the PPI surface to directly modulate the target interaction.

Structural characterization of binding interfaces aids in small molecules design because the atomic level resolution structure provides specific chemical features within a binding site. High resolution X-ray crystal structures require diffraction-quality crystals which flexible proteins such as KIX and its endogenous partners struggle to form. Structural information for the yeast and human MED15

KIX as well as CBP KIX complexes come primarily from solution NMR techniques.

In the case of selectivity, the GACKIX motifs in the master coactivators CBP and p300 contain over 90% sequence homology so specifically targeting these regions with small molecules is difficult. The discovery of specific small molecule modulators of activator-coactivator PPIs is rendered especially challenging by the level of redundancy in the protein interaction network.¹⁵²⁻¹⁵⁵

Small molecule probes need high affinity to efficiently engage the target and to indirectly reduce off-target effects. The KIX-activator interactions are optimized for efficient functional output – not necessarily high-affinity complexes. For instance, single mutations within CREB such as Y134F or KIX L607F increases their affinity for each other which would possibly switch CREB from an inducible to constitutive activator demonstrating the importance of endogenous, concomitant affinity.^{95,156} By using phage display peptides containing eight amino acids underwent directed evolution to identify sequences that bind with high affinity to KIX. Their affinity remains moderate within $16.1 \pm 1.1 \mu\text{M}$.¹⁵⁷ These peptides are short and flanking these sequences may improve the affinity and biological potency. Peptidomimetic strategies to modulate PPIs are available, however for the promiscuous binding between KIX and its partners their applicability for this system remains unexplored.

On the bright side KIX appears amenable to small molecule intervention. The pKID-KIX inhibitor, naphthol-AS-E-phosphate (KG-501) was identified from a screen using ^1H - ^{15}N heteronuclear signal quantum coherence (HSQC) spectra to monitor shifts in isotopically ^{15}N -labeled KIX (Figure 1-11).¹⁵⁸ Structure-activity relationship of KG-501 demonstrated that naphthol-ASE is the most active form in cell culture and disrupts CREB-dependent genes in AML.¹⁵⁹ The AML-treated cells undergo cell-cycle arrest and apoptosis illustrating how naphthol-ASE displays

potential efficacy. CREB-bound loci also show reduced histone 3 lysine 27 (H3K27) acetylation reflecting the loss HAT-based activity from the CREB-CBP interaction.¹⁵⁹ KG-501 also disrupts c-Myb from CBP and p300 and it has the potential to treat leukemia progression because it has promoted differentiation in myeloid leukemia cell culture models.¹⁶⁰ Two natural products, sekikaic acid and lobaric acid natural products, were identified from a fluorescence polarization (FP) screen that targeted the KIX-MLL complex (Figure 1-11).¹⁵²

Since these KIX screens have a very low hit rate fragment-based screening and design is an attractive method to design small molecules for these “target-rich, lead-poor” cases.¹⁶¹ Fragments need to be efficient binders with a higher energy per molecular mass unit and molecular weights between 150 to 250 Da.

Although fragments tend to efficiently bind to a target with a high

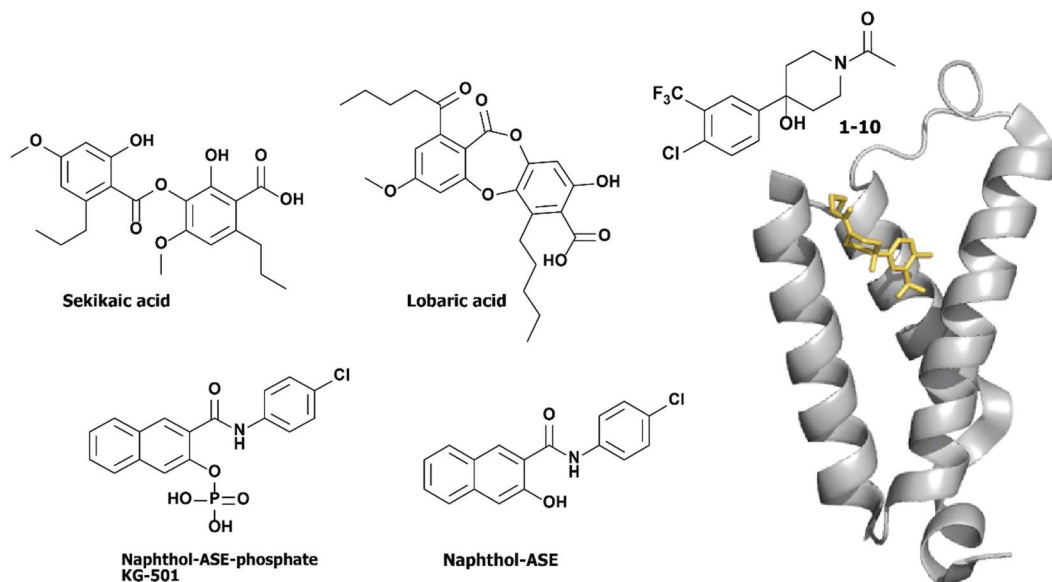


Figure 1-11 Small molecules that bind to KIX The structures for the small molecules are shown. Naphthol-ASE-phosphate and its dephosphorylated derivative inhibit the KIX interactions with pKID and c-Myb. Sekikaic acid and lobaric acid are natural products identified initially as inhibitors of the MLL-KIX interactions. The fragment core of **1-10** is shown. The **1-10** disulfide was identified from an LC-MS Tethering screen. The first crystal structure of KIX was obtained from the complex of KIX L664C mutant (grey) tethered to the **1-10** (yellow). PDB ID: 4I9O

proportion of their atoms their affinities range from mM to 30 μ M. Therefore, the fragment hits require optimization by either growing or linking chemical features on to it. Frequently growing-out the fragments rapidly result in higher affinities reaching the low to nanomolar range. Although linking fragments together results in a drug-like compound optimizing linkers are more difficult to achieve optimal and efficient binding affinity. However, this discussion is premature. First fragments need to be identified which is challenging considering their high μ M to mM affinities.

Techniques such as NMR and x-ray crystallography may detect fragments that target a specific PPI. Protein-observed fluorine NMR spectroscopy (PrOF NMR) is particularly sensitive and has identified fragments that target KIX. Here, 3-fluorotyrosine (3FY) replaces tyrosine naturally located in the MLL-binding site (Y631) and pKID or c-Myb binding site (Y649, Y650, Y658).^{162,163} Residue-specific ¹⁹F shifts indicates a favorable interaction between a fragment and KIX at a specific region and from 508 fragments 3 fragments perturbed the MLL-binding site and 1 fragment hit the pKID-binding site with K_{DS} ranging from around 1.5 to 7 mM. These scaffolds provide a framework to derivatize KIX inhibitors, however additional work is required to turn these fragments into probes.¹⁶³

From these screens KIX remains a difficult target with a range of hit rates (0-0.8%) resulting in ligands with moderate μ M to mM affinities.^{152,163} A site-direct strategy known as Tethering* is the next approach in the quest for the optimal KIX probe. Just to clarify; the optimal KIX probe would differentiate between p300 KIX, CBP KIX and ARC105/MED15 KIX (very selective), and efficiently capture specific

* Tethering was developed by Dr. James Wells and his team at Sunesis Pharmaceuticals and this technique was first reported by Erlanson *et al.*, *Proc. Nat. Acad Sci. USA* 97:9367-9372 (2000).¹⁶⁴ Concepts of disulfide Tethering are also patented (Erlanson, *et. al.* U.S Patent 6,919,178, 2005).¹⁶⁵

conformations of KIX (potent with K_i or IC_{50} within 10^{-7} - 10^{-9} M).¹⁶⁶

Tethering has built a reputation as a method to develop fragments for difficult targets.¹⁶⁷ This strategy requires a cysteine either endogenous or strategically positioned in proximity to the target site of the protein. This target is screened against disulfide-containing fragments using a reducing agent to vary the stringency of the assay (Figure 1-12). Through the thiol-disulfide exchange the fragment with affinity for the target is readily detectable as a covalent mixed disulfide using liquid chromatography mass spectrometry (LC-MS).¹⁶⁷

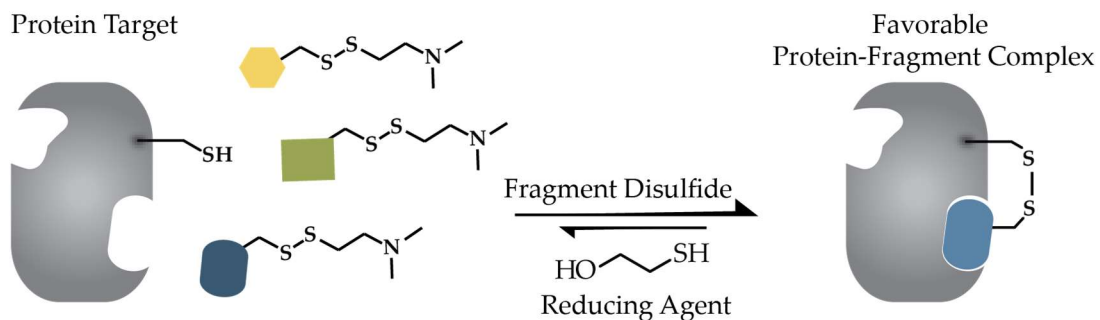


Figure 1-12 Disulfide Tethering scheme The protein target must contain a cysteine which is screened against a library of disulfide fragments. The stringency of the assay is controlled with a reducing agent such as β -mercaptoethanol as shown. The fragment with intrinsic affinity for the target predominantly forms a complex with the target at equilibrium.

Initially Tethering was used to identify fragments that manipulate malleable sites within proteins.¹⁶⁷ In one case, the immune response player, interleukin-2 (IL-2) appears to reveal a defined hot-spot region only upon binding to an inhibitor.¹⁶⁸ A fragment-based screen directed in the receptor binding site revealed a hot-spot region. The strongest hit from the Tethering screen formed the basis of a nanomolar small molecule that binds efficiently through the adaptive surface identified in the screen. Since then other Tethering screens have revealed both flexible sites that adapt to small molecules and novel allosteric regions.¹⁶⁹⁻¹⁷²

For KIX the LC-MS Tethering screen targeted at the MLL-binding site

resulted in promising fragments that capture conformational features of KIX. The KIX L664C mutant tethered to the fragment known as **1-10** forms a stable construct amenable to crystallography.¹⁷³ The fragment sits within the MLL-binding site where hydrophobic region including I611, L628, L607, V635, and Y631 hug the aromatic ring of **1-10**. KIX is highly dynamic so modifying **1-10** based on this structure remains difficult since KIX shifts in regions such as the loop between α_1 and α_2 to accommodate the ligand. From PrOF NMR the **1-10** fragment interacts with KIX with a dissociation constant, K_D of 210 μM .¹⁶² However, the main advantage of these cysteine-reactive fragments is that it targets the cysteine-containing KIX domain. Therefore, introducing a cysteine within CBP means the probes specifically target this coactivator over say p300. While these probes provide the first insight on an atomic level additional applications are required to delineate exact mechanisms for KIX-activator interactions in biological processes. These probes may be the start since they specifically target the cysteine-containing KIX domain of CBP versus p300.

1.10 What does this thesis offer?

This thesis focuses on the challenge to understand how conformational changes within KIX dictates its interactions, mainly using small molecule probes. The site-directed Tethering strategy is initially used to identify fragments that capture the conformations and interactions with KIX *in vitro*. In Chapter 2, a novel FP screen is used to identify novel fragments that disrupt the CREB(pKID)-KIX complex. In Chapter 3, transient-state kinetics of these tethered KIX complexes examines what contributes to these allosteric effects. These probes are prepared for *in cellulo* studies by converting the reversible covalent inhibitor into a series of irreversible covalent co-chaperones. From biochemical studies, the subtle chemical

changes between the fragment and irreversible “warhead” are tested for how they react with KIX *in vitro* (Chapter 4). In summary, this dissertation seeks to identify strategies and probes that examine the mechanism of KIX-activator complexes.

1.11 References

1. Ptashne, M. & Gann, A. *Genes and Signals*. (Cold Spring Harbor Laboratory Press, 2002).
2. Studier, F. W. & Moffatt, B. A. Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *J Mol Biol* **189**, 113-130, (1986).
3. Tabor, S. & Richardson, C. C. A bacteriophage T7 RNA polymerase/promoter system for controlled exclusive expression of specific genes. *Proc Natl Acad Sci USA* **82**, 1074-1078, (1985).
4. Cheetham, G. M., Jeruzalmi, D. & Steitz, T. A. Structural basis for initiation of transcription from an RNA polymerase-promoter complex. *Nature* **399**, 80-83, (1999).
5. Cosma, M. P. Ordered Recruitment: Gene-Specific Mechanism of Transcription Activation. *Mol Cell* **10**, 227-236, (2002).
6. Feklistov, A., Sharon, B. D., Darst, S. A. & Gross, C. A. Bacterial sigma factors: a historical, structural, and genomic perspective. *Annu Rev Microbiol* **68**, 357-376, (2014).
7. Gruber, T. M. & Gross, C. A. Multiple sigma subunits and the partitioning of bacterial transcription space. *Annu Rev Microbiol* **57**, 441-466, (2003).
8. Bae, B., Feklistov, A., Lass-Napiorkowska, A., Landick, R. & Darst, S. A. Structure of a bacterial RNA polymerase holoenzyme open promoter complex. *Elife* **4**, (2015).
9. Kornberg, R. D. Chromatin Structure: A Repeating Unit of Histones and DNA. *Science* **184**, 868-871, (1974).
10. Venter, J. C. *et al.* The Sequence of the Human Genome. *Science* **291**, 1304-1351, (2001).

11. Luger, K., Mader, A. W., Richmond, R. K., Sargent, D. F. & Richmond, T. J. Crystal structure of the nucleosome core particle at 2.8 angstrom resolution. *Nature* **389**, 251-260, (1997).
12. Bednar, J. *et al.* Nucleosomes, linker DNA, and linker histone form a unique structural motif that directs the higher-order folding and compaction of chromatin. *Proc Natl Acad Sci USA* **95**, 14173-14178, (1998).
13. Babu, A. & Verma, R. S. Chromosome Structure: Euchromatin and Heterochromatin. *Int Rev Cytol* **108**, 1-60, (1987).
14. Björklund, S., Almouzni, G., Davidson, I., Nightingale, K. P. & Weiss, K. Global Transcription Regulators of Eukaryotes. *Cell* **96**, 759-767, (1999).
15. Berger, S. L. Histone modifications in transcriptional regulation. *Curr Opin Genet Dev* **12**, 142-148, (2002).
16. Fischle, W., Wang, Y. & Allis, C. D. Histone and chromatin cross-talk. *Curr Opin Cell Biol* **15**, 172-183, (2003).
17. Horn, P. J. & Peterson, C. L. Molecular biology. Chromatin higher order folding--wrapping up transcription. *Science* **297**, 1824-1827, (2002).
18. Hebbes, T. R., Thorne, A. W. & Crane-Robinson, C. A direct link between core histone acetylation and transcriptionally active chromatin. *EMBO J* **7**, 1395-1402, (1988).
19. Wang, Z. *et al.* Genome-wide Mapping of HATs and HDACs Reveals Distinct Functions in Active and Inactive Genes. *Cell* **138**, 1019-1031, (2009).
20. Pazin, M. J. & Kadonaga, J. T. What's up and down with histone deacetylation and transcription? *Cell* **89**, 325-328, (1997).
21. Shahbazian, M. D. & Grunstein, M. Functions of Site-Specific Histone Acetylation and Deacetylation. *Annu Rev Biochem* **76**, 75-100, (2007).
22. Roeder, R. G. The role of general initiation factors in transcription by RNA polymerase II. *Trends Biochem Sci* **21**, 327-335, (1996).

23. Dignam, J. D., Lebovitz, R. M. & Roeder, R. G. Accurate transcription initiation by RNA polymerase II in a soluble extract from isolated mammalian nuclei. *Nucleic Acids Res* **11**, 1475-1489, (1983).
24. Liu, X., Bushnell, D. A. & Kornberg, R. D. RNA polymerase II transcription: Structure and mechanism. *Biochim Biophys Acta* **1829**, 2-8, (2013).
25. Corre, G. *et al.* Stochastic fluctuations and distributed control of gene expression impact cellular memory. *Plos One*, (2014).
26. Cantin, G. T., Stevens, J. L. & Berk, A. J. Activation domain–mediator interactions promote transcription preinitiation complex assembly on promoter DNA. *Proc Natl Acad Sci U S A* **100**, 12003-12008, (2003).
27. Borggreffe, T. & Yue, X. Interactions between subunits of the Mediator complex with gene-specific transcription factors. *Semin Cell Dev Biol* **22**, 759-768, (2011).
28. Merika, M., Williams, A. J., Chen, G., Collins, T. & Thanos, D. Recruitment of CBP/p300 by the IFN β Enhanceosome Is Required for Synergistic Activation of Transcription. *Mol Cell* **1**, 277-287, (1998).
29. Hager, G. L., McNally, J. G. & Misteli, T. Transcription Dynamics. *Cell* **35**, 741-753, (2009).
30. Ansari, A. Z. & Mapp, A. K. Modular design of artificial transcription factors. *Curr Opin Chem Biol* **6**, 765-772, (2002).
31. Krasnov, A. N., Mazina, M. Y., Nikolenko, J. V. & Vorobyeva, N. E. On the way of revealing coactivator complexes cross-talk during transcriptional activation. *Cell & Bioscience* **6**, 1-14, (2016).
32. Fulton, D. L. *et al.* TFCat: the curated catalog of mouse and human transcription factors. *Genome Biol* **10**, R29, (2009).
33. Messina, D. N., Glasscock, J., Gish, W. & Lovett, M. An ORFeome-based analysis of human transcription factor genes and the construction of a microarray to interrogate their expression. *Genome Res* **14**, 2041-2047, (2004).
34. Larivière, L., Seizl, M. & Cramer, P. A structural perspective on Mediator function. *Curr Opin Cell Biol* **24**, 305-313, (2012).

35. Conaway, R. C., Sato, S., Tomomori-Sato, C., Yao, T. & Conaway, J. W. The mammalian Mediator complex and its role in transcriptional regulation. *Trends Biochem Sci* **30**, 250-255, (2005).
36. Liu, J. *et al.* Intrinsic disorder in transcription factors. *Biochemistry* **45**, 6873-6888, (2006).
37. Dunker, A. K., Cortese, M. S., Romero, P., Iakoucheva, L. M. & Uversky, V. N. Flexible nets. The roles of intrinsic disorder in protein interaction networks. *FEBS J* **272**, 5129-5148, (2005).
38. Goodman, R. H. & Smolik, S. CBP/p300 in cell growth, transformation, and development. *Genes & Development* **14**, 1553-1577, (2000).
39. Giles, R. H., Peters, D. J. M. & Breuning, M. H. Conjunction dysfunction: CBP/p300 in human disease. *Trends in Genetics* **14**, 178-183, (1998).
40. Lin, C. H. *et al.* A small domain of CBP/p300 binds diverse proteins: solution structure and functional studies. *Mol Cell* **8**, 581-590, (2001).
41. Demarest, S. J. *et al.* Mutual synergistic folding in recruitment of CBP/p300 by p160 nuclear receptor coactivators. *Nature* **415**, 549-553, (2002).
42. Ruas, J. L. *et al.* Complex Regulation of the Transactivation Function of Hypoxia-inducible Factor-1 α by Direct Interaction with Two Distinct Domains of the CREB-binding Protein/p300. *J Biol Chem* **285**, 2601-2609, (2010).
43. De Guzman, R. N., Wojciak, J. M., Martinez-Yamout, M. A., Dyson, H. J. & Wright, P. E. CBP/p300 TAZ1 domain forms a structured scaffold for ligand binding. *Biochemistry* **44**, 490-497, (2005).
44. Filippakopoulos, P. *et al.* Histone recognition and large-scale structural analysis of the human bromodomain family. *Cell* **149**, 214-231, (2012).
45. Ragvin, A. *et al.* Nucleosome Binding by the Bromodomain and PHD Finger of the Transcriptional Cofactor p300. *J Mol Biol* **337**, 773-788, (2004).
46. Park, S., Martinez-Yamout, M. A., Dyson, H. J. & Wright, P. E. The CH2 domain of CBP/p300 is a novel zinc finger. *FEBS Lett* **587**, 2506-2511, (2013).

47. De Guzman, R. N., Liu, H. Y., Martinez-Yamout, M., Dyson, H. J. & Wright, P. E. Solution structure of the TAZ2 (CH3) domain of the transcriptional adaptor protein CBP1. *J Mol Biol* **303**, 243-253, (2000).
48. Parker, D. *et al.* Phosphorylation of CREB at Ser-133 induces complex formation with CREB-binding protein via a direct mechanism. *Mol Cell Biol* **16**, 694-703, (1996).
49. Ogryzko, V. V., Schiltz, R. L., Russanova, V., Howard, B. H. & Nakatani, Y. The Transcriptional Coactivators p300 and CBP Are Histone Acetyltransferases. *Cell* **87**, 953-959, (1996).
50. Maksimoska, J., Segura-Pena, D., Cole, P. A. & Marmorstein, R. Structure of the p300 histone acetyltransferase bound to acetyl-coenzyme A and its analogues. *Biochemistry* **53**, 3415-3422, (2014).
51. Thompson, P. R. *et al.* Regulation of the p300 HAT domain via a novel activation loop. *Nature Struct Mol Biol* **11**, 308-315, (2004).
52. Taunton, J., Hassig, C. A. & Schreiber, S. L. A mammalian histone deacetylase related to the yeast transcriptional regulator Rpd3p. *Science* **272**, 408-411, (1996).
53. Grunstein, M. Histone acetylation in chromatin structure and transcription. *Nature* **389**, 349-352, (1997).
54. Ness, S. A. Myb protein specificity: evidence of a context-specific transcription factor code. *Blood Cells Mol Dis* **31**, 192-200, (2003).
55. Lu, Q., Hutchins, A. E., Doyle, C. M., Lundblad, J. R. & Kwok, R. P. S. Acetylation of cAMP-responsive Element-binding Protein (CREB) by CREB-binding Protein Enhances CREB-dependent Transcription. *J Biol Chem* **278**, 15727-15734, (2003).
56. Sano, Y. & Ishii, S. Increased Affinity of c-Myb for CREB-binding Protein (CBP) after CBP-induced Acetylation. *J Biol Chem* **276**, 3674-3682, (2001).
57. Yao, T.-P. *et al.* Gene Dosage-Dependent Embryonic Development and Proliferation Defects in Mice Lacking the Transcriptional Integrator p300. *Cell* **93**, 361-372, (1998).

58. Ghosh, A. K. & Varga, J. The transcriptional coactivator and acetyltransferase p300 in fibroblast biology and fibrosis. *J Cell Physiol* **213**, 663-671, (2007).
59. Iyer, N. G., Ozdag, H. & Caldas, C. p300/CBP and cancer. *Oncogene* **23**, 4225-4231, (2004).
60. Arany, Z., Sellers, W. R., Livingston, D. M. & Eckner, R. E1A-associated p300 and CREB-associated CBP belong to a conserved family of coactivators. *Cell* **77**, 799-800, (1994).
61. Birchler, J. A. & Veitia, R. A. Gene balance hypothesis: connecting issues of dosage sensitivity across biological disciplines. *Proc Natl Acad Sci USA* **109**, (2012).
62. Xu, W., Kasper, L. H., Lerach, S., Jeevan, T. & Brindle, P. K. Individual CREB-target genes dictate usage of distinct cAMP-responsive coactivation mechanisms. *EMBO J* **26**, 2890-2903, (2007).
63. Ramos, Y. F. M. *et al.* Genome-wide assessment of differential roles for p300 and CBP in transcription regulation. *Nucleic Acids Res* **38**, 5396-5408, (2010).
64. Smith, J. L. *et al.* Kinetic profiles of p300 occupancy in vivo predict common features of promoter structure and coactivator recruitment. *Proc Natl Acad Sci USA* **101**, 11554-11559, (2004).
65. Kalkhoven, E. CBP and p300: HATs for different occasions. *Biochem Pharmacol* **68**, 1145-1155, (2004).
66. Taatjes, D. J., Marr, M. T. & Tjian, R. Regulatory diversity among metazoan co-activator complexes. *Nature Rev* **5**, 403-410, (2004).
67. Embley, T. M. & Martin, W. Eukaryotic evolution, changes and challenges. *Nature* **440**, 623-630, (2006).
68. Kitami, T. & Nadeau, J. H. Biochemical networking contributes more to genetic buffering in human and mouse metabolic pathways than does gene duplication. *Nat Genet* **32**, 191-194, (2002).
69. Novatchkova, M. & Eisenhaber, F. Linking transcriptional mediators via the GACKIX domain super family. *Curr Biol* **14**, R54-55, (2004).

70. Gonzalez, N. *et al.* A Repressor Protein Complex Regulates Leaf Growth in Arabidopsis. *Plant Cell* **27**, 2273-2287, (2015).
71. Thakur, J. K., Agarwal, P., Parida, S., Bajaj, D. & Pasrija, R. Sequence and expression analyses of KIX domain proteins suggest their importance in seed development and determination of seed size in rice, and genome stability in Arabidopsis. *Mol Genet Genomics* **288**, 329-346, (2013).
72. Gonzalez, N. *et al.* A Repressor Protein Complex Regulates Leaf Growth in Arabidopsis. *Plant Cell* **27**, 2273-2287, (2015).
73. Thakur, J. K. *et al.* A nuclear receptor-like pathway regulating multidrug resistance in fungi. *Nature* **452**, 604-609, (2008).
74. Thakur, J. K. *et al.* Mediator Subunit Gal11p/MED15 Is Required for Fatty Acid-dependent Gene Activation by Yeast Transcription Factor Oaf1p. *The J Biol Chem* **284**, 4422-4428, (2009).
75. Gouet, P., Courcelle, E., Stuart, D. I. & Métoz, F. ESPript: analysis of multiple sequence alignments in PostScript. *Bioinformatics* **15**, 305-308, (1999).
76. Maxwell, K. L. *et al.* Protein folding: Defining a “standard” set of experimental conditions and a preliminary kinetic data set of two-state proteins. *Protein Sci* **14**, 602-616, (2005).
77. Nishioka, K. & Reinberg, D. Switching Partners in a Regulatory Tango. *Science* **294**, 2497-2498, (2001).
78. Chevillard-Briet, M., Trouche, D. & Vandell, L. Control of CBP co-activating activity by arginine methylation. *EMBO J* **21**, 5457-5466, (2002).
79. Xu, W. *et al.* A Transcriptional Switch Mediated by Cofactor Methylation. *Science* **294**, 2507-2511, (2001).
80. Nagulapalli, M., Maji, S., Dwivedi, N., Dahiya, P. & Thakur, J. K. Evolution of disorder in Mediator complex and its functional relevance. *Nucleic Acids Res* **44**, 1591-1612, (2016).
81. Yang, F. *et al.* An ARC/Mediator subunit required for SREBP control of cholesterol and lipid homeostasis. *Nature* **442**, 700-704, (2006).

82. Zhang, Y., Xiaoli, Zhao, X. & Yang, F. The Mediator Complex and Lipid Metabolism. *J Biochem Pharmacol Res* **1**, 51-55, (2013).
83. Liu, Y.-P., Chang, C.-W. & Chang, K.-Y. Mutational analysis of the KIX domain of CBP reveals residues critical for SREBP binding. *FEBS Letters* **554**, 403-409, (2003).
84. Walker, Amy K. *et al.* A Conserved SREBP-1/Phosphatidylcholine Feedback Circuit Regulates Lipogenesis in Metazoans. *Cell* **147**, 840-852, (2011).
85. Oliner, J. D., Andresen, J. M., Hansen, S. K., Zhou, S. & Tjian, R. SREBP transcriptional activity is mediated through an interaction with the CREB-binding protein. *Genes Dev* **10**, 2903-2911, (1996).
86. Dunker, A. K. *et al.* Intrinsically disordered protein. *J Mol Graph Model* **19**, 26-59, (2001).
87. Romero, P. *et al.* Sequence complexity of disordered protein. *Proteins* **42**, (2001).
88. Vacic, V., Uversky, V. N., Dunker, A. K. & Lonardi, S. Composition Profiler: a tool for discovery and visualization of amino acid composition differences. *BMC Bioinformatics* **8**, 211, (2007).
89. Williams, R. M. *et al.* The protein non-folding problem: amino acid determinants of intrinsic order and disorder. *Pac Symp Biocomput*, 89-100, (2001).
90. Tompa, P., Szasz, C. & Buday, L. Structural disorder throws new light on moonlighting. *Trends Biochem Sci* **30**, 484-489, (2005).
91. Zor, T., De Guzman, R. N., Dyson, H. J. & Wright, P. E. Solution Structure of the KIX Domain of CBP Bound to the Transactivation Domain of c-Myb. *J Mol Biol* **337**, 521-534, (2004).
92. Radhakrishnan, I. *et al.* Solution Structure of the KIX Domain of CBP Bound to the Transactivation Domain of CREB: A Model of Activator:Coactivator Interactions. *Cell* **91**, 741-752, (1997).
93. Sugase, K., Dyson, H. J. & Wright, P. E. Mechanism of coupled folding and binding of an intrinsically disordered protein. *Nature* **447**, 1021-1025, (2007).

94. Gabdouliline, R. R. & Wade, R. C. On the protein–protein diffusional encounter complex. *J Mol Recognit* **12**, 226-234, (1999).
95. Shaywitz, A. J., Dove, S. L., Kornhauser, J. M., Hochschild, A. & Greenberg, M. E. Magnitude of the CREB-Dependent Transcriptional Response Is Determined by the Strength of the Interaction between the Kinase-Inducible Domain of CREB and the KIX Domain of CREB-Binding Protein. *Mol Cell Biol* **20**, 9409-9422, (2000).
96. Parker, D. *et al.* Role of secondary structure in discrimination between constitutive and inducible activators. *Mol Cell Biol* **19**, 5601-5607, (1999).
97. Shoemaker, B. A., Portman, J. J. & Wolynes, P. G. Speeding molecular recognition by using the folding funnel: the fly-casting mechanism. *Proc Natl Acad Sci USA* **97**, 8868-8873, (2000).
98. Huang, Y. & Liu, Z. Kinetic advantage of intrinsically disordered proteins in coupled folding-binding process: a critical assessment of the "fly-casting" mechanism. *J Mol Biol* **393**, 1143-1159, (2009).
99. Zhou, H.-X. & Bates, P. A. Modeling protein association mechanisms and kinetics. *Curr Opin Struct Biol* **23**, 887-893, (2013).
100. Arai, M., Sugase, K., Dyson, H. J. & Wright, P. E. Conformational propensities of intrinsically disordered proteins influence the mechanism of binding and folding. *Proc Natl Acad Sci USA* **112**, 9614-9619, (2015).
101. Dogan, J., Jonasson, J., Andersson, E. & Jemth, P. Binding Rate Constants Reveal Distinct Features of Disordered Protein Domains. *Biochemistry* **54**, 4741-4750, (2015).
102. Shammass, S. L., Travis, A. J. & Clarke, J. Remarkably Fast Coupled Folding and Binding of the Intrinsically Disordered Transactivation Domain of cMyb to CBP KIX. *J Phys Chem B* **117**, 13346-13356, (2013).
103. Ferguson, N. & Fersht, A. R. Early events in protein folding. *Curr Opin Struct Biol* **13**, 75-81, (2003).
104. Shammass, S. L., Crabtree, M. D., Dahal, L., Wicky, B. I. M. & Clarke, J. Insights into Coupled Folding and Binding Mechanisms from Kinetic Studies. *J Biol Chem* **291**, 6689-6695, (2016).

105. Giri, R., Morrone, A., Toto, A., Brunori, M. & Gianni, S. Structure of the transition state for the binding of c-Myb and KIX highlights an unexpected order for a disordered system. *Proc Natl Acad Sci USA* **110**, 14942-14947, (2013).
106. Shammass, S. L., Travis, A. J. & Clarke, J. Allostery within a transcription coactivator is predominantly mediated through dissociation rate constants. *Proc Natl Acad Sci U S A* **111**, 12055-12060, (2014).
107. Crabtree, M. D. *et al.* Conserved Helix-Flanking Prolines Modulate Intrinsically Disordered Protein:Target Affinity by Altering the Lifetime of the Bound Complex. *Biochemistry* **56**, 2379-2384, (2017).
108. Brüscheiler, S., Konrat, R. & Tollinger, M. Allosteric Communication in the KIX Domain Proceeds through Dynamic Repacking of the Hydrophobic Core. *ACS Chem Biol* **8**, 1600-1610, (2013).
109. Toto, A., Giri, R., Brunori, M. & Gianni, S. The mechanism of binding of the KIX domain to the mixed lineage leukemia protein and its allosteric role in the recognition of c-Myb. *Protein Sci* **23**, 962-969, (2014).
110. Campbell, K. M. & Lumb, K. J. Structurally distinct modes of recognition of the KIX domain of CBP by Jun and CREB. *Biochemistry* **41**, 13956-13964, (2002).
111. Clark, M. D. *et al.* Molecular Basis for the Mechanism of Constitutive CBP/p300 Coactivator Recruitment by CRTC1-MAML2 and Its Implications in cAMP Signaling. *Biochemistry* **54**, 5439-5446, (2015).
112. Goto, N. K., Zor, T., Martinez-Yamout, M., Dyson, H. J. & Wright, P. E. Cooperativity in transcription factor binding to the coactivator CREB-binding protein (CBP). The mixed lineage leukemia protein (MLL) activation domain binds to an allosteric site on the KIX domain. *J Biol Chem* **277**, 43168-43174, (2002).
113. Arai, M., Dyson, H. J. & Wright, P. E. Leu628 of the KIX domain of CBP is a key residue for the interaction with the MLL transactivation domain. *FEBS Lett* **584**, 4500-4504, (2010).
114. Denis, C. M. *et al.* Structural basis of CBP/p300 recruitment in leukemia induction by E2A-PBX1. *Blood* **120**, 3968-3977, (2012).

115. Denis, C. M. *et al.* Functional redundancy between the transcriptional activation domains of E2A is mediated by binding to the KIX domain of CBP/p300. *Nucleic Acids Res* **42**, 7370-7382, (2014).
116. Amelio, A. L. *et al.* CRTC1/MAML2 gain-of-function interactions with MYC create a gene signature predictive of cancers with CREB-MYC involvement. *Proc Natl Acad Sci USA* **111**, E3260-3268, (2014).
117. Vendel, A. C. & Lumb, K. J. NMR mapping of the HIV-1 Tat interaction surface of the KIX domain of the human coactivator CBP. *Biochemistry* **43**, 904-908, (2004).
118. Vendel, A. C., McBryant, S. J. & Lumb, K. J. KIX-Mediated Assembly of the CBP-CREB-HTLV-1 Tax Coactivator-Activator Complex. *Biochemistry* **42**, 12481-12487, (2003).
119. Cook, P. R., Polakowski, N. & Lemasson, I. HTLV-1 HBZ protein deregulates interactions between cellular factors and the KIX domain of p300/CBP. *J Mol Biol* **409**, 384-398, (2011).
120. Ernst, P., Wang, J., Huang, M., Goodman, R. H. & Korsmeyer, S. J. MLL and CREB Bind Cooperatively to the Nuclear Coactivator CREB-Binding Protein. *Mol Cell Biol* **21**, 2249-2258, (2001).
121. Morgunova, E. & Taipale, J. Structural perspective of cooperative transcription factor binding. *Curr Opin Struc Biol* **47**, 1-8, (2017).
122. Guo, J. & Zhou, H. X. Dynamically Driven Protein Allostery Exhibits Disparate Responses for Fast and Slow Motions. *Biophys J* **108**, 2771-2774, (2015).
123. Guo, J. & Zhou, H. X. Protein Allostery and Conformational Dynamics. *Chem Rev* **116**, 6503-6515, (2016).
124. Cooper, A. & Dryden, D. T. F. Allostery without conformational change. *Eur Biophys J* **11**, 103-109, (1984).
125. Palazzesi, F., Barducci, A., Tollinger, M. & Parrinello, M. The allosteric communication pathways in KIX domain of CBP. *Proc Natl Acad Sci USA* **110**, 14237-14242, (2013).

126. Gianni, S., Dogan, J. & Jemth, P. Coupled binding and folding of intrinsically disordered proteins: what can we learn from kinetics? *Curr Opin Struc Biol* **36**, 18-24, (2016).
127. Wang, F. *et al.* Synergistic Interplay between Promoter Recognition and CBP/p300 Coactivator Recruitment by FOXO3a. *ACS Chem Biol* **4**, 1017-1027, (2009).
128. Lee, C. W., Arai, M., Martinez-Yamout, M. A., Dyson, H. J. & Wright, P. E. Mapping the interactions of the p53 transactivation domain with the KIX domain of CBP. *Biochemistry* **48**, 2115-2124, (2009).
129. Teufel, D. P., Freund, S. M., Bycroft, M. & Fersht, A. R. Four domains of p300 each bind tightly to a sequence spanning both transactivation subdomains of p53. *Proc Natl Acad Sci USA* **104**, 7009-7014, (2007).
130. Wang, F. *et al.* Structures of KIX domain of CBP in complex with two FOXO3a transactivation domains reveal promiscuity and plasticity in coactivator recruitment. *Proc Natl Acad Sci USA* **109**, 6078-6083, (2012).
131. Li, J. J. & Biggin, M. D. Gene expression. Statistics requantitates the central dogma. *Science* **347**, 1066-1067, (2015).
132. Jovanovic, M. *et al.* Dynamic profiling of the protein life cycle in response to pathogens. *Science* **347**, (2015).
133. Li, J. J., Bickel, P. J. & Biggin, M. D. System wide analyses have underestimated protein abundances and the importance of transcription in mammals. *PeerJ* **2**, e270, (2014).
134. Battle, A. *et al.* Impact of regulatory variation from RNA to protein. *Science* **347**, 664-667, (2015).
135. Bayly, R. *et al.* E2A-PBX1 interacts directly with the KIX domain of CBP/p300 in the induction of proliferation in primary hematopoietic cells. *J Biol Chem* **279**, 55362-55371, (2004).
136. Craig, R. W. & Bloch, A. Early decline in c-myc oncogene expression in the differentiation of human myeloblastic leukemia (ML-1) cells induced with 12-O-tetradecanoylphorbol-13-acetate. *Cancer Res* **44**, 442-446, (1984).

137. Todokoro, K. *et al.* Down-regulation of c-myb gene expression is a prerequisite for erythropoietin-induced erythroid differentiation. *Proc Natl Acad Sci USA* **85**, 8900-8904, (1988).
138. Greig, K. T., Carotta, S. & Nutt, S. L. Critical roles for c-Myb in hematopoietic progenitor cells. *Semin Immunol* **20**, 247-256, (2008).
139. Mucenski, M. L. *et al.* A functional c-myb gene is required for normal murine fetal hepatic hematopoiesis. *Cell* **65**, 677-689, (1991).
140. Ell, B. & Kang, Y. Transcriptional control of cancer metastasis. *Trends Cell Biol* **23**, 603-611, (2013).
141. Zhou, Y. & Ness, S. A. Myb proteins: angels and demons in normal and transformed cells. *Front Biosci (Landmark Ed)* **16**, 1109-1131, (2011).
142. Vendel, A. C. & Lumb, K. J. Molecular recognition of the human coactivator CBP by the HIV-1 transcriptional activator Tat. *Biochemistry* **42**, 910-916, (2003).
143. Dukare, S. & Klempnauer, K. H. A conserved patch of hydrophobic amino acids modulates Myb activity by mediating protein-protein interactions. *Biochim Biophys Acta* **1859**, 914-921, (2016).
144. Odoux, A. *et al.* Experimental and molecular dynamics studies showed that CBP KIX mutation affects the stability of CBP:c-Myb complex. *Comput Biol Chem* **62**, 47-59, (2016).
145. Wood, M. A., Attner, M. A., Oliveira, A. M. M., Brindle, P. K. & Abel, T. A transcription factor-binding domain of the coactivator CBP is essential for long-term memory and the expression of specific target genes. *Learn Memory* **13**, 609-617, (2006).
146. Petrif, F. *et al.* Rubinstein-Taybi syndrome caused by mutations in the transcriptional co-activator CBP. *Nature* **376**, 348-351, (1995).
147. Petrij, F. *et al.* Diagnostic analysis of the Rubinstein-Taybi syndrome: five cosmids should be used for microdeletion detection and low number of protein truncating mutations. *J Med Genet* **37**, 168-176, (2000).

148. Bourtchouladze, R. *et al.* A mouse model of Rubinstein-Taybi syndrome: Defective long-term memory is ameliorated by inhibitors of phosphodiesterase 4. *Proc Natl Acad Sci USA* **100**, 10518-10522, (2003).
149. Tanaka, Y. *et al.* Extensive brain hemorrhage and embryonic lethality in a mouse null mutant of CREB-binding protein. *Mech Dev* **95**, 133-145, (2000).
150. Veitia, R. A., Bottani, S. & Birchler, J. A. Gene dosage effects: nonlinearities, genetic interactions, and dosage compensation. *Trends Genet* **29**, (2013).
151. Veitia, R. A. & Birchler, J. A. Models of buffering of dosage imbalances in protein complexes. *Biol Direct* **10**, 42, (2015).
152. Majmudar, C. Y. *et al.* Sekikaic Acid and Lobaric Acid Target a Dynamic Interface of the Coactivator CBP/p300. *Angew Chem* **51**, 11258-11262, (2012).
153. Majmudar, C. Y. & Mapp, A. K. Chemical approaches to transcriptional regulation. *Curr Opin Chem Biol* **9**, 467-474, (2005).
154. Lee, L. W. & Mapp, A. K. Transcriptional Switches: Chemical Approaches to Gene Regulation. *J Biol Chem* **285**, 11033-11038, (2010).
155. Schreiber, G. & Keating, A. E. Protein binding specificity versus promiscuity. *Curr Opin Struct Biol* **21**, 50-61, (2011).
156. Du, K., Asahara, H., Jhala, U. S., Wagner, B. L. & Montminy, M. Characterization of a CREB Gain-of-Function Mutant with Constitutive Transcriptional Activity In Vivo. *Mol Cell Biol* **20**, 4320-4327, (2000).
157. Frangioni, J. V., LaRicca, L. M., Cantley, L. C. & Montminy, M. R. Minimal activators that bind to the KIX domain of p300/CBP identified by phage display screening. *Nat Biotech* **18**, 1080-1085, (2000).
158. Best, J. L. *et al.* Identification of small-molecule antagonists that inhibit an activator: coactivator interaction. *Proc Natl Acad Sci USA* **101**, 17622-17627, (2004).
159. Mitton, B. *et al.* Small molecule inhibition of cAMP response element binding protein in human acute myeloid leukemia cells. *Leukemia* **30**, 2302-2311, (2016).

160. Uttarkar, S. *et al.* Naphthol AS-E Phosphate Inhibits the Activity of the Transcription Factor Myb by Blocking the Interaction with the KIX Domain of the Coactivator p300. *Mol Cancer Ther* **14**, 1276-1285, (2015).
161. Rees, D. C., Congreve, M., Murray, C. W. & Carr, R. Fragment-based lead discovery. *Nat Rev Drug Discov* **3**, 660-672, (2004).
162. Pomerantz, W. C. *et al.* Profiling the dynamic interfaces of fluorinated transcription complexes for ligand discovery and characterization. *ACS Chem Biol* **7**, 1345-1350, (2012).
163. Gee, C. T., Koleski, E. J. & Pomerantz, W. C. K. Fragment Screening and Druggability Assessment for the CBP/p300 KIX Domain through Protein-Observed ¹⁹F NMR Spectroscopy. *Angew Chem* **54**, 3735-3739, (2015).
164. Erlanson, D. A. *et al.* Site-directed ligand discovery. *Proc Natl Acad Sci USA* **97**, 9367-9372, (2000).
165. Erlanson, D. A., Braisted, A. C., McDowell, R. & Prescott, J. Extended tethering approach for rapid identification of ligands US Patent 6,919,178 (2005).
166. Workman, P. & Collins, I. Probing the Probes: Fitness Factors For Small Molecule Tools. *Chem Biol* **17**, 561-577.
167. Erlanson, D. A., Wells, J. A. & Braisted, A. C. Tethering: fragment-based drug discovery. *Annu Rev Biophys Biomol Struct* **33**, 199-223, (2004).
168. Thanos, C. D., Randal, M. & Wells, J. A. Potent small-molecule binding to a dynamic hot spot on IL-2. *J Am Chem Soc* **125**, 15280-15281, (2003).
169. Arkin, M. R. *et al.* Binding of small molecules to an adaptive protein-protein interface. *Proc Natl Acad Sci USA* **100**, 1603-1608, (2003).
170. Hardy, J. A., Lam, J., Nguyen, J. T., O'Brien, T. & Wells, J. A. Discovery of an allosteric site in the caspases. *Proc Natl Acad Sci USA* **101**, 12461-12466, (2004).
171. Buck, E. & Wells, J. A. Disulfide trapping to localize small-molecule agonists and antagonists for a G protein-coupled receptor. *Proc Natl Acad Sci USA* **102**, 2719-2724, (2005).

172. Sadowsky, J. D. *et al.* Turning a protein kinase on or off from a single allosteric site via disulfide trapping. *Proc Natl Acad Sci USA* **108**, 6056-6061, (2011).
173. Wang, N. *et al.* Ordering a dynamic protein via a small-molecule stabilizer. *J Am Chem Soc* **135**, 3363-3366, (2013).

Chapter 2 FP Tethering: a screening technique to rapidly identify compounds that disrupt protein-protein interactions^{1,†}

2.1 Abstract

Tethering is a screening technique for discovering small-molecule fragments that bind to pre-determined sites via the formation of a disulfide bond. Tethering screens traditionally rely upon mass spectrometry to detect the disulfide bond formation, which requires a time-consuming liquid chromatography step. Here we show that Tethering can be performed rapidly and economically using a homogenous fluorescence polarization (FP) assay that detects displacement of a peptide ligand from the protein target as an indirect effect of disulfide formation. We apply this method, termed FP Tethering, to identify fragments that disrupt the protein-protein interaction between the KIX domain of the transcriptional coactivator CBP and the transcriptional activator peptide pKID.

2.2 Introduction to FP-based Tethering screens

Protein-protein interactions (PPIs) underpin all cellular processes, and dysregulation of PPI networks is strongly correlated with human disease.³ For this reason, synthetic molecules that modulate PPIs are highly sought tools. Despite

¹Contents of this chapter are reproduced and adapted from a published article:¹ J. M. Lodge, T. Justin Rettenmaier, J. A. Wells, W. C. Pomerantz and A. K. Mapp, *Med. Chem. Commun.*, 2014, **5**, 370 DOI: 10.1039/C3MD00356F– Reproduced by permission of The Royal Society of Chemistry (RSC).

[†]This research was a collaborative effort. Dr. T. J. Rettenmaier conducted and analyzed the FP-Tethering screens and the confirmatory LC-MS experiments at the University of California, San Francisco. Dr. W. C. Pomerantz contributed to the preliminary screen against the KIX N627C mutant and provided peptides for these studies. Dr. M. E. Breen provided the fragment, **6D11** and the alkylators for the follow-up studies. These compounds were used for the experiments shown in Figure 2-6, and Figure 2-8.

their importance, small molecule PPI modulators are difficult to obtain through either screening or design.^{4,6} This is largely due to the intrinsic features of PPIs, most of which are comprised of significantly larger surface areas (average $\sim 1,949 \pm 760 \text{ \AA}^2$) than typical protein–ligand interfaces and are often flatter with few interaction features.^{7,8} Indeed, PPIs have often been described as ‘undruggable’ due to the challenges associated with identifying small molecules that can effectively engage these binding interfaces.⁹⁻¹¹

Tethering is a screening strategy that circumvents many of the difficulties associated with PPIs and has been successfully used to discover a range of small molecule modulators.^{12,13} It is a fragment discovery method in which a library of disulfide containing fragments (molecular weight $< 300 \text{ Da}$)¹⁴ are screened under reversible conditions against a protein target bearing a native or engineered cysteine adjacent to the binding site of interest (Figure 2-1A).^{15,16} Fragments that interact favorably with the protein target bias the equilibrium towards the mixed disulfide, which is subsequently detected by liquid chromatography-mass spectrometry (LC-MS) (Figure 2-1B).¹⁵ resulting fragment molecules can be converted to non-covalent inhibitors by growing the fragment or used as covalent binders for functional and biophysical studies.¹⁷⁻¹⁹

We previously reported the application of Tethering to discover small molecule ligands for the KIX domain of the master coactivator, CBP.²⁰ Several fragments identified from this screen proved to be excellent inhibitors of the PPI formed between KIX and the transcriptional activator MLL as well as enhancers

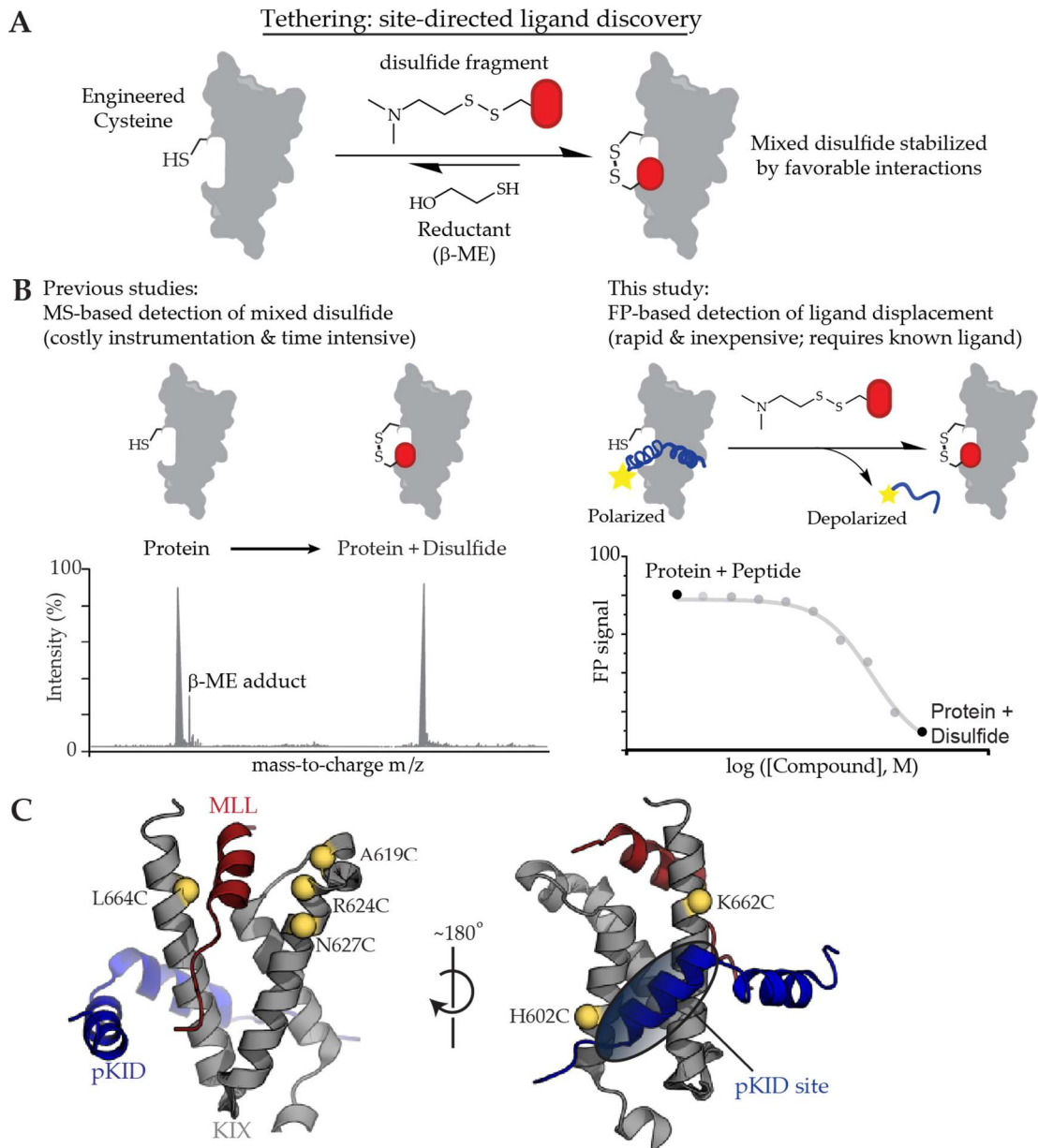


Figure 2-1 The Tethering methodology for site-directed ligand discovery A) A reversible disulfide reaction is set-up such that the mixed disulfide only persists when the ligand makes favorable interactions with the protein. Increasing the β -ME concentration results in a more stringent assay. B) Comparison between traditional liquid chromatography-mass spectrometry (LC-MS)-based Tethering with the novel fluorescence polarization (FP)-based Tethering. With FP the signal drops when the peptide is displaced from its partner due to the formation of a mixed disulfide. C) Cartoon representation of the KIX domain (grey) bound to two coactivator peptides MLL (dark red) and pKID (blue). (PDB ID: 2LXT) Previously, four cysteines lining the MLL-binding site (left) were screened using the LC-MS. Here, two mutants are shown that lined the pKID-binding site. The cysteine at position H602 is well-behaved in comparison to wild-type KIX so it was used to test the FP-based Tethering screen.

of the complex that KIX forms with the transcriptional activation domain of CREB (pKID).^{2,21} Nonetheless, we noted that some of the fragments that bind to KIX efficiently did not affect the interactions between activators and KIX. In retrospect, this is not surprising since the screen detects the disulfide formation, not inhibition of a PPI. We hypothesized that combining Tethering with a competitive inhibition effect would provide a more direct route to PPI modulators (Figure 2-1B). Here we show that fluorescence polarization (FP) Tethering is a rapid method for the direct discovery of PPI modulators and we use this method for the identification of effective inhibitors of the complex formed between KIX and pKID.

2.3 Results

FP is a homogenous, mix-and-read method used to directly measure the fraction of a small fluorescently-labeled peptide tracer that is bound to a larger protein.²² Binding of the tracer to the protein results in high FP signal, but the tracer's fluorescence is rapidly depolarized following displacement by an unlabeled ligand (Figure 2-1B).²³ This type of ligand displacement assay provides a way to measure the affinity between the protein target and an unlabeled ligand. However, FP ligand displacement assays are less sensitive than X-ray crystallography or NMR for detecting the low affinity interactions of fragments, which typically occur in the high micromolar to millimolar range.^{24,25} Since Tethering fragments to a target via a disulfide bond enhances their affinity, we reasoned that these binding events should be detectable by FP.

To test this approach, we first used the results from our earlier Tethering screen against cysteine mutants of the KIX domain. The KIX domain is a small (90 residue) three-helix bundle containing two binding sites that can be occupied simultaneously by different transcriptional activators (Figure 2-1C).²⁶⁻³⁰ These two binding sites are in allosteric communication, with the binding of MLL in the

smaller and deeper of the two sites leading to up to 2-fold enhancement of binding with activators such as pKID (the transcriptional activation domain of CREB) at the second, more shallow binding site.^{27,31} Two disulfide fragments isolated from the original Tethering screen bind well to the KIX N627C mutant, fragments **1A10** and **1C11**, and also were effective inhibitors of MLL binding when covalently associated with KIX N627C.³² Rather than measuring binding inhibition using a fully labeled protein, we reasoned a competitive inhibition signal could indicate the fragment-labeling of the KIX N627C mutant during the Tethering reaction. The inhibition of a fluorescently-labeled MLL peptide was observed in a dose-dependent fashion as a functional consequence of protein labeling (**1A10**: IC₅₀ 68 μM and **1C11**: IC₅₀ 115 μM, see Figure 2-2A). In the presence of 5 mM β-ME and the highest concentration (250 μM) of **1A10** used in the FP competition assay 82% of the KIX N627C mutant was labeled. The ability to increase or decrease the inhibition levels by changing the β-ME concentration affords a control for ruling out fluorescent impurities or other false positives.

Next, 80 disulfide-containing fragments were screened at two different β-ME concentrations against the KIX N627C mutant (Figure 2-2A). The FP Tethering screen identified the same set of ligands that were previously identified in the MS-based screen (Figure 2-2B). Notably the hits that resulted in the largest inhibition of the KIX domain and MLL interaction were also the most potent hits from our MS-based screen. These results suggested that the FP Tethering assay was suitable for high-throughput identification of fragments that disrupt the interaction

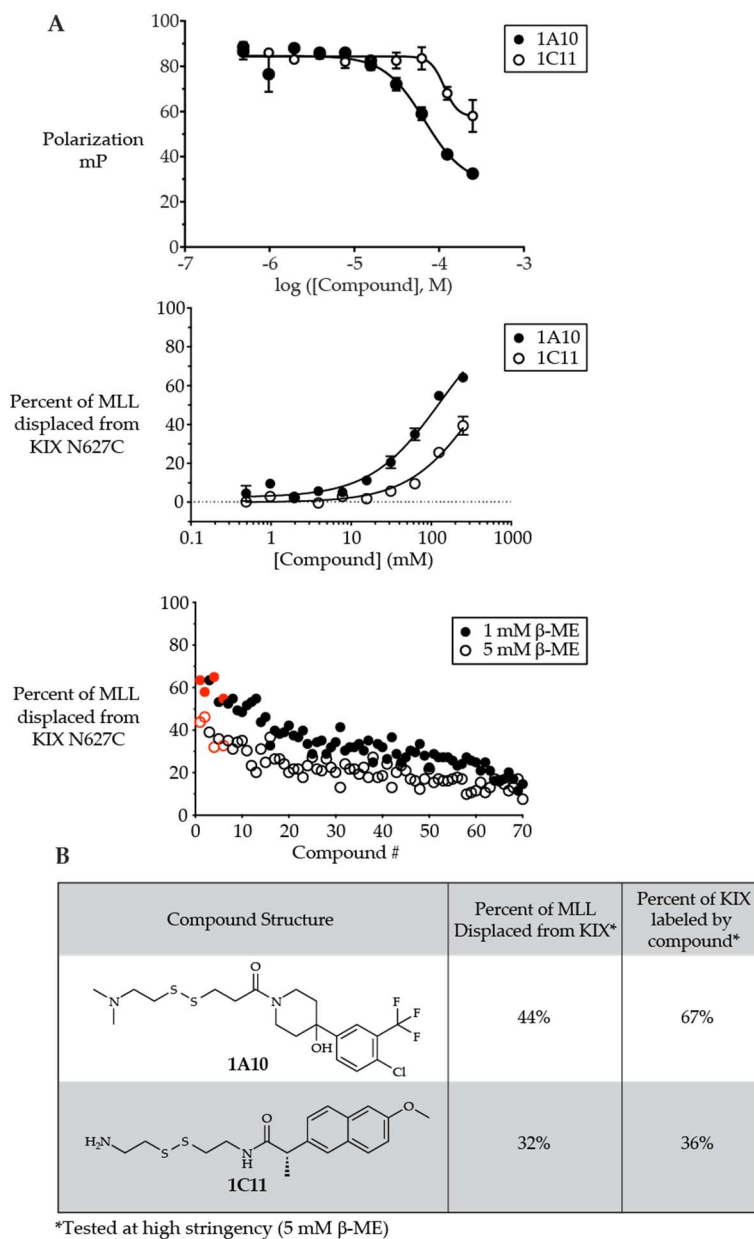


Figure 2-2 FP- and MS-based Tethering identified overlapping ligand sets for the MLL-binding site. A) The FP competition assay was tested by varying the concentrations of the fragments, **1A10** and **1C11** and keeping the concentration of KIX N627C mutant, and fluorescently-labeled MLL peptide constant.² The polarization signal was plotted against the various concentrations of the fragment, **1A10** and **1C11** and analyzed by GraphPad Prism 4.00. The last two traces show the percent of MLL displaced from KIX N627C for the two fragments and for the 80 small-molecule disulfides. The screen was performed at medium and high stringency (1 mM and 5 mM β -ME, respectively). The top 4 fragments that were previously identified by a LC-MS-based Tethering screen are colored in red. B) The structures and properties of the two fragments were rediscovered by FP-based Tethering.

between KIX and its coactivators.

To test the FP Tethering approach in a pure discovery mode, we targeted the second binding site within the KIX domain that is used by pKID in addition to the activator c-Myb. This is the considerably more challenging of the two sites, with a larger surface area (1,480 Å²) and a less well-defined surface.³³ In preparation for the screen, we mutated several residues lining the binding site to cysteine and determined the integrity of the mutants. The helicity and stability of the KIX mutants was assessed by circular dichroism (CD) (Figure 2-3). While the thermal stability difference is similar between wild-type KIX and KIX K662C the cysteine mutation at K662C disrupts the helical propensity of KIX.

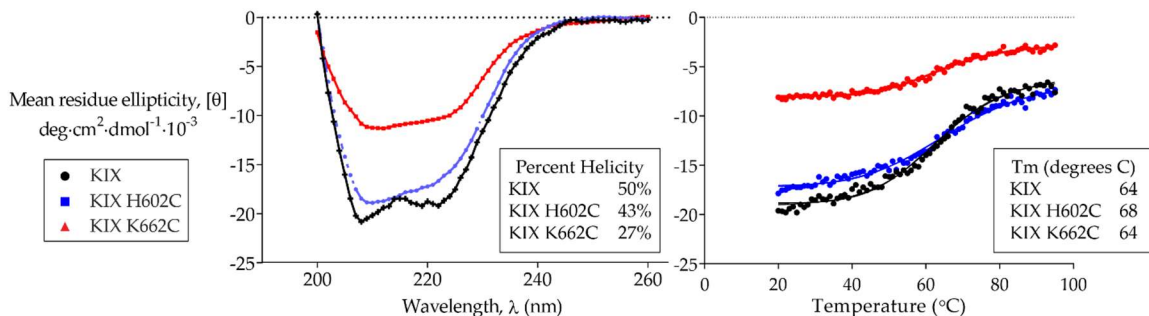


Figure 2-3 The impact of the cysteine mutants within the pKID-site of KIX on the overall KIX structure was assessed by CD. The wavelength scan shows the relative helicity of the KIX constructs (on the left). The scan represents an average of five measurements. The melting temperature curve was measured monitoring the helicity change at 222 nm (on the right). While the melting temperatures have little variations for the KIX constructs, the KIX K662C mutant lost approximately 20% of the native helical structure.

The affinity for pKID, c-Myb, and MLL tracers were also compared to the wild-type KIX protein (Table 2.1 and Figure 2-4). The KIX H602C mutant retained its affinity for the tracers and it was thus chosen for the FP Tethering screen. The competition assay was developed with 25 nM FITC-labeled pKID and 4 μM of KIX H602C. These conditions were sufficient to detect displacement of FITC-labeled pKID tracer from the KIX domain by the unlabeled pKID peptide providing a signal/background ratio of at least 1.5-fold across the different $\beta\text{-ME}$

concentrations. The assay was also precise with coefficient of variation (CV) values for the maximum and minimum signals was between 2 to 3% for the bound and unbound form of the FITC-pKID tracer following the guidelines that. CV values below 15% are acceptable as a measure of precision for the assay.³⁴

Table 2-1 FP results for wild-type KIX and two cysteine mutants The equilibrium dissociation constants (K_D) for the three tracers, MLL, pKID, and c-Myb were compared between two KIX cysteine mutants and wild-type KIX using the FP-direct binding assay. Errors reflect the standard error (SE) of nonlinear fits in GraphPad Prism. The corresponding curves are shown in Figure 2-4.

K_D (μM)	pKID	c-Myb	MLL
Wild-type KIX	0.54 ± 0.07	1.1 ± 0.2	0.26 ± 0.03
KIX H602C	0.22 ± 0.02	2.2 ± 0.2	0.24 ± 0.05
KIX K662C	4.9 ± 0.9	0.52 ± 0.05	0.64 ± 0.04

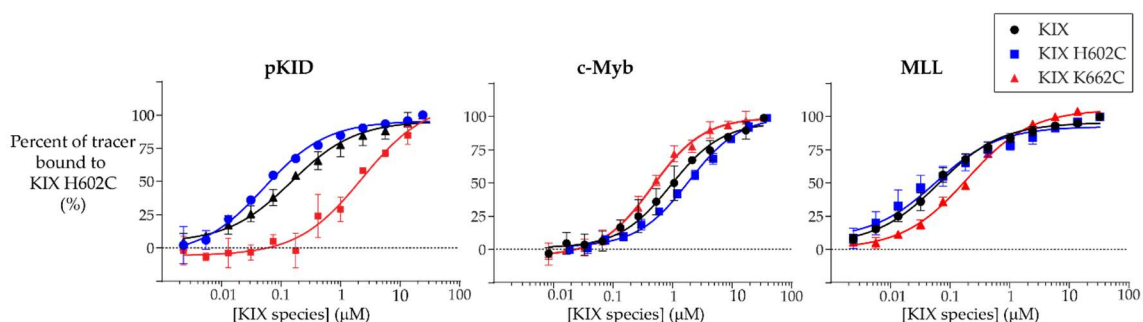


Figure 2-4 The FP Tethering screen requires a compatible, cysteine-containing KIX target. A couple of cysteine mutants were generated in proximity to the pKID-binding site and the mutants were compared to wild-type KIX by FP-direct binding assays with the three different tracers, MLL, pKID, and c-Myb. The FP direct binding curves of KIX and the cysteine mutants are shown for the three different tracers. All measurements were performed in triplicate and the error bars reflect the standard deviation (SD) error of nonlinear fits in GraphPad Prism.

Using this assay, we screened 960 disulfide-containing fragments at three different β -ME concentrations against a fluorescent pKID tracer in complex with KIX H602C (Figure 2-5A). The Z' factors were calculated as 0.69 at 0.2 mM β -ME, 0.63 at 1 mM β -ME, and 0.60 at 5 mM β -ME. Therefore, this the variability of the assay's signal and dynamic range was considered acceptable for a high-throughput screen ($1 > Z' > 0.5$). Fragments that yielded total fluorescence

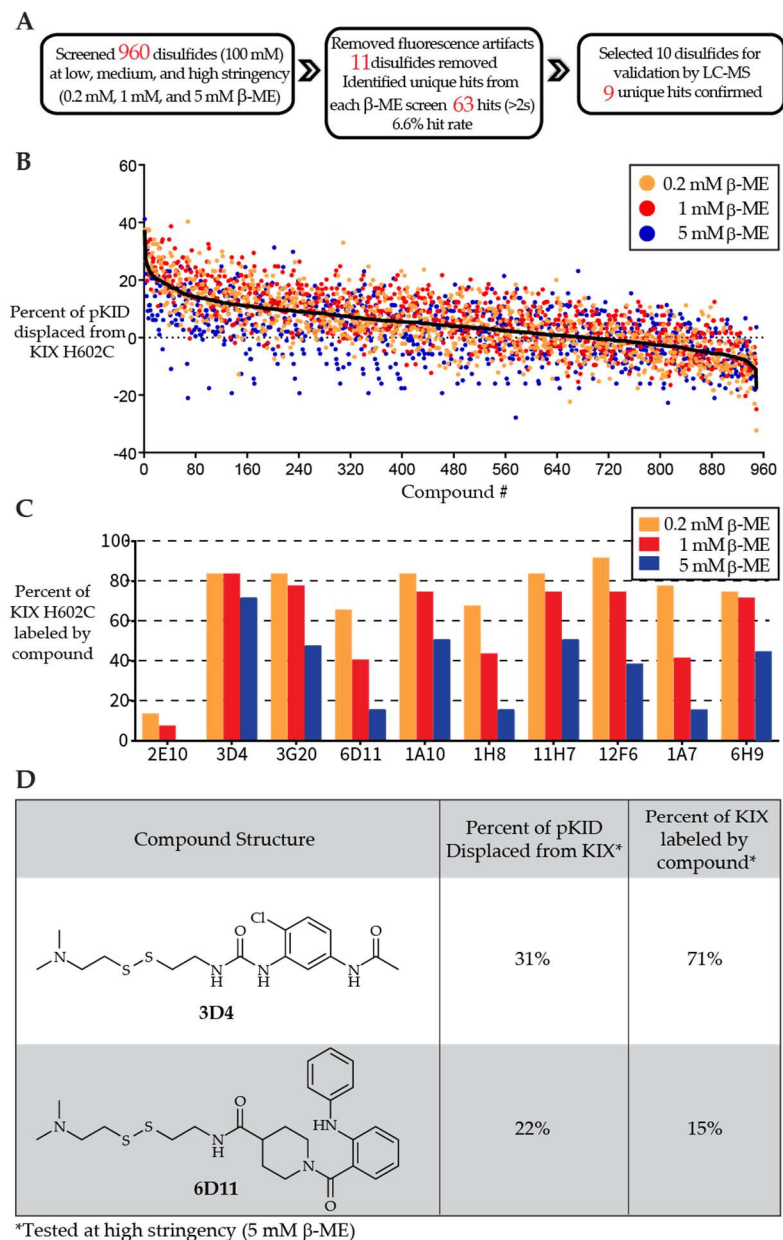


Figure 2-5 FP-based Tethering screen to identify new ligands that disrupt the interaction between KIX and pKID. A) Overview of the workflow for screening, hit selection, and validation. B) The percent of pKID displaced from KIX H602C is shown for 960 small-molecule disulfides at low, medium, and high stringency (0.2 mM, 1 mM, and 5 mM β -ME, respectively). The curve is black represents the average percent inhibition for the β -ME screens. C) The percent of KIX H602C that is labeled by the top 10 small molecule disulfides was measured by LC-MS. The reaction was performed with increasing stringency to assess the relative affinities. D) The structures and properties of ligands that displace pKID from KIX H602C.

intensities (TFI) greater than 150% of the DMSO control wells were flagged as fluorescent artifacts and excluded from further analysis. Next the remaining fragments were ranked by their percent inhibition of pKID binding across the three β -ME concentrations (Figure 2-5B). The average percent inhibition for all the compounds was $5 \pm 14\%$ at 0.2 mM β -ME, $7 \pm 12\%$ at 1 mM β -ME, and $3 \pm 12\%$ at 5 mM β -ME. As expected, the highest stringency of 5 mM β -ME resulted in decreased inhibition relative to the lower β -ME concentrations. The two lower β -ME concentrations, 0.2 mM and 1 mM, deviated from this trend in which the lowest percent inhibition fluctuated between 1 mM β -ME and 0.2 mM β -ME for the different fragments (Figure 2-5B).

From the top hits, 10 disulfides were cherry-picked for validation by LC-MS analysis (Figure 2-5C). These fragments covalently labeled the KIX domain. However, **2E10** was likely a false positive given its high percent inhibition by FP yet a low degree of labeling by MS. Comparison of labeling across the three β -ME concentrations reveals the relative potencies of the fragments. From the ten hits, two fragments, **3D4** and **6D11** were chosen for further characterization. The fragment **3D4** labeled KIX H602C more effectively than any other fragments and maintained a high percent inhibition of pKID even at 5 mM β -ME (31%) (Figure 2-5D). The fragment **6D11** had the highest average percent inhibition of pKID for the three β -ME concentration. Notably, percent labeling did not always correlate with inhibition of pKID binding. While fragment **3D4** labeled 70% of KIX H602C at 5 mM β -ME, this resulted in 24% inhibition of pKID binding. The larger fragment **6D11** labeled KIX H602C only 15% at 5 mM β -ME, but this translated to nearly equivalent inhibition (22%) of pKID binding. This finding suggests that although **6D11** may exhibit weaker overall potency, it is the more efficient inhibitor. Other top hits replaced the diphenylamine moiety in **6D11** for a diphenylmethane or 2-chloro-2-N phenylaniline. These changes did not improve

their ability to displace the pKID tracer, but their ability to bind to the KIX H602C mutant improved slightly. The disulfide fragment library contained several variants of the core scaffolds of **3D4** and **6D11** and a comparison revealed that both structures are sensitive to substituent placement and identity; this is particularly true for **6D11** where modifications of the distal aromatic ring leads to dramatic changes in inhibition (Figure 2-7).

Dose–response curves were determined for **3D4** and **6D11** using LC-MS to determine the percent of KIX H602C mutant bound to the fragment (Figure 2-6A and 2-6E). The concentration at which 50% of the KIX H602C mutant is bound by the fragment (DR_{50}) was lower with the fragment **3D4** than **6D11**. This finding was consistent with single point LC-MS measurements in which the fragment **3D4** was more potent than the fragment **6D11**. Finally, we labeled greater than 90% of KIX H602C with the fragments, **3D4** and **6D11** and used these protein–ligand conjugates for direct binding experiments with the MLL, pKID, and c-Myb tracers.

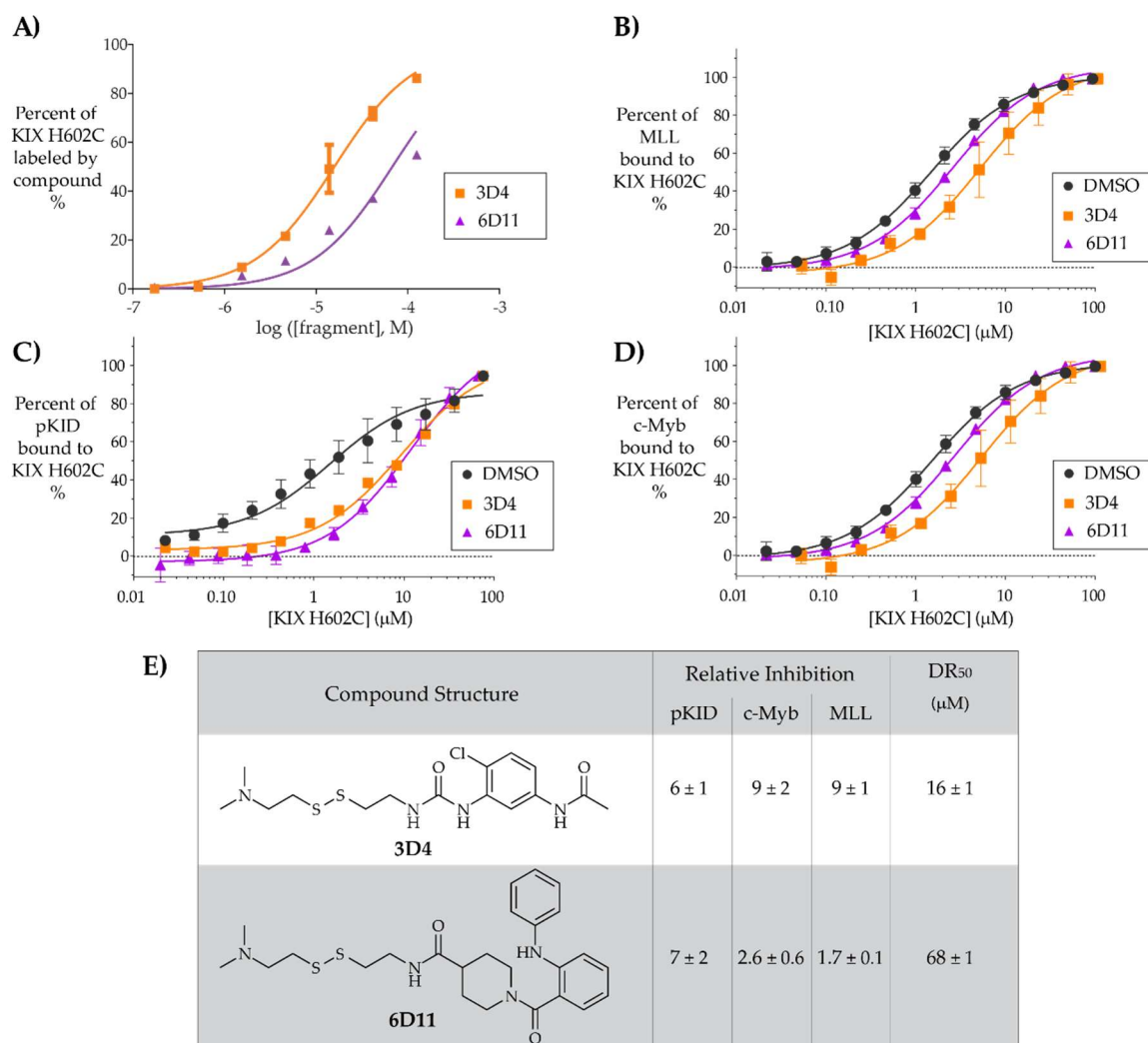


Figure 2-6 Labeling of KIX H602C with 3D4 and 6D11 decreases the affinity for pKID, c-Myb, and MLL. A) The dose-response binding experiments were performed with a constant concentration of KIX H602C and various concentrations of fragment, **3D4** and **6D11**. Each value was duplicated in separate experiments except the top concentration of compound (250 μM) which was only measured once due to a limited quantity of compound. The extent of KIX H602C tethered to the fragments was quantified using LC-MS. The average of the duplicated values is plotted and the error bars indicated SD error. B) The FP direct binding curves are shown for MLL, C) pKID, and D) c-Myb. KIX H602C was quantitatively labeled with either **3D4** or **6D11** and this tethered construct was used to assess the K_D s for the three tracers. All FP experiments were performed in triplicate and the error bars indicate SD error. E) The results are shown for the FP direct binding assays and dose-response curve. The relative inhibition values were assessed by comparing the K_D values from the labeled KIX H602C to the untethered KIX H602C. Error reflects the SE of nonlinear fits calculated in GraphPad Prism.

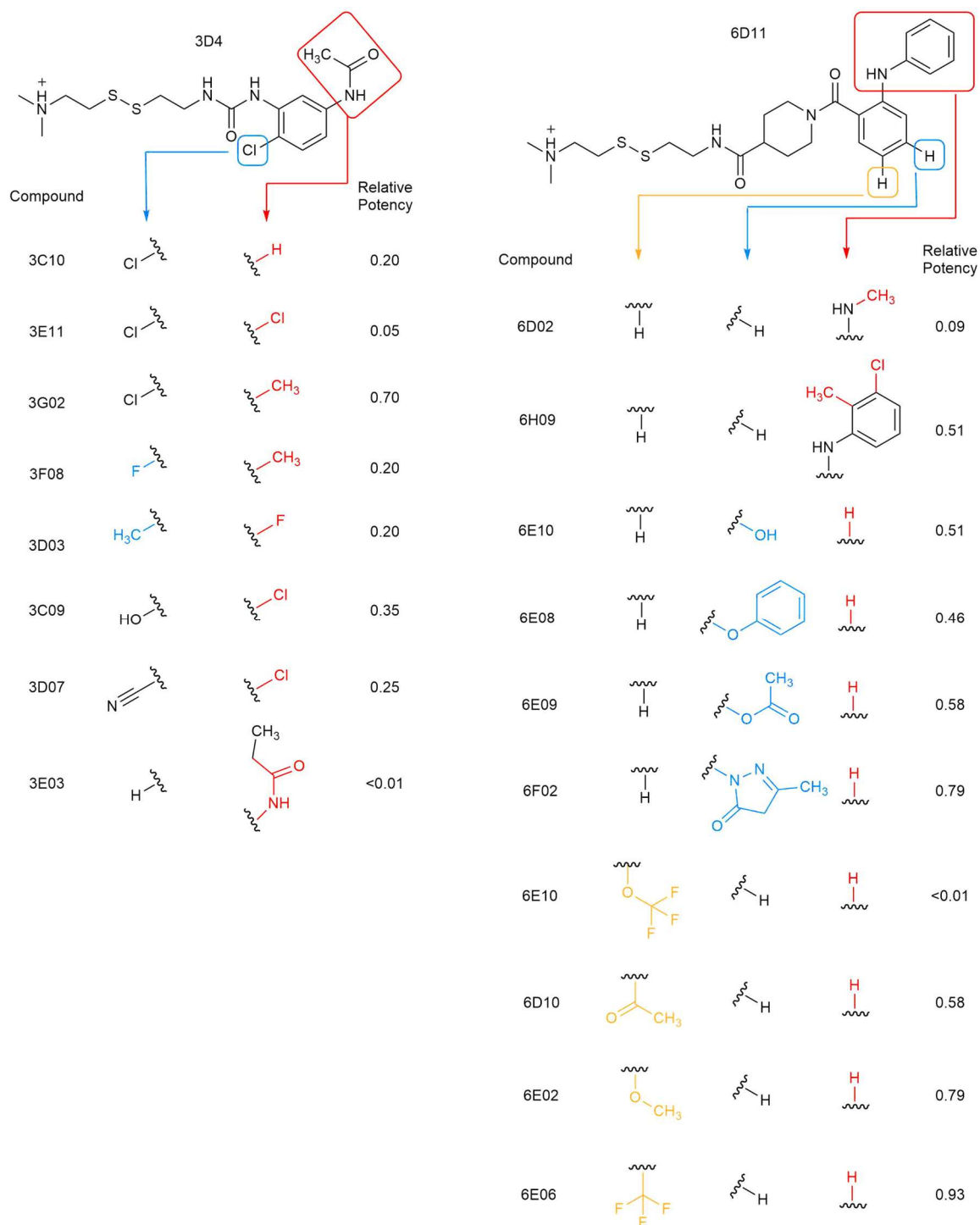


Figure 2-7. 3D4 and 6D11 contain chemical features that supported their discovery in the FP-Tethering screen in comparison to other fragments. A comparison of inhibition for 3D4 and 6D11 and related fragments reveals that both scaffolds are sensitive to substituent placement and identity. Relative potency values are the inhibition value of each fragment measure under stringent (5 mM β -ME) conditions divided by the inhibition value for the parent fragment (3D4 or 6D11).

The equilibrium dissociation constant (K_D) of each tracer for the KIX-fragment conjugate was assessed by a binding isotherm that accounts for ligand depletion (Figure 2-6B-E).³⁵ Surprisingly, **3D4** not only increases the K_D for binding to pKID and c-Myb but also for MLL, which binds on the opposite face of the KIX domain. This finding demonstrates a negative allosteric coupling between the pKID- and MLL-binding sites. The fragment, **6D11** was a moderate inhibitor from the FP-Tethering screen and **6D11** tethered to KIX H02C disrupted pKID from KIX to a similar extent as **3D4**. This fragment perturbed c-Myb and MLL at the distal site to a lesser extent. As probes, these fragments may moderate the assembly of activators that share the same binding site within KIX.

Given that **3D4** and **6D11** disrupts the binding of all three coactivator peptides, irreversible covalent analogues would be useful chemical probes for studying CBP and p300 biology. However, the conversion of disulfide fragments into irreversible co-chaperones presents challenges. The fragments must orient to first bind and then react with the cysteine-containing target. Here we used different irreversible covalent analogues (Figure 2-8). The disulfides reacted more readily with the KIX H602C target than these analogues after one hour in the presence of 1 mM β -ME. These irreversible fragments struggled to fully label the target KIX H602C even under forcing conditions. As observed in the screen, changes within the core **3D4** and **6D11** scaffolds disrupts the fragment's ability to bind or to disrupt the pKID-KIX H602C interaction (Figure 2-7). Therefore, additional linkers and alkylator moieties may be required to develop irreversible

covalent co-chaperones at the pKID-binding site of CBP KIX.

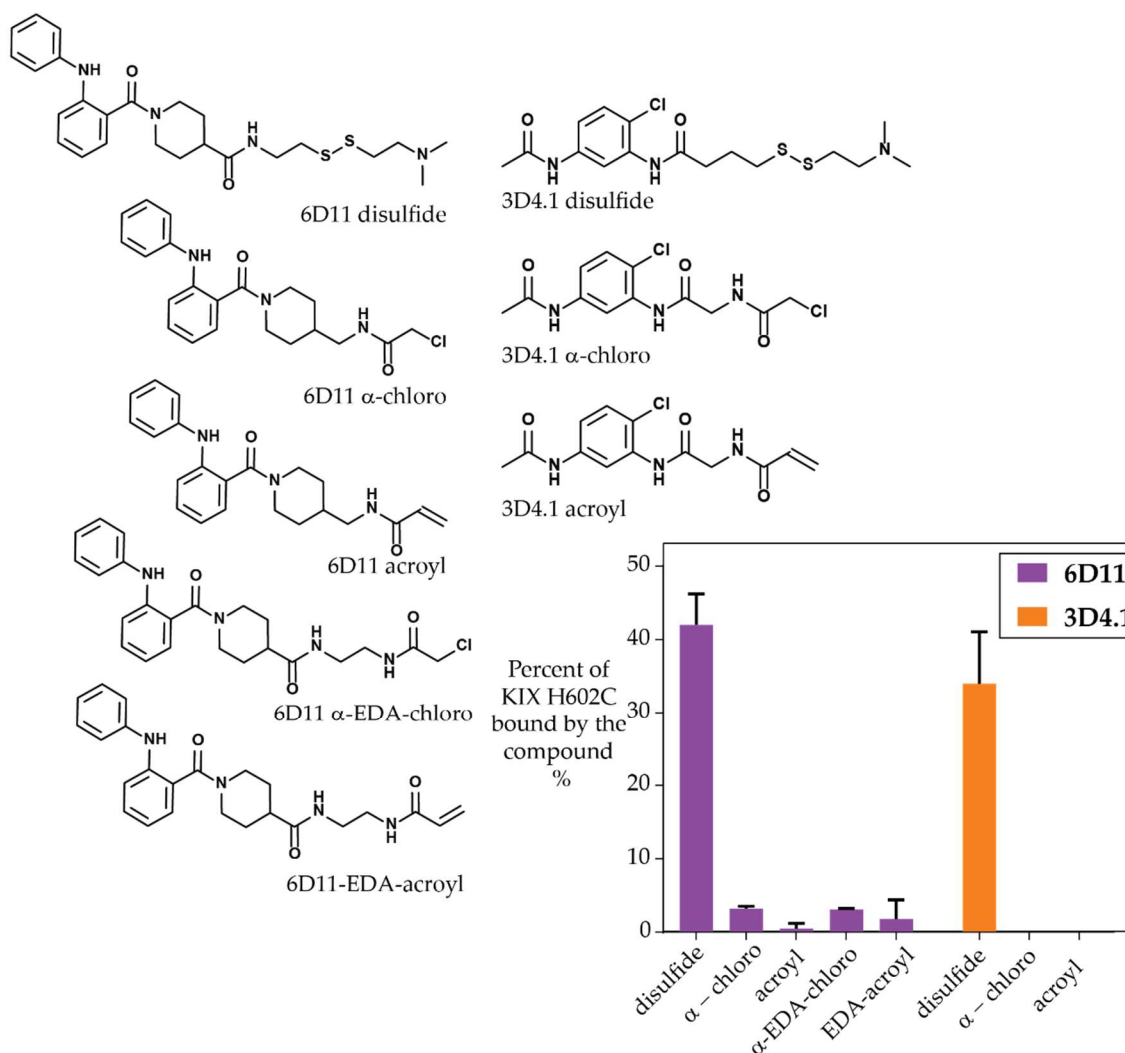


Figure 2-8 The initial 6D11 and 3D4 irreversible derivatives suffer from poor labeling efficiency with the KIX H602C target. The disulfides and their irreversible structures are shown. Two of the 6D11-based compounds contain ethylenediamine (EDA) linkers. In the graph, the extent of KIX H602C bound by the compound was measured by LC-MS after one hour of incubating 5 μ M of KIX H602C with 100 μ M of compound in the presence of 1 mM β -ME. Each value is an average of two separate experiments and the error bars represent the SD error.

2.4 Conclusion

FP Tethering is an efficient and accessible screening format for identifying fragments that disrupt a PPI. This method was successfully benchmarked against a previous LC-MS-based Tethering screen on the MLL-binding site of KIX.

Application of this method in discovery mode against a more challenging surface within KIX,³⁶ the broad interface of the KIX·pKID complex, was also a success. FP Tethering lead to 9 fragments confirmed to displace pKID from the KIX domain by covalently labeling the protein. These molecules represent promising leads for follow-up studies. For instance, **3D4** engages the KIX H602C mutant disrupting interactions at both the pKID or c-Myb binding site and the MLL-binding site. From the previous screen targeting the MLL-binding site KIX L664C tethered to **1A10** also known as **1-10** also directly disrupts MLL and allosterically inhibits pKID.³² Small molecules appear to readily engage in this negative effect from either site within KIX. Previously, KIX L664C tethered to **1A10** was shown to induce a negative cooperativity effect.³² This appears to be readily available to small molecules, but no known, endogenous KIX-associated ligands induce negative allostery within KIX.

The **3D4** and **6D11** disulfides efficiently labeled KIX H602C, however the irreversible derivatives were unable to label KIX H602C to the same extent. With this small fraction of labeled KIX H602C the fragments antagonized the binding of MLL, pKID and c-Myb between 1-5 relative fold inhibition. While these disulfides have potential as effective inhibitors, the probes need to efficiently bind to the target. Excess, unbound small molecule would lead to off-target effects in cellular environments where other reactive cysteines and free thiols are present such as glutathione.

More broadly, these results suggest that Tethering could be combined with any binding assay that is compatible with disulfide-containing molecules and β -ME^{37,38} increasingly the accessibility of this important technology. In retrospect, a parallel LC-MS Tethering screen would have identified fragments that labeled the pKID-binding site. At this stage, we cannot conclude on how Tethering screens result in different disulfide fragment hits depending on whether the screen was

conducted by LC-MS or FP. Tethering through a disulfide bond significantly enhances the affinity of fragments for the protein. One important advantage of Tethering is to rapidly identify disulfide fragments for the protein target that would not, otherwise, be detectable through standard LC-MS or FP fragment-based screens.

2.5 Materials and Methods

2.5.1 *Expression and purification of the KIX domain*

The KIX domain (residues 586-672) of mouse CBP protein fused to a hexahistidine tag and short polar linker was cloned into the bacterial expression pRSETB vector. All cysteine mutants were generated by using site-direct mutagenesis as previously described.^{32,39} The KIX protein was overexpressed from *Escherichia coli* Rosetta2(DE3) pLysS (Novagen) cells in Terrific Broth (37 °C, 250 rpm). When the cell density reached an OD_{600nm} ~0.8, protein expression was induced with 0.1 mM isoprpyl-β-D-thiogalactopyranoside (IPTG) for approximately 12 hours at 25 °C. The cells were harvested by centrifugation and stored at -80 °C. The hexahistidine-tagged KIX protein was affinity purified using Ni-NTA beads (Qiagen) followed by ion-exchange chromatography (Source S column, GE Healthcare) using an FPLC system. The purified protein was stored in 10 mM phosphate, 100 mM NaCl, pH 6.8 at -80 °C.

2.5.2 *Synthesis of fluorescent peptides*

Peptides were synthesized using standard Fmoc solid phase synthesis as previously described.⁴⁰

2.5.3 *Fluorescence polarization binding assay*

In 384-well microplates various concentrations of KIX protein were mixed in binding buffer (10 mM sodium phosphate, 100 mM NaCl, pH 6.8) with a constant concentration of fluorescein isothiocyanate (FITC)-labeled peptides (25

nM) and β -mercaptoethanol to a final volume of 10 μ L in each well. The samples incubated at room temperature for 1 hour. The anisotropy and fluorescence intensity were monitored using the Tecan Genios Pro microplate reader at an excitation wavelength of 485 nm and an emission wavelength at 535 nm. The equilibrium dissociation constants, K_D in Eq. 1 were calculated by fitting the observed anisotropy values as a function of KIX protein concentration to the ligand depletion model assuming single site binding of peptide to protein,

$$A = A_f + (A_b - A_f) \times \frac{([L_T] + K_d + [P_T]) - \sqrt{([L_T] + K_d + [P_T])^2 - 4[L_T][P_T]}}{2[L_T]} \quad (\text{Eq. 1})$$

where $[L_T]$ and $[P_T]$ are the total concentrations of fluorescent peptide and KIX protein, respectively, A is the observed anisotropy, A_b is the maximum anisotropy value for the fully bound peptide, and A_f is the minimum anisotropy value for the free peptide. All nonlinear data analysis was performed with GraphPad Prism 4.0 software (San Jose, CA).

2.5.4 Fluorescence polarization Tethering screens

For the proof of concept screen, KIX N627C (0.5 μ M) was mixed in binding buffer with FITC-labeled MLL peptide (25 nM), β -ME (1 mM or 5 mM), and one of 80 fragment disulfides (160 μ M) in a final volume of 30 μ L. For the discovery screen, KIX H602C (4 μ M) was mixed in binding buffer with FITC-labeled pKID peptide (25 nM), β -mercaptoethanol (0.2 mM, 1 mM or 5 mM), and one of 960 fragment disulfides (125 μ M) to a final volume of 20 μ L. Wells containing DMSO instead of the fragment disulfides and those containing only 25 nM of FITC-labeled pKID were used as controls.

All plates were shaken at room temperature for 1 hour to reach equilibrium. The fluorescence polarization (FP) and total fluorescence intensity (TFI) were measured using the Analyst HT Multi-Plate Reader (Molecular Devices) at an excitation wavelength of 485 nm and an emission wavelength at 530 nm.

Fragments that yielded TFI greater than 150% of the DMSO control were flagged as artifacts and excluded from further analysis. FP readings were transformed percent inhibition, using the following equation.

$$\% \text{ Inhibition} = (\text{FP}_{\text{Bound}} - \text{FP}_{\text{sample}}) / (\text{FP}_{\text{Bound}} - \text{FP}_{\text{free}}) \times 100 \quad (\text{Eq. 2})$$

The statistical parameter Z' factor was calculated by

$$Z' = 1 - \frac{(3\sigma_{\text{max}} - 3\sigma_{\text{min}})}{|\mu_{\text{max}} - \mu_{\text{min}}|} \quad (\text{Eq. 3})$$

where σ_{max} is the standard deviation of the positive controls (maximum signal), σ_{min} is the standard deviation of the negative controls (minimum signal), μ_{max} is the mean of the maximum signal or positive controls, and μ_{min} is the mean of the minimum signal or the negative controls.^{34,41} The signal/background (S:B) ratio was calculated by

$$S:B = \frac{\mu_{\text{max}}}{\mu_{\text{min}}} \quad (\text{Eq. 4})$$

and the coefficient variance was calculated from

$$\%CV = \frac{\sigma}{\mu} \times 100 \quad (\text{Eq. 5})$$

where σ is the standard deviation of the assay signal and μ is the mean of the assay signal. The acceptable limit is set at less than 15%.³⁴

2.5.5 LC-MS validation of fragment hits

KIX H602C (4 μM) was mixed in binding buffer with β -mercaptoethanol (0.2 mM, 1 mM or 5 mM) and each fragment disulfide (125 μM) to a total volume of 25 μL . The plate was shaken for 1 hour at room temperature to reach equilibrium. Each sample was subjected to LC-MS using an LCT Premier ESI-MS (Waters) with an inline C4 protein desalting column (Microtrap). Protein masses were deconvoluted using the Max-Ent algorithm within the MassLynx software. Percent labeling was measured by comparing the peak areas for the labeled versus

unlabeled protein.

2.5.6 Determination of the dose response (DR_{50})

A constant concentration of the KIX H602C mutant and fresh β -ME in 10 mM phosphate buffer, 100 mM NaCl, pH 6.8 were incubated with various concentrations of compound (0.2 μ M -125 μ M) at room temperature for one hour. The samples were spun-down and injected onto an Agilent Q-TOF HPLC-MS instrument. Protein masses were deconvoluted using the Max-Ent algorithm within the Agilent MassHunter Workstation Bioconfirm software. The dose-response (DR_{50}) values were determined with a normalized response from 1 to 100 using the equation

$$y = \frac{100}{1+10^{(\text{Log}DR_{50}-x)}} \quad (\text{Eq. 6})$$

where x is the logarithm of the concentration of the fragment and y is the percent of protein bound to the fragment. For the relative inhibition values the following equation was used;

$$\text{Relative fold inhibition value} = \frac{k_d \text{ for labeled KIX H602C}}{k_d \text{ for untethered KIX H602C}} \quad (\text{Eq. 7})$$

2.5.7 Fragment Tethering

The KIX H602C mutant was incubated with 8 - 10 equivalents of small molecule and fresh β -ME in 10 mM phosphate buffer, 100 mM NaCl, pH 6.8 and incubated overnight. Excess small molecule was removed and small molecule-protein complexes were concentrated using 10 kD molecular weight cutoff concentrators (Vivascience). The extent of labeling was measured by Q-TOF LC-MS (Agilent). Protein complexes that were at least 90% alkylated were used for the FP binding assays. A control samples was included replacing the small molecule with DMSO.

2.5.8 CD wavelength and variable temperature scans

For the sample 300 μ l of 10 μ M KIX protein in 10 mM sodium phosphate,

100 mM NaCl, pH 6.8 was used. The CD cuvette was a Spectrosil far UV quartz, standard rectangular 1 mm cell with stopper (useable range 170 to 2700 nm, Starna® Cells, Inc., 21-Q-1). The CD instrument (JASCO J-175) equipped with a modified water-bath to modulate the temperature scanned the samples for the wavelength from 260 to 195 nm in continuous scanning mode at 500 nm min⁻¹ with 1.0 nm bandwidth, accumulation of 5 scans, and data pitch of 1 nm at 20°C. For data analysis, the sample's CD signal was corrected by subtracting the buffer's signal for each wavelength. The CD signal was cut-off when the dynode energy reached 500 eV or greater. Therefore, the typical scan resulted in analysis from 260 nm to 195-198 nm. The data was converted to mean residue ellipticity (deg·cm²·dmol⁻¹) using the following equation:

$$[\theta] = \frac{\Psi}{1000 n * l * c} \quad (\text{Eq. 8})$$

where Ψ was the CD signal in degrees, l was the path length in centimeters, c was the concentration in decimoles per cm², and n was the number of amides. This KIX construct contained 100 amides. The helical content was determined:

$$\% \text{ helical content} = 100 * \left[Q_{222nm}^{obs} * \frac{-39,5000 * (1 - 2.57)}{n} \right] \quad (\text{Eq. 9})$$

where Q was the ellipticity measured at 222 nm.⁴²

For the thermal melt, the temperature scan was set to read at 222 nm from 20°C to 95°C ramping 1.0 °C·min⁻¹ and collecting data for each 1 °C. The resulting data were subtracted from the data for buffer at 222 nm. The data were fitted using the nonlinear regression model ([Inhibitor] versus response using a variable slope considering four parameters) using GraphPad Prism 4 software.

2.5.9 Alkylation experiments

5 μM of KIX H602C mutant was incubated with 100 μM compound, 1 mM β-ME in 10 mM sodium phosphate buffer, 100 mM NaCl, at pH 6.8. After one hour 1 μL was injected with the 1290 Agilent autosampler onto the 1290 binary pump

(Agilent). The protein was eluted off the Proshell 300 SB-C8 column (Agilent) using a gradient of 95% water containing 0.1% formic acid (Solvent A) and 5% acetonitrile, 0.1% formic acid (Solvent B) at 0.3 ml per min. In the first two minutes 95% of solvent A was sent to waste. Subsequently the LC stream was directed to the MS and solvent B increased over three minutes to 100%. Intact mass analysis was measured by electrospray ionization on a quadrupole time-of-flight (Q-TOF) mass spectrometer (6545 Agilent Q-TOF). The MS-TOF was set to acquire spectra from 400 to 3200 m/z and set-up with an internal reference mass of 922.009798 m/z. The MS fragmentor was set to 300 V and the skimmer was at 75 V. After the protein eluted, the data was analyzed using the Agilent MassHunter Bioconfirm software. The spectra were extracted from the total ion chromatography peak after subtracting a background spectra. The raw data mass spectrum of a protein appeared as a series of multiply charged ions and these were selected for the deconvolution of the protein using the Maximum Entropy algorithm. The percent of protein labeled by the small molecules was calculated by dividing the height of the labeled protein peak over the total KIX H602C species (the sum of the labeled protein, unbound protein, and β -ME bound protein).

2.6 References

1. Lodge, J. M., Rettenmaier, T. J., Wells, J. A., Pomerantz, W. C. & Mapp, A. K. FP Tethering: a screening technique to rapidly identify compounds that disrupt protein-protein interactions. *MedChemComm* **5**, 370-375, (2014).
2. Pomerantz, W. C. *et al.* Profiling the dynamic interfaces of fluorinated transcription complexes for ligand discovery and characterization. *ACS Chem Biol* **7**, 1345-1350, (2012).
3. Berg, T. Modulation of protein-protein interactions with small organic molecules. *Angew Chem* **42**, 2462-2481, (2003).
4. Wells, J. A. & McClendon, C. L. Reaching for high-hanging fruit in drug discovery at protein-protein interfaces. *Nature* **450**, 1001-1009, (2007).

5. Koehler, A. N. A complex task? Direct modulation of transcription factors with small molecules. *Curr Opin Chem Biol* **14**, 331-340, (2010).
6. Dunker, A. K., Silman, I., Uversky, V. N. & Sussman, J. L. Function and structure of inherently disordered proteins. *Curr Opin Struc Biol* **18**, 756-764, (2008).
7. Thompson, A. D., Dugan, A., Gestwicki, J. E. & Mapp, A. K. Fine-tuning multiprotein complexes using small molecules. *ACS Chem Biol* **7**, 1311-1320, (2012).
8. Tompa, P., Schad, E., Tantos, A. & Kalmar, L. Intrinsically disordered proteins: emerging interaction specialists. *Curr Opin Struc Biol* **35**, 49-59, (2015).
9. Hopkins, A. L. & Groom, C. R. The druggable genome. *Nat Rev Drug Discov* **1**, 727-730, (2002).
10. Zinzalla, G. & Thurston, D. E. Targeting protein-protein interactions for therapeutic intervention: a challenge for the future. *Fut Med Chem* **1**, 65-93, (2009).
11. Cimermancic, P. *et al.* CryptoSite: Expanding the Druggable Proteome by Characterization and Prediction of Cryptic Binding Sites. *J Mol Biol* **428**, 709-719, (2016).
12. Braisted, A. C. *et al.* Discovery of a potent small molecule IL-2 inhibitor through fragment assembly. *J Am Chem Soc* **125**, 3714-3715, (2003).
13. Sadowsky, J. D. *et al.* Turning a protein kinase on or off from a single allosteric site via disulfide trapping. *Proc Natl Acad Sci USA* **108**, 6056-6061, (2011).
14. Teague, S. J., Davis, A. M., Leeson, P. D. & Oprea, T. The Design of Leadlike Combinatorial Libraries. *Angew Chem* **38**, 3743-3748, (1999).
15. Erlanson, D. A. *et al.* Site-directed ligand discovery. *Proc Natl Acad Sci USA* **97**, 9367-9372, (2000).
16. Erlanson, D. A., Wells, J. A. & Braisted, A. C. Tethering: fragment-based drug discovery. *Annu Rev Biophys Biomol Struct* **33**, 199-223, (2004).

17. Nonoo, R. H., Armstrong, A. & Mann, D. J. Kinetic template-guided tethering of fragments. *ChemMedChem* **7**, 2082-2086, (2012).
18. Miller, R. M., Paavilainen, V. O., Krishnan, S., Serafimova, I. M. & Taunton, J. Electrophilic fragment-based design of reversible covalent kinase inhibitors. *J Am Chem Soc* **135**, 5298-5301, (2013).
19. Jöst, C., Nitsche, C., Scholz, T., Roux, L. & Klein, C. D. Promiscuity and Selectivity in Covalent Enzyme Inhibition: A Systematic Study of Electrophilic Fragments. *J Med Chem* **57**, 7590-7599, (2014).
20. Wang, N. *et al.* Ordering a dynamic protein via a small-molecule stabilizer. *J Am Chem Soc* **135**, 3363-3366, (2013).
21. Wang, N., Lodge, J. M., Fierke, C. A. & Mapp, A. K. Dissecting Allosteric Effects of Activator-Coactivator Complexes Using a Covalent Small Molecule Ligand. *Proc Natl Acad Sci USA* **11**, 12061-12066, (2014).
22. Lakowicz, J. R. *Principles of Fluorescence Spectroscopy*. (Kluwer Academic/Plenum Publishers 1999).
23. Owicki, J. C. Fluorescence Polarization and Anisotropy in High Throughput Screening: Perspectives and Primer. *J Biomol Screen* **5**, 297-306, (2000).
24. Huang, X. Fluorescence Polarization Competition Assay: The Range of Resolvable Inhibitor Potency Is Limited by the Affinity of the Fluorescent Ligand. *J Biomol Screen* **8**, 34-38, (2003).
25. Kobayashi, M. *et al.* Target immobilization as a strategy for NMR-based fragment screening: comparison of TINS, STD, and SPR for fragment hit identification. *J Biomol Screen* **15**, 978-989, (2010).
26. Campbell, K. M. & Lumb, K. J. Structurally distinct modes of recognition of the KIX domain of CBP by Jun and CREB. *Biochemistry* **41**, 13956-13964, (2002).
27. Goto, N. K., Zor, T., Martinez-Yamout, M., Dyson, H. J. & Wright, P. E. Cooperativity in transcription factor binding to the coactivator CREB-binding protein (CBP). The mixed lineage leukemia protein (MLL) activation domain binds to an allosteric site on the KIX domain. *J Biol Chem* **277**, 43168-43174, (2002).

28. Radhakrishnan, I., Perez-Alvarado, G. C., Dyson, H. J. & Wright, P. E. Conformational preferences in the Ser¹³³- phosphorylated and non-phosphorylated forms of the kinase inducible transactivation domain of CREB. *FEBS Letters* **430**, 317-322, (1998).
29. Radhakrishnan, I. *et al.* Solution Structure of the KIX Domain of CBP Bound to the Transactivation Domain of CREB: A Model of Activator:Coactivator Interactions. *Cell* **91**, 741-752, (1997).
30. Vendel, A. C., McBryant, S. J. & Lumb, K. J. KIX-Mediated Assembly of the CBP-CREB-HTLV-1 Tax Coactivator-Activator Complex. *Biochemistry* **42**, 12481-12487, (2003).
31. De Guzman, R. N., Goto, N. K., Dyson, H. J. & Wright, P. E. Structural Basis for Cooperative Transcription Factor Binding to the CBP Coactivator. *J Mol Biol* **355**, 1005-1013, (2006).
32. Wang, N. *et al.* Ordering a Dynamic Protein Via a Small-Molecule Stabilizer. *J Am Chem Soc* **125**, 3363-3366, (2013).
33. Zor, T., De Guzman, R. N., Dyson, H. J. & Wright, P. E. Solution Structure of the KIX Domain of CBP Bound to the Transactivation Domain of c-Myb. *J Mol Biol* **337**, 521-534, (2004).
34. Inglese, J. *et al.* High-throughput screening assays for the identification of chemical probes. *Nat Chem Biol* **3**, 466-479, (2007).
35. Majmudar, C. Y. *et al.* Sekikaic Acid and Lobaric Acid Target a Dynamic Interface of the Coactivator CBP/p300. *Angew Chem* **51**, 11258-11262, (2012).
36. Gee, C. T., Koleski, E. J. & Pomerantz, W. C. K. Fragment Screening and Druggability Assessment for the CBP/p300 KIX Domain through Protein-Observed ¹⁹F NMR Spectroscopy. *Angew Chem* **54**, 3735-3739, (2015).
37. Buck, E., Bourne, H. & Wells, J. A. Site-specific disulfide capture of agonist and antagonist peptides on the C5a receptor. *J Biol Chem* **280**, 4009-4012, (2005).
38. Buck, E. & Wells, J. A. Disulfide trapping to localize small-molecule agonists and antagonists for a G protein-coupled receptor. *Proc Natl Acad Sci USA* **102**, 2719-2724, (2005).

39. Buhrlage, S. J. *et al.* Amphipathic Small Molecules Mimic the Binding Mode and Function of Endogenous Transcription Factors. *ACS Chem Biol* **4**, 335-344, (2009).
40. Pomerantz, W. C. *et al.* Profiling the Dynamic Interfaces of Fluorinated Transcription Complexes for Ligand Discovery and Characterization *ACS Chem Biol* **7**, 1345-1350, (2012).
41. Zhang, J. H., Chung, T. D. & Oldenburg, K. R. A Simple Statistical Parameter for Use in Evaluation and Validation of High Throughput Screening Assays. *J Biomol Screen* **4**, 67-73, (1999).
42. Chen, Y. H., Yang, J. T. & Chau, K. H. Determination of the helix and beta form of proteins in aqueous solution by circular dichroism. *Biochemistry* **13**, 3350-3359, (1974).

Chapter 3 Dissecting Allosteric Effects of Activator-Coactivator Complexes Using a Covalent Small Molecule Ligand[‡]

3.1 Abstract

Allosteric binding events play a critical role in the formation and stability of transcriptional activator-coactivator complexes. These often involve an intrinsically disordered protein that binds to an accommodating partner. The KIX domain of the master coactivator CBP/p300 is a conformationally dynamic motif that interacts with transcriptional activators at two discrete binding sites that are in allosteric communication. The complexation of KIX with the transcriptional activation domain of Mixed Lineage Leukemia (MLL) leads to an enhancement of binding by the activation domain of CREB (pKID) to the second site. An underlying question has been how this positive cooperativity is achieved. A transient kinetic analysis of the ternary complex formation aided by small molecule ligands that induce positive or negative cooperative binding reveals that positive cooperativity is largely governed by a decrease in the dissociation rate constant, k_{off} . This suggests that the increased binding affinity for the second ligand such as pKID is not due to an allosteric creation of a more favorable binding interface by the first ligand, MLL.

In contrast, negative cooperativity is manifested by alterations in both

[‡]This chapter are adapted and reproduced from a published article: Wang, N. K., Lodge, J. M., Fierke, C.A., and Mapp, A. K. *Proc. Natl. Acad. Sci. U.S.A.* **111**,12061-12066 (2014). This was a collaboration initiated by Dr. Ningkun Wang who conducted experiments including FP-direct binding assays, and stopped-flow kinetics for wt KIX, KIX N627C, and KIX L664C constructs. She also prepared the materials (purified protein, and peptides) required for those experiments. Dr. W. C. Pomerantz and Dr. T. Johnson provided the small molecule, **1-10** for these studies.

the association and dissociation rate constants, k_{on} and k_{off} , respectively. These data lead to a general model for allosteric regulation in transcriptional activator-coactivator complexes.

3.2 Introduction to the cooperativity between activators and coactivators

Protein-protein interactions (PPIs) underscore all cellular processes, and the mechanistic dissection of PPI networks is thus a high priority.¹⁻⁴ A particular challenge is defining the mechanism of PPI formation between intrinsically disordered proteins (IDPs) and their dynamic partners where allosteric changes play a substantive role.⁶⁻⁸ Allosteric communication between distal binding sites is vital for the proper function of regulatory circuits in the cell.^{9,10} The conformationally dynamic KIX domain of the master transcriptional coactivator CBP (CREB-binding protein) undergoes a structural shift and stabilization upon binding to a cognate ligand such as the intrinsically disordered transcriptional activation domain (TAD) of MLL (Figure 3-1).¹¹ The interaction of a second ligand, pKID (phosphorylated Kinase Inducible Domain of CREB) is enhanced up to 2-fold by the presence of MLL and several

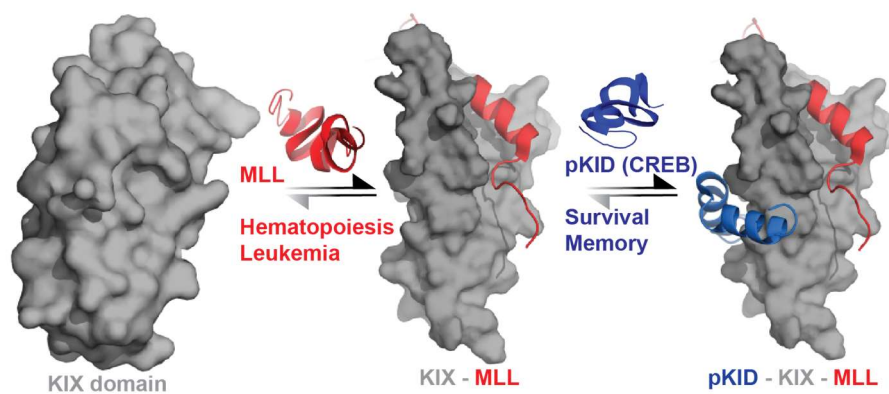


Figure 3-1 Activator-KIX domain interactions The KIX domain (grey) regulates different functions through interactions with activators from MLL(red) and pKID(blue) shown in a cartoon form. The intrinsically disordered regions of transcriptional activators form helical structures upon binding to the KIX domain which contains two discrete binding sites.

elegant studies have documented allosteric communication between the two binding sites.¹²⁻¹⁶ However, the relatively modest affinities (K_{Ds} on the order of 1-25 μM) of the native complexes and the aggregation propensities of the transcriptional activation domains have hampered kinetic dissection of this and other complexes.^{17,18} High concentrations of ligand are required to maintain these fully bound complexes at equilibrium. NMR has shown that at high concentrations, ligands tend to bind nonspecifically to KIX complicating the evaluations between the binary and ternary complexes.^{19,20}

The coactivator CBP exists across metazoans²¹⁻²³ and is a transcription hub that interacts with numerous transcriptional activators using several discrete domains.^{24,25} The KIX domain within CBP is a 90-residue motif that consists of a three helix bundle along with two 3_{10} helices.¹⁴ It interacts with more than ten distinct transcriptional activators,²⁵ via two different binding sites, with the resulting complexes leading to regulation of key functions such as hematopoiesis, cell cycle progression, and memory formation.^{14,16,26-28} Many of the KIX-binding TADs from activators such as pKID, MLL and c-Myb are intrinsically disordered regions that assume a helical structure only upon binding to interaction partners such as KIX.^{26,29,30} Conformational changes in both the TADs and KIX occur upon binding.^{11,12,14,16,28} However, the transient-state kinetic mechanism of two ligands binding cooperatively to the KIX domain has yet to be defined, and there is little information on the mechanism of how different protein ligands elicit differential degrees of allosteric effects.^{13,15} Further insights into these processes are crucial for defining the

communication pathways within CBP interaction networks.

PPIs between IDPs and their binding partners are complex, affecting both the thermodynamic binding affinity of the interaction (dissociation constants, K_{DS}) and kinetic properties. Previously, stopped-flow transient-state kinetic studies were used to examine complexes between activators and coactivators.³¹ Here we hypothesized that a similar strategy could be applied to dissect the allosteric communication between the MLL and pKID binding sites of native KIX. As a complementary tool in this effort we use a covalent small molecule ligand of the KIX domain, fragment **1-10**, that quantitatively forms a stable complex at the MLL binding site and in doing so alters the conformational dynamics of KIX (Figure 3-2A). This small molecule can influence binding at the second site either positively or negatively, depending on the site of covalent attachment, thus facilitating transient kinetic studies of KIX in its complexes with native binding partners to examine both positive and negative cooperativity. We find that the change in the dissociation rate of the TAD•KIX complex is the main driving force for changes in overall binding

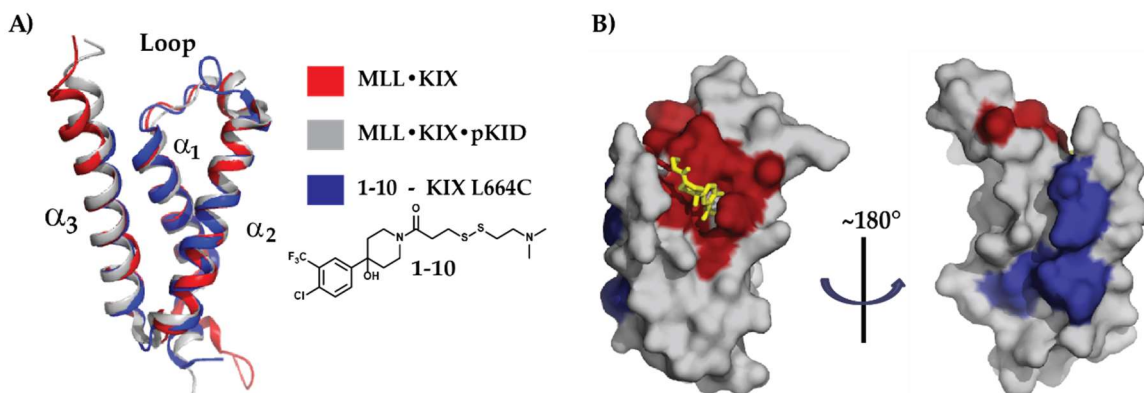


Figure 3-2 Comparing the conformations of KIX (A) Overlay of KIX structures shows **1-10**-KIX L664C⁵, with MLL, and both pKID and MLL. Note the structural shift in the loop and terminal regions upon ligands binding to KIX. (B) The binding sites of MLL (fire-brick red) and pKID (blue) on KIX, with **1-10** depicted in yellow from structure **1-10**-KIX L664C. PDB IDs: 4I9O, 2LXS, 2LXT

affinities in positive cooperative binding events. These results suggest that KIX allosterically regulates its binding events not simply through altering interaction interfaces, but also through stabilizing the energetic state of the ternary complex. Thus, in this IDP protein complex, dissociation rate mainly governs the binding equilibrium.

3.3 Results

The goal in this study was to kinetically dissect the continuum of cooperativity in KIX ternary complex formation. Toward that end, in addition to wild-type (wt) KIX domain in complex with one or both of the native transcriptional activation domains of MLL and pKID, we used several KIX mutants as well as the cysteine-reactive small molecule ligand for KIX, compound **1-10** (Figure 3-2B). Replacement of isoleucine 660 with a valine has been shown to disrupt allosteric communication within KIX, providing a valuable control for these experiments.^{11,12} The two additional KIX mutants are those in which a cysteine residue replaces either N627 or L664, both of which border the MLL binding site. Both the N627C and L664C mutants maintain affinity for MLL and for pKID.^{5,32} Previously, we described the discovery of compound **1-10** through a Tethering screen of KIX L664C.⁵

This small molecule forms a selective covalent complex through formation of a disulfide bond with the cysteine residue and, in doing so, occupies the MLL binding site. Furthermore, equilibrium binding experiments with the **1-10**-KIX L664C complex demonstrated that **1-10** tethered to a cysteine residue at position 664 allosterically inhibits pKID binding. Parallel to the KIX L664C screen we carried out a Tethering screen of N627C and compound **1-10** emerged as a molecule that selectively tethers to that cysteine with high efficiency as well³²; the binding properties of the **1-10**-KIX N627C

complex are defined below. An advantage of 1–10–KIX L664C and 1–10–KIX N627C complexes is that they can be generated quantitatively in a disulfide exchange reaction and then purified, enabling a single species to be used in kinetic assays.

3.3.1 Equilibrium Binding Measurements Define Extent of Cooperativity for Each Complex.

Dissociation constant measurements of wild-type KIX and mutants either in complex with excess MLL or fully labeled with small molecule 1–10

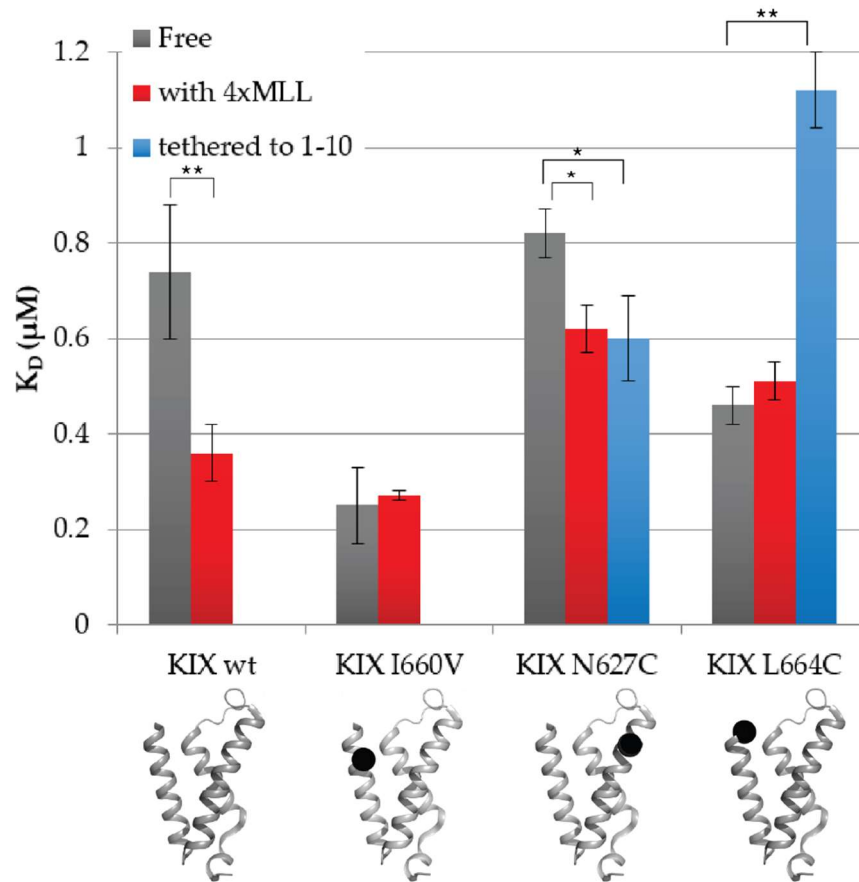
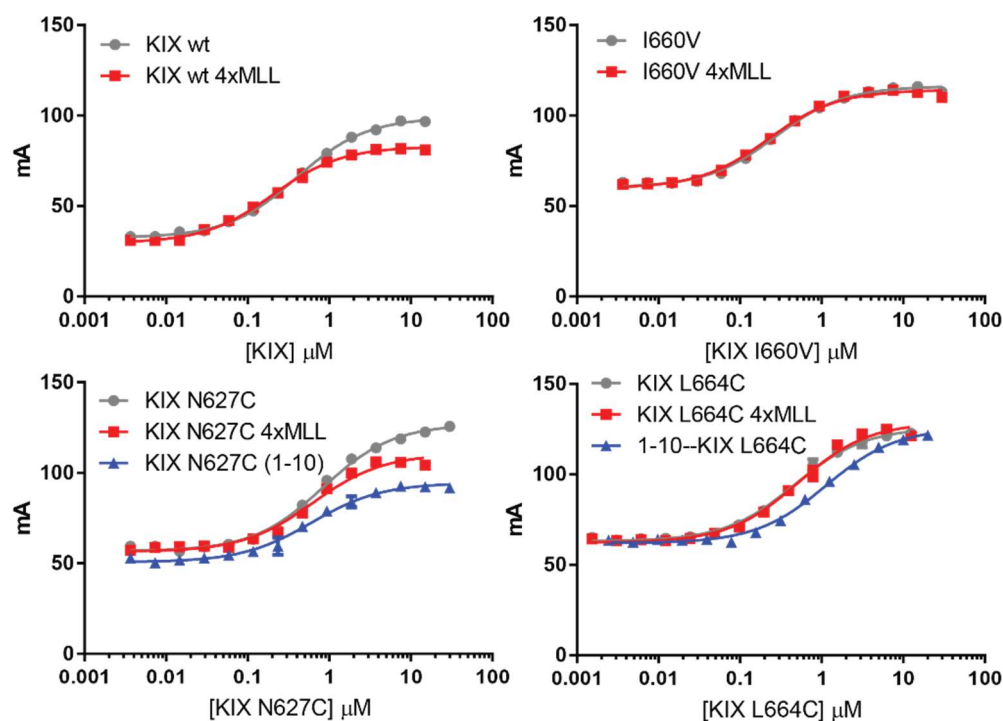


Figure 3-3 Bar graph of relative KD values of pKID•KIX complexes normalized to the KD of pKID and free KIX constructs The KD values were obtained from fluorescence anisotropy assays performed in triplicate. The KIX structure (PDB ID: 4I9O) with the corresponding mutated residue depicted in a black sphere is shown under each KIX construct. Bars below the dotted line reflect positive cooperativity whereas bars above the line reflect negative cooperativity. Errors reflect the standard error (SE) of nonlinear fits in GraphPad Prism propagated for normalization to free KIX constructs. *0.01 < P < 0.05, **P < 0.01; P values are calculated by GraphPad Prism.



pKID K_D , μM	Free	With MLL	With 1-10
KIX wt	0.74 ± 0.06	0.36 ± 0.06	NA
KIX I660V	0.25 ± 0.01	0.27 ± 0.02	NA
KIX N627C	0.82 ± 0.05	0.62 ± 0.08	0.60 ± 0.09
KIX L664C	0.46 ± 0.04	0.51 ± 0.04	1.12 ± 0.08

Figure 3-4 Anisotropy binding curves of complexes of KIX wt, KIX I660V, KIX N627C, and KIX L664C binding to pKID All measurements were performed in triplicate and the error bars indicate the standard deviation (SD) error of the three measurements. The K_D values and SE are shown in the table and they were calculated in GraphPad Prism.

were carried out to define the extent of cooperativity for each of the complexes (Figure 3-3 and Figure 3-4). As expected, KIX I660V precomplexed with excess MLL exhibited no cooperative binding to pKID. However, KIX N627C either precomplexed with excess MLL or fully labeled with small molecule 1-10 reproduced the cooperative binding behavior observed in the wild-type protein. In contrast, when 1-10 was fully tethered to KIX L664C, negative

cooperativity was observed, with a more than twofold loss of pKID affinity.

3.3.3 Transient-State Kinetic Analysis of KIX Complexes.

Subsequently, we examined the transient-state binding kinetics between KIX and pKID by fluorescence stopped flow in the presence and absence of the peptide and small molecule ligand at room temperature (Figure 3-6A). The association rate is determined from the changes in fluorescence intensity after rapidly mixing fluorescently labeled pKID with KIX complexes and the dissociation rate is measured from changes in fluorescence intensity when the labeled-pKID•KIX complexes are mixed with an excess of unlabeled pKID peptide. Fluorescein is tagged at the amino terminus of the pKID peptide, a construct that has been used in previous studies³³⁻³⁵ to show allosteric effects consistent with results from isothermal calorimetry (ITC) assays using unlabeled pKID.¹⁵ One rapid association phase was observed upon mixing pKID with varying concentrations of KIX complexes and the observed association rates exhibit a linear dependence on the KIX complex concentration (Figure 3-5). The slope from the linear regression fit is the rate constant (k_{on}) of the association step. The value of k_{on} for pKID binding to free KIX wt ($k_{on} = 14 \pm 1 \mu\text{M}^{-1}\cdot\text{s}^{-1}$) is of the same magnitude as values obtained by NMR relaxation dispersion in the literature ($6 \mu\text{M}^{-1}\cdot\text{s}^{-1}$)²⁶, confirming that this is a viable method for studying the transient-state interaction between pKID and KIX. We did not observe an additional rapid conformational change step (approximately 200 s^{-1}) reported in the literature due to time resolution constraints of the stopped-flow instrument (dead time $\sim 2 \text{ ms}$). The calculated association and dissociation rate constants measured for the various pKID•KIX complexes are summarized in Table 3-1, and Figure 3-6.

Table 3-1 Association and dissociation rate constants of pKID·KIX complexes Association rate constants are slopes of linear regression fits. Errors reflect the SE of nonlinear fits in GraphPad Prism. For dissociation rate constants, errors are SDs of the average of two separate values.

Construct	Rate Constants	Free	with 4x MLL	with 1-10
KIX wt	k_{on} ($\mu\text{M}^{-1} \text{s}^{-1}$)	14 ± 1	14 ± 1	NA
	k_{off} (s^{-1})	7.0 ± 0.4	5 ± 1	NA
KIX I660V	k_{on} ($\mu\text{M}^{-1} \text{s}^{-1}$)	14 ± 1	11 ± 1	NA
	k_{off} (s^{-1})	5.3 ± 0.5	5.4 ± 0.3	NA
KIX N627C	k_{on} ($\mu\text{M}^{-1} \text{s}^{-1}$)	15 ± 1	14 ± 1	17 ± 2
	k_{off} (s^{-1})	7 ± 1	4 ± 3	4.0 ± 0.8
KIX L664C	k_{on} ($\mu\text{M}^{-1} \text{s}^{-1}$)	23 ± 2	24 ± 2	6 ± 1
	k_{off} (s^{-1})	8.7 ± 0.4	7.7 ± 0.3	13.8 ± 0.6

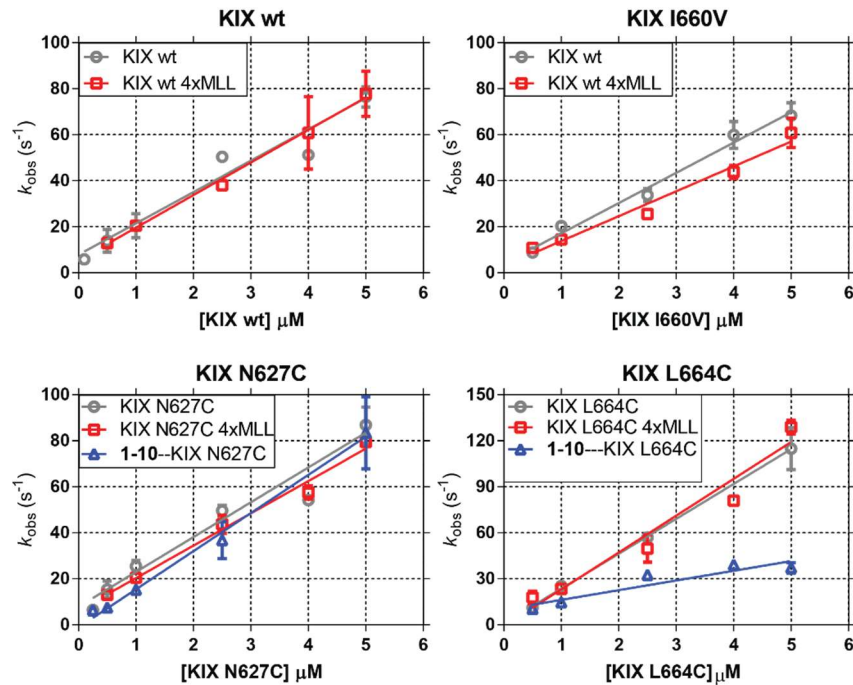


Figure 3-5 Stopped-flow binding kinetics for the observed association rates For measuring the association rate constant excess KIX complex of varying concentrations is rapidly mixed with a constant amount of FITC-pKID peptide (final concentration 25 nM) at 25 °C, and the change in fluorescence intensity is measured in real time. The dependence of k_{obs} on the concentration of excess KIX for formation of pKID•KIX complexes (wild-type KIX, KIX I660V, KIX N627C, and KIX L664C), which suggests this association step is a bimolecular interaction. The linear regression of each slope approximates the association rate constant (k_{on}) of this interaction. Each k_{obs} value is an average of two separate experiments; each experiment is an average of five to eight traces. The error bars represent the SD of the two k_{obs} values.

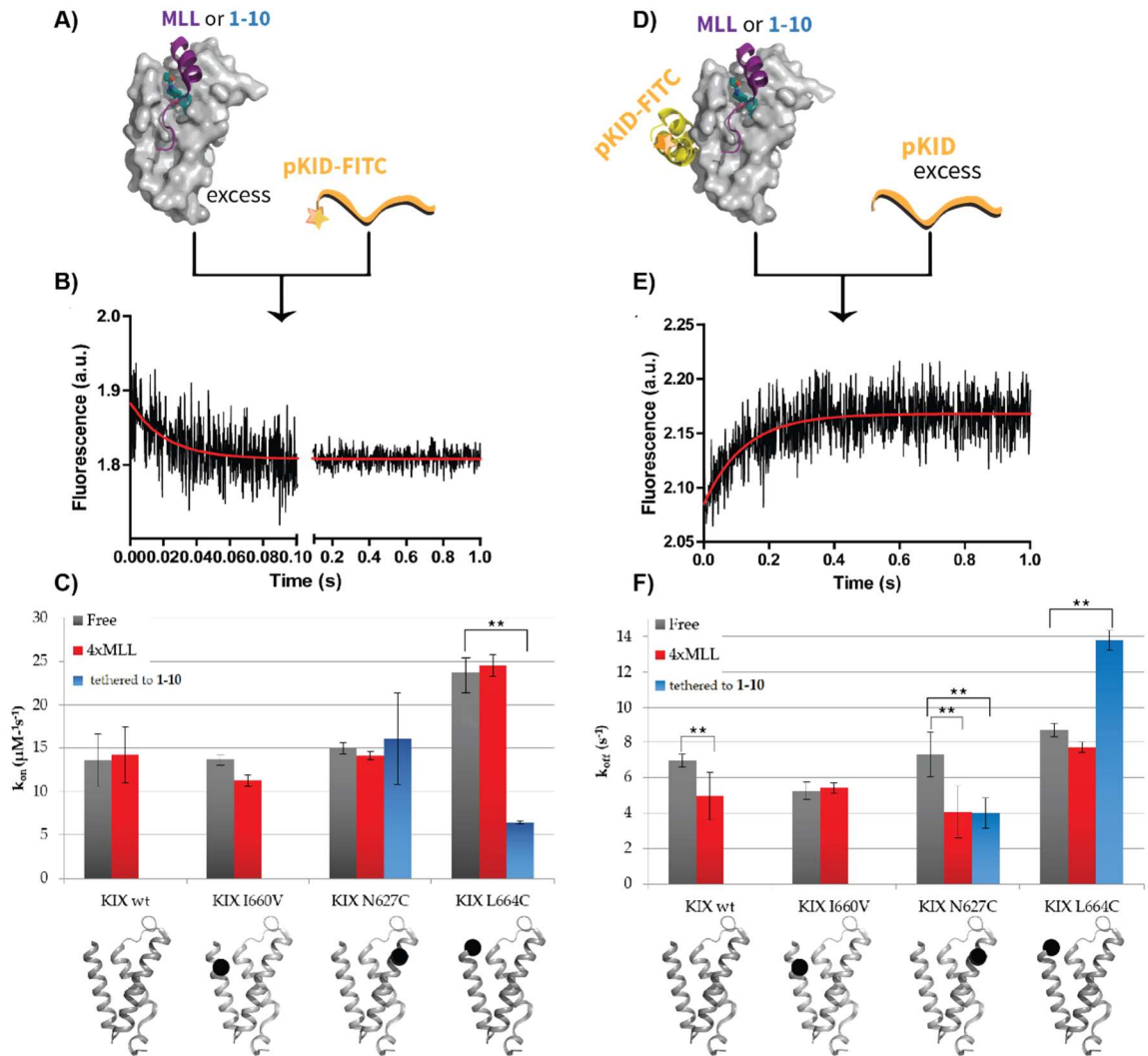


Figure 3-6 Comparing the K_{on} and K_{off} between pKID and KIX complexes (A) The set-up for measuring the association rate constant. An excess KIX complex of varying concentrations is rapidly mixed with a constant amount of FITC-pKID peptide (final concentration 25 nM) at 25 °C. (B) The trace is shown for the change in fluorescence intensity in real time. (C) Bar graph of k_{on} between pKID and KIX complexes. The error bars represent SD of the average of two values from separate experiments. *0.01 < P < 0.05, **P < 0.01; P values are calculated by GraphPad Prism. (D) The schematic shows the experimental set-up for measuring the dissociation rate constant of pKID from KIX pre-complexed with FITC-pKID. This complex was rapidly mixed with an excess of unlabeled pKID peptide. (E) Dissociation traces of pKID from KIX complexes by competition of excess amounts of unlabeled pKID peptide. The red lines represent the fitted curve. All spectra in (B) and (D) were obtained by excitation at 493 nm and emission collected with a 510 (LP) filter. All spectra are an average of 5-8 separate traces. (F) Bar graph of dissociation rate constants of pKID dissociating from KIX complexes. The error bars represent SD of the average of two values from separate experiments. *0.01 < P < 0.05, **P < 0.01; P values are calculated by GraphPad Prism.

The kinetics for the binding of pKID to the various KIX complexes show a positive correlation between the change in the affinity of KIX for pKID (K_D) and the dissociation rate of pKID from the KIX•pKID complex (k_{off}). Specifically, N627C shows positive cooperativity upon addition of either MLL or **1-10**, increasing the binding affinity of pKID to KIX by nearly twofold, whereas the value of k_{off} similarly decreases almost two-fold from 7.3 s^{-1} to 4.1 s^{-1} with no observed effect on the k_{on} , which is in a constant range between $14 \mu\text{M}^{-1}\cdot\text{s}^{-1}$ and $17 \mu\text{M}^{-1}\cdot\text{s}^{-1}$. Therefore, in this system, positive cooperativity is primarily determined by k_{off} . KIX L664C previously showed negative cooperativity upon tethering to **1-10**.⁵ Whereas an increase in k_{off} contributes to this allosteric change as well (from 8.7 s^{-1} to 13.8 s^{-1}), there is also a decrease in k_{on} values from $23 \mu\text{M}^{-1}\cdot\text{s}^{-1}$ to $6 \mu\text{M}^{-1}\cdot\text{s}^{-1}$. The control KIX variant, KIX I660V, in which the allosteric network is interrupted, shows no significant change in either the association or dissociation rate constants with and without the presence of MLL.

3.3.2 Testing for stabilizing effects with KIX mutants

The allosteric mechanism was shown to depend upon the structural stability of the domain in molecular dynamics simulations by Law, S.E. *et al.*³⁶ They identified a combination of six mutations at two residues within the helix α_3 that were predicted to increase KIX's helicity and the affinity for c-Myb mimicking the allosteric effect from the MLL-KIX interaction. A few of these mutants were generated and their overall helicity was assessed by circular dichroism (CD). From these experiments, the helicity appeared to moderately decrease and subsequently pKID and c-Myb did not show an enhanced affinity for KIX over wild-type KIX (Figure 3-7 and 3-8). Previously at lower pH c-Myb was shown to increase its

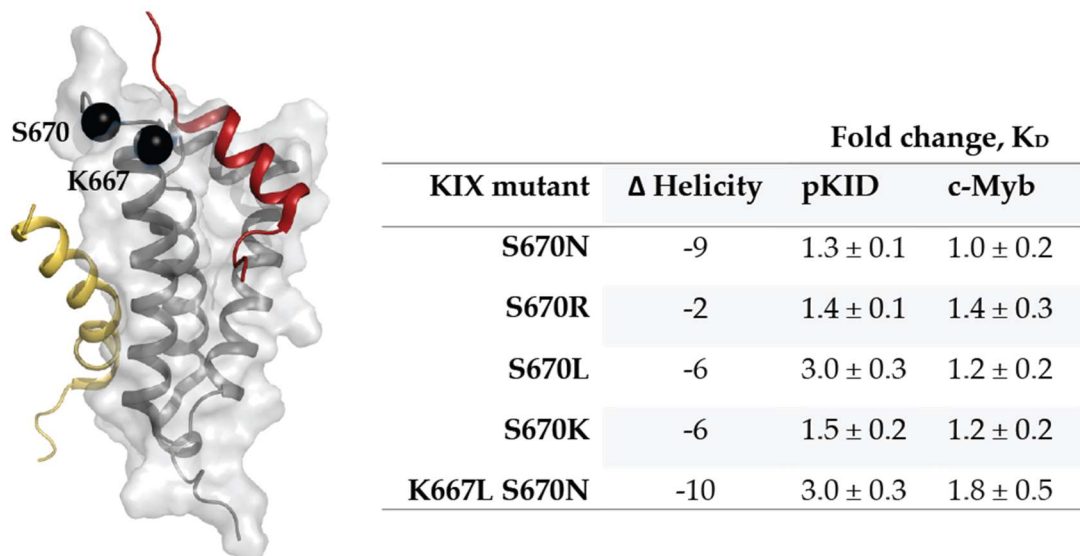


Figure 3-7 Mutants aimed to stabilize KIX and induce positive cooperativity The position of the chosen residues are shown as S670 and K667. These mutants were assessed for their change in helicity by CD and their relative change in K_D for c-Myb and pKID by FP-direct binding assays.

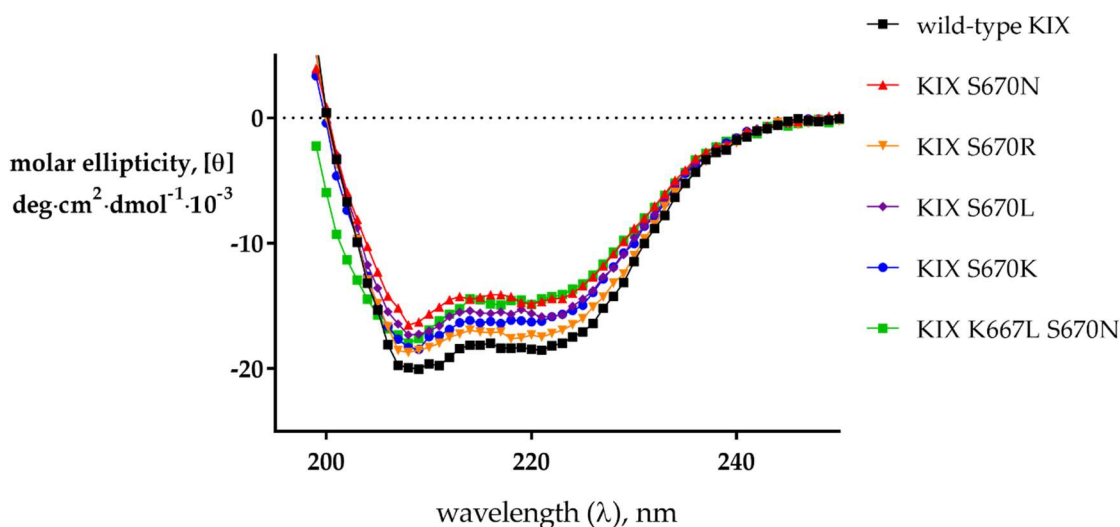


Figure 3-8 Analyzing the helicity of KIX mutants The selected mutants were scanned by CD and the helicity was determined from the molar ellipticity at 222 nm.

affinity for KIX due to more favorable enthalpy.²⁸ By decreasing the pH from 6.8 to 5.5 the affinity of c-Myb for some of these mutants increased by 2-fold (Figure 3-9). However, this enhancement was attributed to a favorable enthalpic contribution while these mutants were proposed to reduce the entropic cost.^{28,36,37} Finally, the predicted helicity changes were minimal (4.2% for KIX S670N to 1.2%

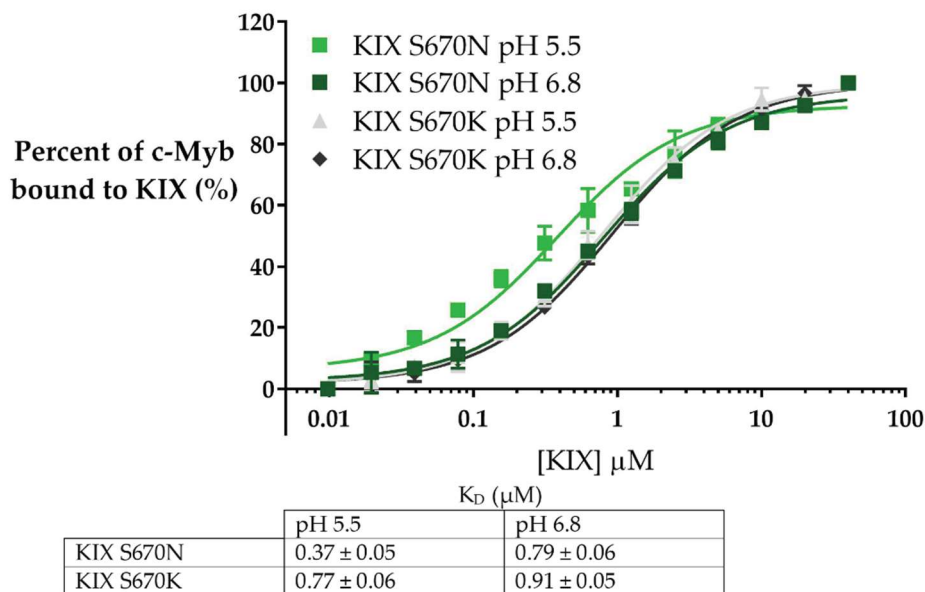


Figure 3-9 Equilibrium binding results for KIX S670N and S670K against c-Myb The pH of the buffer was changed from pH 6.8 to pH 5.5. The binding curve measurements were performed in triplicate and the error bars represent the SD of all three measurements. The K_D values and SE errors are shown in the table and they calculated with nonlinear fits in GraphPad Prism 7.0.

for KIX S670L).³⁶ Consequently, the mutants may have induced subtle changes that were not readily observed with these CD and FP-direct binding assays. Finally, double and triple mutants may contribute to the stability of the third helix and this increase in the stability and helicity of KIX may detected more readily.

3.4 Discussion on cooperativity

Taken together, these data provide insight to the mechanism of ternary activator-coactivator complex formation (Figure 3-10a). In the case of positive cooperativity, the consistent k_{on} values and the correlation of relative k_{off} and K_D values (Figure 3-10b) suggest that this cooperativity is achieved not by lowering the energy barrier for association but by increasing the energy barrier for dissociation. This indicates a scenario where the binding of one ligand to KIX does not significantly change the interaction interface for the second ligand, but when both ligands bind to KIX, this ternary complex occupies an energy minimum that

is more stable than that of the binary complex. Therefore, the major effect of positive cooperativity is to stabilize the ternary complex (Figure 3-10c). That k_{off} plays a major role in the cooperative binding of two activators to KIX is reasonable considering that our measured k_{on} values ($0.7\text{-}2.4 \times 10^7 \text{ M}^{-1}\text{s}^{-1}$, Table 3-1) are close to the predicted maximal diffusion rate of $10^7\text{-}10^9 \text{ M}^{-1}\text{s}^{-1}$ for proteins in a cellular environment.^{30,38-40} Thus when cellular regulation requires an allosteric increase in protein-protein binding affinity, there is little room for increase in the association rate constant (k_{on}). Alternatively, the dissociation rates (k_{off}) are not governed by diffusion limits and can be significantly altered, providing access to a large range of binding affinities. The dissociation rates of transcriptional activators from DNA or small molecule ligands from proteins^{41,42} have been shown to govern the residence time of proteins and ligands in particular functional states and/or

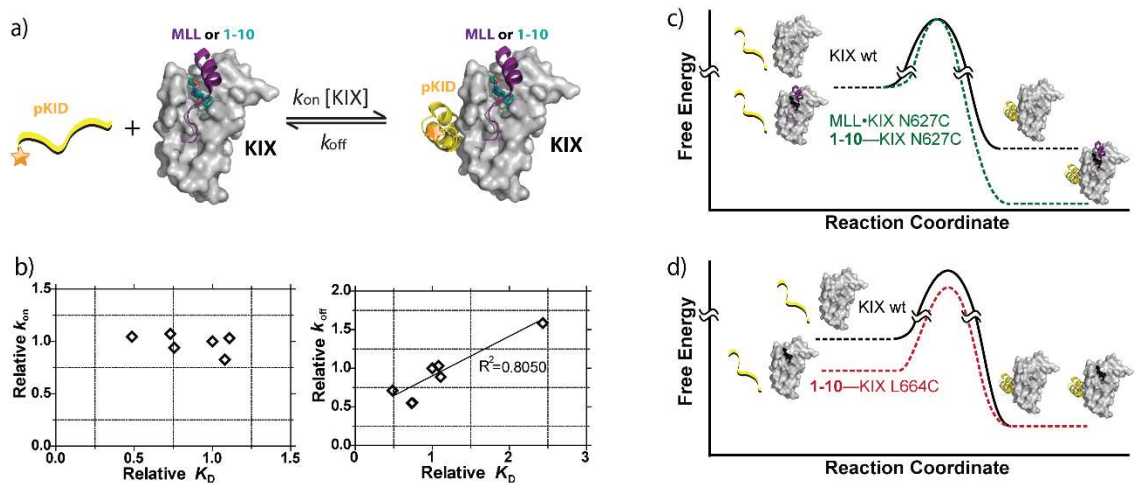


Figure 3-10 Summary for the k_{on} and k_{off} analysis (a) A mechanism scheme for pKID association with KIX complexed with peptide (in purple, MLL) or small molecule (in cyan, 1-10) ligand, where the k_{on} is a bimolecular process. The orange star on pKID depicts the position of the fluorescein label, which is not at the primary site of interaction. (b) Cooperativity correlation plot of relative k_{on} vs. K_D (Left) and k_{off} vs. K_D values (Right) (normalized to values for free KIX binding to pKID for each KIX construct). These plots illustrate that there is little correlation between k_{on} and K_D values whereas k_{off} values linearly correlate with K_D values exhibiting an R^2 value of 0.8. (c) A schematic energy diagram of the proposed mechanism of positive cooperativity through stabilizing the final ternary complex (resulting in increase of k_{off}). (d) A schematic energy diagram of the proposed mechanism of negative cooperativity through stabilizing the initial binary complex (resulting in both decrease in k_{on} and increase in k_{off}).

locations. It is possible that in this case the residence time of various TADs with the coactivator CBP is used to regulate the degree of transcriptional activation and/or histone acetyl transferase activity in cellular pathways.

The analysis of negative cooperative binding with KIX (**1-10** tethered to KIX L664C) reveals a more complicated scenario. In this instance both k_{on} and k_{off} are affected, with k_{on} decreasing 4-fold and k_{off} increasing 1.6-fold (Table 3-1). In a case when negative cooperativity is required in physiological regulation, there is ample room for the decrease of k_{on} from diffusion-controlled limits. We noted from the chemical perturbation data that the residues perturbed by **1-10** upon tethering to KIX L664C are similar to those altered by the binding of sekikaic acid, a natural product small molecule that also allosterically decreases pKID binding affinity⁴³. In the case of both **1-10** and sekikaic acid, the residues 608-611 and 629-631, which flank the loop region in helices α_1 and α_2 , exhibit significant alterations. As we have identified two small molecule ligands of KIX, **1-10** and sekikaic acid, that both show negative cooperativity with pKID binding through standard screens, it is possible that negative allosteric regulation of KIX binding occurs with native ligands, a scenario currently under investigation.

As demonstrated here, mutants and small molecules that lock the allosteric network within KIX are useful to understand the binding mechanisms for these conformationally dynamic complexes. The proposed mutants that promote positive cooperativity would also aid in understanding its function. They were chosen based on their ability to increase the overall helicity as shown by the calculation of KIX.^{36,44,45} Because the allosteric enhancement depends on the stability of KIX these mutants were expected to enhance the binding affinity for c-Myb or pKID. From our work, so far, this helicity and enhanced affinity was not observed. However, the predicted changes were subtle - enhancing the overall helicity of KIX by no more than 4.2%. Therefore, the mutants may require different

experimental techniques to judge their effects. Previously other NMR and top-down electron capture dissociation of gas-phase structures of KIX led to the proposal that this third helix is more stable than the first two.⁴⁶⁻⁴⁸ This may reflect how the stability of this helix is important for the formation of KIX-activator complexes.

3.5 Conclusion

The transient kinetic experiments described here were significantly influenced by the availability of a covalent small molecule ligand for KIX. Not only do covalent ligand-protein complexes offer the advantage of an isolable, homogeneous sample for study, they enable us to access a continuum of cooperativity modes, significantly expanding the breadth of the analysis. Given the ease with which such covalent small molecules can be discovered via Tethering, we anticipate the broader use of such ligands in mechanistic studies of conformationally dynamic protein complexes that comprise critical cellular functions.⁴⁹⁻⁵¹

3.6 Methods and supplemental experimental information

3.6.1 Preparation of protein and peptides

Cloning, expression and purification of the KIX domain of CBP (residues 586-672) and the KIX mutants were carried out as previously described.^{52,53} Fragment **1-10** was tethered to KIX mutants by incubating 10 equivalents of **1-10** in 50 mM phosphate buffer at pH 6.8 overnight.⁵ Excess small molecule was removed by illustra NAP-5 desalting column (GE Healthcare). The column was equilibrated with 10 ml of buffer and 0.25 or 0.5 ml of the sample was added, and eluted with 0.7 or 1.5 ml into storage buffer (10 mM Sodium Phosphate, 100 mM NaCl, 0.01 % NP-40, 10 % glycerol, pH 6.8) by gravity flow. The extent of tethering was verified by Q-TOF LC-MS (Agilent Technologies). When at least 95% of the

KIX cysteine mutant was labeled by the small molecule, the sample was flash frozen in liquid nitrogen and stored at -80°C. Peptides were synthesized by Fmoc-based solid phase synthesis methods, purified by HPLC, and verified by LC-MS as previously described.⁵³ The sequences were as follows:

(unlabeled) pKID: Ac-βA-TDSQKRREILSRRPS(Phos)YRKILNDLSSDAPG

FITC-pKID: FITC-βA-TDSQKRREILSRRPS(Phos)YRKILNDLSSDAPG

FITC MLL19: FITC-βA-ADAGNILPSDIMDFVLKNTF

FITC c-Myb25: FITC-βA-KEKRIKELELLLMSTENELKGQQALW

3.6.2 Fluorescence anisotropy assays

The fluorescence anisotropy assays were performed in triplicate with a final sample volume of 10 μL in a low volume, non-binding, black, 384-well plate (Corning), and read using a Tecan Genios Pro plate reader with polarized excitation at 485 nm and emission intensity measured through a parallel and perpendicularly polarized 535 nm filter. FITC (Fluorescein isothiocyanate)-labeled peptides were diluted in storage buffer to a concentration of 25 nM. Then 10 μL of the peptide solution was added to a series of 50 μL solutions of various KIX concentrations in storage buffer to obtain the final concentrations of up to 20 μM. The samples were incubated for 30 min at room temperature before the degree of fluorescence anisotropy was measured (Tecan Genios Pro). Anisotropy data were corrected for the change in fluorescence intensity using Equation 1, where “ f_b ” is the fraction bound, “ R ” is the ratio between fluorescence intensity of 100% bound ligand and free ligand, “ r ” is the measured anisotropy value, “ r_F ” is the anisotropy

value of free ligand and “ r_B ” is the anisotropy value of 100% bound ligand.

$$f_B = \frac{r - r_F}{R(r_B - r) + r - r_F} \quad \text{Eq. 1}$$

A binding isotherm that accounts for ligand depletion (assuming a 1:1 binding model of peptide to KIX) was fit to the observed anisotropy values as a function of KIX concentration to obtain the apparent equilibrium dissociation constant, K_D :

$$y = c + (b - c) \times \frac{[(K_D + a + x) - \sqrt{(K_D + a + x)^2 - 4ax}]}{2a} \quad \text{Eq. 2}$$

where “ a ” and “ x ” are the total concentrations of fluorescent peptide and KIX, respectively, “ y ” is the observed anisotropy at any KIX concentration, “ b ” is the maximum observed anisotropy value, and “ c ” is the minimum observed anisotropy value. Each data point is an average of three independent experiments with the indicated error (standard deviation). Data analysis was performed using GraphPad Prism 5 software.

An analogous experiment with varying molecular equivalents of MLL to KIX was also carried out to determine that at least 4 molecular equivalence of MLL to KIX is required to elicit maximum degree of positive cooperativity.

3.6.3 Fluorescence stopped-flow kinetic assays

Stopped-flow experiments were performed on a KinTek model SF-2001 stopped-flow equipped with a 75W Xe arc lamp in two-syringe mode. FITC was excited at 493 nm and its emission was monitored at wavelengths > 510 nm using a long-pass filter (Corion).

All kinetic traces reported are an average of five to eight independent determinations. All traces displayed a monophasic association and fit to a single exponential equation (Eq. 3). Below, the single exponential function was fit to the transient kinetic time courses, $F(t)$ as in Eq. 3, to obtain the fluorescence amplitude (A) and the observed rate constant, k_{obs} ; $F(0)$ is the initial fluorescence intensity and

t is the time.:

$$F(t) = A[1 - \exp(-k_{obs}t)] + F(0) \quad (\text{Eq. 3})$$

Analysis of the time courses was performed using Kintek software, and the reported errors are the asymptotic standard errors. The dependence of the observed rates on KIX concentration was analyzed using GraphPad Prism 4.0 software.

Control Experiments: 25 nM of pKID-FITC was rapidly mixed with either KIX storage buffer (10 mM Sodium Phosphate, 100 mM NaCl, 0.01 % NP-40, 10 % glycerol, pH 6.8) (Figure 3-11a) to control for potential photobleaching of the fluorophore or 0.1 mg/ml of BSA in KIX storage buffer (Figure 3-11b) to control for effects of nonspecific binding and crowding of protein. Both control experiments revealed no change in fluorescence intensity over 120 seconds.

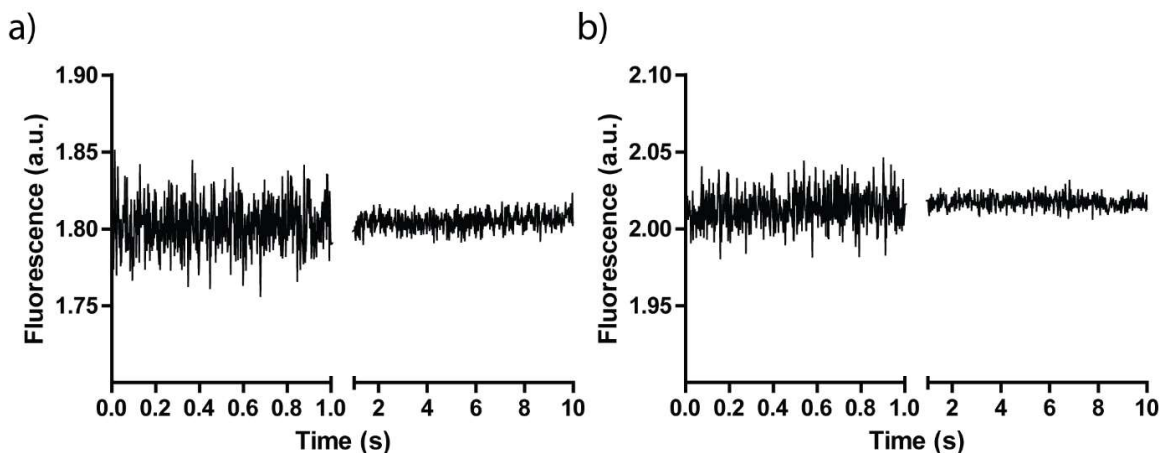


Figure 3-11 Stopped-flow traces of control experiments: (a) 25 nM pKID-FITC rapidly mixed with KIX storage buffer; (b) 25 nM pKID-FITC rapidly mixed with 0.1 mg/mL BSA in KIX storage buffer. Both spectra were obtained by excitation at 493 nm and emission was collected with a 510 (LP) filter. Both spectra are an average of five to eight separate traces.

Association Experiments: 25 nM (final concentration after mixing) of pKID-FITC in KIX storage buffer was rapidly mixed with varying concentrations (0.1-5 μ M, final concentration after mixing) of KIX complexes (free protein, pre-complexed with 4 equivalents of MLL peptide, or pre-tethered with **1-10**) in KIX storage buffer at 25°C. The measured time domains (0.1 and 1 second or 0.05 and 0.5 second) were selected to enhance the data analysis as they were closest to the predicted best time-frame by the fits in the Kintek software.

Dissociation Experiments: 25 nM (final concentration after mixing) of pKID-FITC in KIX storage buffer was pre-equilibrated with 500 nM KIX complex (final concentration after mixing) and rapidly mixed with 12.5 μ M (500 molar equivalents, final concentration after mixing) of unlabeled pKID peptide in KIX storage buffer at 25°C. The results are summarized in Table 3-1. The Y-intercepts from (k_{obs} vs KIX concentration) approximately correlate with the off-rates fitted in the dissociation experiments. Three of the Y-intercept values had a larger error of fitting, rendering them with too high a degree of uncertainty to compare with experimental dissociation values.

3.6.4 CD analysis of KIX mutants

For the sample 300 μ l of 10 μ M KIX protein in 10 mM sodium phosphate, 100 mM NaCl, pH 6.8 was used. The CD cuvette was a Spectrosil far UV quartz, standard rectangular 1 mm cell with stopper (useable range 170 to 2700 nm, Starna® Cells, Inc., 21-Q-1). The CD instrument (JASCO J-175) equipped with a modified water-bath to modulate the temperature scanned the samples for the wavelength from 260 to 195 nm in continuous scanning mode at 500 nm min⁻¹ with 1.0 nm bandwidth, accumulation of 5 runs, and data pitch of 1 nm at 20°C. For data analysis, the sample's CD signal was corrected by subtracting the buffer's signal for each wavelength. The CD signal was cut-off when the dynode energy reached

500 volts or greater. Therefore, the typical scan resulted in analysis from 260 nm to 195-198 nm. The data was converted to mean residue ellipticity (deg·cm²·dmol⁻¹) using the following equation:

$$[\theta] = \frac{\Psi}{1000 n * l * c} \quad (\text{Eq. 4})$$

where Ψ was the CD signal in degrees, l was the path length in centimeter, c was the concentration in decimoles per cm², and n was the number of amide bonds (100 amides in KIX). The helical content was determined using the following equation⁵⁴:

$$\% \text{ helical content} = 100 * \left[[\theta] * \frac{-3,5000 * (1 - 2.57)}{n} \right] \quad (\text{Eq. 5})$$

For the thermal melt, the temperature scan was set to read at 222 nm from 20°C to 95°C ramping 1.0 °C·min⁻¹ and collecting data for each 1 °C. The resulting data were subtracted from the data for the equivalent scans of buffer only at 222 nm. The data were fit using the nonlinear regression model response using a variable slope considering four parameters) using GraphPad Prism 7 software (La Jolla, CA).

3.7 References

1. Bonetta, L. Protein-protein interactions: Interactome under construction. *Nature* **468**, 851-854, (2010).
2. Nooren, I. M. & Thornton, J. M. Diversity of protein-protein interactions. *EMBO J* **22**, 3486-3492, (2003).
3. Thompson, A. D., Dugan, A., Gestwicki, J. E. & Mapp, A. K. Fine-tuning multiprotein complexes using small molecules. *ACS Chem Biol* **7**, 1311-1320, (2012).
4. Garner, A. L. & Janda, K. D. Protein-protein interactions and cancer: targeting the central dogma. *Curr Top Med Chem* **11**, 258-280, (2011).
5. Wang, N. *et al.* Ordering a dynamic protein via a small-molecule stabilizer. *J Am Chem Soc* **135**, 3363-3366, (2013).

6. Schreiber, G. & Keating, A. E. Protein binding specificity versus promiscuity. *Curr Opin Struct Biol* **21**, 50-61, (2011).
7. Kim, P. M., Sboner, A., Xia, Y. & Gerstein, M. The role of disorder in interaction networks: a structural analysis. *Mol Syst Biol* **4** (2008).
8. Tsai, C.-J., Ma, B. & Nussinov, R. Protein–protein interaction networks: how can a hub protein bind so many different partners? *Trends Biochem Sci* **34**, 594-600, (2009).
9. Ferreon, A. C. M., Ferreon, J. C., Wright, P. E. & Deniz, A. A. Modulation of allostery by protein intrinsic disorder. *Nature* **498**, 390-394, (2013).
10. Garcia-Pino, A. *et al.* Allostery and intrinsic disorder mediate transcription regulation by conditional cooperativity. *Cell* **142**, 101-111, (2010).
11. Brüschweiler, S. *et al.* Direct Observation of the Dynamic Process Underlying Allosteric Signal Transmission. *J Am Chem Soc* **131**, 3063-3068, (2009).
12. Brüschweiler, S., Konrat, R. & Tollinger, M. Allosteric Communication in the KIX Domain Proceeds through Dynamic Repacking of the Hydrophobic Core. *ACS Chem Biol* **8**, 1600-1610, (2013).
13. Cook, P. R., Polakowski, N. & Lemasson, I. HTLV-1 HBZ protein deregulates interactions between cellular factors and the KIX domain of p300/CBP. *J Mol Biol* **409**, 384-398, (2011).
14. De Guzman, R. N., Goto, N. K., Dyson, H. J. & Wright, P. E. Structural Basis for Cooperative Transcription Factor Binding to the CBP Coactivator. *J Mol Biol* **355**, 1005-1013, (2006).
15. Goto, N. K., Zor, T., Martinez-Yamout, M., Dyson, H. J. & Wright, P. E. Cooperativity in transcription factor binding to the coactivator CREB-binding protein (CBP). The mixed lineage leukemia protein (MLL) activation domain binds to an allosteric site on the KIX domain. *J Biol Chem* **277**, 43168-43174, (2002).
16. Radhakrishnan, I. *et al.* Solution Structure of the KIX Domain of CBP Bound to the Transactivation Domain of CREB: A Model of Activator:Coactivator Interactions. *Cell* **91**, 741-752, (1997).

17. Arai, M., Dyson, H. J. & Wright, P. E. Leu628 of the KIX domain of CBP is a key residue for the interaction with the MLL transactivation domain. *FEBS Lett* **584**, 4500-4504, (2010).
18. Huth, J. R. *et al.* Design of an expression system for detecting folded protein domains and mapping macromolecular interactions by NMR. *Protein Sci* **6**, 2359-2364, (1997).
19. Gianni, S., Morrone, A., Giri, R. & Brunori, M. A folding-after-binding mechanism describes the recognition between the transactivation domain of c-Myb and the KIX domain of the CREB-binding protein. *Biochem Biophys Res Commun* **428**, 205-209, (2012).
20. Giri, R., Morrone, A., Toto, A., Brunori, M. & Gianni, S. Structure of the transition state for the binding of c-Myb and KIX highlights an unexpected order for a disordered system. *Proc Natl Acad Sci USA* **110**, 14942-14947, (2013).
21. Chan, H. M. & La Thangue, N. B. p300/CBP proteins: HATs for transcriptional bridges and scaffolds. *J Cell Sci* **114**, 2363-2373, (2001).
22. Shi, Y. & Mello, C. A CBP/p300 homolog specifies multiple differentiation pathways in *Caenorhabditis elegans*. *Genes Dev* **12**, 943-955, (1998).
23. Bordoli, L., Netsch, M., Lüthi, U., Lutz, W. & Eckner, R. Plant orthologs of p300/CBP: conservation of a core domain in metazoan p300/CBP acetyltransferase-related proteins. *Nucleic Acids Res* **29**, 589-597, (2001).
24. Giordano, A. & Avantaggiati, M. L. p300 and CBP: partners for life and death. *J Cell Physiol* **181**, 218-230, (1999).
25. Thakur, J. K., Yadav, A. & Yadav, G. Molecular recognition by the KIX domain and its role in gene regulation. *Nucleic Acids Res* **42**, 2112-2125, (2014).
26. Sugase, K., Dyson, H. J. & Wright, P. E. Mechanism of coupled folding and binding of an intrinsically disordered protein. *Nature* **447**, 1021-1025, (2007).
27. Wang, F. *et al.* Structures of KIX domain of CBP in complex with two FOXO3a transactivation domains reveal promiscuity and plasticity in coactivator recruitment. *Proc Natl Acad Sci USA* **109**, 6078-6083, (2012).

28. Zor, T., De Guzman, R. N., Dyson, H. J. & Wright, P. E. Solution Structure of the KIX Domain of CBP Bound to the Transactivation Domain of c-Myb. *J Mol Biol* **337**, 521-534, (2004).
29. Ansari, A. Z., Reece, R. J. & Ptashne, M. A transcriptional activating region with two contrasting modes of protein interaction. *Proc Natl Acad Sci USA* **95**, 13543-13548, (1998).
30. Shammass, S. L., Travis, A. J. & Clarke, J. Remarkably Fast Coupled Folding and Binding of the Intrinsically Disordered Transactivation Domain of cMyb to CBP KIX. *J Phys Chem B* **117**, 13346-13356, (2013).
31. Wands, A. M. *et al.* Transient-state Kinetic Analysis of Transcriptional Activator·DNA Complexes Interacting with a Key Coactivator. *J Biol Chem* **286**, 16238-16245, (2011).
32. Lodge, J. M., Rettenmaier, T. J., Wells, J. A., Pomerantz, W. C. & Mapp, A. K. FP Tethering: a screening technique to rapidly identify compounds that disrupt protein-protein interactions. *MedChemComm* **5**, 370-375, (2014).
33. Parker, D. *et al.* Phosphorylation of CREB at Ser-133 induces complex formation with CREB-binding protein via a direct mechanism. *Mol Cell Biol* **16**, 694-703, (1996).
34. Parker, D. *et al.* Analysis of an Activator:Coactivator Complex Reveals an Essential role for Secondary Structure in Transcriptional Activation. *Mol Cell* **2**, 353-359, (1998).
35. Parker, D. *et al.* Role of secondary structure in discrimination between constitutive and inducible activators. *Mol Cell Biol* **19**, 5601-5607, (1999).
36. Law, S. M., Gagnon, J. K., Mapp, A. K. & Brooks, C. L. Prepaying the entropic cost for allosteric regulation in KIX. *Proc Natl Acad Sci USA* **111**, 12067-12072, (2014).
37. Korkmaz, E. N., Nussinov, R. & Haliloglu, T. Conformational control of the binding of the transactivation domain of the MLL protein and c-Myb to the KIX domain of CREB. *PLoS Comput Biol* **8**, e1002420, (2012).
38. Berg, T. Modulation of protein-protein interactions with small organic molecules. *Angew Chem* **42**, 2462-2481, (2003).

39. Rogers, J. M., Steward, A. & Clarke, J. Folding and Binding of an Intrinsically Disordered Protein: Fast, but Not 'Diffusion-Limited'. *J Am Chem Soc* **135**, 1415-1422, (2013).
40. Schreiber, G., Haran, G. & Zhou, H. X. Fundamental Aspects of Protein-Protein Association Kinetics. *Chem Rev* **109**, 839-860, (2009).
41. Carroll, M. J. *et al.* Evidence for dynamics in proteins as a mechanism for ligand dissociation. *Nat Chem Biol* **8**, 246-252, (2012).
42. Lickwar, C. R., Mueller, F., Hanlon, S. E., McNally, J. G. & Lieb, J. D. Genome-wide protein-DNA binding dynamics suggest a molecular clutch for transcription factor function. *Nature* **484**, 251-255, (2012).
43. Majmudar, C. Y. *et al.* Sekikaic Acid and Lobaric Acid Target a Dynamic Interface of the Coactivator CBP/p300. *Angew Chem* **51**, 11258-11262, (2012).
44. Munoz, V. & Serrano, L. Elucidating the folding problem of helical peptides using empirical parameters. *Nat Struct Biol* **1**, 399-409, (1994).
45. Munoz, V. & Serrano, L. Elucidating the folding problem of helical peptides using empirical parameters. III. Temperature and pH dependence. *J Mol Biol* **245**, 297-308, (1995).
46. Grutsch, S., Bruschweiler, S. & Tollinger, M. NMR Methods to Study Dynamic Allostery. *PLoS Comput Biol* **12**, e1004620, (2016).
47. Tollinger, M. *et al.* An isolated helix persists in a sparsely populated form of KIX under native conditions. *Biochemistry* **45**, 8885-8893, (2006).
48. Breuker, K., Bruschweiler, S. & Tollinger, M. Electrostatic stabilization of a native protein structure in the gas phase. *Angew Chem* **50**, 873-877, (2011).
49. Sadowsky, J. D. *et al.* Turning a protein kinase on or off from a single allosteric site via disulfide trapping. *Proc Natl Acad Sci USA* **108**, 6056-6061, (2011).
50. Erlanson, D. A. *et al.* Site-directed ligand discovery. *Proc Natl Acad Sci USA* **97**, 9367-9372, (2000).
51. Hardy, J. A., Lam, J., Nguyen, J. T., O'Brien, T. & Wells, J. A. Discovery of an allosteric site in the caspases. *Proc Natl Acad Sci USA* **101**, 12461-12466, (2004).

52. Buhrlage, S. J. *et al.* Amphipathic Small Molecules Mimic the Binding Modes and Function of Endogenous Transcription Factors. *ACS Chem Biol* **4**, 335-344, (2009).
53. Pomerantz, W. C. *et al.* Profiling the dynamic interfaces of fluorinated transcription complexes for ligand discovery and characterization. *ACS Chem Biol* **7**, 1345-1350, (2012).
54. Chen, Y. H., Yang, J. T. & Chau, K. H. Determination of the helix and beta form of proteins in aqueous solution by circular dichroism. *Biochemistry* **13**, 3350-3359, (1974).

Chapter 4 Covalent chemical co-chaperones for the CBP KIX domain[§]

4.1 Abstract

As part of the coactivator CBP, the KIX activator-binding domain regulates diverse processes from development to cognitive function. It has been an attractive target for small molecule probes to understand how context-dependent activator-KIX complexes regulate cellular functions, especially in diseases ranging from leukemia to memory disorders. However, the dynamic and transient PPIs formed by KIX are difficult targets for small molecules. An additional complication is that activator-binding motifs such as KIX are found in multiple coactivators, making specificity difficult to attain. In this study, we demonstrate that the strategy of Tethering can be used to rapidly develop highly specific covalent modulators of the dynamic PPIs between activators and coactivators. These serve as both orthosteric and allosteric modulators, enabling the tunable assembly or disassembly of the activator-coactivator complexes formed between the KIX domain and its cognate activator binding partners. These molecules maintain their function and selectivity even in human cell lysates and in bacterial cells and thus will ultimately be highly useful probes for cellular studies.

[§] This chapter was a collaborative effort. Dr. C.Y. Majmudar, and Dr. W.C. Pomerantz constructed the KIX cysteine mutants. Dr. C.Y. Majmudar and Dr. W.C. Pomerantz ran the LC-MS Tethering screen at the University of California, San Francisco with Prof. J.A. Wells and Dr. J.D. Sadowsky. Dr. J.C. Clayton synthesized the irreversible **1-10** small molecules referred to as **1-10a-f**. Dr. C.Y. Majmudar helped conduct experiments to assess their effect on KIX N627C and these results are shown found in in Table 4-1. Dr. C.Y. Majmudar helped plan and run the selectivity experiments in Figure 4-3. Dr. J.C. Clayton helped prepare samples for the selectivity experiments in Figure 4-7c. Dr. J.C. Clayton helped with the cell viability experiments found in Figure 4-9 and Dr. C.H. Doss supervised these experiments.

4.2 Introduction

KIX is one of several conformationally dynamic domains found in the master coactivators CBP and p300.^{1,2} It interacts with more than 15 transcriptional activators at two distinct interfaces.² NMR studies of KIX bound with native transcriptional activation domains (TADs) have shown that the endogenous partners such as MLL and c-Jun interact at a relatively deeper and smaller site (an area of approximately 900 Å²)^{3,4} while the transcriptional activation domains of CREB (pKID) and c-Myb interact with a shallower and larger surface on the opposite side.⁵ The two interfaces are allosterically connected such that binding of the transcriptional activation domain of MLL enhances the interaction of c-Myb or pKID at the other binding site by approximately 2-fold.^{4,6} This allosteric enhancement is thought to play a key role in recruitment of p300 or CBP to gene promoters.^{4,7,8}

The activator-KIX complexes participate in fundamental processes such as hematopoiesis and memory formation.⁹⁻¹¹ The discovery of small molecule modulators for those processes has been of high priority in the selection, screening and top-down approaches to find KIX inhibitors.¹²⁻¹⁶ One early success was Naphthol AS-E, a molecules that disrupts oncogenic responses in cancer cell models with the CREB-KIX and MYB-KIX complexes as its intended targets.^{12,17,18} As this examples illustrates, small molecules that target KIX have the potential to delineate activator-coactivator functions on a phenotypic level. However, even the most specific modulators for KIX have limited selectivity due to the level of redundancy in the protein interaction network.^{15,19,20} The binding surfaces that coactivators use to interact with activators are often similar. For example, the activator p53 interacts with the GACKIX motif as well as three other domains in CBP and p300, illustrating fundamental similarities in these binding surfaces.²¹⁻²⁵

Finally, the GACKIX motif within CBP has been found in at least four eukaryotic coactivators, including the closely related p300 but also in unrelated transcriptional regulators, such as ARC105/MED15, and RECQL5.²⁶⁻²⁸ Therefore, the unusually specific inhibitor sekikaic acid, a natural product that exhibits high specificity for the KIX domain, has the undesirable potential to target other coactivators with this motif in a cellular setting.¹⁵ This severely limits the utility of these probe molecules to dissect individual activator-KIX interactions and to understand their role in normal and pathological processes.²⁷ Probes that also complement an interface in KIX are likely to accommodate other regions within co-regulators.

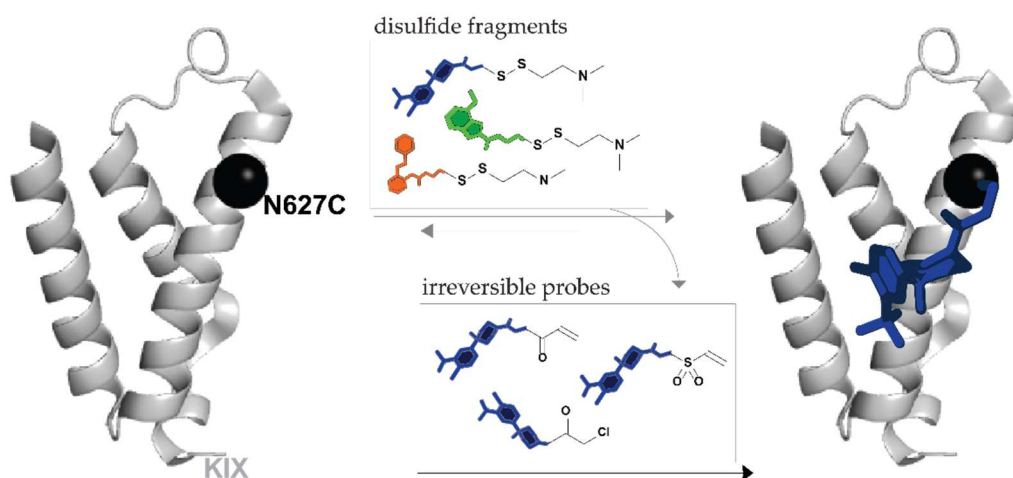


Figure 4-1 Irreversible Tethering scheme The standard Tethering scheme involves in the formation of a mixed disulfide with the target when the fragment favorably interacts with the regions surrounding cysteine residue. By replacing the fragment's disulfide with an alkylating moiety, the probes are compatible with the reducing environment of a cell.

Recently we reported that the covalent fragment discovery strategy of Tethering yields covalent chemical co-chaperones for KIX that stabilize distinct conformations of this plastic motif and, in doing so, modulate its ability to form binary and ternary complexes.²⁹ While useful for structural and *in vitro* biophysical

studies, these molecules react with KIX via disulfide exchange under reversible conditions and are thus not usable in more complex environments such as those of cells and cell lysates. Here we show, that despite the shallow and poorly defined binding surfaces in GACKIX, irreversible covalent co-chaperones highly selective for their cognate binding site can be readily accessed by replacement of disulfides in the Tethering hits with more reactive moieties (Figure 4-1). Further, the co-chaperones show high selectivity for their cognate binding site, even in the presence of many potential reactive partners. Finally, changes in the thiophile enable tuning of the assembly behavior of KIX, leading to allosteric enhancement or inhibition of binding with a subset of partners.

4.3 Results and Discussion

4.3.1 *Converting 1-10 into irreversible analogues*

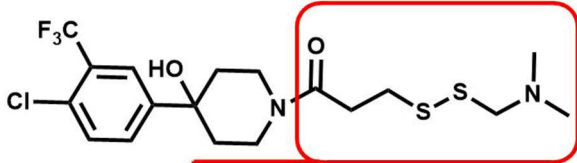
The KIX domain does not contain well-defined pockets for small molecules, but rather binding surfaces that are relatively featureless. These KIX binding surfaces adapt upon interacting with their binding partners so the conversion of the disulfide fragments into irreversible probes can be challenging. The irreversible probes need orient the fragment at the selected point of Tethering and effectively modulate the cysteine-containing KIX domain interactions.

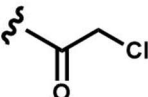
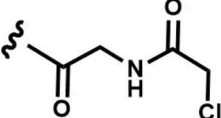
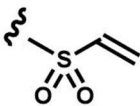
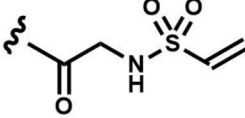
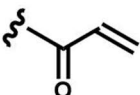
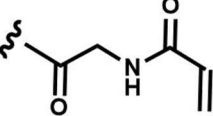
We sought a fragment that is well-established to impact the KIX interactions. The fragment **1-10** has previously been shown by the Mapp lab to act as an effective modulator of GACKIX-activator interactions by both disrupting MLL binding and enhancing interactions between KIX and pKID by 2-fold. Thus, **1-10** was chosen as the scaffold for irreversible inhibitor development (Table 4-1). To initiate the investigation, the disulfide moiety of **1-10** was replaced with three distinct thiophiles with or without a glycine linker to produce **1-10a-f**. These molecules were then assessed for dose-dependent alkylation of the KIX N627C

mutant and, as shown in Table 4-1, the DR₅₀ values varied significantly with the nature and linkage of thiophilic group. The least effective reactive group in this series is the α,β -unsaturated amide, with the labeling of the protein occurring only under forcing conditions. Both the α -chloro amide and the vinyl sulfonamide series were effective labelers, with DR₅₀ values ranging from 4.6 - 150 μ M after one hour.

Any impact of these fragments on KIX N67C interactions was considered

Table 4-1 Evaluating the 1-10 irreversible probes The structure of the 1-10 disulfide and the irreversible derivatives are shown. The DR₅₀ values were determined by varying the concentration of compound and after one hour evaluating how much of KIX N627C was labeled using LC-MS. The fold inhibition values were obtained from FP experiments by comparing the dissociation constants (K_{DS}) for the unlabeled (DMSO) KIX N627C construct with the K_{DS} for the labeled KIX N627C- 1-10 alkylator complex for both the MLL and pKID tracers. All FP measurements were performed in triplicate and errors reflect the standard deviation (SD) error.



Compound		Dose Response DR ₅₀ (μ M)	Fold Inhibition MLL	pKID
1-10a		25	12 \pm 1	0.9 \pm 0.1
1-10b		150	13 \pm 1	1.4 \pm 0.2
1-10c		6.8	17 \pm 2	1.5 \pm 0.2
1-10d		4.6	17 \pm 2	0.65 \pm 0.08
1-10e		>500	14 \pm 2	0.88 \pm 0.09
1-10f		>500	5.9 \pm 0.6	0.66 \pm 0.07

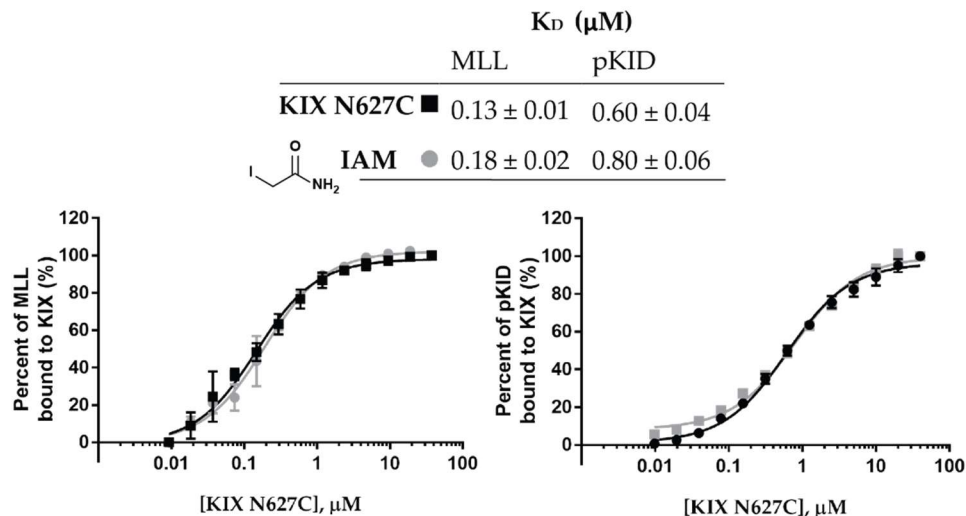


Figure 4-2 Effect of the alkylating agent, IAM on KIX N627C The interactions between KIX N627C labeled by IAM and its endogenous partners, MLL and pKID remained unaffected by FP direct binding assays. All measurements were performed in triplicate and the error bars indicate SD of the three measurements. The K_D values and SE are shown in the table above from the nonlinear fits in GraphPad Prism 7.0.

unique because the labeling KIX N627C with the alkylator, iodoacetamide (IAM), had minimal impact on the affinity of the endogenous ligands (Figure 4-2). When covalently attached to KIX each of these structures, **1-10a-d**, competitively blocked MLL from interacting with its cognate binding site analogous to the parent disulfide **1-10**. More remarkable are the effects at the distal binding site, with fragments **1-10b** and **1-10c** allosterically *inhibiting* the binding of pKID and **1-10d** *enhancing* the binding nearly 2-fold, comparable to the effects observed with the native ligands MLL, pKID and c-Myb. E2A binds to the MLL-binding site of KIX and the alkylators **1-10a**, **1-10c**, and **1-10d** also disrupt this interaction, but to a lesser extent than MLL (Figure 4-3). These alkylators also have negligible effects on c-Myb which binds to the pKID-site. Therefore, the linker and alkylation moiety, appears to differentially impact the assembly state of KIX.

4.3.2 Assessing target engagement of two **1-10** irreversible probes

A further examination of the reactivity of the covalent modifiers revealed a remarkable degree of selectivity for the MLL-binding site within KIX. None of the

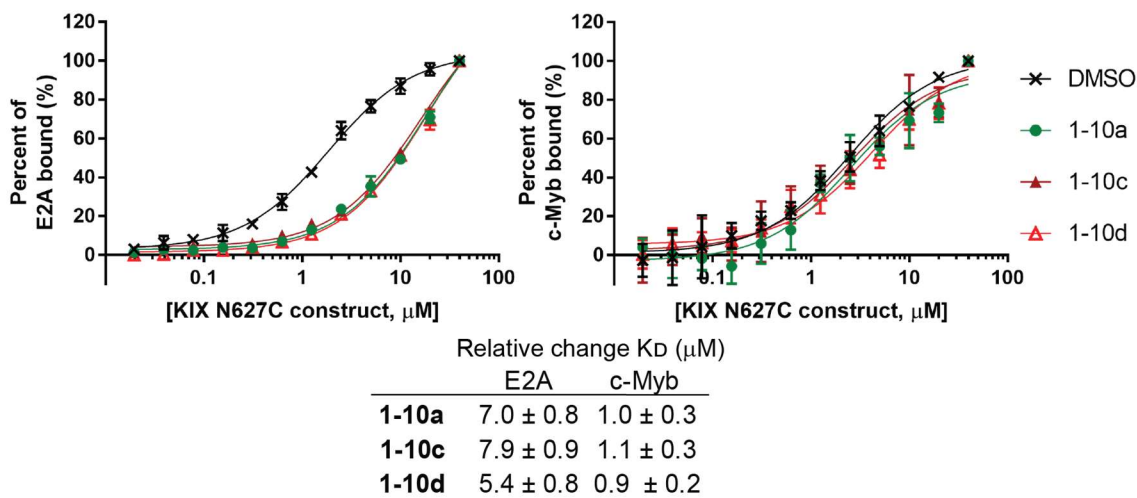


Figure 4-3 Other TADs and their interactions with alkylated KIX N627C The FP-direct binding assays were conducted with KIX N627C bound to either **1-10a**, **1-10c**, and **1-10d** against the E2A and c-Myb tracers which share binding sites with MLL and pKID, respectively. Each value represents an average of three separate measurements and the error bars show the SD error of nonlinear fits in GraphPad Prism 7.0 software. The table shows the relative change in K_D from the unlabeled (DMSO) KIX N627C control versus the alkylated KIX N627C complexes and the error reflects the SD error.

molecules (**1-10a-f**) react with the KIX domain in the absence of a cysteine residue (native KIX). In addition, covalent modification selectively occurs only in the MLL-binding site, even when cysteine residues are introduced around other activator binding sites. Figure 4-4A shows covalent modifiers **1-10a** and **1-10d**, both of which are reactive electrophiles, only form covalent bonds with KIX when a cysteine is present at the cognate (MLL) binding site even when cysteines are introduced in other activator binding sites within KIX. In a more complex environment *E. coli* expressing KIX N627C protein was dosed with the irreversible probes for one hour and the resulting purified protein was analyzed by Q-TOF LC-MS to detect the extent of labeling by the irreversible probes (Figure 4-4B). Covalent modifier **1-10a** required high (250 μ M or more) concentrations for effective labeling to occur, suggesting high cross-reactivity with other protein or small molecule thiols and/or a slower reaction with the KIX N627C target. In contrast, molecule **1-10d** covalently labeled KIX at concentrations as low as 15.6

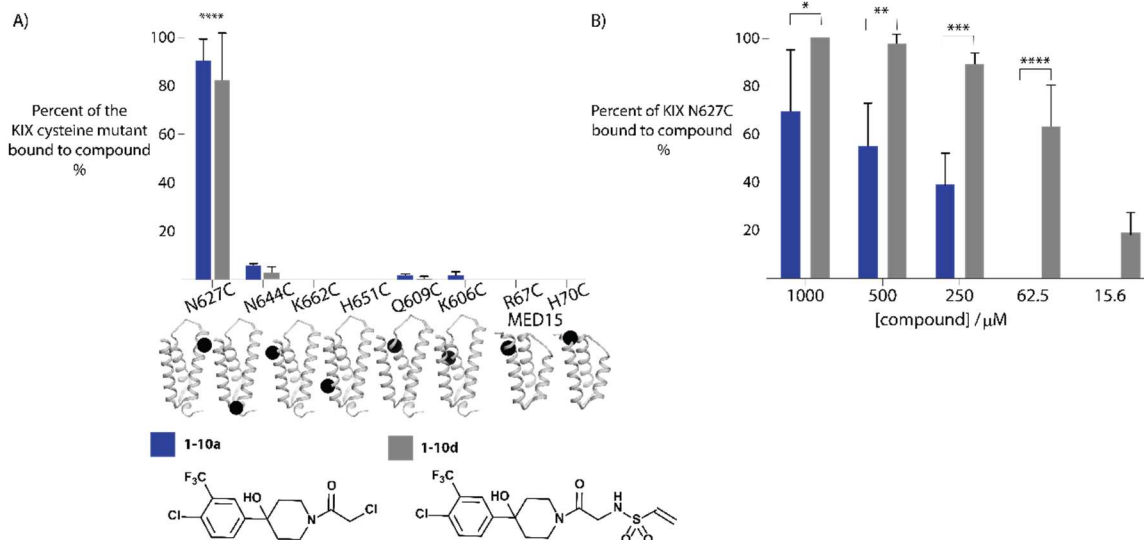


Figure 4-4 Selectivity of 1-10 irreversible alkylators for KIX N627C A) 1-10a and 1-10d show preference for the KIX N627C mutant located near the MLL-binding site. N644C is located in the loop and the remaining mutants, K662C, H651C, Q609C and K606C are located within the pKID-binding site. The error bars represent standard error of the average of two separate experiments. B) The 1-10 alkylators were dosed into *E. coli* during the expression of KIX N627C protein and after nickel affinity purification, the extent of labeling was assessed by LC-MS and each reaction was repeated three times. The error bars represent standard error from three separate experiments. *0.01 < P < 0.05, **0.05 < P < 0.001, ***0.001 < P < 0.001, ****P < 0.001; P values are calculated by GraphPad Prism 7.0.

μM , close to the *in vitro* DR₅₀ value and displays strong selectivity for the KIX domain. The affinity of the core scaffold of 1-10 (no electrophilic moiety) has been measured at approximately 250 μM ; thus, it is unlikely 1-10d at concentrations under 10 μM will produce effects due to non-covalent binding interactions.

The concern surrounding irreversible probes is that they may target critical cellular components in mammalian cells, prolonging adverse effects due to their reactivity. We assessed the cell viability of these alkylators in HeLa cells (Figure 4-5). The 1-10a α -chloro and 1-10c sulfonamide derivatives were more toxic at concentrations approaching 100 μM while the corresponding glycine-linked versions, 1-10b and 1-10d were tolerated after three hours. This demonstrates that

cells handle these probes well, perhaps due to the lack of nonspecific and toxic targets.

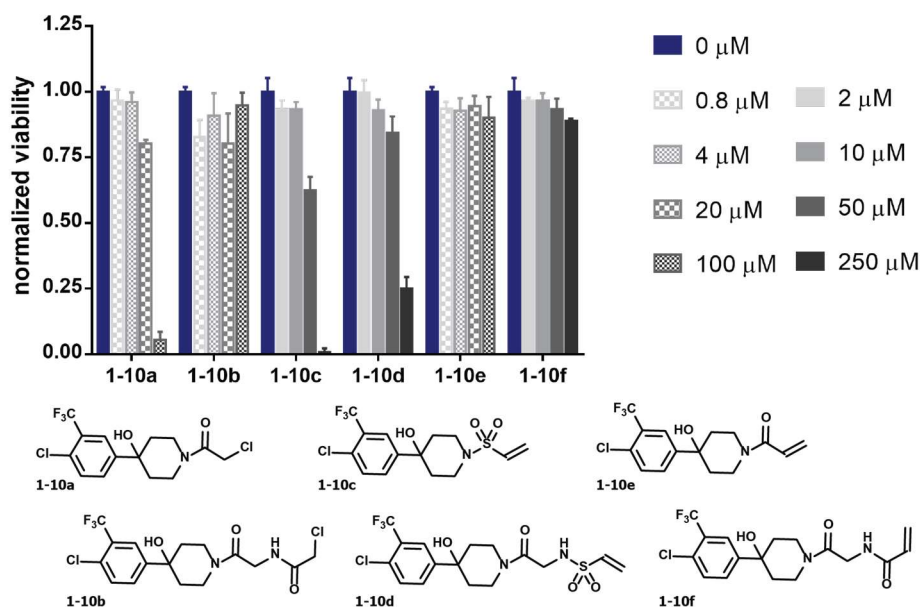
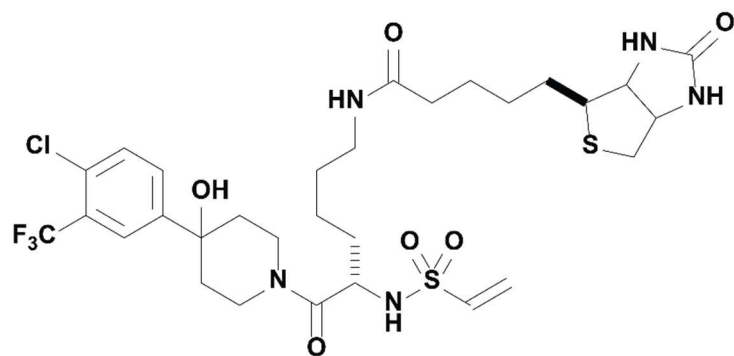


Figure 4-5 Cell viability for the 1-10 alkylators A WST-1 cell proliferation assay was used to measure the viability of HeLa cells in the presence of the **1-10** alkylators. Each measurement was duplicated with the values representing the average of all the measurements and the error bars represent the standard error.

Finally, we wanted to visualize the potential targets of these probes in a cellular environment. We did this in cells by adding the purified KIX N627C mutant to HEK293T lysate and then dosed in the biotinylated **1-10d** probe. Biotin-containing components were pulled-out of the lysate using neutravidin resin and analyzed by western blot with streptavidin-HRP (Figure 4-6). Once again, the **1-10d** alkylator displayed significant selectivity for the target.

4.4 Conclusion

The purpose of targeting the KIX domain is to understand what roles its interactions with activators complexes play in transcriptional events. However, it is difficult to distinguish the role of the KIX domain in CBP versus p300 function due to their significant homology (90%). From our results, the irreversible **1-10** derivatives displayed a preference for the cysteine-containing KIX N627C mutant.



biotinylated 1-10d vinyl sulfonamide probe

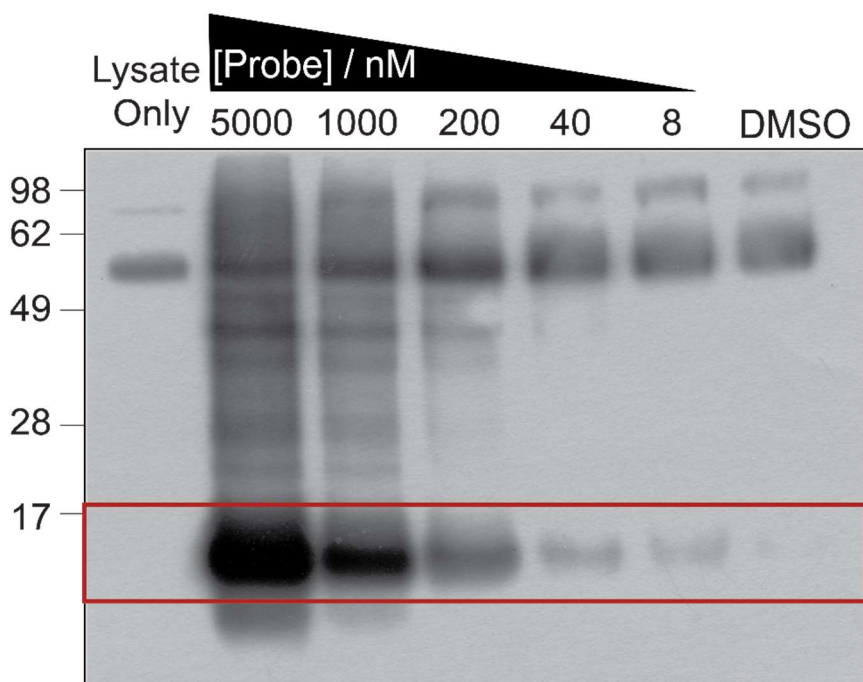


Figure 4-6 Visualizing specificity with the biotinylated 1-10d probe The structure of the biotinylated **1-10d** probe is shown above. It was added with purified KIX N627C (25 nM) and 100 μ g of HEK293T lysate. Any biotin-containing proteins purified on neutravidin agarose beads were visualized by western blot using chemiluminescence with Streptavidin- HRP. The bands within the red box represent the expected mass for KIX N627C (12 kDa).

They selectively enhance or disrupt interactions between KIX N627C and its activators only when the cognate cysteine residue is present. Since the binding of these probes are irreversible, nonspecific targets can drastically weaken their fitness as cellular probes.³⁰⁻³⁴ Our preliminary evidence shows that **1-10d** engages the KIX N627C target with no prominent, off-targets in different complex

environments.

Other covalent fragment screens start their search for alkylator-containing fragments rather than disulfide fragment.³⁵⁻³⁷ Our data reveal that the reactivity of the fragment regulates its ability to both engage KIX and to modulate interactions with KIX. If we ran the same LC-MS screen against a library containing the acrylamide-based fragments instead of disulfide fragments then **1-10** with intrinsic affinity for KIX may have been missed due to its suboptimal reactivity with KIX N627C ($DR_{50} > 500 \mu\text{M}$). By starting with the disulfide fragments the screen depends upon reaching equilibrium conditions while the stringency of the screen is controlled through the concentration of reducing agent.³⁸ Then a favorable fragment resides with its target and makes it less dependent on the kinetics or reactivity of its tail. However, the conversion from the mixed disulfide to the irreversible thiolophile still requires optimization. Based on our data one fragment from the reversible disulfide Tethering screen provided a framework to produce a range of irreversible modulators.

The next step is to apply these probes in a cellular system; however the cognate cysteine mutation is required to achieve selective target engagement. A similar irreversible Tethering scheme has previously been adapted to target an oncogenic cysteine mutation (G12C) within Ras that impairs GTP hydrolysis.³⁹ The purpose of those irreversible probes was also to selectively target the cysteine mutant form of Ras over the wild-type in cancer cells. In our method, an additional step is required to introduce the target cysteine mutant. For now, we have showed for the first time that Tethering can produce selective irreversible probes for PPIs.

4.5 Methods

4.5.1 Protein Expression and Purification

As previously described⁴⁰, the DNA sequence encoding the KIX domain

from mouse CBP residues 586-672 was cloned into the bacterial expression pRSETB vector with an additional hexahistidine tag and a short polar linker fused to the N-terminus of KIX resulting in protein with the sequence (tag and linker residues are shown in lower case):

mrgshhhhhhgmasGVRKGWHEHVVTQDLRSHLVHKLQVAIFPTDPAALKDRR
MENLVAYAKKVEGDMYESANSRDEYYHLLAEKIYKIQKELEEKRRSRL

The cysteine mutants at N627, R644, K662, H651, Q609, and K606 of CBP and R67 and H70 of MED15 KIX were generated using site-directed mutagenesis as previously described.⁴¹

The CBP KIX protein was expressed in Rosetta2(DE3) pLysS *Escherichia coli* (Novagen). Cells were grown to an OD_{600nm} of 0.8 - 1.0 (37 °C, 250 rpm), induced with 0.25 mM IPTG for 4 hours at 25 °C, harvested by centrifugation, and stored at -80 °C. The His₆-tagged CBP KIX protein was affinity purified using a batch method with Ni-NTA beads (QIAGEN) following the manufacturer's instructions and eluted with 400 mM imidazole. KIX was polished up with ion-exchange chromatography on a Source S column (GE Healthcare) in 50 mM phosphate buffer, 1 mM DTT, pH 7.2 eluting with the NaCl gradient from 0 to 1 M. Purified protein was buffer exchanged into 10 mM sodium phosphate, 100 mM NaCl, pH 6.8 using a PD-10 column (GE Healthcare) and stored at -80 °C.

The MED15 KIX cDNA encoding the amino acids 1 through 78 was synthesized by GenScript USA, Inc. and cloned into the pET-15b plasmid (Novagen, EMD Millipore) using the Nde1 and Xho1 cloning sites. The resulting recombinant ARC105/MED15 KIX protein sequence contained the N-terminal hexa-histidine tag and a thrombin cleavage site as shown in lower case letters; mrgshhhhhhssglvprgsHMDVSGQETDWRSTAFRQKLVSQIEDAMRKAGVAHS
KSSKDMESHVFLKAKTRDEYLSLVARLIHFRDIHNKKSQASV

The ARC105/MED15 KIX (1-78) protein was also expressed in

Rosetta2(DE3) pLysS *Escherichia coli* (Novagen) except after the cultures reached an OD_{600nm} between 0.6-1 and then the temperature of the incubator was dropped to 20 °C. After one hour KIX expression was induced through the lac operon with the addition of 125 µM IPTG for 12-18 hours at 20°C, 250 rpm. The cells were harvested by centrifugation for 15 minutes at 6,000 rpm in an Avanti Beckman Coulter High-speed Centrifuge with a JLA 8.1000 rotor or at 7,903 xg in a Sorvall™ LYNX™ Superspeed Centrifuge with a Fiberlight F6-6x1000 LEX Carbon Fiber Rotor (ThermoFisher Scientific), collected in a 50 ml falcon tube, and stored at -80°C.

For purification of the MED15 KIX cysteine mutants, a bacterial cell pellet expressed in one liter of terrific broth was thawed on ice and then suspended in approximately 25 ml of lysis buffer (50 mM sodium phosphate, 300 mM NaCl, 5 mM imidazole, 1 mM β-ME pH 7.2, cOmplete™, EDTA-free Protease Inhibitor Cocktail Tablet (Roche)). The cells were lysed by sonication at 50% amplitude on ice using a 6 mm tip with pulsing cycles of 3 seconds on and 6 seconds off for at least 3 minutes of total pulsing time. The soluble lysate was collected by centrifugation for 30 minutes at 9,500 rpm (9,299 x g) in an Allegra X-22R centrifuge (Beckman Coulter) with a C0650 fixed angle rotor, and then incubated with 2 ml of suspended Ni-NTA agarose resin (Qiagen) for 1 to 2 hours rotating on a rotisserie-style tube revolver (Fisher Scientific). The resin was washed 5 times by adding 5 ml of wash buffer (50 mM sodium phosphate, 300 mM NaCl, 30 mM imidazole, 1 mM β-ME, pH 7.2), spinning down the resin at 2,500 x g for 2 minutes, and gently pouring off the buffer. The nickel-bound protein was eluted by incubating the resin with 1 ml of elution buffer (50 mM sodium phosphate, 300 mM imidazole, 1 mM β-ME, pH 7.2) for 30 minutes to one hour, and collecting the elution for a total of 4 ml of elution buffer per ml of Ni-NTA agarose resin. The elutions were concentrated to about at least 5 ml and diluted up to 12 ml with 10

mM sodium phosphate, 100 mM NaCl, 10% glycerol, 0.01% NP-40 pH 6.8. Then 600 μ l of 10x thrombin cleavage buffer was added (20 mM Tris-HCl pH 8.4, 150 mM NaCl, 2.5 mM CaCl₂) with 10 μ l of Thrombin (Restriction grade, Novagen® 69671). The protein incubated overnight on the rotisserie at 4°C. Then this protein-cleavage cocktail incubated with 2 ml Ni-NTA agarose resin for one hour to remove remaining histidine-tagged KIX protein. The cleaved protein was isolated by pouring off the flow through from the resin and washing the resin once with 5 ml of 50 mM sodium phosphate buffer at pH 6.8. This protein was further purified on the AKTA FPLC Purifier (GE Healthcare) with the strong cation exchanger Source 15S media (GE Healthcare) packed to a 17-ml column volume. After loading through the 50-ml super loop the column was washed for 1.5 column volumes with buffer A (50 mM sodium phosphate, pH 6.8 1 mM DTT). The protein was eluted from the column with a gradient of 0% to 60% of buffer B (50 mM sodium phosphate, pH 6.8, 1 mM DTT, 1 M NaCl) over four column volumes. Major peaks typically eluted around 20% and 30% of buffer B with thrombin-cleaved KIX observed as the major species in the first peak and any remaining thrombin-tagged KIX eluting in the second peak.

Using a 3 kDa MWCO Amicon Ultra-15 Centrifuge Filter Units (EMD Millipore) the ARC105/MED15 KIX was concentrated and buffer exchanged by adding 5 to 10 ml of buffer (10 mM sodium phosphate, 100 mM sodium chloride, 10% glycerol, 0.01% NP-40 pH 6.8) after each spin at 3,500 rpm, 4 °C for 15 minutes in a swinging bucket rotor. The buffer was added in increments to the filter until the final added buffer volume was at least three times the final volume of protein. The protein was flash frozen in liquid nitrogen as 250 to 500 μ l aliquots and stored at -80 °C. The concentration of the protein was measured using absorbance at 280 nm on a NanoDrop 1000 Spectrophotometer (Thermo Scientific). The extinction coefficient for ARC105/MED15 KIX was measured using ProtPram with the

following equation

$$\varepsilon = (\#Tyr) \times \varepsilon(Tyr) + (\#Trp) \times \varepsilon(Trp) + (\#Cys) \times \varepsilon(Cys) \quad (\text{Eq. 1})$$

where # is the number of amino acid residues present in the sequence and ε is the extinction coefficient. The extinction coefficients for tyrosine, tryptophan and cysteine measured in water are $\varepsilon(Tyr)$ is $1,490 \text{ M}^{-1} \text{ cm}^{-1}$, $\varepsilon(Trp)$ is $5,500 \text{ M}^{-1} \text{ cm}^{-1}$, $\varepsilon(Cys)$ is $125 \text{ M}^{-1} \text{ cm}^{-1}$. The background absorbance for the protein was subtracted for the absorbance value at 340 nm and the buffer absorbance at 280 nm. From the absorbance at 280 nm the concentration of KIX was calculated using Beer's Law

$$\text{Absorbance (protein)} = \varepsilon \times l \times c \quad (\text{Eq. 2})$$

where l is the path length and c is the concentration in M (mol L^{-1}). When necessary the sample was diluted in buffer to maintain an absorbance below 1.0 unit. The ARC105/MED15 KIX protein was verified by Q-TOF LC-MS (Agilent) analysis and the purity of the protein was assessed by SDS-PAGE.

4.5.2 Peptide Synthesis and Purification

All peptides were synthesized by standard N-9 Fluorenylmethoxycarbonyl (Fmoc) solid phase synthesis methods⁴² as previously described.⁴⁰ The peptide sequences written as single letter amino acid abbreviations are as follows:

MLL	$\beta\text{A-DAGNILPSDIMDFVLKNTP-CONH}_2$
pKID	$\beta\text{A-DSQKRREILSRRPS(Phos)YRKILNDLSSDAPG-CONH}_2$
c-Myb	$\beta\text{A-KEKRIKELELLLMSTENELKGQQALW-CONH}_2$
E2A	$\beta\text{A-KEKRIKELELLLMSTENELKGQQALW-CONH}_2$

The abbreviation, βA represents beta-alanine and S(Phos) is phosphoserine. The fluorescent fluorescein isothiocyanate (FITC) tag was added at the amino-terminus of the peptide before the β -alanine residue. The C-terminus of the

peptides contained an amide group (-CONH₂).

4.5.3 Determination of the dose response (DR₅₀)

For the DR₅₀ values, various concentrations (500-0.2 μM) of the fragments were incubated with 5 μM KIX mutant protein (at 1 mM β-ME) for 45 minutes at RT and incubated at 4°C for at least 15 minutes prior to analysis. The percent of protein tethered to fragment molecules was determined by Q-TOF LC-MS (Agilent).⁴⁴ The concentration of fragment molecule required for 50% maximum Tethering (DR₅₀) was determined by data analysis in GraphPad Prism software 4.00, fitting to Equation 3, where x is the log of fragment molecule concentration and y is the normalized response from 1 to 100 (percent of protein tethered to fragment molecule) (Figure 4-7a).

$$y = \frac{100}{1+10^{(\text{Log}DR_{50}-x)}} \quad (\text{Eq. 3})$$

4.5.4 Alkylation reaction of KIX

The KIX N627C mutant was incubated with 5 to 10 equivalents of small molecule in 10 mM phosphate buffer, 100 mM NaCl, pH 6.8 overnight at room temperature. Excess small molecule was removed and small molecule-protein complexes were concentrated using 10 kDa molecular weight cutoff concentrators (Vivascience). The extent of labeling was measured by Q-TOF LC-MS (Agilent). Protein complexes that were at least 95% alkylated were flash frozen in liquid nitrogen and stored at -80 °C.

4.5.5 Fluorescent anisotropy assays

The fluorescent anisotropy was measured in triplicate with a final sample volume of 10 μl in a low volume, non-binding, black, 384-well plate (Corning). For each experiment 25 nM of fluorescently labeled peptide tracers, FITC-MLL and FITC-pKID were incubated with various concentrations of the small molecule-KIX mutant complexes in binding buffer (10 mM phosphate, 100 mM NaCl at pH 6.8)

for 30 min. at RT. The plates were read using a Tecan Genios Pro plate reader with polarized excitation at 485 nm and emission intensity measured through a parallel and perpendicularly polarized 535 nm filter. A binding isotherm that accounts for ligand depletion (assuming a 1:1 binding model of peptide to KIX) was fit to the observed anisotropy values as a function of KIX to obtain the apparent equilibrium dissociation constant, K_D :

$$A = A_f + (A_b - A_f) \times \frac{([L_T] + K_d + [P_T]) - \sqrt{([L_T] + K_d + [P_T])^2 - 4[L_T][P_T]}}{2[L_T]} \quad (\text{Eq. 4})$$

where $[L_T]$ and $[P_T]$ are the total concentrations of fluorescent peptide and KIX protein, respectively, A is the observed anisotropy, A_b is the maximum anisotropy value for the fully bound peptide, and A_f is the minimum anisotropy value for the free peptide. All nonlinear data analysis was performed with GraphPad Prism 4.0 software (San Jose, CA). The K_{DS} and FP direct binding curves for KIX N627C and the irreversible probes against the MLL and pKID tracer are shown in Figure 4-7.

When the total intensity changed due to a difference in signal between the fluorophore's bound and free species then the anisotropy values needed to be corrected.⁴⁵ The anisotropy, r was corrected with an average of free and bound fluorophore:

$$r = f_F \times r_F + f_B \times r_B \quad (\text{Eq. 5})$$

where f_F and f_B were the fraction contributions of the free and bound forms of the fluorophore.^{46,47} The fraction of bound peptide, f_B was calculated from

$$f_B = \frac{r - r_F}{(r - r_F) + R(r_B - r)} \quad (\text{Eq. 6})$$

where $R = I_B/I_F$ was the ratio of total intensities for the bound and free peptide species.

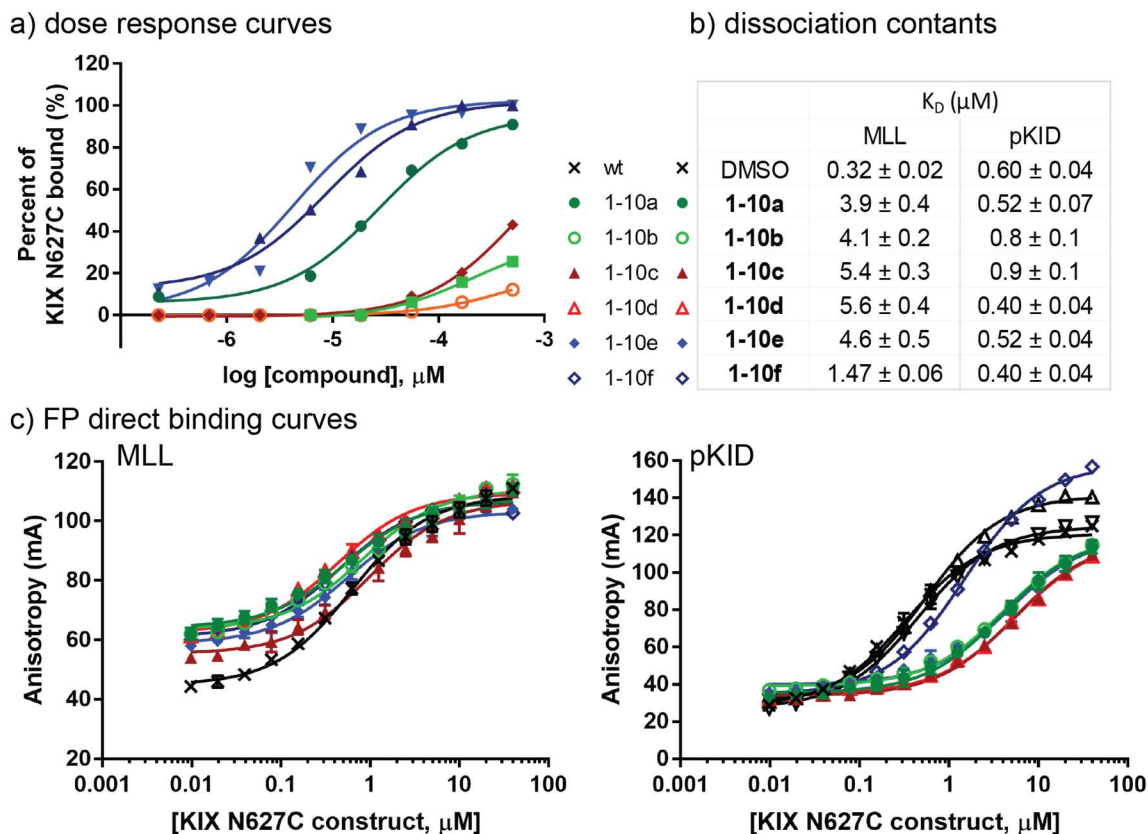


Figure 4-7 Summary of DR_{50} and FP experiments to compare the effect of the irreversible probes with KIX N627C a) The dose response curves are shown for the different 1-10 alkylators. b) The dissociation constants, K_D for MLL and pKID are shown for alkylated and unbound (DMSO) KIX N627C. The error reflects SE from the nonlinear fit in GraphPad Prism 7.0. c) The FP direct binding curves correspond to the K_D values above. All measurements were performed in triplicate. The values represent the average and the error bars represent the SD error.

4.5.6 Labeling KIX cysteine mutants

5 μM of cysteine mutants at KIX N627C, R664C, K662C, H651C, Q609C, and K606C KIX as well as R67C and H70C MED15 KIX stored in 10 mM sodium phosphate, 100 mM NaCl pH 6.8 were incubated with 100 μM of small molecule in DMSO (at 1 mM β -ME) for 1 hour at RT. The extent of labeling was measured

by Q-TOF LC-MS (Agilent). Each labeling reaction was duplicated.

4.5.7 Alkylation reaction in growing *E. coli*

The protein expression system for the KIX N627C protein was expressed, as described above. Three hours after induction with 0.25 mM IPTG the culture was concentrated by centrifugation in which the cell pellet from 50 ml of culture was suspended in 1 ml of media. For a 1 ml alkylation reaction, 10 μ L of compound in DMSO was added to concentrated cells to obtain concentrations of 100 μ M, 250 μ M, 50 μ M, and 12.5 μ M. The mixtures were incubated for 1 hour at 25 $^{\circ}$ C (250 rpm). The cell pellets were washed three times with 10 mM phosphate, 100 mM NaCl at pH 6.8 and stored at -80 $^{\circ}$ C. Purification was carried out as previously described.⁴⁸ The elutions were buffer exchanged into 10 mM phosphate, 100 mM NaCl at pH 6.8 and concentrated using 10 kilodalton molecular weight cutoff concentrators (Vivascience). The samples were analyzed by Q-TOF LC-MS (Agilent). The extent of labeling was determined by comparing the peak intensity of small molecule-KIX N627C complex versus unlabeled KIX N627C.

4.5.8 Cell viability assay

Hela cells were maintained in Dulbecco's Modified Eagle Medium (DMEM) supplemented with 10% fetal bovine serum albumin (FBS). In a 96-well plate 4,000 cells were plated per well. The next day they were treated for three hours with the indicated concentrations of compound and the control vehicle was dosed with 1% v/v DMSO. After three hours, the media was swapped with fresh DMEM containing 10% FBS with 10% WST-1 reagent and the plate was read at an absorbance set to 440 nm after one to two hours and the data was normalized against the DMSO (1% v/v) vehicle control.

4.5.9 Neutravidin pull-down assay with the biotinylated **1-10** probe

HEK293T cells were maintained in DMEM supplemented with 10% FBS.

The cells were lysed in 750 μ l of lysis buffer (150 mM NaCl, 50 mM Tris (pH 8.0), 0.1% Nonidet P-40 (NP-40)) containing halt protease inhibitor cocktail (Thermo Scientific) and soluble lysate was isolated by centrifugation (12,000 rpm, 4 °C, 15 min.) HEK 293T lysate (100 ng total protein) and 25 nM of purified KIX N627C were incubated with DMSO or various concentrations of the biotinylated **1-10d** probe (5 μ M, 1 μ M, 200 nM, 40 nM and 8 nM) at RT for 1 hour. Following incubation with NeutrAvidin agarose resin (50 μ L, Thermo Scientific) for 1 hour at 4 °C the beads were washed twice with 1 ml of wash buffer (10 mM phosphate, 100 mM sodium chloride, 10% glycerol, 0.1% NP-40, pH 7.2) and the resin-bound complexes were eluted by boiling in NuPAGE® LDS Sample Buffer (Invitrogen) containing DTT. The samples (15 μ L) were resolved on 12% SDS polyacrylamide gel by electrophoresis. The proteins were transferred onto a PVDF membrane and incubated with streptavidin conjugated to hydrogen peroxidase (HRP) enzyme (ab7403, Abcam) at 1:10,000 dilution in 10 mM PBS, 0.2% Tween-20. The membrane was developed using Amersham ECL™ Prime Western Blotting Detection Reagent (GE Healthcare) and captured on X-ray film.

4.6 References

1. Dyson, H. J. & Wright, P. E. Role of Intrinsic Protein Disorder in the Function and Interactions of the Transcriptional Coactivators CREB-Binding Protein (CBP) and p300. *J Biol Chem* (2016).
2. Thakur, J. K., Yadav, A. & Yadav, G. Molecular recognition by the KIX domain and its role in gene regulation. *Nucleic Acids Res* **42**, 2112-2125, (2014).
3. Campbell, K. M. & Lumb, K. J. Structurally distinct modes of recognition of the KIX domain of CBP by Jun and CREB. *Biochemistry* **41**, 13956-13964, (2002).

4. De Guzman, R. N., Goto, N. K., Dyson, H. J. & Wright, P. E. Structural Basis for Cooperative Transcription Factor Binding to the CBP Coactivator. *J Mol Biol* **355**, 1005-1013, (2006).
5. Zor, T., De Guzman, R. N., Dyson, H. J. & Wright, P. E. Solution Structure of the KIX Domain of CBP Bound to the Transactivation Domain of c-Myb. *J Mol Biol* **337**, 521-534, (2004).
6. Goto, N. K., Zor, T., Martinez-Yamout, M., Dyson, H. J. & Wright, P. E. Cooperativity in transcription factor binding to the coactivator CREB-binding protein (CBP). The mixed lineage leukemia protein (MLL) activation domain binds to an allosteric site on the KIX domain. *J Biol Chem* **277**, 43168-43174, (2002).
7. Toto, A., Giri, R., Brunori, M. & Gianni, S. The mechanism of binding of the KIX domain to the mixed lineage leukemia protein and its allosteric role in the recognition of c-Myb. *Protein Sci* **23**, 962-969, (2014).
8. Brüscheiler, S., Konrat, R. & Tollinger, M. Allosteric Communication in the KIX Domain Proceeds through Dynamic Repacking of the Hydrophobic Core. *ACS Chem Biol* **8**, 1600-1610, (2013).
9. Wood, M. A., Attner, M. A., Oliveira, A. M. M., Brindle, P. K. & Abel, T. A transcription factor-binding domain of the coactivator CBP is essential for long-term memory and the expression of specific target genes. *Learn Memory* **13**, 609-617, (2006).
10. Sandberg, M. L. *et al.* c-Myb and p300 regulate hematopoietic stem cell proliferation and differentiation. *Dev Cell* **8**, 153-166, (2005).
11. Pattabiraman, D. R., Sun, J., Dowhan, D. H., Ishii, S. & Gonda, T. J. Mutations in multiple domains of c-Myb disrupt interaction with CBP/p300 and abrogate myeloid transforming ability. *Mol Cancer Res* **7**, 1477-1486, (2009).
12. Best, J. L. *et al.* Identification of small-molecule antagonists that inhibit an activator: coactivator interaction. *Proc Natl Acad Sci USA* **101**, 17622-17627, (2004).
13. Li, B. X. & Xiao, X. Discovery of a Small-Molecule Inhibitor of the KIX-KID Interaction. *Chembiochem* **10**, 2721-2724, (2009).

14. Gee, C. T. *et al.* Protein-observed (19)F-NMR for fragment screening, affinity quantification and druggability assessment. *Nat Protoc* **11**, 1414-1427, (2016).
15. Majmudar, C. Y. *et al.* Sekikaic Acid and Lobaric Acid Target a Dynamic Interface of the Coactivator CBP/p300. *Angew Chem* **51**, 11258-11262, (2012).
16. Frangioni, J. V., LaRiccia, L. M., Cantley, L. C. & Montminy, M. R. Minimal activators that bind to the KIX domain of p300/CBP identified by phage display screening. *Nat Biotech* **18**, 1080-1085, (2000).
17. Mitton, B. *et al.* Small molecule inhibition of cAMP response element binding protein in human acute myeloid leukemia cells. *Leukemia* **30**, 2302-2311, (2016).
18. Uttarkar, S. *et al.* Naphthol AS-E Phosphate Inhibits the Activity of the Transcription Factor Myb by Blocking the Interaction with the KIX Domain of the Coactivator p300. *Mol Cancer Ther* **14**, 1276-1285, (2015).
19. Majmudar, C. Y. & Mapp, A. K. Chemical approaches to transcriptional regulation. *Curr Opin Chem Biol* **9**, 467-474, (2005).
20. Schreiber, G. & Keating, A. E. Protein binding specificity versus promiscuity. *Curr Opin Struct Biol* **21**, 50-61, (2011).
21. Teufel, D. P., Freund, S. M., Bycroft, M. & Fersht, A. R. Four domains of p300 each bind tightly to a sequence spanning both transactivation subdomains of p53. *Proc Natl Acad Sci USA* **104**, 7009-7014, (2007).
22. Wang, F. *et al.* Synergistic Interplay between Promoter Recognition and CBP/p300 Coactivator Recruitment by FOXO3a. *ACS Chem Biol* **4**, 1017-1027, (2009).
23. Wang, F. *et al.* Structures of KIX domain of CBP in complex with two FOXO3a transactivation domains reveal promiscuity and plasticity in coactivator recruitment. *Proc Natl Acad Sci USA* **109**, 6078-6083, (2012).
24. Krois, A. S., Ferreon, J. C., Martinez-Yamout, M. A., Dyson, H. J. & Wright, P. E. Recognition of the disordered p53 transactivation domain by the transcriptional adapter zinc finger domains of CREB-binding protein. *Proc Natl Acad Sci USA* **113**, E1853-1862, (2016).

25. Lee, C. W., Arai, M., Martinez-Yamout, M. A., Dyson, H. J. & Wright, P. E. Mapping the interactions of the p53 transactivation domain with the KIX domain of CBP. *Biochemistry* **48**, 2115-2124, (2009).
26. Novatchkova, M. & Eisenhaber, F. Linking transcriptional mediators via the GACKIX domain super family. *Curr Biol* **14**, R54-55, (2004).
27. Yang, F. *et al.* An ARC/Mediator subunit required for SREBP control of cholesterol and lipid homeostasis. *Nature* **442**, 700-704, (2006).
28. Kassube, S. A., Jinek, M., Fang, J., Tsutakawa, S. & Nogales, E. Structural mimicry in transcription regulation of human RNA polymerase II by the DNA helicase RECQL5. *Nature* **20**, 892-899, (2013).
29. Wang, N., Lodge, J. M., Fierke, C. A. & Mapp, A. K. Dissecting Allosteric Effects of Activator-Coactivator Complexes Using a Covalent Small Molecule Ligand. *Proc Natl Acad Sci USA* **11**, 12061-12066, (2014).
30. González-Bello, C. Designing Irreversible Inhibitors—Worth the Effort? *ChemMedChem*, (2015).
31. Jöst, C., Nitsche, C., Scholz, T., Roux, L. & Klein, C. D. Promiscuity and Selectivity in Covalent Enzyme Inhibition: A Systematic Study of Electrophilic Fragments. *J Med Chem* **57**, 7590-7599, (2014).
32. Schwartz, P. A. *et al.* Covalent EGFR inhibitor analysis reveals importance of reversible interactions to potency and mechanisms of drug resistance. *Proc Natl Acad Sci USA* **111**, 173-178, (2014).
33. Singh, J., Petter, R. C., Baillie, T. A. & Whitty, A. The resurgence of covalent drugs. *Nat Rev Drug Discov* **10**, 307-317, (2011).
34. Johnson, D. S., Weerapana, E. & Cravatt, B. F. Strategies for discovering and derisking covalent, irreversible enzyme inhibitors. *Fut Med Chem* **2**, 949-964, (2010).
35. Mah, R., Thomas, J. R. & Shafer, C. M. Drug discovery considerations in the development of covalent inhibitors. *Bioorg Med Chem Lett* **24**, 33-39, (2014).
36. Kathman, S. G., Xu, Z. & Statsyuk, A. V. A fragment-based method to discover irreversible covalent inhibitors of cysteine proteases. *J Med Chem* **57**, 4969-4974, (2014).

37. Nonoo, R. H., Armstrong, A. & Mann, D. J. Kinetic template-guided tethering of fragments. *ChemMedChem* **7**, 2082-2086, (2012).
38. Erlanson, D. A. *et al.* Site-directed ligand discovery. *Proc Natl Acad Sci USA* **97**, 9367-9372, (2000).
39. Ostrem, J. M., Peters, U., Sos, M. L., Wells, J. A. & Shokat, K. M. K-Ras(G12C) inhibitors allosterically control GTP affinity and effector interactions. *Nature* **503**, 548-551 (2013).
40. Buhrlage, S. J. *et al.* Amphipathic Small Molecules Mimic the Binding Modes and Function of Endogenous Transcription Factors. *ACS Chem Biol* **4**, 335-344, (2009).
41. Pomerantz, W. C. *et al.* Profiling the dynamic interfaces of fluorinated transcription complexes for ligand discovery and characterization. *ACS Chem Biol* **7**, 1345-1350, (2012).
42. Chan, W. & White, P. *Fmoc Solid Phase Peptide Synthesis: A Practical Approach*. (OUP Oxford, 2000).
43. Erlanson, D. A., Wells, J. A. & Braisted, A. C. Tethering: fragment-based drug discovery. *Annu Rev Biophys Biomol Struct* **33**, 199-223, (2004).
44. Sadowsky, J. D. *et al.* Turning a protein kinase on or off from a single allosteric site via disulfide trapping. *Proc Natl Acad Sci USA* **108**, 6056-6061, (2011).
45. Turconi, S. *et al.* Real Experiences of uHTS: A Prototypic 1536-Well Fluorescence Anisotropy-Based uHTS Screen and Application of Well-Level Quality Control Procedures. *J Biomol Screen* **6**, 275-290, (2001).
46. Malencik, D. A. & Anderson, S. R. Peptide binding by calmodulin and its proteolytic fragments and by troponin C. *Biochemistry* **23**, 2420-2428, (1984).
47. Lakowicz, J. R. *Principles of Fluorescence Spectroscopy*. (Kluwer Academic/Plenum Publishers 1999).
48. Wang, N. *et al.* Ordering a dynamic protein via a small-molecule stabilizer. *J Am Chem Soc* **135**, 3363-3366, (2013).

Chapter 5 Conclusions and Future Directions

5.1 Abstract

The GACKIX activator-binding motif plays a key role in the transcription by interacting with different activators to regulate the expression of genes. The functions of activator·GACKIX complexes are diverse and include cellular growth, hematopoiesis, synaptic plasticity, and lipid homeostasis. Chemical tools are sought to dissect the individual functions of activator·coactivator complexes and to provide a potential therapeutic model for diseases as diverse as leukemia and Alzheimer's disease. This small domain has been found in CBP and its homolog p300 as well as the other unrelated coactivators such as ARC105/MED15. Therefore, selective small molecules that distinguish between these KIX domains such as CBP versus p300 remains a challenge. The site-directed strategy known as Tethering was applied to match orthogonal disulfide-containing fragments with the cysteine-modified KIX domain. These probes were designed to selectively target CBP KIX versus p300. Finally, the activator-KIX complexes are transient and defining so how KIX responds to the demands of multiple activators remains unclear. Chemical and biochemical tools were also used to impact the conformations of KIX. These KIX constructs enable us to define how the ternary complex depends on the stabilization of the ternary complex. Additional work remains to develop additional probes for these activator·coactivator complexes and to understand their function in cells.

5.2 Conclusions

This dissertation focuses on the development of analytical tools and chemical probes for the study of the interactions between the KIX domain of CBP and its activator binding partners including c-Myb, MLL, and pKID. This work began by developing orthosteric modulators for the MLL- and CREB(pKID)-binding site of KIX. Effective small molecule inhibitors (500 Da or less) for PPIs may have a daunting task when they must compete with large peptides (~2,000 Da) to disrupt the interactions. Therefore, PPIs were initially viewed as “undruggable”^{1,2} – a researcher may have preferred to look for the proverbial needle in a haystack than find small molecule PPI inhibitors. In the last decade, small molecule development for PPIs has gained traction from being “undruggable” to often being “difficult”. Prior to this work small molecules that target CBP KIX lacked potency being within the mid-micromolar range. Since the KIX domain is highly conserved especially between the coactivators CBP and p300 these probes also remain limited in their ability to selectively distinguish CBP KIX versus p300 KIX prompting the need for alternative attack strategies. This dissertation work demonstrated how the Tethering strategy developed site-selective probes for CBP KIX (Figure 5-1).

In Chapter 2 small molecules were identified using a novel fluorescence polarization (FP)-based Tethering screen to identify inhibitors of the CREB(pKID)·KIX interaction.³ A liquid chromatography mass spectrometry (LC-MS) Tethering screen identified the disulfide fragment known as **1-10** which bound within the MLL-binding site to the KIX N627C mutant and directly inhibited MLL binding. The **1-10** disulfide displayed unique capabilities unforeseen for this fragment. The KIX N627C tethered to **1-10** mimicked the KIX-MLL complex by engaging the allosteric networking to increase its affinity for

pKID by two-fold. However, **1-10** tethered to KIX L664C disrupted interactions with pKID. In addition to the I660V KIX mutant that nullified the allosteric network these KIX constructs provided a framework with which to understand how the allosteric alterations impact the formation from a binary KIX-MLL to the ternary KIX-MLL-pKID complex using transient kinetic stopped-flow experiments (Chapter 3). In Chapter 4 the **1-10** disulfide was converted into irreversible compounds to produce compatible cellular probes. The strategy involved the addition of electrophiles with or without a glycine linker to the **1-10** fragment. The linkers and electrophilic moieties with **1-10** ranged in their abilities to engage the target and impact both sites within KIX N627C. For instance, the **1-10 vinyl sulfone** had a slight disruption with the pKID interaction. While the **1-10 glycine-linked vinyl sulfonamide** enhanced the pKID interaction with the KIX N627C mutant 2-fold. These two derivatives disrupted the formation of the MLL complex to the same degree. These results show that subtle changes in contacts

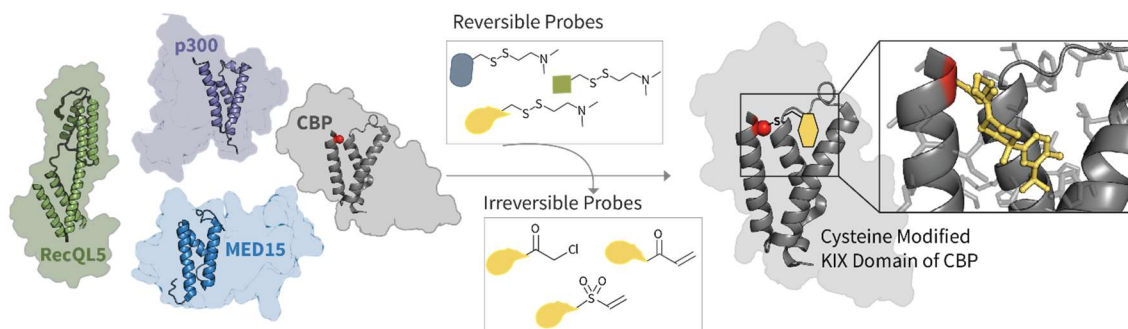


Figure 5-1 The Tethering-based approach to develop selective probes for CBP KIX Eukaryotes contain at least four GACKIX motifs in different transcriptional regulators as shown on the far left: RecQL5 (green), p300 (purple), CBP (grey), and MED15 (blue). By identifying reversible, disulfide-containing fragments and converting them into irreversible covalent probes, these probes are expected to selectively target the cysteine-containing KIX domain of CBP over the other KIX-containing proteins in cellular models. PDB IDs: 2LXT, 2GUT, 2K0N, 4BK0

between KIX and its ligand alter its allosteric capabilities.

5.3 Future directions

5.3.1 Expanding on the Tethering screen

In Chapter 2 a novel FP-Tethering screen was developed to directly target the pKID-binding site of KIX. Different screening platforms can broaden the applicability of Tethering to other systems. For instance, a fluorescence resonance energy transfer (FRET)-based screen should also work as a homogenous and high-throughput Tethering screen.⁴ FRET occurs through nonradiative (dipole-dipole) energy transfer from a donor to an acceptor fluorophore when they are around 1-10 nm apart. Time-resolved FRET (TR-FRET) reduces auto-fluorescent artefacts because of the time delay of the signal. The FP Tethering screen must remove fluorescent artefacts that were more than 150% of the total fluorescent intensity of the controls, but these artefacts are less likely to occur with a TR-FRET signal. A high-throughput screen utilized both FP and FRET (F²) signals to produce a robust screening strategy.⁵

The fragments identified from these assays depend on the screening platform (Figure 5-2). They could expand to evaluate the interactions between KIX and its other partners. For instance, c-Myb shares an overlapping site with pKID. Therefore, running two parallel FP screens against pKID and c-Myb could measure how fragments impact these interactions. A similar type of FP screen identified fragments that target an interaction between BRCA1 and a phosphorylated peptide except the same peptide was labeled with either fluorescein or rhodamine.⁶ The main advantages of the dual-fluorophore screen is to validate hits reducing common artefacts including false positives from

fragments with intrinsic fluorescence.⁶

With this dual FP scenario, the assay with c-Myb requires optimization. The affinity of KIX and c-Myb is high (K_D $1.1 \pm 0.2 \mu\text{M}$, as shown in Chapter 2), and detrimental is its low dynamic range (~ 30 anisotropy units) and high signal-background ratio (S:B) using the minimal recognition motif of c-Myb (25 amino acids) with an amino-terminal fluorescein tracer. Therefore, this assay requires optimization starting with changes in the c-Myb tracer. A longer 32-amino acid c-

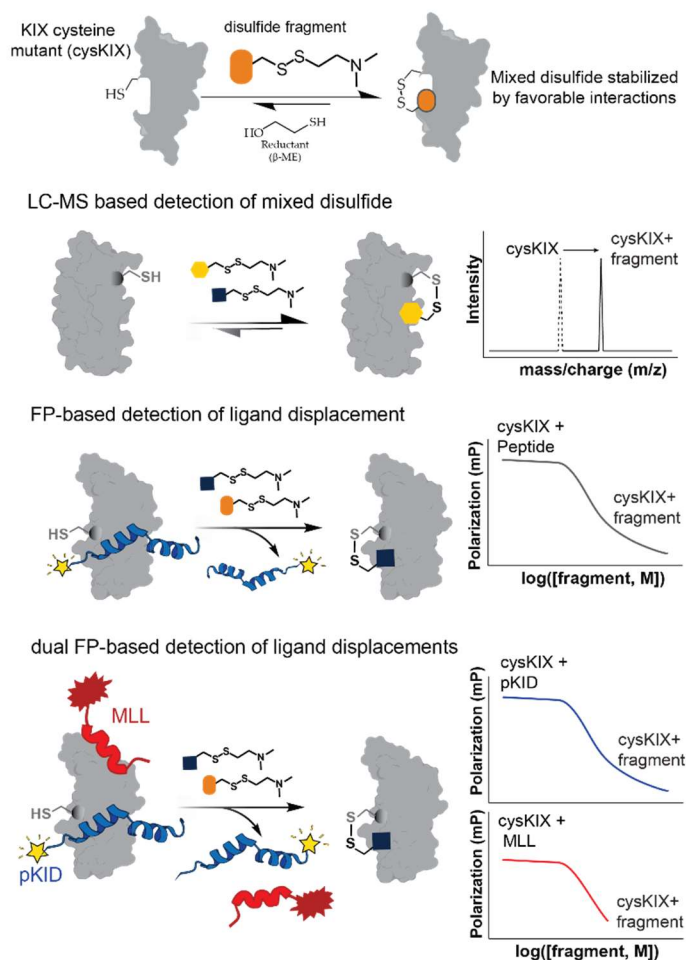


Figure 5-2 Site-directed Tethering scheme Fragments with intrinsic affinity for the KIX cysteine mutant (cysKIX) are identified through the formation of a stable mixed disulfide. LC-MS based Tethering screens directly measure the formation of fragment-cysKIX complex by a shift in the mass of cysKIX. The added mass is equivalent to the mass of the fragment. FP-based Tethering screens require a fluorescently labeled peptide. When the fragment binds to cysKIX the FP signal drops due to the displacement of the peptide from cysKIX. A dual FP Tethering screen is proposed to measure the orthosteric and allosteric effect of a favorable fragment-cysKIX complex.

Myb construct improves its affinity for KIX (K_D $0.23 \pm 0.06 \mu\text{M}$ by ITC).⁷ Finally, changing the fluorescent dye can potentially improve the robustness of this assay. Fluorophores with high extinction coefficients are considered bright (high maximum of absorbance and extinction coefficient), but the emission efficiency (quantum yield) of the dye depends on the environment. For a dual FP screen, the same peptide labeled with the red-shift tetramethylrhodamine (TAMARA) fluorophore exhibited a larger FP dynamic range than the green-shift fluorescein dye. Therefore, the FP assay against c-Myb could rapidly improve with minor changes.⁶

Small molecule probes for KIX often either allosterically disrupt or enhance interactions at the second site. However, few screens detect the impact of fragments at both sites.^{8,9} While examining different fluorophores, an assay could focus on both sites within KIX. This assay could, for instance, involve MLL-labeled with a red-shift rhodamine fluorophore and a green-shift fluorescein-labeled pKID peptide. The two dyes must act independently and they shouldn't interfere with the native interactions between the peptides and KIX. The plate reader also needs compatible optical filters to run two different tracers in the same screen. This type of FP assay that rapidly detects binding events at both sites in the primary screen can lead to potentially interesting allosteric modulators (Figure 5-2).

So far, the LC-MS and FP-based Tethering screens have been used (Figure 5-2).^{3,10,11} Both screens have advantages and disadvantages. For instance, the LC-MS screen directly detects favorable fragment-complexes but it is time-consuming and expensive. The FP screen is rapid, but it requires a fluorescently-labeled partner. The drop in a FP signal should represent the same events observed in the LC-MS screen – a fragment *binds* to the cysteine within KIX. Practically, the FP screen ranks the most active or best inhibitors for the target complexes while the LC-MS screen provides the efficient binders. The FP Tethering screen was tested

against a pKID tracer with KIX H602C. Perhaps, in hindsight a parallel LC-MS analysis would provide a comparison for the fragments that result from using these two different platforms. The parallel Tethering screens require additional time for the LC-MS analysis, however by reading the FP signal first the same samples can then be directly injected onto the LC-MS system. This process reduces the material consumption if the reagents and buffers are compatible with both FP and LC-MS analysis.

The pKID-binding site has a large surface area and lower affinity interactions than the MLL-binding site so it has previously been described as less amenable to small molecules.⁸ Therefore, it is likely advantageous to target other cysteines within the pKID-binding site. The caveat of Tethering as a site-directed approach is that identifying effective small molecules may require extensive screens probing at different regions of the interface. Since KIX is conformationally flexible, it may appear featureless and broad, but it could contain cryptic small molecule binding regions. These sites only appear in the presence of a ligand. Computational models are available to predict the locations of these sites, but additional screens would be needed to test these potential sites.¹² Once identifying cryptic sites, the fragments would efficiently bind to this region – a characteristic that would certainly overcome the current limitations of developing irreversible probes for this pKID-binding site. When converting the fragments, **6D11** and **3D4**, into irreversible probes, their binding efficiencies were poor in comparison to the disulfide fragments (Figure 5-3). Even with less than half of KIX H602C labeled by these derivatives they retained their ability to disrupt the interactions with KIX. Unfortunately, as probes they need to efficiently recognize their target otherwise they react with other free thiols in cells. Therefore, designing probes from

fragments that engage the target and elicit a desirable function improve on the development of irreversible probes.

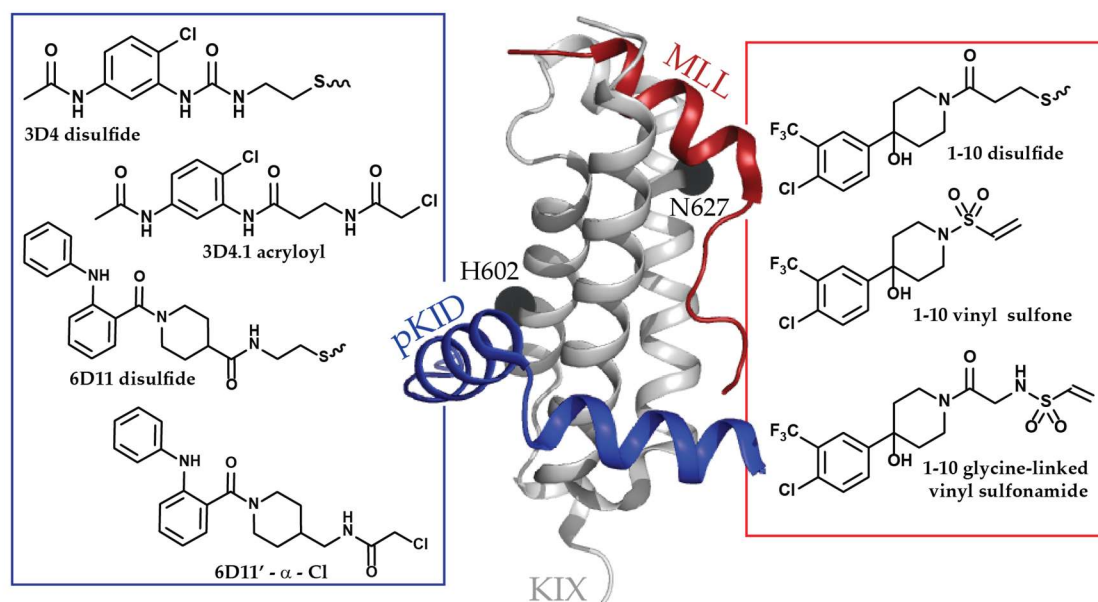


Figure 5-3 Development of irreversible probes for KIX The disulfide fragments, **3D4** and **6D11** bind to KIX H602C within the pKID-binding site. Examples of irreversible probes based on these fragments are shown with their activity from KIX H602C. The disulfide **1-10** was converted into a series of alkylators with two derivatives shown. PDB ID: 2AGH

5.3.2 Developing covalent probes

In Chapter 4 covalent probes were developed to target the engineered nucleophilic thiol cysteine mutant of KIX N627C. This conversion requires optimization with different reactive groups to improve the efficiency of covalent bond formation which leads to functional and efficient probes (Figure 5-4).¹³ For instance, electrophiles, such as vinyl amide, α -chloro ketone and vinyl sulfonamide derivatives range in reactivity.¹⁴⁻¹⁶ The alkylators need to balance this so they efficiently engage the target above all other free thiols.¹⁷⁻¹⁹ An assay that measures the reactivity of the alkylator and the reversible affinity of the fragments ranks the function of the irreversible probes.²⁰ This involves a time-dependent assay that measures the irreversible inhibition (k_{inact}) with the reversible binding

affinity of the fragment (K_i).²¹

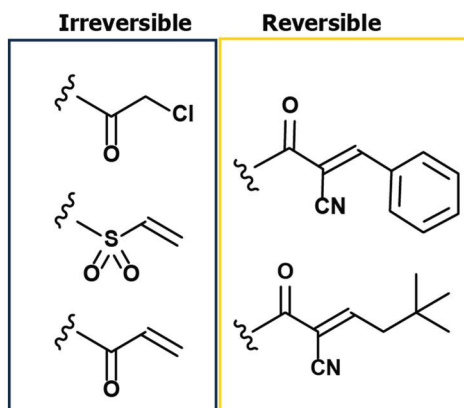


Figure 5-4 Examples of reactive warheads On the left, the blue box encloses irreversible functional groups. In the yellow box, functional groups that react through a reversible bond with a cysteine target are shown.

Other reversible covalent electrophiles would reduce intrinsic reactivity of the electrophile with the aim to minimize any off-target effects. While other cyanoacrylamide groups were explored with **1-10** their complexes were undetectable by LC-MS. However, these covalent modifications are advantageous because if they react with other free thiols the bond is rapidly reversible. Therefore, these reversible electrophiles may be useful to develop cellular probes from the disulfide fragment (Figure 5-3).^{22,23}

This leads to the main concern that the covalent probes need to undergo extensive selectivity assessments.²⁴ In cells, irreversible alkylators may react with glutathione and target reactive cysteines in proteins.²⁵ For instance, 518 human kinases have an accessible reactive cysteine.²² Therefore, providing rapid assays against the reactive cysteines within cellular models would reduce the investments required to test these probes. Activity-based probes such as attaching biotin to the probe also identify any off-targets. After cells are treated with this compound then biotinylated protein targets are isolated. Western blots quantitatively show if other biotin-containing proteins are isolated using a

detection reagent such as streptavidin conjugated to horseradish peroxidase.²⁶ If these biotinylated proteins are isolated, digested into peptides then any potential targets of the biotin-tagged probe can be identified through a LC-MS-based proteomics experiment.^{27,28} For downstream cellular analyses when probes cause a biological response in cells then more confident hypotheses are drawn because any potential non-specific targets of a probe are known and minimal or ideally, non-existent.²⁹

5.3.3 *Alternative method to identify probes for homologous domains*

In Chapter 4 the irreversible Tethering strategy led to the development of probes to selectively target the KIX domain within CBP versus other coactivators such as p300. The ‘bump in the hole’ strategy represents another method that designs selective probes against highly similar domains. With the bump-and-hole approach, a binding site is reconstructed using site-directed mutagenesis. Then a fragment with intrinsic affinity for the domain is modified with large, bulky groups that accommodate the modified target over the apo protein.²⁴ Compatible modifications are made between the protein and fragment until a specific probe results for the mutant protein. For more difficult targets this requires “larger” holes (more dramatic mutations); but these mutations must also conserve the structure and function of the protein domain.³⁰⁻³² This technique has adapted a BET bromodomain inhibitor, I-BET762/JQ1 selective for BET bromodomains possessing a distinct leucine or alanine mutation versus the entire BET subfamily.³³ This method could also produce selective probes for the KIX domain. For instance, the initial fragments such as **3D4** and **6D11** bind to region near the cysteine mutant KIX H602C. By mutating around this region in KIX and altering the fragments selective small molecules may be developed for KIX. Additionally, in CBP KIX nearby the 602 residue is a serine, but in p300 KIX this residue is an asparagine so

the modifications to the fragments **3D4** or **6D11** could exploit this difference to selectively target one KIX motif over another (Figure 5-5).³⁴

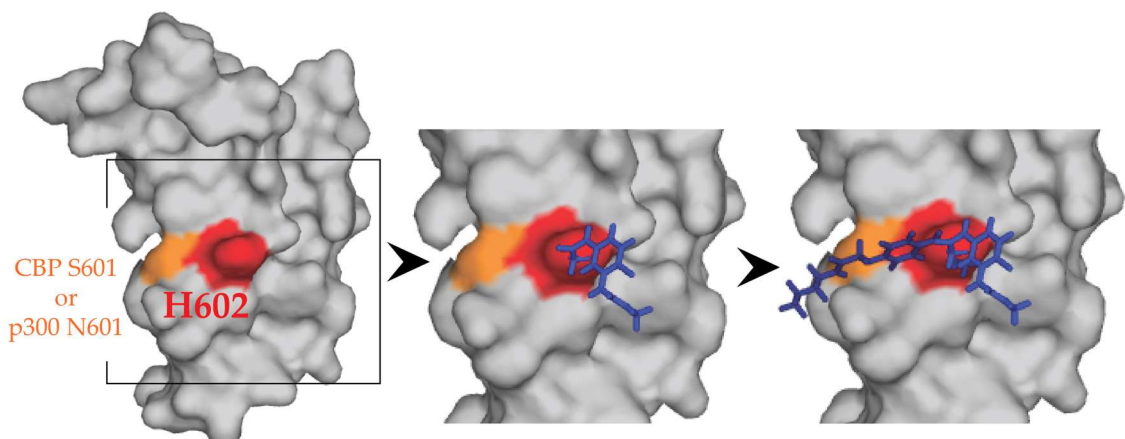


Figure 5-5 Bumb-and-hole approach A proposed scheme with KIX (grey surface) H602C binds to a the known ligand, **6D11** represented as blue sticks that binds to KIX H602C. This ligand binds in proximity to a single amino acid difference between CBP and p300 (orange region) and this difference could be used to generate a selective small molecule. PDB ID: 4I9O

5.3.4 *Venturing to probe the cooperativity imparted by KIX*

In Chapter 3 the ternary complex formation from the binary MLL-KIX to MLL-KIX-pKID was examined using different constructs of KIX that conveyed no, negative, and positive cooperativity and transient state kinetics.³⁵ Allostery is a complex process that involves changes in how a partner binds in the absence and presence of a second ligand.³⁶ The mutants and covalent probes produce homogenous populations of KIX and these complexes are useful to understand how the second partner binds to an otherwise nonnative, but functional form of KIX. Therefore, identifying additional ligands and mutants that target conformations of KIX could be used to aid in our understanding of the underlying allosteric mechanism of KIX and its partners.

Fragments that bind to KIX alter the conformational states by either stabilizing or destabilizing KIX. These changes shift the thermodynamic equilibrium between the bound and native state of KIX which are detectable by measuring the melting temperature (T_m) through “thermal shift” or “thermofluor”

DSF assays. It detects the unfolding process through a fluorescence signal of an environmentally sensitive fluorescent dye. SYPRO® orange, for instance, emits light from the hydrophobic surface of protein and it is quenched as the protein unfolds exposing the dye to aqueous solution.³⁷ The melting temperature, T_m reflects the temperature when 50% of the protein remains folded. This is measured from the point of inflection of the fluorescence signal against the temperature. DSF is a high throughput technique heating as fast as 2 degrees C per minute and it is used to identify ligands that stabilize proteins. The material consumption is moderate requiring a volume of 10-20 μ l for a 384-well plate. Multiple fragments can be screened per well. This combinatorial screening approach reduces nonspecific interactions due to competing fragments. If the T_m shifts, then the compounds from the well are separated out in a secondary screen. This reduces the intensive material consumption and time required for DSF screens. This type of screen may provide novel fragments for the coactivator domain targets because it primarily detects stabilization or destabilization of the target. This can later show how these effects impact binding events and functions.³⁸

In Chapter 3, mutations within the third α -helix of KIX attempted to stabilize its structure and to increase the affinity of KIX for c-Myb. Structurally, KIX is marginally stable and this stability is hypothesized to regulate its function and ability to induce cooperativity between the two activator-binding interfaces.³⁹⁻
⁴¹ Stability results by limiting the dynamics or thermal fluctuations within the protein, but specific factors remain difficult to isolate and to predict. Large structural shifts result in minimal stability changes while small, subtle changes may significantly stabilize a structure.⁴² Therefore, stable conformational states of KIX that impact its interactions remain difficult to capture. An artificial selection technique could experimental identify mutations that stabilize it. For instance, a shotgun mutagenesis library can be made with a parental protein expression

vector that is randomly mutated to create a library of clones with each well representing an individual point mutation.⁴³ Two or three mutants are generated per position often reflecting a conservative and more drastic change. In a well-based format, the mutants are processed through the cloning process to purified protein. Then the protein can either be screened by a thermal melt assay or by a binding assay. In a high-throughput binding assay the protein expressed with a tag (hexahistidine or biotin) is easily isolated on compatible plates. This format enables a rapid screen with methods such as surface plasmon resonance or biolayer interferometry that require an immobilized protein to measure its affinity against different binding partners.⁴⁴

Alternatively, an endogenous screen in cells detects robust mutants that maintain function and cell viability.⁴⁵ For this type of screen a CRISPR/Cas9 platform introduces double-stranded breaks (DSB) using library of guide RNAs (sgRNA) to target regions across the target gene.^{46,47} The breaks induce the non-homologous repair pathway which produces small insertions or deletions (indels) at the DSB.⁴⁸ While these indels initially appear random they are reproducible across experimental replicates, and cell lines. Viable genes are sequenced and then the protein is re-assessed *in vitro* for how these mutations effect the function or structure of the target protein. Due to subtle changes within KIX this type of screen would likely benefit from a robust signal such as designing a reporter (FRET or luminescence) assay within the cells.⁴⁹ The mutational robustness of KIX may reflect how it selects its partners. For instance, CBP and p300 KIX bind to a variety of activators while ARC105/MED15 KIX selectively bind to sterol response element binding proteins (SREBPs) to regulate metabolism.⁵⁰ This type of *in cellulo* screen may reveal how the GACKIX motif evolves to selectively bind to activators versus impart promiscuous binding profiles through alternative conformational

states.

5.3.5 Examining coactivators and its complexes

The PPIs between activator and coactivators are reduced to a minimal transcriptional activation domain (TAD) from the activator and the coactivator domain partner to study them using biochemical methods *in vitro*. In the full-length coactivator, domains are separated by flexible regions that allow them to function independently.⁵¹ Little evidence is available to understand how the domains orient in the full-length coactivator and whether a domain's interactions are regulated by the others.⁵² The domains within CBP/p300 have been extensively characterized so it is time to link these complexes together starting with more extensive structural characterization of the coactivators. The first hurdle it to isolate these large proteins and insect cells enable the expression of large, challenging proteins (Figure 5-6).⁵³ The initial structural characterization of the coactivator can proceed through techniques such as electron microscopy (EM)⁵⁴ and ion mobility mass spectrometry (IM-MS). With IM-MS collisional cross-

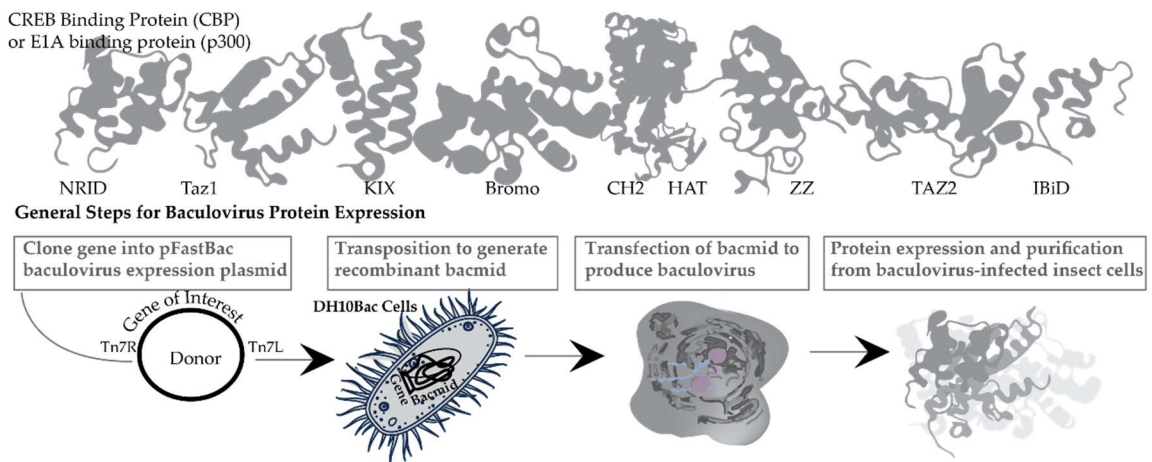


Figure 5-6 Structural characterization of the coactivator, CBP or p300 Above represents the domains of CBP and p300. The respective positions of these domains are taken from the protein sequence and individually the structures for the domains have been characterized. The overall structure for the coactivators has not yet been characterized. The general steps for protein expression in insect cells are outlined. This type of procedure is useful to successfully isolate large and/or complex proteins.

section analysis measures the average rotational orientation or size. This large coactivator makes it difficult to reconstruct a model, but could be aided with EM structures and the expansion of computational tools. Perhaps, by using an ensemble of known ligands, these extensive experiments and computational predictions can show how these domains orient with respect to one another. Additional mass spectrometry tools including cross-linking, electron capture dissociation and other bottom-up approaches can analyze peptide cleavages to reconstruct structures.⁵⁵ This type of molecular detail provides insights into how the coactivator functions as a hub to interact with hundreds of activators.⁵⁶

5.3.6 Expanding into cellular systems

The long-standing goal is to understand how these activator-coactivator complexes function in a cellular environment. *For in vitro* experiments, highly pure proteins, peptides, and small molecules are mixed in buffered solutions for binding interaction studies. In cells, these PPIs occur in heterogenous and crowded environments yet detecting PPIs in these complex environment remains challenging. Additional methods are expanding the ability to probe these PPIs *in cellulo*. As previously stated the covalent probes are tested for their ability to bind to the target. These probes often affect the conformational stability of KIX which also modulates its interactions. The cellular thermal shift assay (CTSA) is analogous to the DSF assay described in section 5.3.3, but it measures the relative stability of a protein extracted from cells. In this approach, cells are heated causing unstable protein to precipitate. Then the target protein is isolated from the soluble fraction and quantified by immunoblotting. Probes that target the protein or mutations within the protein are introduced into cells and their effects on protein stability are measured through the thermal shifts (the amount of protein against the increasing temperature). The limit for this technique is the ability to selectively

isolate the target protein, usually through a specific antibody. KIX is a small domain making up 12 kDa of the total 265.2 kDa of the coactivator CBP. Therefore, it is unlikely that immunoblotting for CBP detects the stability differences within the KIX domain and even less likely that a specific antibody can detect the KIX domain of CBP over p300. However different probes that target other domains within CBP or p300 can be used in combination with the probes for KIX and then the stability differences may be substantial enough to detect. Mass spectrometry techniques may also be developed to quantify stability differences of these coactivators.

This represents how transcriptional regulators are difficult to study in cells because the individual complexes make small, but critical contribution to transcriptional processes. These complexes are also dynamic meaning they are in constant flux drifting on-and-off their interacting partners. Therefore, these interactions are harder to detect from a heterogenous cell population. In a cellular system, for instance, the changes in biomolecules from proteomics, lipidomics and metabolomics are detected globally through LC-MS-based analyses. At the genome level, a population's genome is sequenced either at localized regions or on a large scale through deep sequencing. Finally, gene transcripts are quantified through target-based approaches (quantitative PCR) or the complete gene expression profile is monitored through RNA sequencing (RNAseq). Each of these techniques has their advantages and limitations. The main disadvantage is that these measurements represent averages from large population of different cells so the subtle differences in transcriptional processes are undetectable. By analyzing each cell individually these networks and their effects may be drawn out.

Single-cell approaches simplify the molecular processes but they are difficult to conduct due to their small scale. For instance, amplification of genomic regions through PCR on minute samples are prone to introduce error from one

genomic sample than from multiple samples. However, improvements have been made. Microfluidics, for instance, is contributing to these methods through the development of droplet microfluids and single-cell printers that gently isolate cells for downstream analysis. Flow cytometry and cell microscopy are coming together to rapidly monitor specific fluorescently-tagged proteins on a smaller scale. Mass cytometry is also a quantitative approach to rapidly detect cellular features. From the point-of-view from one cell, there are fewer spatial and temporal configurations that the target domain can adopt. In the future, the single cell analysis may aid our understanding of how these individual activator-coactivator interactions function.

5.4 Final remarks

In conclusion, this work has developed selective probes for the CBP KIX domain – a highly homologous domain across other transcriptional regulations including its homolog p300. From these studies, **1-10**, for instance can be a useful probe to study the KIX interactions on both MLL-and CREB(pKID)-responsive genes. The probes developed in this dissertation provide the framework to develop potent and selective probes for other highly homologous targets.

This work also delved into how dynamic coactivator domain regulates its interactions through an allosteric network. The activation of transcription has been shown to depend on the binding mechanism between activator-coactivator complexes and subtle transitions convey important functional responses. For instance, this work demonstrated that positive cooperativity may result from an extended interaction time with CBP/p300 possibly resulting in an increase in the transcriptional output. However, additional work is required to understand how

PPIs such as KIX and its activators function in cells (Figure 5-6).

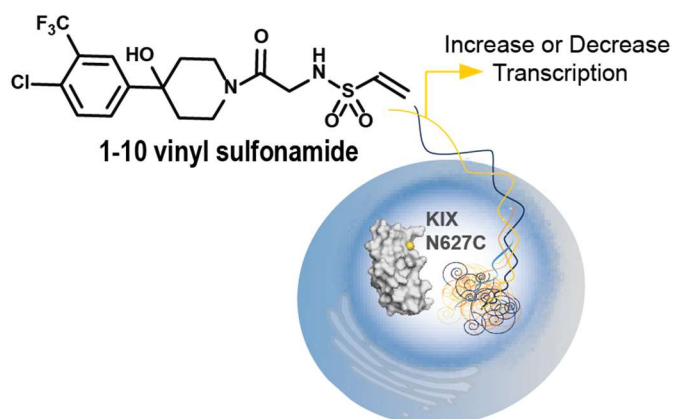


Figure 5-6 The irreversible probe 1-10 vinyl sulfonamide was developed to selectively target the coactivator containing the KIX N627C mutant in cells.

5.5 References

1. Surade, S. & Blundell, Tom L. Structural Biology and Drug Discovery of Difficult Targets: The Limits of Ligandability. *Chemistry & Biology* **19**, 42-50, (2012).
2. Hantschel, O., Grebien, F. & Superti-Furga, G. Targeting allosteric regulatory modules in oncoproteins: “Drugging the Undruggable”. *Oncotarget* **2**, (2011).
3. Lodge, J. M., Rettenmaier, T. J., Wells, J. A., Pomerantz, W. C. & Mapp, A. K. FP Tethering: a screening technique to rapidly identify compounds that disrupt protein-protein interactions. *MedChemComm* **5**, 370-375, (2014).
4. Inglese, J. *et al.* High-throughput screening assays for the identification of chemical probes. *Nat Chem Biol* **3**, 466-479, (2007).
5. Du, Y. *et al.* A dual-readout F2 assay that combines fluorescence resonance energy transfer and fluorescence polarization for monitoring bimolecular interactions. *Assay Drug Dev Technol* **9**, 382-393, (2011).
6. Simeonov, A. *et al.* Dual-fluorophore quantitative high-throughput screen for inhibitors of BRCT–phosphoprotein interaction. *Anal Biochem* **375**, 60-70, (2008).

7. Arai, M., Sugase, K., Dyson, H. J. & Wright, P. E. Conformational propensities of intrinsically disordered proteins influence the mechanism of binding and folding. *Proc Natl Acad Sci USA* **112**, 9614-9619, (2015).
8. Gee, C. T., Koleski, E. J. & Pomerantz, W. C. K. Fragment Screening and Druggability Assessment for the CBP/p300 KIX Domain through Protein-Observed ¹⁹F NMR Spectroscopy. *Angew Chem* **54**, 3735-3739, (2015).
9. Gee, C. T. *et al.* Protein-observed (¹⁹F)-NMR for fragment screening, affinity quantification and druggability assessment. *Nat Protoc* **11**, 1414-1427, (2016).
10. Wang, N. *et al.* Ordering a dynamic protein via a small-molecule stabilizer. *J Am Chem Soc* **135**, 3363-3366, (2013).
11. Pomerantz, W. C. *et al.* Profiling the dynamic interfaces of fluorinated transcription complexes for ligand discovery and characterization. *ACS Chem Biol* **7**, 1345-1350, (2012).
12. Cimermancic, P. *et al.* CryptoSite: Expanding the Druggable Proteome by Characterization and Prediction of Cryptic Binding Sites. *J Mol Biol* **428**, 709-719, (2016).
13. Ostrem, J. M., Peters, U., Sos, M. L., Wells, J. A. & Shokat, K. M. K-Ras(G12C) inhibitors allosterically control GTP affinity and effector interactions. *Nature* **503**, 548-551 (2013).
14. Reddick, J. J., Cheng, J. & Roush, W. R. Relative rates of Michael reactions of 2'-(phenethyl)thiol with vinyl sulfones, vinyl sulfonate esters, and vinyl sulfonamides relevant to vinyl sulfonyl cysteine protease inhibitors. *Org Lett* **5**, 1967-1970, (2003).
15. Li, M., Wu, R. S., Tsai, J. S. & Salamone, S. J. Evaluation of vinylsulfamides as sulfhydryl selective alkylation reagents in protein modification. *Bioorg Med Chem Lett* **13**, 383-386, (2003).
16. Kathman, S. G., Xu, Z. & Statsyuk, A. V. A fragment-based method to discover irreversible covalent inhibitors of cysteine proteases. *J Med Chem*

- 57, 4969-4974, (2014).
17. Miller, R. M. & Taunton, J. Targeting protein kinases with selective and semipromiscuous covalent inhibitors. *Methods Enzymol* **548**, 93-116, (2014).
 18. Long, M. J. C. & Aye, Y. Privileged Electrophile Sensors: A Resource for Covalent Drug Development. *Cell Chem Biol* **24**, 787-800, (2017).
 19. González-Bello, C. Designing Irreversible Inhibitors—Worth the Effort? *ChemMedChem* (2015).
 20. Kwarcinski, F. E., Fox, C. C., Steffey, M. E. & Soellner, M. B. Irreversible inhibitors of c-Src kinase that target a nonconserved cysteine. *ACS Chem Biol* **7**, 1910-1917, (2012).
 21. Krippendorff, B.-F., Neuhaus, R., Lienau, P., Reichel, A. & Huisinga, W. Mechanism-Based Inhibition: Deriving KI and kinact Directly from Time-Dependent IC₅₀ Values. *J Biomol Screen* **14**, 913-923, (2009).
 22. Bradshaw, J. M. *et al.* Prolonged and tunable residence time using reversible covalent kinase inhibitors. *Nat Chem Biol* **11**, 525-531, (2015).
 23. Miller, R. M., Paavilainen, V. O., Krishnan, S., Serafimova, I. M. & Taunton, J. Electrophilic fragment-based design of reversible covalent kinase inhibitors. *J Am Chem Soc* **135**, 5298-5301, (2013).
 24. Bishop, A. *et al.* Unnatural Ligands for Engineered Proteins: New Tools for Chemical Genetics. *Annu Rev Biophys* **29**, 577-606, (2000).
 25. Weerapana, E. *et al.* Quantitative reactivity profiling predicts functional cysteines in proteomes. *Nature* **468**, 790-795, (2010).
 26. Taipale, M. *et al.* Quantitative Analysis of Hsp90-Client Interactions Reveals Principles of Substrate Recognition. *Cell* **150**, 987-1001, (2012).
 27. Ong, S.-E., Foster, L. J. & Mann, M. Mass spectrometric-based approaches

- in quantitative proteomics. *Methods* **29**, 124-130, (2003).
28. Speers, A. E. & Cravatt, B. F. Chemical Strategies for Activity-Based Proteomics. *Chembiochem* **5**, 41-47, (2004).
 29. Workman, P. & Collins, I. Probing the Probes: Fitness Factors For Small Molecule Tools. *Chemistry & Biology* **17**, 561-577.
 30. Arrowsmith, C. H., Bountra, C., Fish, P. V., Lee, K. & Schapira, M. Epigenetic protein families: a new frontier for drug discovery. *Nat Rev Drug Discov* **11**, 384-400, (2012).
 31. Runcie, A. C., Chan, K. H., Zengerle, M. & Ciulli, A. Chemical genetics approaches for selective intervention in epigenetics. *Curr Opin Chem Biol* **33**, 186-194, (2016).
 32. Kung, A. *et al.* A Chemical-Genetic Approach to Generate Selective Covalent Inhibitors of Protein Kinases. *ACS Chemical Biology* **12**, 1499-1503, (2017).
 33. Baud, M. G. J. *et al.* A bump-and-hole approach to engineer controlled selectivity of BET bromodomain chemical probes. *Science* **346**, 638-641, (2014).
 34. Kozakov, D. *et al.* Ligand deconstruction: Why some fragment binding positions are conserved and others are not. *Proc Natl Acad Sci USA* **112**, E2585-E2594, (2015).
 35. Wang, N., Lodge, J. M., Fierke, C. A. & Mapp, A. K. Dissecting Allosteric Effects of Activator-Coactivator Complexes Using a Covalent Small Molecule Ligand. *Proc Natl Acad Sci USA* **11**, 12061-12066, (2014).
 36. Tsai, C.-J. & Nussinov, R. A Unified View of "How Allostery Works". *PLoS Comput Biol* **10**, e1003394, (2014).
 37. Biggar, K. K., Dawson, N. J. & Storey, K. B. P07. *Cryobiology* **69**, 197, (2014).
 38. Troilo, F. *et al.* The Folding Pathway of the KIX Domain. *ACS Chem Biol* **12**,

- 1683-1690, (2017).
39. Law, S. M., Gagnon, J. K., Mapp, A. K. & Brooks, C. L. Prepaying the entropic cost for allosteric regulation in KIX. *Proc Natl Acad Sci USA* **111**, 12067-12072, (2014).
 40. Tollinger, M. *et al.* An isolated helix persists in a sparsely populated form of KIX under native conditions. *Biochemistry* **45**, 8885-8893, (2006).
 41. Brüscheiler, S., Konrat, R. & Tollinger, M. Allosteric Communication in the KIX Domain Proceeds through Dynamic Repacking of the Hydrophobic Core. *ACS Chem Biol* **8**, 1600-1610, (2013).
 42. Sikosek, T. & Chan, H. S. Biophysics of protein evolution and evolutionary protein biophysics. *J Royal Soc Interface* **11**, (2014).
 43. Melamed, D., Young, D. L., Miller, C. R. & Fields, S. Combining Natural Sequence Variation with High Throughput Mutational Data to Reveal Protein Interaction Sites. *PLoS Genetics* **11**, e1004918, (2015).
 44. Ciulli, A. in *Protein-Ligand Interactions: Methods and Applications* (eds Mark A. Williams & Tina Daviter) 357-388 (Humana Press, 2013).
 45. Shi, J. *et al.* Discovery of cancer drug targets by CRISPR-Cas9 screening of protein domains. *Nat Biotechnol* **33**, 661-667, (2015).
 46. Jinek, M. *et al.* RNA-programmed genome editing in human cells. *Elife* **2**, e00471, (2013).
 47. Shalem, O., Sanjana, N. E. & Zhang, F. High-throughput functional genomics using CRISPR-Cas9. *Nat Rev Genet* **16**, 299-311, (2015).
 48. van Overbeek, M. *et al.* DNA Repair Profiling Reveals Nonrandom Outcomes at Cas9-Mediated Breaks. *Cell* **63**, 633-646, (2016).
 49. Truong, K. & Ikura, M. The use of FRET imaging microscopy to detect protein-protein interactions and protein conformational changes in vivo.

Curr Opin Struct Biol **11**, 573-578, (2001).

50. Yang, F. *et al.* An ARC/Mediator subunit required for SREBP control of cholesterol and lipid homeostasis. *Nature* **442**, 700-704, (2006).
51. Dyson, H. J. & Wright, P. E. Role of Intrinsic Protein Disorder in the Function and Interactions of the Transcriptional Coactivators CREB-Binding Protein (CBP) and p300. *J Biol Chem*, (2016).
52. Sormanni, P. *et al.* Simultaneous quantification of protein order and disorder. *Nat Chem Biol* **13**, 339-342, (2017).
53. Brown, W. C. *et al.* New ligation-independent cloning vectors compatible with a high-throughput platform for parallel construct expression evaluation using baculovirus-infected insect cells. *Prot Express Purif* **77**, 34-45, (2011).
54. Yi, P. *et al.* Structure of a Biologically Active Estrogen Receptor-Coactivator Complex on DNA. *Cell* **57**, 1047-1058, (2015).
55. Yu, C., Kandur, W., Kao, A., Rychnovsky, S. & Huang, L. Developing new isotope-coded mass spectrometry-cleavable cross-linkers for elucidating protein structures. *Anal Chem* **86**, 2099-2106, (2014).
56. Mosca, R., Ceol, A. & Aloy, P. Interactome3D: adding structural details to protein networks. *Nat Meth* **10**, 47-53, (2013).

Appendix I Unfinished business: Addressing the cellular requirements for the irreversible Tethering strategy

I.1 Abstract

The concerted effort between DNA-bound activators and coactivators mediate the activation of specific genes. However, these individual activator-coactivator interactions and their contributions to transcription remain unclear. In Chapter 4 we developed irreversible probes to selectively target the KIX domain of CBP and its interactions with activators including MLL, CREB (pKID), and c-Myb. Probes that selectively target a KIX N627C cysteine mutant were developed using an irreversible Tethering scheme, but the cellular systems must contain the mutation within KIX to use these probes *in cellulo*. We proposed to introduce or generate this mutation using the CRISPR/Cas9 genome editing techniques.

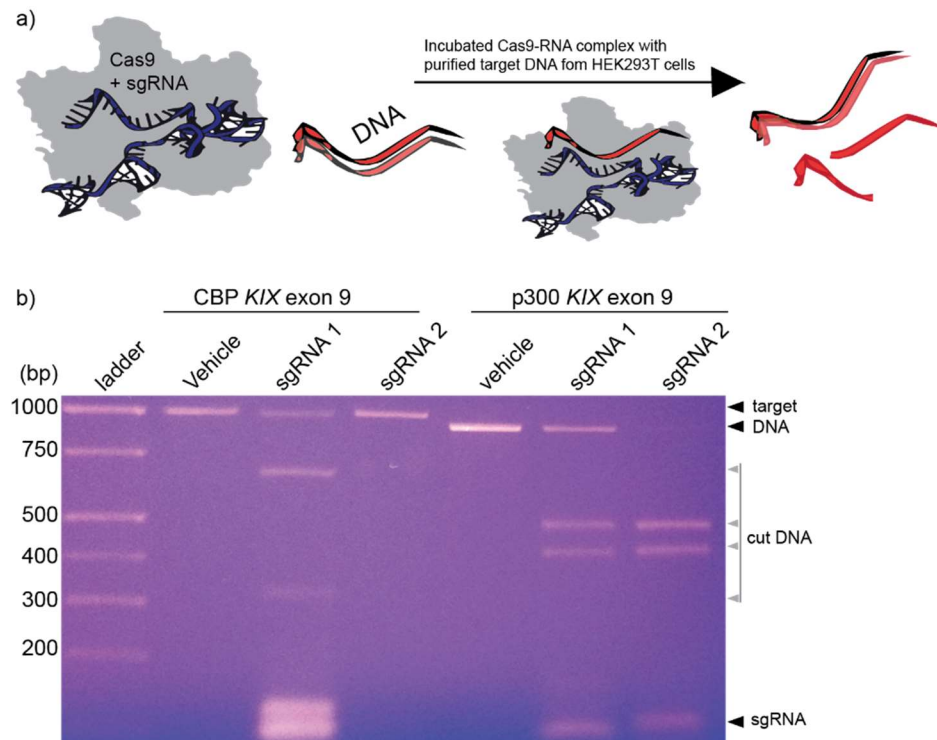
I.2 Introduction

The human genome is as important as the regulation of its genes. Genome engineering tools that directly and specifically target regions of the genome are valuable to understand details about gene expression and regulation.¹ In 2012 a CRISPR/Cas9 system modified specific regions of the genome in mammalian cells.²⁻⁴ The type II Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) Type II system involves the co-expression of the CRISPR-associated endonuclease (spCas9) from *Streptococcus pyogenes* and guide RNA (gRNA).^{5,6} The Cas9 ribonuclease in complex with a guide RNA strand specifically recognizes DNA through a protospacer motif (PAM) and creates a double-stranded break (DSB).^{2,7-10} Homology directed repair utilizes a single-stranded template to

generate the desired insertions, deletions, or mutations directly into the target region.^{9,11-13} We proposed to generate the necessary KIX cysteine mutant using this technique.

I.3 Results

Single guide RNAs (sgRNAs) were designed to induce a DSB as close to the cysteine mutant as possible within CBP or p300 *KIX*.⁵ They cleaved their target



DNA based on an *in vitro* Cas9-sgRNA cleavage assay (Figure II-1). However, the

Figure I-1 In vitro endonuclease cleavage assay to target the KIX gene a) The scheme of this assay involves incubating Cas9 with the RNA to form the ribonucleotide-protein complex and then adding the DNA with the target sequence. b) The endonuclease reactions were visualized on a 2% agarose gel. The PCR products amplified from the HEK293T genome contained the target sequences for the CBP and p300 *KIX* gene. For p300 *KIX* both sgRNAs cleaved the target sequence but sgRNA 2 was more efficient than sgRNA 1. For the CBP *KIX* sgRNA 1 recognized and cleaved the target. The sgRNA 2 likely degraded because its band was missing at 100 bp.

⁵ As the DSB moves further than 7 bases away from the mutation site the efficiency of HDR starts to decrease.

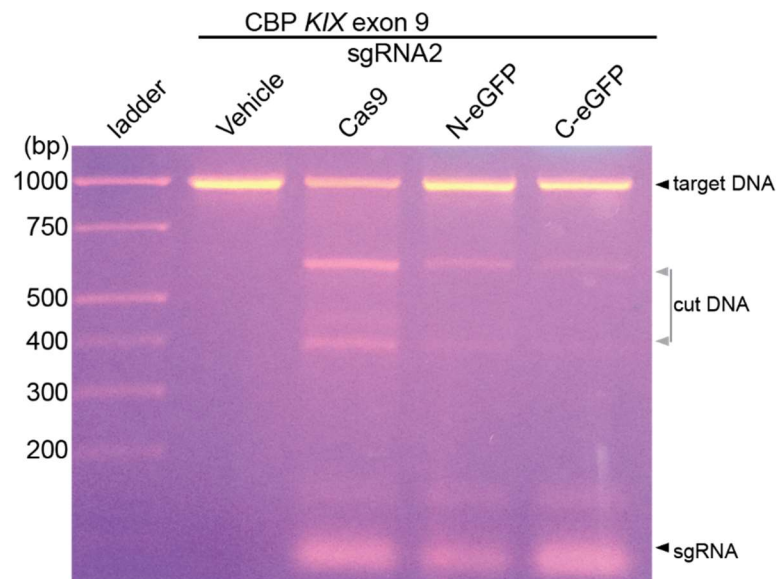


Figure I-2 Comparing the cas9 constructs using the *in vitro* endonuclease cleavage assay The three different Cas9 proteins (without eGFP, an amino-terminal eGFP and carboxyl-terminal eGFP) incubated with sgRNA2. The DNA target was an amplified region of the CBP *KIX* gene and added to the Cas9-sgRNA complexes. The Cas9-tagged eGFP constructs cleaved the target DNA, but Cas9 alone was more efficient.

cleavage wasn't evident in cells either by the visualizing the amplified DNA regions with the DNA gel and T7 endonuclease I (T7EI) cleavage assay or by Sanger sequencing.^{9,14,15} To monitor Cas9 delivery into cells N- and C-terminal enhanced green fluorescent protein (eGFP)-tagged Cas9 constructs were introduced using commercially available lipid-based transfection reagents (Lipofectamine 2000, Thermo Fischer Scientific) or through electroporation.¹⁶⁻¹⁸ By flow cytometry the eGFP constructs entered a small fraction of cells providing a strong signal, but cell viability was poor. While the eGFP constructs retained their ability to cleave their targets *in vitro* their efficiency was lower than the untagged Cas9 endonuclease (Figure I-2). Therefore, this process requires additional

optimization in both its design and implementation.

I.4 Discussion and Future Directions

Most notably, the sgRNA design is critical to the success of the genome editing experiment. The Cas9 and sgRNAs are generally more efficient at targeting the first couple of exons within a gene.¹⁹ Two sgRNAs can also work in tandem to cut at different regions of an exon and increase the efficiency of knocking out the gene. These types of experiments have also been designed. With this scenario upon targeting the first exons of *CBP* or *p300* the respective coactivator would be reintroduced using a mammalian expression vector encoding the coactivator and KIX cysteine mutant. Other RNA-guided nucleases have also been applied for genome engineering purposes. These nucleases are designed to vary how or where the target is cleaved and they are often driven by different protospacer adjacent motif (PAM) preferences. While more prediction tools are being developed, difficult DNA targets are successfully cleaved by comparing many different guides to find the most efficient or fragile DSB region. For spCas9 the guide sequence is around 19 bases and the target strand must contain the sequence, NGG on the 3' side where N is any base pair and G is guanine.⁴ The PAM for *Acidaminococcus* and *Lachnospiraceae* Cpf1, for instance, is TTN on the 5' side of the target where T represents thymine. Cpf1 also generates a staggered cut with a 5' overhang. These types of cleavages are also seen with transcription activator like-effector nucleases (TALENs) and they allow for precise genome edits. With blunt DSB the cells must be dividing to introduce the mutation through HDR and some tinkering with these cells increase the editing efficiency such as using NHEJ (Ku) pathway inhibitors or synchronizing cell division and timing the delivery of the Cas9-sgRNA components. With the overhangs, the desired edit occurs through

non-homologous repair end joining (NHEJ) throughout the cell cycle.

As we observed spCas9 is fussy about its delivery and function in mammalian cells. It is a bacterial enzyme so it is not designed for the complex mammalian cell. The first improvements to Cas9 have made it more specific. Cas9 has a very high affinity for DNA (dissociation constant, K_D of 1.2 ± 0.1 nM) and it remains localized to DNA searching for a matching sequence.¹³ Mutations within Cas9 reduce its high affinity for the DNA target to decrease off-target cleavages. When Cas9 lingers in the nucleus the number of off-target effects increase. To reduce these potential off-target cleavages, mammalian expression vectors for Cas9 contain an inducible promoter to limit the expression of Cas9. Finally, a more soluble Cas9 protein increases the nuclease activity in cells as claimed by different companies through undisclosed modifications. Excitement surrounded the discovery of Cpf1 because it is significantly smaller and easier to introduce into mammalian cells.²⁰⁻²²

Finally, the validation of the mutation as well as off-targets are important to properly and responsibly employ these genome-editing strategies. Since these approaches have rapidly evolved over the last few years the validations fall to the discretion of individual research groups and their peer reviewers. Some argue that modified cells should be verified by deep sequencing of the entire genome and/or of the expressed genes (RNA-seq) while others find this excessive. So far, the minimal requirements are to provide sequencing results for the genomic regions that are related to any genotypic and phenotypic conclusions from the study. Straight-forward assays detect insertions or deletions (indels) triggered by RNA-guided DNA nucleases use the T7 endonuclease I or analogous Surveyor nuclease which cleave mismatches or bulges in double-stranded DNA.¹⁵ A robust and clean PCR product from the region of interest is critical to obtain reliable results. Simple sanger sequencing of the PCR product also validates the presence of indels. High

throughput cloning kits such as the T-Blunt PCR Cloning Kit (SolGent) or TOPO® TA kit (Invitrogen) insert the regions of interest into a vector, these insertions are screened through colonies, and positive colonies can then be sequenced.

Although CRISPR/Cas9 genome editing techniques extend into organisms, they are broadly used in mammalian cell culture. Mycoplasma testing results is one quality-control standard that may be required by peer-reviewed journals. Mycoplasma is a well-known contaminant found in cell cultures which can go unnoticed due to its small size. Their infection of cells led to misconstrued results in certain pathways.²³⁻²⁶ Therefore, researcher must test their cells for mycoplasma when they submit research containing data from cell models.

As these genome engineering techniques advance it will likely become easier to directly generate the cysteine mutation and to also target other co-regulators with our irreversible tethering strategy.

I.5 Materials and Methods

I.5.1 Protocol for Cas9 purification

A pET-based plasmid containing the *spCas9* gene from *Streptococcus pyogenes* (Addgene pMJ806, plasmid #39312, deposited by Dr. Jennifer Doudna's Lab)² was transformed in the *E. coli* strain Rosetta™ (DE3) pLysS competent cells (Novagen, Merck Millipore) by means of heat-shock or with *Mix & Go Z* competent cells (Zymo Research). For heat shock transformations 1 to 0.1 µg of plasmid was mixed with 50 µl of competent cells. The cells were incubated on ice for 10 minutes, placed on a heat block set at 42 °C for 45 seconds, and set on ice for another 2 minutes. For recovery, the cells were incubated with 950 µl of SOC shaking at 250 rpm, 37 °C for 1 hour. For Z-competent cells 1 to 0.1 µg of plasmid was incubated with 50 µl of with *Mix & Go Z*-competent cells on ice for 30 minutes. Then 500 µl of SOC was added and the transformation mixture incubated for 1-2 hours at 37

$^{\circ}\text{C}$ with shaking at 250 rpm. The cells were collected by centrifugation at top speed (14 rcf) for 1 minute in a benchtop microcentrifuge. The cell pellet was suspended in 100 μl of SOC, they were plated on a LB agar plate containing 25 $\mu\text{g ml}^{-1}$ kanamycin, and the plate incubated overnight (16 to 18 hours) at 37 $^{\circ}\text{C}$.

For expression of Cas9 protein, 50 ml of LB containing 25 $\mu\text{g}\cdot\text{ml}^{-1}$ kanamycin and 34 $\mu\text{g}\cdot\text{ml}^{-1}$ chloramphenicol was inoculated with one white colony in a 250-ml baffled flask (Bellco Glass Inc.) and grown overnight at 37 $^{\circ}\text{C}$, 250 rpm. The culture was visible cloudy after 18 hours and 10 ml of this starter culture was used to initiate growth in a 2 L baffled flask containing 1 L of Terrific broth (24 g yeast extract, 12 g tryptone, 4% glycerol, 100 mM potassium phosphate buffer) and 25 $\mu\text{g}\cdot\text{ml}^{-1}$ kanamycin and 34 $\mu\text{g}\cdot\text{ml}^{-1}$ chloramphenicol. The cultures incubated at 37 $^{\circ}\text{C}$, 250 rpm until the $\text{OD}_{600\text{nm}}$ reached between 0.6-1 then the temperature was reduced to 20 $^{\circ}\text{C}$. After 1 hour, a final concentration of 125 μM IPTG was added to induce protein expression and left to shake for 12-18 hours. The cells were harvested by centrifugation for 15 minutes at 6,000 rpm in a swing bucket rotator (Beckman Coulter, JLA 8.1000). The cell pellet was transferred to a 50-ml falcon tube and stored at -80 $^{\circ}\text{C}$.

The pellet was thawed on ice and resuspended in lysis buffer (50 mM Tris-HCl, pH 8.0, 1 M NaCl, 20 % glycerol, 10 mM TCEP) containing cOmpleteTM, EDTA-free Protease Inhibitor Cocktail Tablet (Roche). The cells were lysed on ice by sonication (3 seconds pulse-on, 6 seconds pulse-off at 6 W output) for 10 minutes or until the solution appeared clarified and homogenous. The soluble lysate was obtained by centrifugation at 20,000 g for 30 minutes and incubated with Nickel-NTA agarose resin (Qiagen) at 4 $^{\circ}\text{C}$ for 1 hour to capture His-tagged Cas9. The resin was washed with 20 resin volumes of wash buffer (50 mM Tris-HCl, 1 M NaCl, 20 % glycerol, 5 mM imidazole, 10 mM TCEP pH 8.0 at 4 $^{\circ}\text{C}$). Cas9 was eluted with 50 mM Tris-HCl (pH 8.0), 0.1 M NaCl, 20 % glycerol, 10 mM TCEP

and 500 mM imidazole.

Alternatively, the hexahistidine-tagged Cas9 protein was isolated using an AKTA Pure 25 L FPLC (GE Healthcare Life Sciences) system with a HisTrap HP 5 ml (GE Healthcare Life Sciences) column. After sonicating and spinning down the cellular debris the soluble lysate underwent a second round of sonication and clarification to chew up remaining debris and nucleic acid. This also made it easier to pass the lysate through a 0.45 μm filter. After loading the sample with the S9 sample pump onto the FPLC column, it was washed with lysis buffer until the absorbance was stable (± 0.1 mAu) at the baseline for two minutes or until the total wash volume reached 50 ml. The protein was eluted with a gradient of 0% lysis buffer to 100% elution buffer (50 mM Tris-HCl, 250 mM NaCl, 20 % glycerol, 10 mM TCEP, 300 mM imidazole, pH 8.0) over 12 column volumes. The peak fractions were simultaneously concentrated and buffer exchanged into TEV cleavage buffer (20 mM HEPES-KOH, pH 7, 150 KCl, 10% glycerol, 1 mM DTT, 1 mM EDTA) using a centrifugal filter (10 – 30 kilodalton molecular weight cut-off, Amicon Ultra Millipore). The protein concentration was estimated by measuring the absorbance at 280 nm and diluted to 1 mg·ml⁻¹ with the TEV cleavage buffer and 0.5 mg TEV protease was added per 50 mg of protein. The cleavage progressed overnight or at least 18 hours, with gently mixing at 4 °C. The protein was spun-down at 3,000 x g for about two minutes because slight precipitation was visible. The soluble, filtered Cas9 was loaded onto the FPLC equipped with a the 5-mL HiTrap SP HP column (GE Healthcare)⁶ and washed with buffer A (20 mM HEPES-KOH, 5% glycerol, 100 mM KCl, pH 7, 1 mM DTT). The bound protein

⁶ Alternative FPLC methods were used during the ion exchange purification step. A 5-ml HiTrap Q HP column was placed on top of the cation exchange column and removed after washing out the unbound protein. Alternatively, for cation exchange the Source 15S Tricorn (17 mL column volume, GE Healthcare) resulted in pure protein using a similar run.

eluted using a gradient from 0 to 60% buffer B containing 1 mM KCl over 12 column volumes (60 ml). Cas9 typically eluted by 50% of buffer B. Cas9 elutes out as two prominent peaks and some fractions contained nucleic acids and other contaminants. When the ratio of the absorbance at 260 nm and 280 nm (Abs_{260nm}/Abs_{280nm}) was less than 0.6 then these fractions were pooled. The concentration of protein was determined from the absorbance at 280 nm using the extinction coefficients (ϵ) below. Aliquots were flash frozen in liquid nitrogen and stored at -80 °C. Properties of the Cas9 constructs are as follows:

eGFP-Cas9-NLS	MW 187169.91 Da
	$\epsilon = 142340 \text{ M}^{-1}\text{cm}^{-1}$
Cas9-NLS-eGFP	MW 187169.91 Da
	$\epsilon = 142340 \text{ M}^{-1}\text{cm}^{-1}$
Cas9-NLS	MW 160448 Da
	$\epsilon = 120450 \text{ M}^{-1}\text{cm}^{-1}$

NLS denotes the nucleoplasmin nuclear localization sequence and eGFP is the enhanced green fluorescent protein. The NLS single amino acid code is KRPAATKKAGQAKKKK. It was inserted at the carboxyl-terminus of the *spCas9* gene using standard Phusion site-directed mutagenesis following the manufacturer's instructions (Thermo Fischer). The *eGFP* tag was taken from the pMCSG18 plasmid (courtesy of Dr. R. Pricer) and incorporated into the pMJ806-Cas9 vector with the In-Fusion HD cloning kit following the manufacturer's protocols (Clontech Laboratories Inc., Mountain View, CA).

I.5.2 Preparation and purification of RNA

The sgRNA was prepared following the GeneArt™ Precision gRNA Synthesis Kit (Invitrogen, ThermoFisher Scientific) with modifications. First the DNA template was assembled by PCR assembly to create the gRNA DNA with

the T7 promoter. Primers were designed based-up the variable 18-20 DNA sequence upstream of the PAM.² The forward primer includes the T7 promoter sequence followed by the first 16-20 nucleotides of the target sequence and the reverse primer includes the first 15 nucleotides of the crRNA/tracrRNA constant region DNA sequence followed by 19-20 nucleotides of the target sequence reverse complement. An example of the variable primers is as follows;

gRNA forward 5'- TAATACGACTCACTATAGCAACTGTCGGAGCTTC -3'

gRNA reverse 5'- TTCTAGCTCTAAAACCGTAGAAGCTCCGACAGTT- 3'

The red letters represent the 20-nucleotide target sequence. The primers were ordered as desalted and dry standard DNA (Sigma) on 0.025 μ mole scale and resuspended in sterile, DEPC-treated water to make stocks. The DNA template was generated following the GeneArt™ Precision gRNA Synthesis Kit (Invitrogen, ThermoFisher Scientific). The PCR product didn't require purification.

For the *in vitro* transcription, the reaction was prepared in the following order; 2 μ l nuclease-free water, 8 μ l of NTP mix (100 mM each of ATP, GTP, CTP, UTP), 6 μ l of DNA template (~ 0.5 μ g), 2 μ l of 10x T7 reaction buffer, 2 μ l of T7 RNA Polymerase mix (HiScribe™ T7 High Yield RNA Synthesis Kit, NEB). It incubated for 4 hours at 37 °C and the template DNA was digested with 1 μ l of DNase I (1 U/ μ l, NEB) for 15 minutes.

For the RNA purification, a modified solid-phase reversible immobilization (SPRI) method was developed with Sera-Mag SpeedBeads™ Carboxyl Magnetic Beads (50 mg/ml, Fisher #09-981-123).²⁷ The beads were prepared in advance by washing 100 μ l of beads three times with 1 ml TE buffer (10 mM Tris-HCl, 1 mM EDTA pH 8.0) mixing well after each addition and isolating the beads on a magnetic stand prior to removing the supernatant. After the final wash the beads were resuspended in 100 μ l of TE and stored in 5 ml of a PEG-NaCl solution (10 mM Tris-HCl, pH 8.0, 1 mM EDTA, pH 8.0, 1 M NaCl, 20% (w/v) PEG-8000, 0.05%

Tween 20) in the dark at 4 °C for up to 3 months. For gRNA purification, the T7 RNA reaction was brought up to 150 µl with 70% ethanol and incubated with 100 µl of the modified beads and incubated for 5 to 10 minutes at RT. The beads were collected on a magnetic stand for 2 minutes, and the supernatant was removed and white particulate appeared on the beads. The beads were washed twice with 500 µl of 70% ethanol. The beads were left to air dry for 15 minutes and the RNA was eluted with 20 µl of nuclease-free water. It was then denatured at 95 °C for 5 to 10 minutes and allowed to refold slowly at room temperature for one hour. The RNA was stored at -20°C and the concentration was determined using the absorbance at 260 nm. After one month, the quality of the RNA was checked using the *in vitro* cleavage assay.^{28,29}

1.5.3 *In vitro* cleavage assay

The HEK 293T cells were released with trypsin, quenched with 8 ml of PBS, collected at 200 xg for 2 minutes, and washed twice with 10 ml of PBS. For the DNA extraction, 50 µl of Quick Extraction solution (Epicenter, Madison, WI) per 50,000 to 2 x10⁶ total cells and incubated for 20 minutes at 65 °C, 20 minutes at 95 °C, and stored at -20 °C. The concentration of DNA was determined using Nanodrop (Thermo Fisher Scientific, Waltham, MA). The PCR reaction was set-up by mixing 2 µl of extracted DNA (50 to 150 ng total genomic DNA), 25 µl of AmpliTaq Gold® 360 Master Mix (Applied Biosystems by Life Technologies, Carlsbad, CA), 1 µl GC Enhancer, 1 µl of 10 µM forward primer, 1 µl of 10 µM reverse primer, and water to bring the final volume to 50 µl. The PCR reaction was carried-out in a thermocycler set to run at 95 °C for 5 minutes, 30 cycles (95 °C for 30 seconds, 65 °C for 45 seconds, 72 °C for one minute) and one cycle at 72 °C for 5 minutes. The annealing step varied depending on the melting temperature DNA primers. The PCR reaction was clean-up using a silica-based column following the

manufacturer's protocol (Qiagen, QIAprep® miniprep kits).

For the cleavage assay a 1:1 ratio of Cas9 protein (5 µM) was mixed gently with sgRNA in the presence of 2 mM MgCl₂ and incubated for 10 minutes at room temperature. The cleavage reaction started with the addition of the DNA target (50 - 100 ng) and proceeded for up to one hour at 37 °C and its was quenched quenching with 1 mM EDTA. Cas9 was digested with 1 µl of Proteinase K (New England Biolabs) for 20 minutes at room temperature. The samples were mixed with 6xNEB loading dye (2.5% Ficoll®-400, 11 mM EDTA, 3.3 mM Tris-HCl, 0.017% SDS, 0.015% bromophenol blue pH 8.0 at 25 °C). The control sample without sgRNA and cleavage samples ran on 2% agarose TBE DNA gel containing 0.5 µg/ml of ethidium bromide dye in TBE buffer (Tris-boric acid-EDTA buffer: 100 mM Tris base, 100 mM boric acid, 2.5 mM sodium ethylenediaminetetraacetic acid) for 30 minutes at 120 volts or until the loading dye ran along at least half of the gel. The DNA was visualized under UV light.³⁰

I.6 References

1. Wright, A. V., Nunez, J. K. & Doudna, J. A. Biology and Applications of CRISPR Systems: Harnessing Nature's Toolbox for Genome Engineering. *Cell* **164**, 29-44, (2016).
2. Jinek, M. *et al.* A Programmable Dual-RNA-Guided DNA Endonuclease in Adaptive Bacterial Immunity. *Science* **337**, 816-821, (2012).
3. Mali, P. *et al.* RNA-guided human genome engineering via Cas9. *Science* **339**, 823-826, (2013).
4. Sternberg, S. H. & Doudna, J. A. Expanding the Biologist's Toolkit with CRISPR-Cas9. *Mol Cell* **58**, 568-574, (2015).
5. Cong, L. *et al.* Multiplex Genome Engineering Using CRISPR/Cas Systems.

- Science* **339**, 819-823, (2013).
6. Shalem, O. *et al.* Genome-Scale CRISPR-Cas9 Knockout Screening in Human Cells. *Science* **343**, 84-87, (2014).
 7. Jinek, M. *et al.* Structures of Cas9 endonucleases reveal RNA-mediated conformational activation. *Science* **343**, 1247997, (2014).
 8. Haurwitz, R. E., Jinek, M., Wiedenheft, B., Zhou, K. & Doudna, J. A. Sequence- and structure-specific RNA processing by a CRISPR endonuclease. *Science* **329**, 1355-1358, (2010).
 9. Ran, F. A. *et al.* Genome engineering using the CRISPR-Cas9 system. *Nat. Protocols* **8**, 2281-2308, (2013).
 10. Ran, F. A. *et al.* Double Nicking by RNA-Guided CRISPR Cas9 for Enhanced Genome Editing Specificity. *Cell* **154**, 1380-1389, (2013).
 11. Elliott, B., Richardson, C., Winderbaum, J., Nickoloff, J. A. & Jasin, M. Gene Conversion Tracts from Double-Strand Break Repair in Mammalian Cells. *Mol Cell Biol* **18**, 93-101, (1998).
 12. Zhang, J.-P. *et al.* Efficient precise knockin with a double cut HDR donor after CRISPR/Cas9-mediated double-stranded DNA cleavage. *Genome Biology* **18**, 35, (2017).
 13. Richardson, C. D., Ray, G. J., DeWitt, M. A., Curie, G. L. & Corn, J. E. Enhancing homology-directed genome editing by catalytically active and inactive CRISPR-Cas9 using asymmetric donor DNA. *Nat Biotechnol* **34**, 339-344, (2016).
 14. Yu, C., Zhang, Y., Yao, S. & Wei, Y. A PCR Based Protocol for Detecting Indel Mutations Induced by TALENs and CRISPR/Cas9 in Zebrafish. *PLoS ONE* **9**, e98282, (2014).
 15. Vouillot, L., Thélie, A. & Pollet, N. Comparison of T7E1 and Surveyor Mismatch Cleavage Assays to Detect Mutations Triggered by Engineered

Nucleases. *G3 (Bethesda)* **5**, 407-415, (2015).

16. Liang, X. *et al.* Rapid and highly efficient mammalian cell engineering via Cas9 protein transfection. *J Biotechnol* **208**, 44-53, (2015).
17. Kouranova, E. *et al.* CRISPRs for Optimal Targeting: Delivery of CRISPR Components as DNA, RNA, and Protein into Cultured Cells and Single-Cell Embryos. *Human Gene Therapy* **27**, 464-475, (2016).
18. Gehl, J. Electroporation: theory and methods, perspectives for drug delivery, gene therapy and research. *Acta Physiologica Scandinavica* **177**, 437-447, (2003).
19. Hu, Johnny H., Davis, Kevin M. & Liu, David R. Chemical Biology Approaches to Genome Editing: Understanding, Controlling, and Delivering Programmable Nucleases. *Cell Chem Biol* **23**, 57-73, (2016).
20. Leenay, R. T. *et al.* Identifying and Visualizing Functional PAM Diversity across CRISPR-Cas Systems. *Mol Cell* **62**, 137-147, (2016).
21. Fonfara, I., Richter, H., Bratovič, M., Le Rhun, A. & Charpentier, E. The CRISPR-associated DNA-cleaving enzyme Cpf1 also processes precursor CRISPR RNA. *Nature* **532**, 517-521, (2016).
22. East-Seletsky, A. *et al.* Two distinct RNase activities of CRISPR-C2c2 enable guide-RNA processing and RNA detection. *Nature* **538**, 270-273, (2016).
23. Uphoff, C. C. & Drexler, H. G. in *Basic Cell Culture Protocols* (eds Cheryl D. Helgason & Cindy L. Miller) 1-13 (Humana Press, 2013).
24. Nikfarjam, L. & Farzaneh, P. Prevention and detection of Mycoplasma contamination in cell culture. *Cell J* **13**, 203-212, (2012).
25. Geraghty, R. J. *et al.* Guidelines for the use of cell lines in biomedical research. *Br J Cancer* **111**, 1021-1046, (2014).
26. Drexler, H. G. & Uphoff, C. C. Mycoplasma contamination of cell cultures: Incidence, sources, effects, detection, elimination, prevention.

Cytotechnology **39**, 75-90, (2002).

27. Rohland, N. & Reich, D. Cost-effective, high-throughput DNA sequencing libraries for multiplexed target capture. *Genome Research*, (2012).
28. Anders, C. & Jinek, M. in *Methods Enzymol* Vol. Volume 546 (eds A. Doudna Jennifer & J. Sontheimer Erik) 1-20 (Academic Press, 2014).
29. Lin, S., Staahl, B. T., Alla, R. K. & Doudna, J. A. Enhanced homology-directed human genome engineering by controlled timing of CRISPR/Cas9 delivery. *Elife* **3**, e04766, (2014).
30. Guschin, D. Y. *et al.* A rapid and general assay for monitoring endogenous gene modification. *Methods Mol Biol* **649**, 247-256, (2010).

Appendix II Where bacteria grow, phage strive to be.

II.1 Abstract

For biochemical, genetic, and molecular research laboratories bacteria are manipulated to generate biological products including proteins and DNA. In these settings bacteriophage (from Greek meaning “bacteria eater”) infections are a lurking problem. Bacteriophages outnumber bacteria 10 to 1 and infect approximately 10^{23} to 10^{25} bacteria cells per second in aquatic environments.³⁻⁷ Therefore, any laboratory growing bacteria for research or industrial purposes is an attractive environment for phage. Unfortunately, pathogenic phage that target bacteria such as *Escherichia coli* can be detrimental to the progress in these laboratories especially when the phage infection goes unnoticed.⁸ Even small isolated infections resulting in low-to-no protein expression may lead researchers to spend time optimizing their constructs, expression conditions, or they may examine the toxicity of the recombinant protein. Meanwhile the phage is rapidly multiplying.⁹ The purpose of this communication is to define symptoms of phage-infected bacteria in research laboratories and to recommend preventative and decontamination procedures.

II.2 The physical features of bacteriophages

The basic definition of bacteriophage is a virus that uses bacteria to replicate itself. Phages are extremely small ranging between 50 to 200 nanometers (nm) in length and the width of the head is between 50-90 nm.⁸ Therefore, they are only visualized using specialized techniques such as transmission electron microscopy.³ The majority of phages consist of a head attached to a tail. The head

or capsid protects its genome from the environment. The tail acts as a transport tube to pass its genetic material into the host cell.^{10,11} The phage genomes range in size, structure, and genetic material from either single-stranded or double-stranded DNA or RNA. The genetic material contains genes for transcriptional regulation, replication machinery, lysis and structural capsid proteins. However, this description is grossly abbreviated. Bacteriophages are predicted to be the most abundant organisms ($>10^{30}$ phages) on Earth in diverse environments including the deep sea, acidic hot springs, alkaline lakes, polar lakes, sewage, and the Sahara Desert.¹²⁻¹⁵ As of 2007 $>5,000$ phages have been structurally characterized and approximately 100 new phages are identified each year.¹⁶ Phage and bacteria influence the diversity of each other through antagonistic coevolution.¹⁷ Phages present selective pressures on bacteria. Bacteria mutate or evolve slightly faster than phage so bacteria develop resistance or defense against infection. However, phage counteract and develop mechanisms to circumvent the bacterial defenses.

Bacteria are provided with certain fitness factors from phage.^{1,18,19} For instance, prophage find that bacteria should survive stressors such as antibiotics – it is in their best interest.²⁰ Phage can pass on antibiotic resistance genes through specific transduction.²¹ Transduction affects bacterial genes through a process known as lateral gene transfer (LGT) whereby genetic material is incorporated from an unrelated organism.²² When foreign genetic material enters into a bacterium, the genes are incorporated into the recipient's genome and the expression benefits the bacterium. LGT requires the right conditions to work. The genetic elements risk integrating into essential genes producing inviable bacterium. Lateral gene transfer plays a role in antibiotic resistance, but these elements are often mobile.^{1,23} The genome sizes of prokaryotes are limited; as bacteria acquire genes non-essential genes are lost.²⁴ This also means that bacteria may evade infection from one type of phage, but loose immunity against a

previous type of phage.

II.3 The life cycle(s) of bacteriophages

Phage survive and propagate through six general steps (Figure II-1). For infection most phages target bacteria with receptor binding proteins located on their tails that chemically complement surface receptors such as polysaccharides and lipopolysaccharides on bacteria.^{25,26} Once hitting their target, phages adsorb to

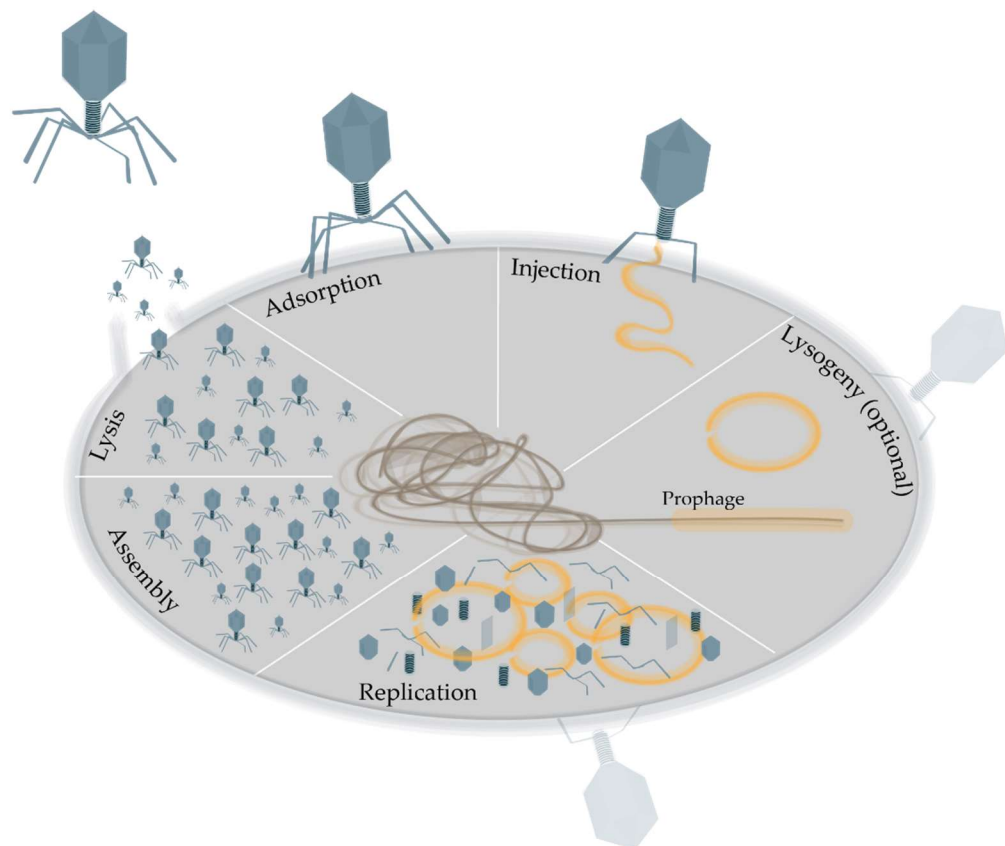


Figure II-1 Phage life cycle is generally described in six steps for phage that are temperate and virulent. 1) Adsorption. Phage recognize their target through receptors 2) Injection. They inject genetic material into the bacterium host. 3) Lysogeny. Depending on the phage type and environment they may conserve their genetic material in a dormant state. 4) Replication. During the lytic cycle phage hijack bacterial transcriptional machinery. 5) Assembly. New phage particles form. 6) Lysis. As they assemble, enzymes breakdown the bacterial cell wall and release new progeny.

host cells, and eject their genomic material into the bacterium.^{27,28} This transduction process is the rate-limiting step for phage infection and so far phage

have lacked the ability to improve upon their strategy. The next step in a phage infection depends on whether the phage is virulent or temperate. For lytic phages, they proceed to produce phage progeny and then escape from the bacteria cell to start new infections. Once the phage genetic material is inserted into the cell, the bacterial machinery is hijacked to replicate phage genomic material and to produce phage proteins. New phage particles assemble and build-up in the cell. Enzymes such as lysozyme degrade the peptidoglycan layer of bacterial cell wall. The proteins responsible for lysis are usually expressed in the late-stage of the infection and eventually cause the cell to burst and to release new phage particles. Filamentous phages (M13, Pf3) are unique because they can escape the bacteria through the cell membrane leaving them intact.²⁹ Phages with complex, double-stranded DNA genomes (>15 genes) such as T1 have holin-endolysin systems.³⁰ Holin is a hydrophobic protein that inserts into the bacterial membrane forming lesions.³¹ Endolysin is a peptidoglycan hydrolase that attacks the cellular wall. In the presence of holin and endolysin bacteria cells lyse within seconds. The burst size (30-300 phages) and latent time (20-100 minutes) required for lysis depends on the type of phage and the environment.^{32,33} For instance, T-odd phages have latency phases from approximately 15 to 45 minutes and burst sizes of 100. The phage fitness depends upon the lysis time and burst size which phage tries to balance. For instance, phage with fast lysis times release few particles. They can potentially infect more bacteria, but only if they efficiently recognize and infect their target. For T1 phage about 1% of the new particles successfully infect a competent bacterial cell.

Temperate phages proceed through a lysogenic cycle. After adsorption, it's genetic material exists as a plasmid or incorporates into the host's chromosome resulting in a prophage. Lysogenic bacteria pass the prophages on to daughter cells. The prophage may lose the ability to excise from the bacterial genome

resulting in cryptic prophages. The cryptic phages are acquired more slowly than the differentiation of bacterial strains because bacteria from the same strain can contain different sequences within prophage sequences. The prophage decay into cryptic phages occurs through mutations within essential genes or introduction of stop codons.¹ Not all phages such as the coliphage T7 generate prophage.

Lysogeny is a dormant state often viewed as a long-term survival mechanism.³⁴ When the bacteria experience adverse conditions including exposure to UV, chemicals, or stress then the dormant phages enter the hostile, lytic cycle. During an initial phage infection, when the bacteria colonies are numerous so the lytic life cycle allows the phage to propagate rapidly. However, as the bacteria population decreases the phage have fewer hosts. Instead of eradicating the bacteria and risking extinction themselves, phage enters the lysogenic life cycle for self-preservation.

The transition from lysogenic to lytic cycles is an important process. Lysogenic phage needs to transition to the lytic cycle so they often hijack the bacterial SOS genes. The SOS genes are responsible for the bacterial response to a crisis such as DNA damage. Lysogenic phage uses SOS to excise itself from the bacterial genome and cell. Recently phages have also been shown to communicate amongst themselves reporting on the infection environment. Specifically, the *Bacillus subtilis* is prone to infection by a phage virus known as phi3T.³⁵ The phi3T phage encodes a protein AimP that is enzymatically cleaved to release a 6-residue peptide fragment. This fragment known as arbitrium is released from the bacteria cell and taken-up by a neighboring bacterium through the oligopeptide permease-transporter protein. As the levels of arbitrium increase in bacteria then the

probability of the phage entering the lysogenic cell cycle increases.³⁵

II.4 How to identifying a phage infection?

Of course, good laboratory hygiene is the best preventative against phage infection. Bacteria cultures should be conscientiously contained, bleached, and



Figure II-2 Preventing phage infections Above is an example where flasks and tubes containing residual bacteria cultures have been left in a sink for an extended period in a high-risk area due to past known phage-infected bacteria cultures.

promptly disposed of (Figure II-2). The most common sources of contamination include pieces of equipment that contact bacterial cultures at any time including such as pipettes, incubators, filters, room aeration systems, buffers, culture flasks, and electroporation devices. Phage -infections are observed more frequently in warm, humid climates rather than in dry, cold ones. In areas with significant rain, airborne pathogens are grounded so phages infections are less common in rainy

environments.⁸ Even with good hygiene phage infections occur. They depend upon bacteria to propagate so it takes only one phage particle to create a widespread infection.³⁶ Extensive and chronic phage infections result in slower bacterial growth rates and significant lysis of cells. If the population of infected bacteria cells is low, then the bacteria cultures grow at an expected rate and appear normal. As the cells lyse and release latent phage, the culture rapidly turns clear due to the bacterial cell lysis. Even if only a few bacteria cells are infected the percentage can rapidly increase to 100% within 30 minutes depending on the nature of the infection.⁸ A distinctive foam on cultures develops due to the presence of bacteria proteins. A healthy, aerated bacterial culture also develops foam; however, the foam dissipates after the culture rests for a few minutes. Phage-infected bacteria cultures also decrease their production of carbon dioxide, and maintain constant alkaline pH rather than decreasing in pH.⁹ Sting-like material is visible in the culture. When infected cultures are harvested the bacterial pellets appear small, loose, and sloppy (Figure II-3). Some infections may result in minimal lysis, but disrupt the metabolism of bacteria causing slower growth rates and decreasing protein production. With some prophages the bacteria grow more slowly and decrease production of desired materials. These infections are problematic because they are difficult to attribute to phage infections. Researchers may question the health of their competent bacteria stocks or the protein-expression plasmid constructs to the slow growth and to the low protein production, respectively.

Alternatively, phage may provide bacteria with additional fitness factors and the lysogenic bacteria grows faster. On solid agar plates, bacterial lawns may form phage plaques.^{37,38} Although bacteria may harbor phage visible formation of plaques is depends on the type of phage, the adsorption rate, lysis timing, and burst sizes.³⁹⁻⁴¹

II.5 Cleaning-up after phage

In general phage is difficult to destroy. The capsid proteins protect the genomic material and even without the capsid the genomic material can infect competent *E coli* cells. Phage range in structure and genomic material so some phages are easier to eradicate than others. Phage such as ecoliphage T1 survive



Figure II-3 Indicators of a phage infection The Rosetta™ (DE3) pLysS *E. coli* (Novagen, Merck Millipore) cells were grown in 1 L of terrific broth for 24 hours and collected at 7,900 xg for 15 minutes. The cellular debris sloshing off the side of the tube and from the pellet indicates lysis due to phage-infected bacteria

many years in dry, hot environments. Working with bacteria next to a Bunsen burner is controversial as a sterile technique to prevent phage infections. Some phages survive the brief exposure to heat and the updraft from the burner makes phage such as desiccated T1 airborne. Ultra-violet light irradiation is the most powerful method to destroy the phage's genetic material. Otherwise heat treatment (200 °C for at least 8 hours) is an effective treatment to destroy common phages.

As previously, mentioned phages are extremely small. Therefore, sterile filtering liquids through a 0.2 µm pore-size filters aren't effective methods to eliminate phage from media and other biological solutions.⁸ These filters are useful to separate phage from bacteria which can't pass through these filters.

Disinfectants can destroy phage by penetrating the capsid and damaging the genomic material, however they should be used with caution. Since phages vary in structure and genomic material their sensitivity to disinfectants vary.⁴² Some phages are prone to aggregation so a fraction of them survive chemical treatments. Basically, any biocide needs to be tested for its effectiveness against the phage-of-interest. For instance, tailed phages from the *Myoviridae* and *Siphoviridae* family containing double-stranded DNA ranging in 50 kbp or 25-170 kb, respectfully range in their susceptibility to disinfectants. Treatment of 75% ethanol is effective against *Myoviridae*-type phage, but harmless against *Siphoviridae* phage.⁴³ The chlorine-releasing disinfectant agent, sodium hypochlorite at 25000 ppm is ineffective at destroying either *Myoviridae* or *Siphoviridae*. However, sodium hypochlorite does appear to target *Pseudomonas aeruginosa* PAO1 phage F116, *Lactobacillus casei* and *Lactobacillus paracasei* phages at <5ppm or 800 ppm, respectfully. Virkon (1%) rapidly inactivated *Myoviridae* and *Siphoviridae* upon contact and destroyed them within 10 minutes.⁴³ Virkon is the

most effective disinfectant for broad-spectrum deactivation of viruses.

Phage infections may also be inhibited with chemical treatments.^{44,45} Vitamins including vitamin E and K at 10^{-5} M has shown to provide >90% inactivation. Ascorbic acid is also selectively toxic to phage, but not bacteria by forming single-stranded breaks in DNA-containing phage such as coliphage T7 after 20 minutes of exposure.^{46,47}

Since phages are limited to recognizing features on the membrane of bacteria for infection bacteria strains may be varied to prevent pathogenic phage from taking over production. Dairy industries often vary the type of bacteria used to make cheese and other cultures.⁴⁸ Therefore, phage with specific receptor binding proteins are unable to infect the new type of bacteria. For research laboratories, strains of bacteria are available that harbor deletions of outer membrane proteins recognized by T1 and T5 phages including deletions of *FhuA* or *tonA*. The *FhuA* deletions confers resistance to T1, T5, and $\phi 80$. For instance, T1, T5, and $\phi 80$ coliphages recognize *FhuA*, an outer membrane ferrichrome iron receptor of *E. coli* to adsorb to the host so without this receptor phage cannot bind to *E. coli*. The *tsx* deletions confers resistance against phage T6 and colicin K. These bacteria have been genotyped and tested so they are arguably well-suited for industrial and research purposes.⁴⁹

Very dense bacteria cultures likely contain genetic heterogeneity due to bacteria mutation rates. At least endogenous bacteria mutation rates are faster than phage so phage must adapt to bacterial defenses. Mutant bacteria propagate and during a phage infection they may survive. Bacteria survivors often contain a point mutation on the cell wall or in structural proteins which disrupt phage's ability to recognize and to infect them. These mutations are often random.⁵⁰

These viable mutants may be used to prevent further phage infection, however with unknown mutation(s) the bacteria may have other unknown and

potentially undesirable properties. This solution also tends to only work temporarily. Although bacteria tend to evolve more quickly, phage also evolve with their tail fiber genes adapting with highly variable genes and higher mutation rates.⁵¹⁻⁵³

Although this strain should be confirmed by genotyping through pulse field gel electrophoresis (PFGE) with restriction enzymes to confirm that the strain has not undergone major deletion or rearrangement in comparison to the parent strain.^{54,55} Bacteria that come into contact with weakened phage particles by exposing the phage to UV can defend themselves. They either incorporate segments of the phage genome to induce their immune-response CRISPR/Cas systems. The number of bacteria with these immune responses are limited. They also divert energy from producing the desired proteins and biomaterials in high quantity.

Other measures have been developed to prevent phage infections. Bacterial strains harboring temperate, lytic phages prevent infection because bacteria infected with a lysogenic phage are resistant to infection by the same phage. Prophages have superinfection exclusion systems whereby proteins are anchored or associated with the bacterial membrane to prevent other phage components from accessing the bacterium. However, the protein production and growth rates of these prophage-containing bacteria remain compromised. Although some *E. coli* containing prophages may grow more quickly than non-lysogenic *E. coli*. Also, bacterial growth conditions require optimization to prevent the development of lytic phage. Growing bacteria at 37 °C means their growth rates are high (~2 doublings per hour) during an exponential growth phase. Slower growth at 20 °C is less likely to induce lysogenic phage into the lytic cycle. Slower bacterial growth also reduces the production of lytic phage.

In small laboratories containing phage infections should be straight-

forward if protocols are established and followed to disinfect areas that bacteria may encounter. In contrast a phage infection can be catastrophic to the production at large bioprocessing plant. A phage-based detection assay may aid in detecting small phage infections and preventing the spread of infections. Although some methods require the use of the bacterial host to propagate the virus.

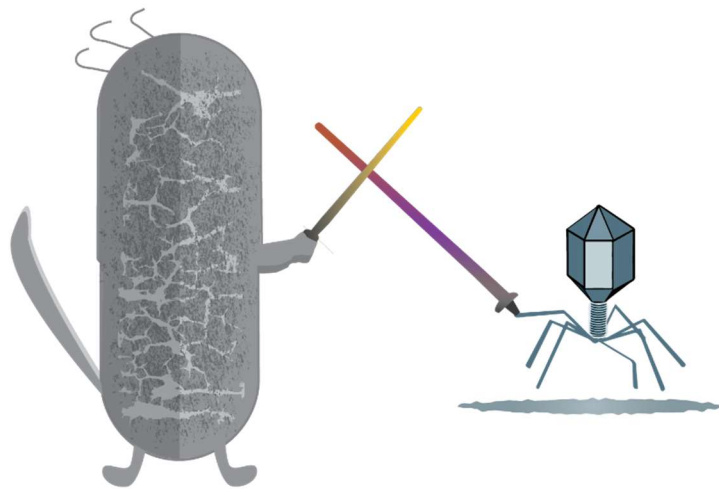


Figure II-4. Bacteria are infected by phage, but they adapt strategies to fight off phage infection. Meanwhile phage develop mechanisms to circumvent bacterial defenses. The antagonistic coevolution relationship between phage and bacteria is referred to as the phage-host arms race which started > 3 billion years ago.^{1,2}

II.6 The good qualities of bacteriophage

The phage and bacteria arms race is the oldest and, in my opinion, a fascinating antagonist co-evolution story which is in constant development (Figure II-4). Despite detesting the phage's disruption on bacteria, the coexistence of bacteria and phage has greatly improved our environment by impacting areas such as pharmacy, biotechnology, food production, and genome engineering.

Biotherapy and human health. Phage were first observed by Frederick Twort when he observed that an unknown agent was disrupting bacterial growth. Then Felix d'Herelle in 1915 reported on an "anti-microbe" that was destroying *dysentery bacillus* in patients.⁵⁶ Early work with phage looked at its potential as a

therapy; however, phage therapy posed efficacy, selection, and immunogenic challenges. Meanwhile broad spectrum antibiotics were developed that eradicated most pathogenic bacteria. Although phage therapy has been used in Eastern Europe, other Western countries remain skeptical. They are after all viruses and pose many challenges and risks. For instance, virulent phages trigger lateral gene transfer and may release toxins triggering autoimmune responses. With the rise of antibiotic resistant infections, the development of phage-based therapies has revitalized with new phages being identified and tested.⁵⁷⁻⁵⁹ Phages kill one species or a strain of bacteria limiting their effectiveness.⁶⁰ By modifying phage therapy approaches, phage strategies have the potential to kill or to weaken antibiotic-resistant bacteria.⁵⁹ For instance, the phage ϕ APCEc03 targets a broad range of *E. coli* strains infecting 9 out of the 16 tested strains. Phage ϕ APCEc0 targets two bacterial genera including three *E. coli* strains and a *Shigella sonnei* strain, responsible for diarrheal diseases. A phage cocktail including ϕ APCEc03 and ϕ APCEc0 isolated from human fecal samples targets a broader spectrum of bacteria and completely disturbs *E. coli* growth in a biofilm by 4-fold after 24 hours. Phage-based therapies have been inefficient because bacteria survive the infection likely due to bacteriophage-insensitive mutants (BIMs). After 48 hours *E. coli* began to recover in the presence of phage cocktail such as ϕ APCEc03 and ϕ APCEc0.⁶¹ In this case combining the ciprofloxacin antibiotic the phage cocktail efficiently reduces *E. coli* growth and limits the emergence of antibiotic- and phage-resistant cells. Over the last few years the emergence of novel antibiotics phage therapy has the potential to increase the effectiveness of the existing antibiotics. In another case the vancomycin-resistant enterococci bacteria are difficult to control in hospitals, but this bacterium is selectively targeted by phage. However, pathogenic bacteria also become resistant to the phage overtime. An effective antisense therapy efficiently targets vancomycin-resistant infections by

targeting vancomycin-resistant genes in bacteria.⁶⁰ The treatment is difficult to deliver. Phage provide a structure to deliver to carry genetic material into bacteria. These synthetic biology approaches would use phage-based treatments as a supplement to antibiotics. The phage constructs target or deliver bacterial factors that enhance an antibiotic's effectiveness and prevents antibiotic-resistance. These therapies require additional improvements such as producing non-replicative phage to prevent the release of toxins and autoimmune responses. The selectivity and delivery may also become more effective by displaying various phage tail proteins.⁶² Finally, these therapies require testing different strains and bacterial gene targets before applying them as human-based strategies.⁶⁰

The co-existence of phage and bacteria is also important to the human microbiota. The healthy human microbiome contains up to 10^{13} bacteria per gram and consequently phage effect the bacterial environment either negatively by challenging bacteria or positively by providing bacteria with fitness factors. Healthy adults contain the well-known *Siphoviridae*, *Myoviridae*, *Podoviridae*, and *Microviridae* as well as many diverse, and novel viruses.⁶³ Diet and environment form the diverse phageome. The microbiota starts developing at birth with the mother.

Differences in the microbiota exist in patients with chronic diseases such as Crohn's disease, irritable bowel syndrome and ulcerative colitis. The phageome differs dramatically between healthy and unhealthy individuals. Crohn's disease patients have a 24-fold higher proportion of virion-like particles than healthy patients in mucosa biopsies. The phage varies between locations too so the ileum, colon, and bowel contain different populations. Probiotics and diets promote a healthy microbiota, but not everyone responds to this treatment. Selective pressures exerted by natural occurring phages may be the missing factor for these treatments. However the phageome's effect on human health is still under

investigation.⁶¹

Diagnosics. The phage-host system is rapid and sensitive so they are exploited to detect the presence of bacterium for diagnostic purposes.⁶⁴ These assays provide the presence and identity of bacteria. For instance, phage is manipulated so during an infection they are biotinylated *in vivo*. If the streptavidin-coated beads capture biotinylated-phage particles then the specific bacterium is present in a complex sample are detected by isolating the phage particles on streptavidin-coated beads. This platform is highly efficient since the rapid phage bursts increase the limit detection (10^2 CFU ml⁻¹ with the model T7 and *E coli*).⁶⁵ Typical phage-system assays for bacteria use plaque assays involving bacterial isolation and cultivation that require 24-36 hours to complete.

These diagnostic tools are especially useful to detect possible attractive bioterrorism agents. Category A bacterial pathogens include *Yersinia pestis* and *Bacillus anthracis* undergo a rapid clinical course with high mortality rates. The *B. anthracis* spores are inhaled or ingested by contaminated food or water supply. Anthrax symptoms are similar to other ailments, but the disease progresses rapidly. If not treated within 24 hours, it is fatal. *Y. pestis* comes in different forms; classical bubonic (lymph nodes), pneumonic (lungs) and septicemic (blood vessels).⁶⁶ Recombinant reporter phage assays provide rapid and specific approaches to potentially save lives. A recombinant "light-tagged" reporter phage assay uses phage-specific for *Y. pestis* and *B. anthracis* with integrated *vibrio haveyi luxAB* genes into their genome. When the reporter phage recognizes the target, it confers a bioluminescent phenotype to recipient cells. The signal is rapidly measured in a microplate luminometer and by using 96-well microtiter plate the assay becomes high-throughput.⁶⁷ This assay is also used to detect whether an antibiotic effectively kill the bacterium.⁶⁸

Phage display applications. Filamentous phage (M13 or fd) and T4, T7,

and lambda phage inspired the development a biotechnique known as phage display.⁶⁹ Variable peptides or protein epitopes are expressed on the viron's surface by inserting the corresponding DNA sequence into the phage's genome. Then these phage particles are screened against a target protein. The nonspecific binders are washed away and the selected markers are identified by amplifying the marker region. Drug discovery, antigen delivery, biomarkers, epitope regions and other applications have advanced using phage display. Most notably drugs have been successfully developed using phage display including Humulin®(Lilly) for diabetes, Vasotec®(enalapril maleate, Merck Shar & Dohme) for hypertension, and Prialt® (ziconotide acetate, Elan Pharms) for severe chronic pain.⁶⁹

Food production. Unlike the biological research and industries where phage wreak havoc on bacterial-based production bacteria can disrupt food production at any stage from the farm to the consumer. These bacteria appear in livestock, on equipment and surfaces, and on raw meats and fresh produce. Consumers are at risk of exposure to more than 200 known diseases of which many are attributed to bacteria including *Salmonella*, *Campylobacter*, pathogenic *E. coli* and *Listeria monocytogenes*.⁷⁰ Different products have been approved by the FDA that contain natural phage cocktails to target harmful bacteria. For instance, in 2005 Agriphage™ (OmniLytics Inc.) treats bacterial spot caused by *Xanthomonas campestris* or bacterial speck caused by *Pseudomonas syringae* on crops. In 2006 ListShield™ was formally approved to prevent *L. monocytogenes* in certain foods containing meat and poultry. In 2011 EcoShield™ was first used to target *E. coli* O157:H7 on red meat began before it was ground into hamburgers. SalmoFresh™ can be applied to poultry products and other foods to eliminate *Salmonella enterica*. Although “general designated as safe” the concern remains that resistant strains emerge over time so phage cocktails are varied and modified in an effort to

proactively address potential resistant bacteria.

Genome engineering. Bacteria contain a unique immune system to fight off phage. A distinct clustered regularly interspaced short palindromic repeats (CRISPRs) loci were discovered in *E. coli* along with CRISPR-associated (Cas) proteins. The CRISPR/cas family is diverse. Mainly a bacterium acquires new spacers that match genetic regions of foreign phage or plasmid DNA. The spacers are clustered by 21 to 48-bp DNA repeats of similar length. Then upon another infection the CRISPR/Cas system recognizes the invading DNA or RNA sequence and destroys it with an endonuclease. The Cas9 proteins are highly selective and efficient at targeting DNA sequences so they present a powerful opportunity to manipulate biological systems at the DNA level.⁷¹

The CRISPR/Cas genome engineering applications was first applied by the dairy industry where they need to prevent phage-induced destruction of their products such as cheese and yogurt. Bacteriophage insensitive mutants (BIMs) from the *Streptococcus thermophiles* lactic acid bacteria are analyzed to replace existing bacteria strains and to prevent a hostile phage takeover in the cheese and yogurt production. Naturally occurring BIMs contain CRISPR loci as a defense against phage and continue to be used to aid the dairy industry. Since the CRISPR/cas system has fascinated the scientific community in the last five years, the applications and potential for this system have been already extensively reviewed.

II.7 Conclusion

In some laboratories phage is hated and its presence force researchers to clean-up or find other ways to combat its' bacteria-killing nature. Phage and bacteria have co-existed for a long time and neither is likely to become extinct. Therefore, we should embrace them and learn how to live around them because

we have certainly lived a better life with both of them.

II.8 References

- 1 Brussow, H., Canchaya, C. & Hardt, W. D. Phages and the evolution of bacterial pathogens: from genomic rearrangements to lysogenic conversion. *Microbiol Mol Biol Rev* **68**, 560-602, table of contents, (2004).
- 2 Pedulla, M. L. *et al.* Origins of Highly Mosaic Mycobacteriophage Genomes. *Cell* **113**, 171-182, (2003).
- 3 Brussow, H. & Hendrix, R. W. Phage genomics: small is beautiful. *Cell* **108**, 13-16 (2002).
- 4 Whitman, W. B., Coleman, D. C. & Wiebe, W. J. Prokaryotes: The unseen majority *Proc Natl Acad Sci USA* **95**, 6578-6583 (1998).
- 5 Hendrix, R. W. Bacteriophages: evolution of the majority. *Theor Popul Biol* **61**, 471-480 (2002).
- 6 Suttle, C. A. Marine viruses--major players in the global ecosystem. *Nat Rev Microbiol* **5**, 801-812, (2007).
- 7 Breitbart, M. & Rohwer, F. Here a virus, there a virus, everywhere the same virus? *Trends Microbiol* **13**, 278-284, (2005).
- 8 Primrose, S. B. in *Applied Molecular Genetics* 1-10 (Springer Berlin Heidelberg, 1990).
- 9 Jones, D. T., Shirley, M., Wu, X. & Keis, S. Bacteriophage infections in the industrial acetone butanol (AB) fermentation process. *J Mol Microbiol Biotechnol* **2**, 21-26 (2000).
- 10 Eiserling, F. A. in *Comprehensive Virology Volume 13: Structure and Assembly: Primary, Secondary, Tertiary, and Quaternary Structures* (eds Heinz Fraenkel-Conrat & Robert R. Wagner) 543-580 (Springer US, 1979).
- 11 Samson, J. E., Magadan, A. H., Sabri, M. & Moineau, S. Revenge of the

- phages: defeating bacterial defences. *Nat Rev Micro* **11**, 675-687, (2013).
- 12 Weinbauer, M. G. Ecology of prokaryotic viruses. *FEMS Microbiol Rev* **28**, 127-181, (2004).
 - 13 Prigent, M., Leroy, M., Confalonieri, F., Dutertre, M. & DuBow, M. S. A diversity of bacteriophage forms and genomes can be isolated from the surface sands of the Sahara Desert. *Extremophiles* **9**, 289-296, (2005).
 - 14 Rice, G. *et al.* Viruses from extreme thermal environments. *Proc Natl Acad Sci USA* **98**, 13341-13345, (2001).
 - 15 Yau, S. *et al.* Virophage control of antarctic algal host-virus dynamics. *Proc Natl Acad Sci USA* **108**, 6163-6168, (2011).
 - 16 Ackermann, H.-W. 5500 Phages examined in the electron microscope. *Archives of Virology* **152**, 227-243, (2007).
 - 17 Hambly, E. & Suttle, C. A. The virosphere, diversity, and genetic exchange within phage communities. *Curr Opin Microbiol* **8**, 444-450, (2005).
 - 18 Weitz, J. S. *et al.* Phage-bacteria infection networks. *Trends Microbiol* **21**, 82-91, (2013).
 - 19 Weinbauer, M. G. & Rassoulzadegan, F. Are viruses driving microbial diversification and diversity? *Environ Microbiol* **6**, 1-11 (2004).
 - 20 Hargreaves, K. R., Kropinski, A. M. & Clokie, M. R. J. What Does the Talking?: Quorum Sensing Signalling Genes Discovered in a Bacteriophage Genome. *PLOS ONE* **9**, e85131, (2014).
 - 21 Wagner, P. L. & Waldor, M. K. Bacteriophage Control of Bacterial Virulence. *Infection and Immunity* **70**, 3985-3993, (2002).
 - 22 Keen, E. C. *et al.* Novel "Superspreader" Bacteriophages Promote Horizontal Gene Transfer by Transformation. *MBio* **8**, (2017).
 - 23 Williams, H. T. Phage-induced diversification improves host evolvability.

BMC Evol Biol **13**, 17, (2013).

- 24 Ochman, H., Lawrence, J. G. & Groisman, E. A. Lateral gene transfer and the nature of bacterial innovation. *Nature* **405**, 299-304 (2000).
- 25 Rakhuba, D. V., Kolomiets, E. I., Dey, E. S. & Novik, G. I. Bacteriophage receptors, mechanisms of phage adsorption and penetration into host cell. *Pol J Microbiol* **59**, 145-155 (2010).
- 26 Luria, S. E. Mutations of bacterial viruses affecting their host range. *Genetics* **30** (1945).
- 27 Drulis-Kawa, Z., Majkowska-Skrobek, G., Maciejewska, B., Delattre, A.-S. & Lavigne, R. Learning from Bacteriophages - Advantages and Limitations of Phage and Phage-Encoded Protein Applications. *Curr Protein Pept* **13**, 699-722, (2012).
- 28 Maluf, N. K. & Feiss, M. Virus DNA translocation: progress towards a first ascent of Mount Pretty Difficult. *Mol Microbiol* **61**, 1-4, (2006).
- 29 Rakonjac, J., Bennett, N. J., Spagnuolo, J., Gagic, D. & Russel, M. Filamentous bacteriophage: biology, phage display and nanotechnology applications. *Curr Issues Mol Biol* **13**, 51-76 (2011).
- 30 Roberts, M. D., Martin, N. L. & Kropinski, A. M. The genome and proteome of coliphage T1. *Virology* **318**, 245-266, (2004).
- 31 Loessner, M. J. Bacteriophage endolysins — current state of research and applications. *Curr Opin Microbiol* **8**, 480-487, (2005).
- 32 Shao, Y. & Wang, I. N. Bacteriophage adsorption rate and optimal lysis time. *Genetics* **180**, 471-482, (2008).
- 33 Wang, I.-N. Lysis Timing and Bacteriophage Fitness. *Genetics* **172**, 17-26, (2006).
- 34 Stewart, F. M. & Levin, B. R. The population biology of bacterial viruses:

- Why be temperate. *Theo Popul Biol* **26**, 93-117, (1984).
- 35 Erez, Z. *et al.* Communication between viruses guides lysis–lysogeny decisions. *Nature* **541**, 488-493, (2017).
- 36 Kleczkowski, A. & Kleczkowski, J. The ability of single phage particles to form plaques and to multiply in liquid cultures. *J Gen Microbiol* **5**, (1951).
- 37 Kaplan, D. A., Naumovski, L., Rothschild, B. & Collier, R. J. Appendix: a model of plaque formation. *Gene* **13**, (1981).
- 38 Krone, S. M. & Abedon, S. T. in *Bacteriophage Ecology* (ed S. T. Abedon) (Cambridge University Press, 2008).
- 39 Gallet, R., Kannyo, S. & Wang, I.-N. Effects of bacteriophage traits on plaque formation. *BMC Microbiology* **11**, 1-16, (2011).
- 40 Hyman, P. & Abedon, S. T. Practical methods for determining phage growth parameters. *Methods Mol Biol* **501**, 175-202, (2009).
- 41 Łoś, J. M., Golec, P., Węgrzyn, G., Węgrzyn, A. & Łoś, M. Simple Method for Plating Escherichia coli Bacteriophages Forming Very Small Plaques or No Plaques under Standard Conditions. *Applied Microbiology* **74**, 5113-5120, (2008).
- 42 Maillard, J.-Y. in *Russell, Hugo & Ayliffe's Principles and Practice of Disinfection, Preservation & Sterilization* 272-323 (Blackwell Publishing Ltd, 2008).
- 43 Halfhide, D. E., Gannon, B. W., Hayes, C. M. & Roe, J. M. Wide variation in effectiveness of laboratory disinfectants against bacteriophages. *Lett Appl Microbiol* **47**, 608-612, x (2008).
- 44 Murata, A. & Kitagawa, K. Mechanism of Inactivation of Bacteriophage J1 by Ascorbic Acid. *Biosci Biotechnol Biochem* **37**, 1145-1151, (1973).
- 45 Richter, H. E. & Loewen, P. C. Rapid inactivation of bacteriophage T7 by

ascorbic acid is repairable. *Biochim Biophys Acta* **697**, 25-30 (1982).

- 46 Lho, I. H. *et al.* Phage-inactivating Effect of Iron(II)-Ascorbate Complex. *Biosci Biotechnol Biochem* **56**, 720-724, (1992).
- 47 Murata, A., Suenaga, H., Hideshima, S., Tanaka, Y. & Kato, F. Hydroxyl Radical as the Reactive Species in the Inactivation of Phages by Ascorbic Acid. *Biosci Biotechnol Biochem* **50**, 1481-1487, (1986).
- 48 Marcó, M. B., Moineau, S. & Quiberoni, A. Bacteriophages and dairy fermentations. *Bacteriophage* **2**, 149-158, (2012).
- 49 Pedruzzi, I., Rosenbusch, J. P. & Locher, K. P. Inactivation in vitro of the Escherichia coli outer membrane protein FhuA by a phage T5-encoded lipoprotein. *FEMS Microbiol Letts* **168**, 119-125, (2006).
- 50 Cuppels, D. A., Vidaver, A. K. & Van Etten, J. L. Resistance to Bacteriophage ϕ 6 by Pseudomonas phaseolicola. *J Gen Virol* **44**, 493-504, (1979).
- 51 Gallet, R., Shao, Y. & Wang, I. N. High adsorption rate is detrimental to bacteriophage fitness in a biofilm-like environment. *BMC Evol Biol* **9**, (2009).
- 52 Scanlan, P. D., Hall, A. R., Lopez-Pascua, L. D. C. & Buckling, A. Genetic basis of infectivity evolution in a bacteriophage. *Molecular Ecology* **20**, 981-989, (2011).
- 53 Paterson, S. *et al.* Antagonistic coevolution accelerates molecular evolution. *Nature* **464**, 275-278, (2010).
- 54 Lingohr, E., Frost, S. & Johnson, R. P. Determination of bacteriophage genome size by pulsed-field gel electrophoresis. *Methods Mol Biol* **502**, 19-25, (2009).
- 55 Serwer, P. *et al.* Pulsed field agarose gel electrophoresis in the study of morphogenesis: packaging of double-stranded DNA in the capsids of bacteriophages. *Electrophoresis* **14**, 271-277 (1993).
- 56 Duckworth, D. H. "Who discovered bacteriophage?". *Bacteriol Rev* **40**, 793-

802 (1976).

- 57 Górski, A., Targońska, M., Borysowski, J. & Weber-Dąbrowska, B. The Potential of Phage Therapy in Bacterial Infections of the Eye. *Ophthalmologica* **223**, 162-165 (2009).
- 58 Kingwell, K. Bacteriophage therapies re-enter clinical trials. *Nat Rev Drug Discov* **14**, 515-516, (2015).
- 59 Merrill, C. R., Scholl, D. & Adhya, S. L. The prospect for bacteriophage therapy in Western medicine. *Nat Rev Drug Discov* **2**, 489-497 (2003).
- 60 Lu, T. K. & Collins, J. J. Engineered bacteriophage targeting gene networks as adjuvants for antibiotic therapy. *Proc Natl Acad Sci USA* **106**, 4629-4634, (2009).
- 61 Dalmaso, M. *et al.* Three New Escherichia coli Phages from the Human Gut Show Promising Potential for Phage Therapy. *PLoS One* **11**, e0156773, (2016).
- 62 Yosef, I., Goren, M. G., Globus, R., Molshanski-Mor, S. & Qimron, U. Extending the Host Range of Bacteriophage Particles for DNA Transduction. *Cell* **66**, 721-728.e723, (2017).
- 63 Mills, S. *et al.* Movers and shakers. *Gut Microbes* **4**, 4-16, (2013).
- 64 van der Merwe, R. G., van Helden, P. D., Warren, R. M., Sampson, S. L. & Gey van Pittius, N. C. Phage-based detection of bacterial pathogens. *Analyst* **139**, 2617-2626, (2014).
- 65 Wang, Z., Wang, D., Chen, J., Sela, D. A. & Nugen, S. R. Development of a novel bacteriophage based biomagnetic separation method as an aid for sensitive detection of viable Escherichia coli. *Analyst* **141**, 1009-1016, (2016).
- 66 Greenfield, R. A. & Bronze, M. S. Prevention and treatment of bacterial diseases caused by bacterial bioterrorism threat agents. *Drug Discov Today*

- 8, 881-888, (2003).
- 67 Schofield, D. A., Molineux, I. J. & Westwater, C. 'Bioluminescent' reporter phage for the detection of Category A bacterial pathogens. *J Vis Exp*, e2740, (2011).
- 68 Schofield, D. A. *et al.* Bacillus anthracis diagnostic detection and rapid antibiotic susceptibility determination using 'bioluminescent' reporter phage. *J Microbiol Methods* **95**, 156-161, (2013).
- 69 Hamzeh-Mivehroud, M., Alizadeh, A. A., Morris, M. B., Bret Church, W. & Dastmalchi, S. Phage display as a technology delivering on the promise of peptide drug discovery. *Drug Discov Today* **18**, 1144-1157, (2013).
- 70 Sulakvelidze, A. Using lytic bacteriophages to eliminate or significantly reduce contamination of food by foodborne bacterial pathogens. *J Sci Food Agric* **93**, (2013).
- 71 Deveau, H. *et al.* Phage Response to CRISPR-Encoded Resistance in *Streptococcus thermophilus*. *J Bacteriol* **190**, 1390-1400, (2008).