

Innovations in Targeting Dynamic Proteins with Small Molecule Modulators

by

Julie M. Garlick

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Chemistry)
in the University of Michigan
2021

Doctoral Committee:

Professor Anna K. Mapp, Chair
Professor Ryan C. Bailey
Assistant Professor Kristin S. Koutmou
Professor David H. Sherman

Julie M. Garlick

garlickj@umich.edu

ORCID iD: 0000-0001-8903-3471

© Julie M. Garlick 2021

Acknowledgements

This past year has been a whirlwind. While I never imagined the final year of my PhD experience would be during a pandemic, reflecting now on all of my time at Michigan, it has really been an incredible experience. During my PhD, I have grown so much as a person and a scientist, and that is in part due to the extensive network of people who have supported, encouraged, and inspired me to get to where I am today.

First, I have to thank my advisor, Professor Anna Mapp. I truly look up to her as a scientist and as a person. Her mentorship and support have been pivotal, and influenced me in the most positive ways. I learned so much from Anna that I know will continue to help me throughout my career. She helped me develop my confidence as a scientist, allowing me to try new experiments and encouraging me to take my research into creative directions. Anna truly cares about the well-being of her graduate students, both personally and scientifically, and I feel fortunate to have been her student.

I also would like to thank my committee: Professor Ryan Bailey, Professor David Sherman and Professor Kristin Koutmou, for their feedback and guidance over these years. I can say from the time you agreed to be my dissertation committee, years ago when I was still a precandidate, I have always looked forward to our meetings. It was very enjoyable, dare I say fun, discussing my research with you, and I learned an incredible amount from our conversations.

My first experience with chemical biology was in Dr. Jordan Meier's lab at the National Cancer Institute. Before my experience as a postbac in the Meier lab, I really wasn't sure what career path I was going to take. His mentorship, along with the mentorship of the incredible postdocs in his lab, were what influenced me to pursue a PhD and develop a passion for research. Jordan encouraged me to apply to the University of Michigan, so I literally would not be here without his guidance!

A huge amount of gratitude goes to my mentor when I joined the Mapp lab, Dr. Steven Sturlis. I really can't thank you enough for guiding young Julie through some of the most challenging years of being a graduate student - I think you probably helped me revise my candidacy

exam 20 plus times. You were always happy to talk random science, help me analyze data, teach me experiments, tell me to go to the dentist, talk about life, give advice, and the list just goes on. You were an awesome lab manager, phenomenal mentor, and great friend. And of course, thanks for norstictic acid. My second chapter wouldn't be possible without all your hard work.

Thank you to everyone in the Mapp lab for all of the great memories, and an extra thanks to everyone who let me be your softball captain: Ballsports and Benjamin's, forever champions! I know this past year with all the regulations and shift work it hasn't quite been the same, but being lab mates with you all has been a true joy. Speaking of "joy", thanks to Dr. Stephen Joy. Stephen has been a great lab manager these past few years, and has helped me countless times whether it be lending peptides for experiments or troubleshooting wonky equipment. Thank you to all the Mapp lab alumni that I had the privilege of overlapping with and contributed to my growth as a scientist, particularly Dr. Matt Beyersdorf, Dr. Rachel Pricer, Dr. Meg Breen, and Dr. Clint Regan. I saved one more Mapp lab shout out for last - my fantastic friend and colleague Amanda Peiffer. You are such an amazing scientist and I couldn't imagine a better way to top off a PhD than working with you this past year on the Tmprss2 project. It has been a privilege to be coworkers, and there is little I enjoy more than talking science with you. Also, you are so creative and make amazing scientific figures – many of the figures in this thesis are or were inspired by Peiffer originals (so thank you!).

I feel incredibly fortunate to have amazing parents. Mom and Dad, thank you for the continuous love and support. You both have always been there for me and pushed me to be my best. Mom, thank you for always being there to talk, and Dad thank you for all the guidance and advice. Love you both.

To my furry and feathered companions, thank you for being adorable and bringing so much happiness even during the most stressful of times.

And a finally, a heartfelt thank you to my fiancé and partner in crime, Nick Bozzonetti. Every day with you is bright and you make me feel so loved. You have always been in my corner and supported me in my career as a scientist. I will forever feel blessed that you came to Ann Arbor to go through this adventure with me. You bring so much happiness to my life and I can't wait to get married in a few months – love you!

There have been so many people who positively influenced me and helped me get to this point that weren't mentioned here, including teachers, friends, current and past colleagues. So, to everyone else, know that I am grateful for the impact you had on my journey.

Table of Contents

Acknowledgements	ii
List of Figures	viii
List of Tables	xii
List of Appendices	xiii
List of Abbreviations	xiv
Abstract	xviii
 CHAPTER 1. Selective Modulation of Dynamic Protein Complexes	
1.1 Abstract	1
1.2 Introduction	1
1.3 Overview of dynamic proteins and their interactions	2
1.4 Controlling the interactions and function of dynamic proteins: Lessons from Nature	6
1.5 Achieving selectivity with chemical probes	9
1.6 Techniques for identification of selective dynamic protein complex modulators	13
1.7 Emergent technologies and alternative approaches to selectivity	18
1.8 Dissertation Summary	19
1.9 References	20
 CHAPTER 2. Targeting the Dynamic Coactivator Med25 with a Selective Small Molecule Probe	
2.1 Abstract	32
2.2 Introduction	32
2.3 Results and Discussion	

	High throughput screening of Med25 leads to identification of covalent depsidone inhibitors	36
	Evaluating the in vitro selectivity of norstictic acid	39
	Elucidation of NA binding site and inhibitory mechanism	41
	Engagement of norstictic acid with Med25 in cells	44
2.4	Conclusions	49
2.5	Materials & Methods	50
2.6	References	61

CHAPTER 3. High-Throughput Method to Identify Selective Modulators of Activator•Coactivator Interactions

3.1	Abstract	66
3.2	Introduction	67
3.3	Results and Discussion	
	Development of duplex fluorescence polarization assay with Med25 AcID and CBP KIX	69
	Pilot screen of small library using duplex assay	73
	Additional selectivity profiling of lead compounds from pilot screen	77
	Optimization of duplex assay for IBiD	80
3.4	Conclusions	81
3.5	Materials & Methods	81
3.6	References	84

CHAPTER 4. Targeting the Transmembrane Serine Protease TMPRSS2 to Prevent SARS-CoV-2 Infection

4.1	Abstract	89
4.2	Introduction	89
4.3	Results and Discussion	

	Recombinant expression of TMPRSS2 Peptidase S1 domain in <i>E. coli</i>	93
	Activity of TMPRSS2 peptidase domain	95
	Analysis of covalent TMPRSS2 inhibitors	98
	Combined in silico and biochemical screening leads to identification of novel inhibitors	101
	Curated libraries based on guanidinium functionality	105
	Covalent inhibitors of TMPRSS2	106
4.4	Conclusions	108
4.5	Materials & Methods	109
4.6	References	113
CHAPTER 5. Conclusions and Future Directions		
5.1	Conclusions	119
5.2	Future Directions	121
	Targeting Med25 with NA and beyond	121
	Elucidation of chemical scaffolds selective for ABDs	124
	Targeting TMPRSS2: profiling mechanisms and identifying selective scaffolds	125
5.3	References	129
	Appendices	135

List of Figures

Figure 1.1	Comparison of PPI structural order and druggability	2
Figure 1.2	The master coactivator CREB binding protein (CBP) analyzed as a representative dynamic protein	4
Figure 1.3	The chaperone HSP70 analyzed as a representative dynamic protein	5
Figure 1.4	Masking as a mechanism to dictate interaction and function of dynamic proteins	7
Figure 1.5	Allostery as a mechanism to regulate dynamic proteins	9
Figure 1.6	Mass spectrometry as a technique for identification of dynamic protein modulators	14
Figure 1.7	Differential scanning fluorimetry (DSF) as a technique for identification of dynamic protein modulators	17
Figure 2.1	Transcriptional activation is mediated by coactivator proteins such as those found in the tail of the Mediator complex	34
Figure 2.2	Coactivator and hub protein Med25 is a subunit of the Mediator coactivator complex.	35
Figure 2.3	Norstictic acid is discovered as the result of a high throughput screen for inhibitors of the AcID•VP16(465-490) PPI	37
Figure 2.4	Norstictic acid is identified as a covalent inhibitor	38
Figure 2.5	Salicylaldehyde covalently modifies, but does not inhibit, Med25 AcID	39
Figure 2.6	Norstictic acid is selective for AcID PPIs	40
Figure 2.7	Comparison of norstictic acid reactivity between AcID and KIX	41
Figure 2.8	Mutation of dynamic loop lysine residues to arginine has minimal impact on ETV5 binding	42
Figure 2.9	Mutation of lysine residues in H2 loop to arginine impacts NA binding and inhibition	43
Figure 2.10	Molecular dynamics of AcID with NA modification at K519	44

Figure 2.11	Thermal shift assays profiling NA engagement with endogenous Med25	45
Figure 2.12	Co-immunoprecipitation of ETV5 with Med25 is inhibited by NA in HeLA cells	46
Figure 2.13	Isobologram showing synergy between NA and Lapatinib	46
Figure 2.14	NA inhibits Med25 ETV5 interaction in triple negative breast cancer cell line VARI068	47
Figure 2.15	NA inhibits Med25 ATF6 interaction mammalian cells line	49
Figure 3.1	Coactivator•activator interactions are classically a challenge to target selectively	68
Figure 3.2	Overview of first generation duplex screen proteins Med25 AcID and CBP KIX	70
Figure 3.3	Description of first generation duplex screen fluorescence polarization assay	71
Figure 3.4	Direct binding experiments under duplex conditions accurately represent interaction of the cognate ABD•TAD pairs	72
Figure 3.5	Duplex assay can accurately characterize the selectivity of known inhibitors lobaric acid, norstictic acid, and garcinolic acid	73
Figure 3.6	Duplex assay can accurately characterize the selectivity of known inhibitors lobaric acid, norstictic acid, and garcinolic acid	74
Figure 3.7	Lead compounds selected for follow up screening	75
Figure 3.8	Inhibition curves of representative duplex screen hits	76
Figure 3.9	Competition FP looking at inhibition of CH1 and IBiD by duplex screen lead compounds	78
Figure 3.10	Selectivity profiling of fumarprotocetraric acid	79
Figure 3.11	Comparison of FPA and NA	80
Figure 3.12	Development of duplex assay utilizing Med25 AcID and CBP IBiD	80
Figure 4.1	TMPRSS2 aids in SARS-CoV-2 viral infection	91
Figure 4.2	The molecules that have commonly been cited as TMPRSS2 inhibitors	92
Figure 4.3	Proposed construct for TMPRSS2(256-492) peptidase expression in <i>E. Coli</i> .	94

Figure 4.4	Expression and purification of TMPRSS2 peptidase domain	95
Figure 4.5	Inhibition curve of TMPRSS2 activity obtained with increasing [FPR-cmk]	96
Figure 4.6	Boc-QAR-AMC as a substrate for TMPRSS2 and Trypsin	97
Figure 4.7	Development of biochemical TMPRSS2 activity assay	98
Figure 4.8	Camostat, nafamostat, FOY 251, and gabexate covalently modify Trypsin	99
Figure 4.9	Decreasing IC ₅₀ values for camostat with increased incubation time for trypsin and TMPRSS2 is consistent with covalent inhibition.	100
Figure 4.10	TMPRSS2 homology model	102
Figure 4.11	Docked poses of camostat, nafamostat, and gabexate in the TMPRSS2 active site	103
Figure 4.12	Profiling of clinically approved drugs identified as top hits from computational screening	104
Figure 4.13	Inhibition of Trypsin by pentamidine, propamidine, and debrisoquine	105
Figure 4.14	SxtG fragment screening results	106
Figure 4.15	Enamine library screening results	108
Figure 5.1	Western blot comparing Med25 localization in HeLa and VARI068	122
Figure 5.2	Potential PTMs within Med25	123
Figure 5.3	NA leads to dose dependent decrease in AcID thermal stability	124
Figure 5.4	Enhancing potency and selectivity of TMPRSS2 targeting fragments	126
Figure 5.5	Common TMPRSS2 variants known to elicit a decrease enzymatic function	127
Figure B.1	Use of a reporter dye is a key component of DSF	143
Figure B.2	Initial melting temperature calculations for Med25 AcID and CBP KIX using DSF	145
Figure B.3	CD thermal melt of Med25 AcID	146
Figure B.4	DSF data for Med25 AcID and known ligands Garcinolic acid and Norstictic acid	147
Figure B.5	Concentration dependence of NA on AcID T _m	148

Figure B.6	Protein gel based thermal shift assay using purified Med25 AcID and small molecules	149
Figure B.7	Protein gel based thermal shift assay using purified Med25 AcID and NA	150
Figure B.8	DSF data with Med25 AcID and TAD peptides	150
Figure B.9	Protein gel based thermal shift assay using purified Med25 AcID and TAD peptides	151
Figure B.10	DSF data with CBP KIX and ligands	153
Figure B.11	DSF with the apo KIX ternary systems	154
Figure B.12	DSF with the apo KIX and KIX _{I660V} ternary systems	155
Figure B.13	DSF with the KIXL664C and small molecule stabilizer 1-10	155

List of Tables

Table 1.1	Representative allosteric small molecule modulators published in recent years	10
Table 3.1	Summary of KIX and AcID inhibition by duplex screen lead molecules	77
Table C.1	Soft-core potentials used in flexible receptor docking	165

List of Appendices

Appendix A.	Characterization of Synthesized Peptides	135
Appendix B.	Differential Scanning Fluorimetry to Profile Dynamic Coactivators and Their Complexes	142
Appendix C.	Techniques Used for <i>in silico</i> Study of TMPRSS2	162

List of Abbreviations

ABD	Activator binding domain
ACE2	Angiotensin-converting enzyme 2
AcID	Activator interaction domain
ACTR	Activator for thyroid hormone and retinoid receptors
AMC	Aminomethyl coumarin
ATF6	Activation transcription factor 6
ATP	Adenosine triphosphate
BRD	Bromodomain
CBP	CREB binding protein
CETSA	Cellular thermal shift assay
CH1	Cysteine histidine rich 1
CID	Collision induced dissociation
CIU	Collision-induced unfolding
co-IP	Coimmunoprecipitation
CRISPR	Clustered regularly interspaced short palindromic repeats
DARTS	Drug affinity responsive target stability
DBD	DNA binding domain
DMEM	Dubelco's modified eagle medium
DMSO	Dimethylsulfoxide
DNA	Deoxyribonucleic acid
DSF	Differential scanning fluorimetry
DTT	Dithiothreitol
EGFR	Epidermal growth factor receptor
ER	Endoplasmic reticulum
Ets	E-twenty six
ETV	Ets translocation variant
FBS	Fetal bovine serum

FITC	Fluorescein isothiocyanate
FOXO	Forkhead box transcription factor class O
FP	Fluorescence polarization
FPA	Fumarprotocetraric acid
FRET	Forster resonance energy transfer
GA	Garcniolic acid
GAPDH	Glyceraldehyde 3-phosphate dehydrogenase
GDP	Guanosine diphosphate
GFP	Green fluorescent protein
GPCR	G-protein coupled receptors
GST	Glutathione transferase
GTP	Guanosine triphosphate
HAT	Histone acetyltransferase
HIF1 α	Hypoxia inducible factor 1 alpha
HPLC	High-performance liquid chromatography
Hsp	Heat shock protein
HSPA5	Heat shock protein family A member 5
HSQC	Heteronuclear single quantum coherence spectroscopy
HTS	High-throughput screening
IBiD	IRF-3 binding domain
IC ₅₀	Half-maximal inhibitory concentration
IDP	Intrinsically disordered protein
IPTG	Isopropyl- β -D-1-thiogalactopyranoside
K _d	Dissociation constant
KIX	Kinase inducible domain interaction domain
K _M	Michaelis constant
LBD	Ligand binding domain
LDLR	Low density lipoprotein receptor
LE	Ligand efficiency
MD	Molecular dynamics
Mdm2	Murine double minute 2

Med15	Mediator subunit 15
Med25	Mediator subunit 25
MLL	Multi-lineage leukemia
MMP	Matrix metalloproteinase
mRNA	Messenger RNA
MW	Molecular weight
NA	Norstictic acid
NBD	Nucleotide binding domain
NFκB	Nuclear factor kappa-light-chain-enhancer of activated B cells
Ni-NTA	Nickel-Nitrilotriacetic Acid
NMR	Nuclear magnetic resonance spectroscopy
NR	Nuclear receptor
PA	Psoromic acid
PAR2	Protease activated receptor 2
PCR	Polymerase chain reaction
PDB	Protein DataBank
PK1	Pyruvate dehydrogenase kinase 1
PEA3	Polyoma enhancer activator 3
pKID	Phosphorylated kinase inducible domain
PKM2	Pyruvate kinase M2
PPAR	Peroxisome proliferator-activated receptor
PPI	Protein-protein interaction
PrOF	Protein-observed fluorine NMR
PROTAC	Proteolysis targeting chimera
PTM	Posttranslational modification
qPCR	Quantitative polymerase chain reaction
RMSF	Root-mean-square fluctuations
RNA	Ribonucleic acid
RPM	Revolutions per minute
RFU	Relative fluorescence units
S protein	Spike protein

SARS	Severe acute respiratory syndrome
SH	Src Homology
SRCR	Scavenger receptor cysteine-rich
STAT	Signal transducer and activator of transcription
SUMO	Small ubiquitin like modifier
Swi/Snf	Switch/Sucrose non-fermentable
TAD	Transcriptional activation domain
TAZ	Transcriptional adapter zinc finger
TEAD	Transcriptional enhanced associate domain
T _m	Melting temperature
TMPRSS2	Transmembrane protease, serine 2
TR	Texas red
VP16	Herpes simplex virus protein 16
VWA	Von Willebrand Factor Type A
WT	Wild-type
YAP	Yes-associated protein

Abstract

For decades, small molecule modulators of protein function have been vital for advancing both understanding of basic biology and therapeutic discovery; however, there are classes of proteins for which small molecule discovery has lagged. The majority of the proteins studied in this work are transcriptional coactivators, proteins that have great potential as therapeutic targets, yet are often classified as “undruggable” because of their atypical structure and mechanism of action. Coactivators are highly malleable and their interactions typically occur over broad surface areas with only moderate affinity, making them difficult to target. Recent advances in protein-protein interaction (PPI) inhibitor discovery, exploiting allostery and dynamic substructures within target proteins, has led to some success, but selectivity remains a challenge. Selectivity is pivotal for candidate probe molecules due to the extensive interaction network of these dynamic hubs. The following dissertation addresses the challenges of selectivity and presents small molecules that target coactivator interactions.

A large portion of this work is dedicated to targeting the activator interaction domain (AcID) of the coactivator Med25. This domain contains a unique protein fold and is not required for basal transcription, but interacts with activators to regulate genes implicated in disease. Previous work from the Mapp lab led to identification of promising AcID PPI inhibitors. Here, this work was continued with the compound norstictic acid (NA). NA demonstrated high selectivity for AcID interactions and, using biochemical techniques in combination with molecular dynamics simulations, the mechanism of inhibition was elucidated: covalent modification of lysine residues within a dynamic loop leads to both orthosteric and allosteric inhibition of activator binding. NA proved useful for studying Med25 in a cellular context, engaging with full length protein from cell extracts and inhibiting the interaction between endogenous Med25 and transcriptional coactivators. Ultimately NA was used to probe the interaction of Med25 and ETV5, a transcription factor linked to metastasis, in a metastatic breast cancer cell line. Dosing with NA was able to decrease expression of the Med25•ETV5 regulated gene MMP-2, similarly to what was observed with KO of Med25.

Inspired by the iterative screening approach used to discover NA, a new small molecule screening method for PPIs was developed, the first method for coactivator targets that directly incorporates selectivity at the primary level. A fluorescence polarization assay that simultaneously monitors multiple activator•coactivator interactions is presented. This method enables assessment of both selectivity and potency of candidate inhibitors in a single screen. A duplex assay containing AcID and CBP KIX has been optimized and a pilot screen has been conducted. The pilot screen was able to categorize compounds as Med25 AcID-selective, CBP KIX-selective, or dual inhibitors. Representative compounds from each subset were evaluated in secondary screening, leading to identification of novel inhibitors of both proteins. Ultimately, this approach is applicable to other coactivator•activator complexes.

The final focus of this dissertation is development of a biochemical screen for the human serine protease TMPRSS2, which plays a key role in SARS-CoV-2 viral infection. TMPRSS2 is a difficult protein target because expression and purification is challenging, thus complicating inhibitor screening and *in vitro* studies. An expression and purification procedure to isolate active TMPRSS2 protease domain from *E. Coli* was developed. Biochemical methods were used to characterize the mechanism of known inhibitors, and, through integration with computational methodologies, novel TMPRSS2 targeting chemical scaffolds were identified.

CHAPTER 1

Selective Modulation of Dynamic Protein Complexes^{*}

1.1 Abstract

Dynamic proteins perform critical roles in cellular machines, including those that control proteostasis, transcription, translation, and signaling. Thus, dynamic proteins are prime candidates for chemical probe and drug discovery but difficult targets because they do not conform to classical rules of design and screening. Selectivity is pivotal for candidate probe molecules due to the extensive interaction network of these dynamic hubs. Recognition that the traditional rules of probe discovery are not necessarily applicable to dynamic proteins and their complexes, as well as technological advances in screening, have produced remarkable results in the last 2-4 years. Particularly notable are the improvements in target selectivity for small molecule modulators of dynamic proteins, especially with techniques that increase the discovery likelihood of allosteric regulatory mechanisms.

This chapter focuses on approaches to small molecule screening that appear to be more suitable for highly dynamic targets and have the potential to streamline identification of selective modulators. First, dynamic proteins are defined and discussed, followed by a review of key principles that govern selectivity in binding and function for these proteins and their complexes. Additionally, notable recent success stories in small-molecule modulation of these proteins, examples that either intentionally or serendipitously mimic natural regulatory mechanisms of dynamic proteins, are highlighted. Finally, we discuss emerging strategies that facilitate the discovery of selective chemical modulators of conformationally dynamic proteins.

1.2 Introduction

Chemical modulation of protein targets with small molecules has been fundamental in the field of chemical biology and drug discovery. Chemical modulators, as opposed to genetic or biological,

^{*} The majority of this chapter is reproduced from Garlick J.M., Mapp A.K. Selective Modulation of Dynamic Protein Complexes. *Cell Chemical Biology* **2020**, 27, (8), 986-997 (DOI: 10.1016/j.chembiol.2020.07.019)

can allow for fine-tuned exploration of a specific protein in vitro in order to better understand that protein in a broader biological context.^{1,2} From a drug discovery perspective, small molecule chemical probes can be key players in the validation of new molecular targets for a therapeutic intervention.³ However, the path to a chemical probe is not well paved for all protein targets. In particular, conformationally dynamic proteins and hub proteins (often one in the same), present novel challenges to probe discovery that remain unaddressed. Thus, it is not surprising that proteins within this class are often deemed “undruggable” and there are few quality chemical probes.^{4,5} Thorough analysis of the current research with regards to selective chemical modulation of dynamic protein targets is necessary to identify broadly applicable strategies that can facilitate probe discovery for these potential drug targets.

1.3 Overview of dynamic proteins and their interactions

Conformationally dynamic proteins underpin all biological processes, from signaling to gene expression.^{6–8} These proteins often form short-lived protein-protein interactions (PPIs) that allow cellular machines to assemble and disassemble as needed to regulate key pathways and events.⁹ Efforts to decipher the principles that govern PPI formation in general have been extensive.^{10–12} However, most work has been focused on PPIs between stable, well-defined partners and the emergent rules governing assembly, molecular recognition, and modulation of PPIs are not generally applicable to dynamic proteins, leading to classification as “undruggable” (Figure 1.1).

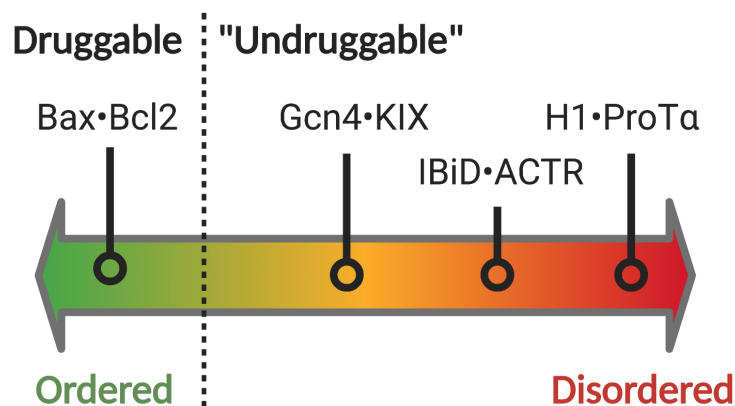


Figure 1.1. Comparison of PPI structural order and “druggability.” An example of a stable, well defined interaction is that of the BH3 domain of Bax and Bcl2. These proteins are both ordered before binding, and interact at a well-defined interface with a high affinity (nM).¹³ This has allowed for multiple inhibitors of clinical relevance to be developed.^{14,15} The interactions of transcriptional activator Gcn4 with the KIX domain of the coactivator Med15 is moderately ordered, but still a challenge for drug discovery with current methods. KIX is folded, but part of the larger Med15 protein which contains multiple regions

of disorder as well as other PPI interfaces. Gcn4 is disordered until binding, and binding occurs with moderate affinity over a larger interface. Both the interaction between CBP/p300 IBD and ACTR as well as Human linker Histone H1 and ProTα are examples extremely challenging PPI targets. IBD and ACTR are both intrinsically disordered and unstructured until binding occurs.^{16,17} H1 and ProTα are also intrinsically disordered, however they continue to remain unstructured even upon interaction, forming an extremely fuzzy complex.¹⁸ Studies of the PPI interface have revealed that it is so large, barely any specific binding site on either protein can be identified.^{19,20} Figure created using BioRender.

While no single definition can be used to identify dynamic proteins, useful metrics for successful classification include: 1) the structure and mobility of the protein and/or the individual domains; 2) the number of functional roles the protein can fulfill; and 3) the number binding interactions the protein can participate in. Dynamic proteins tend to be multifunctional, participating in many pathways and performing multiple cellular roles, either through the presence of multiple domains or by the formation of dynamic transient complexes that vary in function. The functional multiplicity can in part be attributed to the number of binding partners the protein engages with, which oftentimes leads to conformational changes either within a single domain or domain rearrangement as an entire protein to elicit variable functional effects. Members of this protein class tend to have high flexibility and structural plasticity.²¹ Malleable structures allow them to perform context-dependent regulatory functions and interact with a multitude of binding partners. For example, the master coactivator CBP/p300 (Figure 1.2) contains six conformationally dynamic domains that interact with hundreds of transcriptional activators, which are linked by intrinsically disordered regions, thus increasing the overall flexibility of the protein.²²

High degree of disorder

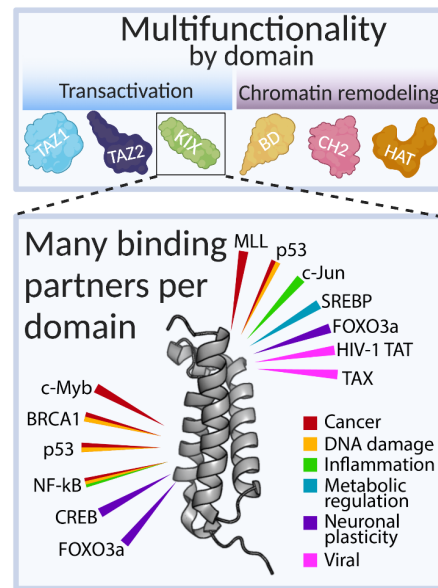
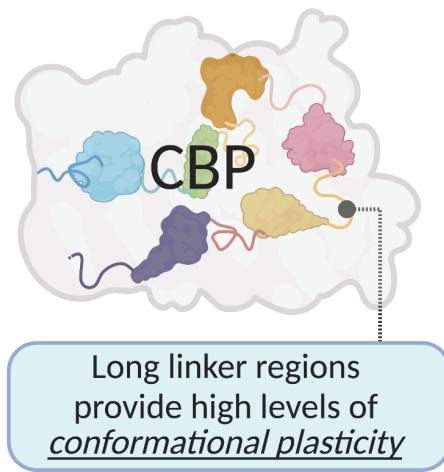


Figure 1.2. The master coactivator CREB binding protein (CBP) analyzed as a representative dynamic protein. CBP is highly disordered, with more unstructured and dynamic regions than structured domains.²² The protein has high conformational plasticity, as its domains are connected by long unstructured linkers. The individual domains within CBP allow it to perform multiple functions. For example, CBP has four activator binding domains, KIX, TAZ1, TAZ2, and IBiD, that interact with transcriptional activators to regulate gene expression. It also contains a histone acetyltransferase domain (HAT) and bromodomain (BD) that it utilizes for perform chromatin remodeling functions.²³ Each of these individual domains has its own network of interactions, summing up to 100s of potential PPIs made by CBP. The KIX domain is shown here as an example – it interacts with a suite of 15+ activator proteins, all which can be implicated in a variety of diseases.²⁴ Figure created using BioRender.

Dynamic proteins are often observed in higher order cellular machines with a functional need to rapidly form and exist only transiently, such as the multiprotein complexes involved in chromatin remodeling²⁵ or protein folding.²⁶ Thus, dynamic proteins are often hubs, interacting with subunits within a complex as well as other proteins and ligands.²⁷ The composition of a dynamic protein complex can dictate enzymatic subunit activity or even change the function of the complex. For example, combinatorial assembly of Hsp70 (Figure 1.3) allows the complex to switch between a folding system and a degradation complex.²⁸ Transcription is another process that relies on dynamic proteins. It is important that individual components of the machinery, such as activators, coactivators, general transcription factors, and RNA polymerase, are able to assemble efficiently at one promotor and then move on to mediate expression of another gene.²⁹

There are extensive reviews regarding the function and composition of these dynamic machines.^{30–}

33

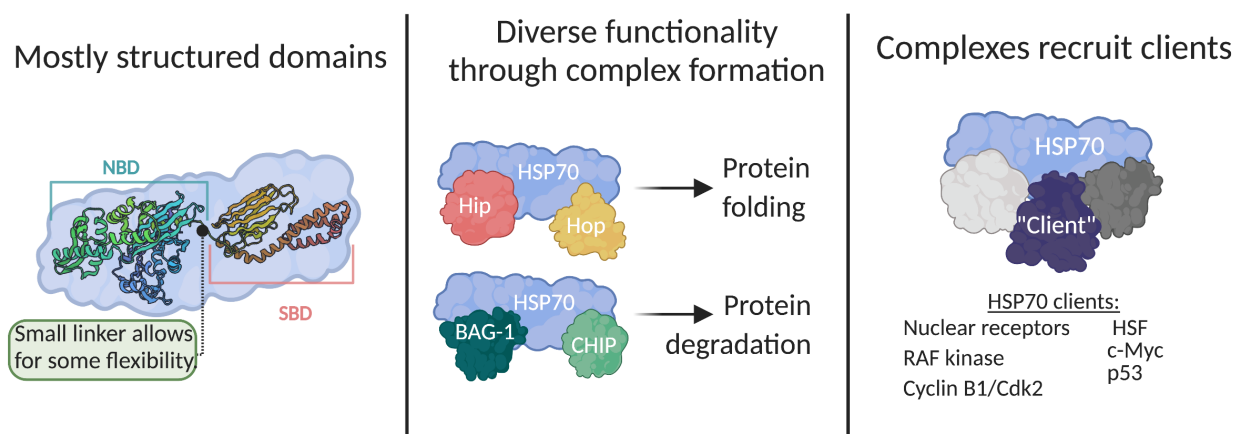


Figure 1.3. The chaperone HSP70 analyzed as a representative dynamic protein. HSP70 contains much more secondary structure than CBP. It is comprised of two main domains: the nucleotide binding domain (NBD) and substrate binding domain (SBD). These two domains are connected by a small dynamic linker.³⁴ Substrate binding can induce domain rearrangement to elicit differential functional effects. HSP70 itself functions as an ATPase, however it is able to perform many diverse cellular functions by forming different multiprotein complexes. For example, Hsp70 in complex with cochaperones Hip and Hop acts as a protein folding complex, while Hsp70 in complex with BAG-1 and CHIP functions as a degradation complex.²⁸ Finally, Hsp70 complexes also interact with many other proteins. Many regulatory proteins are known to be controlled via transient association with Hsp70. Thus, Hsp70 is a potential point of therapeutic intervention for conditions such as cancer, autoimmune and neurodegenerative diseases.³⁵ Figure created using Biorender.

The qualities that make dynamic proteins and their complexes effective cellular machines also make them challenging to target. Due to conformational flexibility, it is often difficult to obtain the structural and biophysical data useful for the rational design and/or optimization of chemical modulators. The PPIs of dynamic proteins are typically lower affinity, utilize a relatively large surface area, and have minimal topology, thus representing a significant challenge for orthosteric inhibition.^{36,37} High throughput screening (HTS) with dynamic protein targets can also be challenging due to the difficult expression and isolation of the full-length proteins. Thus, in many cases, it is necessary to utilize simplified systems such as isolated domains that may not recapitulate the structure, dynamics, and/or activity of the full-length complexes. Finally, selective targeting of a particular PPI or specific binding interface is a challenge because dynamic proteins typically interact with a multitude of binding partners, which is a challenge that will be focused upon in this dissertation.^{5,38,39}

Notably, dynamic proteins are often dysregulated in disease and dynamic multiprotein complexes often contain specific subunits that are attractive therapeutic targets. For example, the tumor suppressor BRCA1, an important target for breast cancer prevention and therapy, associates with the Swi/Snf chromatin remodeling complex.^{40,41} CBP/p300, p53, and other transcription factors are all also dynamic proteins that are potential targets for a variety of diseases.^{42,43} While dynamic PPIs do not conform to typical “rules” used for drug discovery, in recent years significant achievements have been made when it comes to modulating some of these challenging targets.

1.4 Controlling the interactions and function of dynamic proteins: Lessons from Nature

As hubs associated with many complex cellular machines, dynamic proteins are precisely regulated. These regulatory processes often control the access of binding surface(s) within the dynamic protein either directly or indirectly. This can occur through masking of binding surfaces via intra- or intermolecular interactions, through induced conformational changes of the protein, and, most commonly, combinations of both mechanisms.

Masking of binding surfaces within dynamic proteins occurs through intra- and intermolecular complexes that have a variety of functional outcomes. Masking interactions are common in transcriptional activators, where the highly disordered transcriptional activation domain is sequestered by a high affinity interaction until needed. Classic and well-studied examples include the Gal4•Gal80^{44,45} and p53•Mdm2⁴⁶ complexes that provide temporal control of the transcription factors. Masking of dynamic proteins and/or protein domains is similarly important in many other cellular contexts. The scaffolding 14-3-3 protein, for example, regulates the subcellular localization and function of Caspase-2 through stabilizing two dynamic regions in individual subunits. 14-3-3 masks both the nuclear localization sequence of pro-Caspase2 and the dimerization interface, inhibiting pro-Caspase2 activation⁴⁷ (Figure 1.4). 14-3-3 also regulates FOXO transcription factors by masking their nuclear localization sequences.⁴⁸ In all of these examples, the masking interaction provides fine temporal control of the dynamic protein. Obstruction of key binding surfaces in a dynamic protein is also associated with spatial control. For example, until the unfolded protein response is activated in a cell, the transcription factor ATF6 is sequestered at the ER membrane. Once the pathway is stimulated, ATF6 is cleaved from the ER membrane, enabling its translocation to the nucleus, where its potential binding partners reside.⁴⁹

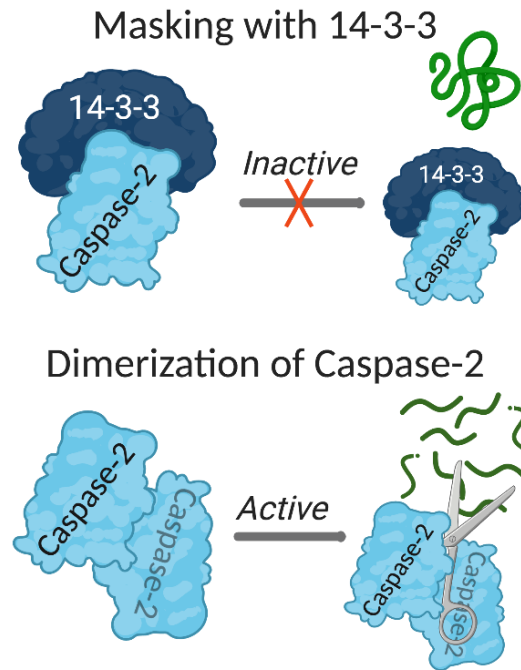


Figure 1.4. Masking as a mechanism to dictate interaction and function of dynamic proteins. Masking particular sequences or binding surfaces on proteins provides control over function, localization, and ligand binding. 14-3-3 regulates activity of Caspase-2 by masking its dimerization interface. Without 14-3-3 bound, Caspase-2 dimerizes and proteolytically cleave other proteins.⁴⁷ Figure created using BioRender.

Emerging data regarding the composition and function of biomolecular condensates indicates that masking of dynamic proteins likely plays a crucial role in the compartmentalization that such condensates afford. More recent examples of compartmentalization focus on phase separation, which increases the strength and specificity of interactions that otherwise appear promiscuous, while still allowing the proteins involved to retain their dynamic nature.⁵⁰ The transcription factors TAZ, OCT4, GCN4, and the estrogen receptor, have been shown to form phase-separated condensates with Mediator via its coactivator subunits.⁵¹ TAZ has also been observed to form nuclear condensates with its DNA binding cofactor TEAD4 as well as the coactivator BRD4.⁵² In all of these cases, unmasked transcriptional activation domains appear necessary for condensate formation. However, there is also some evidence to the contrary, suggesting that there is actually no correlation between phase separation and regulation of gene expression, rather alternative mechanisms such as cooperative chromatin binding and formation of well-defined multivalent protein complexes are the driving forces enhancing transcriptional activation.⁵³

Posttranslational modifications (PTMs) and distal binding events that influence structure are also mechanisms of binding surface control. PTMs occur at many PPI interfaces to promote or inhibit binding. For example, phosphorylation of the coactivator p300 at S89 inhibits its interaction with peroxisome proliferator-activated receptor γ (PPAR- γ) and retinoic acid receptor (RAR).⁵⁴ PTMs can also alter stability, folding, or conformation as a mechanism to regulate dynamic protein interactions.⁵⁵ PTMs, such as ubiquitination, acetylation, phosphorylation, and glycosylation, have been observed to be allosteric regulators.⁵⁶ For example, N-linked glycosylation of interleukin-7 receptor α allosterically enhances binding to human IL-7 nearly 300-fold.⁵⁷ In another example, a conserved cysteine in the palmitate binding pocket of TEAD proteins undergoes palmitoylation that allosterically stabilizes the TEAD-YAP interaction.⁵⁸ Nature often utilizes a combination of PTMs to fine-tune protein function, allowing a single protein to perform diverse cellular roles. A recent study of Hsp90, for example, identified a group of conserved PTMs that globally mediates dynamics and allosteric communication in the Hsp90 structures.⁵⁹

Allosteric crosstalk via inter and intra-molecular interactions can lead to broader conformational changes, redistribution of conformers, and alterations in kinetic/thermodynamic barriers that also regulate protein interfaces.^{60,61} For example, the binding of a cognate ligand or cofactor can lead to structural changes that influence protein function.⁶² A classic example of this is G-protein coupled receptors (GPCRs). Ligand binding results in a conformational change in the GPCR, leading to activation of its associated G protein.⁶³ PPIs can play a similar role. For example, the ubiquitinase BAP must interact with another protein, ASXL2, via its nonenzymatic ULD domain to allow ubiquitin to bind and be cleaved via its hydrolase (UCH) domain. Thus, allosteric changes induced in BAP1 via ASXL2 interaction with the ULD domain are critical for its enzymatic function. Biophysical studies suggest that a loop within the UCH domain of BAP1 is stabilized by ASXL2 interaction and this allows for ubiquitin to bind⁶⁴ (Figure 1.5).

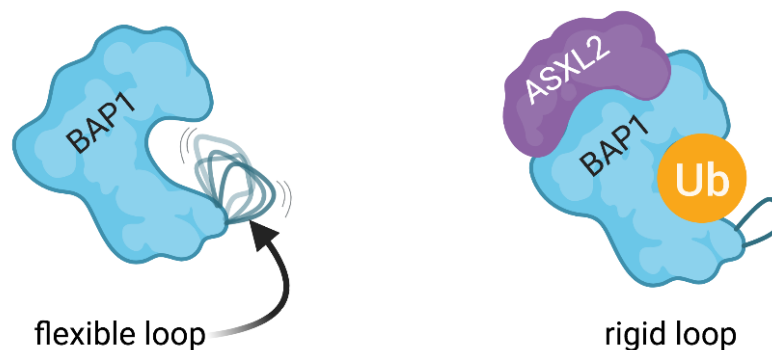


Figure 1.5. Allostery as a mechanism to regulate dynamic proteins. Allostery, via intra or intermolecular interactions, regulates structure and function of proteins. The Polycomb group protein ASXL2 interacts with the deubiquitinase BAP1, stabilizing a distal dynamic loop and allowing ubiquitin to bind and be hydrolyzed. Without ASXL2 interaction, BAP1 is unable to perform its enzymatic function.⁶⁴ Figure created using BioRender.

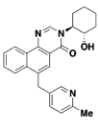
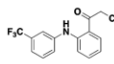
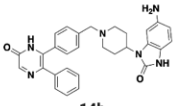
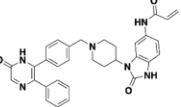
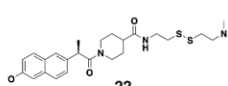
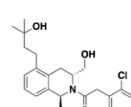
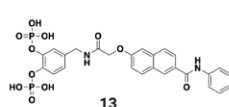
Allosteric sites often reside within the most flexible regions of a protein. Dynamic proteins and the complexes they form rely heavily on unstructured regions such as loops and linkers to influence interactions. Loops are often directing components of selectivity and linkers play an important role in crosstalk between multidomain proteins. Linkers also can play a key role in multiprotein complex assembly and intradomain interactions.⁶⁵ A flexible glycine rich linker region in NF κ B allows for the formation of an interaction between the p50 and SWI6/ANK domains to regulate intracellular transport.⁶⁶ Allosteric sites tend to be less conserved between closely related proteins, subfamily members, or homologues.^{67–69} Thus, allosteric inhibitors are a potential route to selective targeting of challenging PPIs as well as control over multiprotein, dynamic complexes. Perhaps more than any other class of proteins, nature suggests that dynamic proteins are prime targets for allosteric regulation.

1.5 Achieving selectivity with chemical probes

Assessing the native mechanisms that dictate interaction between dynamic proteins and their binding partners can illuminate possible routes or methods to target such challenging interactions. Natural regulation of dynamic proteins suggests that a possible route to successful modulators of their interactions is via allosteric sites. Allosteric modulators have the ability to shift the conformational ensemble of proteins to favor the formation of specific complexes or toggle active/inactive conformers without the need to directly target the PPI interface. This strategy is particularly attractive when it comes to targeting large surface area interactions because orthosteric

modulation, particularly with small molecules, is challenging due to the lack of topological characteristics that allow for high affinity ligand binding. Allosteric modulation can be achieved - even for these challenging targets- with small molecules that have more drug-like properties, such as lower MW (<600), greater structural rigidity, and hydrophobicity.³⁶ Another advantage that comes with targeting allosteric regions of dynamic proteins is that they tend to be less conserved between the same protein family and homologous proteins, therefore, suggesting an avenue for selective modulation.^{69–71} Table 1.1 summarizes the allosteric compounds, all published since 2018, that will be discussed in this section and spotlights their selectivity profiles.

Table 1.1. Representative allosteric small molecule modulators published in recent years

Compound	Target	Affinity / Activity	Selectivity
 MK-7622	Muscarinic acetylcholine receptor M1 (GPCR)	IP = 19 nM	No potentiation or agonism of M2-M4 muscarinic subtypes in CHO cell overexpression model (Beshore et al., 2018)
 TED-347	TEAD family proteins	EC ₅₀ = 3.5 ± 0.3 μM (TEAD4) EC ₅₀ = 4.3 ± 0.3 μM (TEAD2) *FP assay using YAP peptide	Does not inhibit other PPIs via FP assay (urokinase receptor and voltage-gated calcium channel Cav2.2) (Bum-Erdene et al., 2019)
 14b	Akt2	IC ₅₀ = 1237 ± 142 nM	10 fold selective over Akt1 >20 fold selective over Akt3 (Quambusch et al., 2019)
 16a	Akt1	IC ₅₀ = 39 ± 5 nM	334 fold selective over Akt2 >500 fold selective over Akt3 (Quambusch et al., 2019)
 22	Med25	Reduces <i>k_{off}</i> of activators by 25%	No selectivity data (Henderson et al., 2018)
 LY3154207	Dopamine receptor D1	EC ₅₀ = 3.0 ± 0.3 nM	>2700 fold selective over D5 receptor; >1000 fold selective over 40 other targets from the standardized Lilly selectivity panel (Hao et al., 2019)
 13	STAT5b	<i>K_i</i> = 44 ± 1 nM	55 fold selectivity over STAT5a (Gräb et al., 2019)

Allosteric regulators have demonstrated high selectivity for protein targets of many other classes, suggesting that similar results are possible with dynamic proteins. For example, the ability to selectively target kinases, GPCRS, and bromodomains (BRDs) has improved with our understanding of allostery in these systems.^{72–74} Extensive research on selectivity within kinase

families has shown that molecules that bind outside the catalytic domain display high selectivity on both a family and subtype level while still functioning as potent inhibitors of enzymatic activity.⁷² Allosteric modulation of GPCRs and the impact on selectivity is also well documented.^{75,76} For example, a group from Merck recently published work on MK-7622 a novel positive allosteric modulator of the muscarinic acetylcholine receptor M1 for the treatment of Alzheimer's disease. This compound was highly selective for the M1 receptor, with no potentiation or agonism of M2-M4 subtypes in an overexpression model conducted in CHO cells.⁷⁷ Furthermore, compounds that are able to distinguish between the most homologous BRDs, BRD4/7/9, take advantage of subtle conformational differences in non-conserved regions distal to the acetyl lysine binding site.^{78,79}

Covalent small molecule modulators have been successful at allosterically modifying dynamic proteins in a manner that mimics covalent PTMs found in nature. In a study informed by a naturally occurring PTM, Bum-Erdene et al. targeted a conserved cysteine residue in TEAD family proteins known to undergo palmitoylation in order to stabilize interaction with YAP. Starting with the FDA-approved compound flufenamic acid, shown to weakly bind TEAD2 non-covalently near the thiol of the conserved cysteine, a chloromethylketone moiety was incorporated to make it a covalent inhibitor. This compound, TED-347, allosterically inhibited YAP-TEAD PPIs both in vitro and in mammalian cells ($K_i = 10 \mu\text{M}$). TED-347 also demonstrated selectivity for the TEAD family over other related proteins.⁸⁰ The recently published compound MGH-CP1 covalently modifies and inhibits TEAD family proteins similarly to TED-347, but with increased potency. MGH-CP1 inhibits autopalmitylation of TEAD2/4 but does not affect autopalmitylation of ZDHCC family palmityl acetyltransferases, suggesting that the compound is selective for TEAD proteins.⁸¹ The authors also find that TEAD2 adopts similar conformations when bound to palmitylate or MGH-CP1, further suggesting that mimicking regulation by naturally occurring PTMs can be a successful strategy for selectively targeting dynamic proteins. In another example, covalent allosteric inhibitors were designed to selectively target isoforms of protein kinase Akt. There are three different isoforms of Akt, each with a unique intracellular location and function; thus, highly selective probe compounds are needed to study these proteins. The covalent compound borussertib alkylates a cysteine residue located within an interdomain pocket between the PH domain and kinase domain of Akt proteins, irreversibly stabilizing the inactive conformation. Quambusch et al. observed slight differences around this allosteric pocket between isoforms and

designed a library of compounds based on the pyrazinone scaffold of borussertib that could engage these subtle sequence changes. With this approach, they were able to identify compounds with high selectivity for Akt1 and Akt2.⁸² This study highlights paths to attain high selectivity and potency with allosteric modulators.

In the previous example, the structural differences exploited by selective small molecules at the allosteric site of Akt isoforms mainly occurred within a loop formed by residues 259-273. Often the most dynamic components of a protein structure, such as loops, linkers, and flexible helices, are effective sites for allosteric modulation and improved selectivity. Recent work targeting the dynamic coactivator Mediator subunit Med25 also highlights this point. Henderson et al. show that a small molecule, Compound 22, can allosterically regulate binding of transcriptional activators to Med25 via covalent modification of a distal cysteine residue. This cysteine residue is located adjacent to regions predicted to be highly flexible via structural modelling. Additionally, compound 22 influences the flexible substructures within Med25, also known to be perturbed by interactions with transcriptional activation domains.⁸³ Thus, by targeting dynamic regions of a protein, even large PPI interfaces can be inhibited.

Another example illustrating that targeting loop modules can improve inhibitor selectivity is the discovery of LY3154207, a potent inhibitor of the Human Dopamine receptor, D1, presented by Lilly. Classical approaches to targeting D1 receptors have mainly been orthosteric, and candidate modulators bind similarly to the natural substrate dopamine.⁸⁴ However, as D1 and D5 receptors have high structural similarity, this approach leads to selectivity issues. The authors found that LY3154207 and its analogs bound to a novel binding site in the intracellular loop 2 (ICL2), allosterically inhibiting binding of dopamine. Excitingly, this inhibitor was highly selective for the D1 receptor over the closely related D5 receptor (>2700-fold). LY3154207 demonstrated over 1000-fold selectivity for the D1 receptor when tested against a panel of 40 additional targets.⁸⁵ In another recent example targeting a transcription factor, fosfosal, a well-documented clinical prodrug, was found to inhibit the SH2 domain of STAT5b over its homologue STAT5a. Mutational analysis showed this selectivity was dictated by residues within an adjacent linker domain rather than within the SH2 domain itself.⁸⁶ This work illustrates the role that the STAT linker region in SH2 domain function, and targeting the linker with small molecules is thus an avenue for selective inhibition of STAT PPIs. This same group has identified additional

molecules selective for both STAT5a and STAT5b utilizing the same concept (Elumalai et al., 2015; Gräb et al., 2019).

1.6 Techniques for identification of selective dynamic protein complex modulators

The examples discussed above demonstrate the power and promise of allosteric regulators for modulating dynamic proteins, particularly in terms of selectivity. However, in the majority of these examples, the mechanism of action was not the result of the discovery method but rather serendipitous. Here we outline emerging techniques and strategies that facilitate screens more likely to produce allosteric and ultimately selective inhibitors.

Inhibitor identification from well-structured proteins allows for easier structural characterization and lead compound optimization. Dynamic proteins, by definition alone, provide a steep challenge, with many of the common experimental techniques requiring improvements before being applied to proteins that are highly mobile and lack definitive structure. Thus, advancements in structural biology and biophysics are key when it comes to understanding selectivity in dynamic protein complexes. Recent improvements in techniques such as mass spectrometry,⁸⁹ NMR,⁹⁰ and cryogenic electron microscopy (cryo-EM)^{91,92} have allowed a high-resolution view of multiprotein complexes and aided in our understanding of composition as well as identification of key PPIs. For example, Khattabi et al. reported a 5.9 Å cryo-EM structure mapping the entire mammalian Mediator complex (~4 MDa). This included, for the first time, mapping of the exchangeable and conformationally dynamic proteins within the tail region of the complex.⁹³ Many of the tail proteins are the primary targets of transcription factors and these results provide key insights into how to target this class of proteins successfully. More recently and even higher resolution, less than 4 Å, structure of human Mediator bound to the transcriptional preinitiation complex has been reported, providing key insights into many key protein interactions.⁹⁴ In another exciting example, time resolved cryoEM was used to obtain a near-atomic-resolution view of the conformational changes that drive and regulate subunit assembly, initiation factor dissociation, and fMet-tRNA positioning during the formation of the 70S elongation-competent complex in bacteria.⁹⁵ This type of approach allows dissection of the specific timing and order of conformational changes contributing to the mechanism and regulation of large and dynamic multiprotein systems. These examples highlight how improved

understanding of dynamic protein complexes using these techniques will allow researchers to identify more promising regions to target.

Increased complexity of an assay system is characteristic of techniques that allow for enrichment of selective hits from screening. Grey box screening is a prominent example. In this method, multiprotein complexes are reconstituted in vitro and subjected to HTS with the goal of identifying compounds that effect biochemical properties, such as enzymatic activity, of the complex. This approach has identified specific chemical modulators of dynamic proteins within complexes such as Hsp70,^{96,97} Hsp90,⁹⁸ and regulators of G-protein signaling.⁹⁹ This method has been thoroughly reviewed in the literature.^{33,39} Mass spectrometry approaches can also allow for multiplexing of protein assays. Collision-induced unfolding (CIU) can distinguish individual protein and protein complex ions through their distinct unfolding pathways in the gas phase, allowing for analysis of multiprotein and protein-ligand complexes.¹⁰⁰ Combining CIU with collision induced dissociation (CID) mass spec analysis (Figure 1.6) distinguished ATP competitive from allosteric kinase inhibitors of the tyrosine kinase Abl.¹⁰¹ There are also many other examples utilizing mass spectrometry to analyze intact multiprotein complexes.⁸⁹ Application of these techniques to dynamic targets in high throughput screens can streamline identification of selective modulators. Molecules displaying selectivity, as seen with the examples in the previous section, often act allosterically, and thus characterization of their effects on target proteins will provide insights into allosteric regions to focus on in future studies.

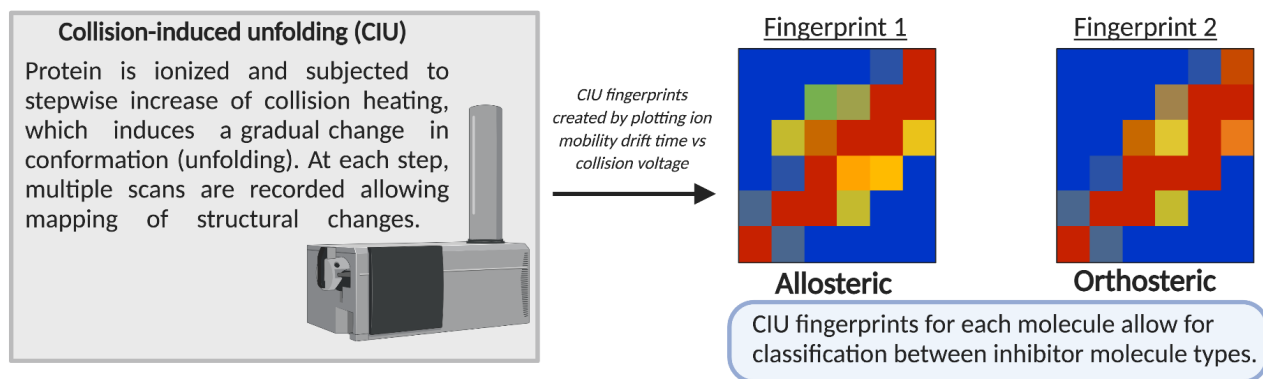


Figure 1.6. Mass spectrometry as a technique for identification of dynamic protein modulators. Collision induced unfolding (CIU) can be utilized as a medium-high throughput approach to determine binding profiles protein•ligand complexes. If the CIU profile is known for a certain type of protein•ligand complexes, for example an allosteric signature or orthosteric signature, a screen can be conducted and molecules can be sorted by distinct fingerprints.¹⁰² Figure created using BioRender.

Classical approaches, such as Forster resonance energy transfer (FRET) and fluorescence polarization (FP), can be utilized for HTS of dynamic protein targets where allosteric sites are unknown. To assess the selectivity of lead compounds, secondary screening against a selected panel of alternate targets is necessary.¹⁰³ This iterative screening process is highlighted in work from Majmudar et al. Hits from a FP-based HTS for inhibitors of CBP KIX were subjected to multiple rounds of secondary screening against increasingly relevant PPIs, from non-transcriptional PPIs to DNA-binding domain•DNA interactions to related coactivator•activator interactions. This led to the identification of the CBP KIX selective molecule lobaric acid.¹⁰⁴ Library complexity is also key for screening dynamic proteins. Diversity oriented synthesis and DNA Encoded Libraries (DEL) have allowed for the generation of huge billion-trillion compound libraries that can be screened in a one pot format.^{105–107} Other notable library developments with the ability to impact dynamic protein probe discovery include diverse natural product focused libraries, protein-protein interaction focused libraries, and cheminformatic based design and optimization.^{108–110}

Targeted fragment approaches have been successful in discovering allosteric sites and modulators of diverse proteins. Fragment based screening does not rely on identification of a single high affinity hit, rather it provides the opportunity to diversify lead scaffolds into directed libraries or link together lower affinity binders, acting at unique sites on the target protein, to construct a more potent compound.^{111,112} The Cravatt lab has pioneered cell-based approaches to fragment screening, providing the advantage of being able to study a protein of interest in its native environment.^{113,114} This would be particularly useful for dynamic proteins that are challenging to work with in vitro. Tethering is a site-directed fragment technique that utilizes a disulfide moiety within each fragment to facilitate localization to protein binding sites via covalent bond formation with adjacent cysteine residues. Tethering has enabled the identification of allosteric modulators for many dynamic protein targets, from enzymes such as PDK1 and Ras to coactivators such as CBP and Med25.^{83,115–117} In a recent example targeting the scaffolding protein 14-3-3 PPI, tethering was used to identify orthosteric stabilizers of the 14-3-3•ER α complex. These stabilizers were able to enhance the interaction by up to 40-fold while demonstrating selectivity for ER α over other 14-3-3 interacting proteins.¹¹⁸

Screening methods that provide mechanistic insight into small molecule binding have proven highly useful for conformationally dynamic proteins, providing the ability to select for molecules

that influence desired regions of the target. NMR-based methods, including Protein Observed Fluorine (PrOF) and $^1\text{H}^{15}\text{N}$ HSQC can provide information on both binding site location and impact on protein dynamics and structure.^{119–121} Thus, one can select for compounds during screening that show specific alterations in structure and/or dynamics. Because throughput in NMR based screening methods can be a limitation, this approach is extremely powerful when coupled with computational screening methods. Recent work by Gupta et al. combined computational and NMR screening approaches with the goal of targeting known allosteric sites on mutant and wildtype KRAS. Gupta et al. started from a virtual library of 76 million compounds and identified nine compounds that lead to $^1\text{H}^{15}\text{N}$ HSQC chemical shift perturbations of residues near the functionally responsive switch loop. Although the authors do not analyze the functional effect of their lead compound, E22, they find that it binds nearly 100-fold tighter to GNP-bound Ras compared to GDP bound Ras and hypothesize that the molecule may affect GEF-mediated GDP/GTP exchange, similar to other indole compounds that target KRAS.¹²² Combinations of computational and experimental approaches, as this example illustrates, have and will continue to be important for dynamic protein probe discovery. Notably, computational studies of allostery, with a focus on identification of allosteric sites and discovery of potent allosteric inhibitors, have been extensively performed in recent years with many successes.^{123–125}

Thermal stability assays, such as differential scanning fluorimetry (DSF), are powerful screening options for dynamic protein targets.^{126–128} While DSF is a technique dating back to 1991, it recently has been used to screen some challenging targets such as STAT proteins (STAT1, STAT3, and STAT5),^{129,130} nuclear receptors,¹³¹ and chaperone proteins.^{132,133} Not only is this method simple, inexpensive, and amenable to HTS, it provides insights into the effect of a compound on the thermal stability of the target protein, making it possible to select for compounds with the desired stabilizing or destabilizing effect (Figure 1.7). Thermal stability assays have the added benefit of not requiring knowledge of binding partners. Recently, groups have been working to improve understanding of the theoretical underpinnings of DSF and further interrogate the potential information that can be obtained from DSF melting curves.^{134,135} This work has made it clear that more than just a melting temperature (T_m) emerges from these experiments. For example, the shape of the melt curve can provide some insight into ligand binding even if T_m shifts are not observed. Additionally, innovations to the technique suggest the ability to multiplex. DSF-GFP, a technique pioneered in 2012, involves labelling the target protein of interest with GFP

and using the change in GFP fluorescence over a temperature gradient as a readout rather than solvatochromic dye.¹³⁶ This approach allows for the selective measurement of target protein T_m in the presence of other proteins.

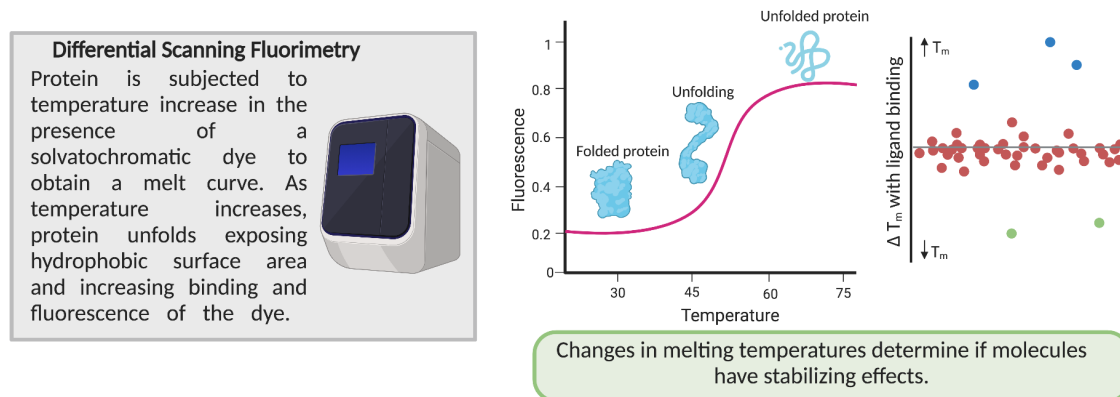


Figure 1.7. Differential scanning fluorimetry (DSF) as a technique for identification of dynamic protein modulators. DSF is a high throughput method for screening proteins. This method monitors protein unfolding as a function of temperature. As the protein unfolds, a solvatochromic dye binds and fluorescence increases. The melting temperature (T_m) of the protein is interpreted as the inflection point of the resulting melt curve.¹²⁸ While allosteric effects of small molecules cannot be determined using this method, their effect on protein stability can be determined. In a HTS, molecules can be sorted by their ability to increase or decrease protein T_m . Figure created using BioRender.

The cellular thermal shift assay (CETSA) is another technique that interrogates changes in protein stability and can also be extremely useful for screening dynamic proteins.¹³⁷ A benefit of this approach is the ability to screen the protein in a more native environment, such as in cell lysates. Recent advances to CETSA techniques have allowed for use in HTS. To date, there have been multiple HT-CETSA approaches published, all focusing on altering protein detection methods to allow for the classic CETSA western blot readout to be adapted to 384 or 1536 well format.¹³⁸ For example, Shaw et al. applied HT-CETSA via AlphaScreen technology to conduct a screen for modulators of BRAF and PARP1 through screening 896 and 6288 compounds, respectively.¹³⁹ In a recent success story with a challenging target, a small molecule, NPD10084, was found to thermally destabilize PKM2, which can function as both a kinase and a transcriptional coactivator. This interaction with PKM2 leads to inhibition of PPIs with STAT3 and beta-catenin as well as altered expression of PKM2 regulated genes. Notably, the authors outlined a novel assay, which they termed 2D gel electrophoresis-based proteome-wide CETSA, or 2DE-CETSA, that allowed for proteome wide screening for small molecule target identification.¹⁴⁰

1.7 Emergent technologies and alternative approaches to selectivity

As can be seen from the outlined examples, it is a particularly exciting time for those who study and/or target dynamic proteins. Discovery methods that are tailored for the distinct biophysical characteristics of dynamic proteins, as well as those that more closely recapitulate the native context, have been highly successful. In addition to these screening techniques, thoroughly interrogating the selectivity of lead compounds on a global scale will be integral in deriving effective biological probes of dynamic protein targets. Techniques such as thermal proteome profiling, activity based protein profiling and Drug Affinity Responsive Target Stability (DARTS) allow determination of potential off target effects for a chemical probe in a cellular context.^{141–143} The resulting generation of more selective and potent small molecule modulators now enable the chemical genetic dissection of critical cellular machines. For example, studies with full length p300 and other various multidomain constructs show BRD inhibitors can affect HAT activity, suggesting interdomain crosstalk does occur.^{144,145} Interdomain crosstalk has been observed in other dynamic systems, particularly in nuclear receptors such as the thyroid hormone receptor and ER α , where the DNA binding via the DNA binding domain (DBD) can be enhanced by changes in the ligand binding domain (LBD).¹⁴⁶ Further exciting studies focusing on the global effects of selective dynamic protein modulators on full length proteins or intact protein complexes are sure to come in the future and improve our understanding of interdomain communication in these systems.

Another interesting avenue for dynamic proteins that has yet to be fully explored is proteolysis targeting chimeras (PROTACs). PROTACs could be particularly advantageous in the field of dynamic protein complexes because the starting scaffold does not need to functionally affect any interactions, as it can be challenging to find modulators with low IC₅₀s or EC₅₀s for certain targets. Additionally, even non-specific inhibitors can be starting points for PROTACS. Recent work has shown specificity can be engineered into a PROTAC starting from a multi-target warhead by exploiting differences in the interface between the protein of interest and the E3 ligase.^{147,148} It is possible that through PROTACS, some of the challenges that come with identifying selective modulators can be averted. With regards to dynamic protein complexes, one could imagine using a PROTAC to completely alter complex composition or skew complexes to favor one exchangeable subunit over others. These and other novel selective mechanisms to chemically target dynamic proteins are sure to be on the horizon.

1.8 Dissertation Summary

The overall goal of the research presented in this thesis is to identify novel probes of challenging dynamic protein targets. Towards this goal, we begin in Chapter 2 by describing the discovery and validation of norstictic acid as the first chemical probe for Med25, a dynamic coactivator and subunit of the transcriptional Mediator complex. In Chapter 3, we build off of the screening approach used in Chapter 2 and present the development of a multiplex fluorescence polarization-based screening method that allows for selectivity to be assessed at the primary screening level. This method is then used to uncover additional novel chemical modulators of both Med25 and the KIX domain of CBP, another dynamic coactivator protein. In Chapter 4, we apply differential scanning fluorimetry to analyze the effects of ligand interactions, both small molecules and native peptides, on coactivator thermal stability. We hypothesize that this system will be uniquely suited to identify allosteric modulators of dynamic protein complexes. Finally, in Chapter 5 we develop a biochemical screen for the human serine protease TMPRSS2, which plays a key role in SARS-CoV-2 viral infection. TMPRSS2 is a challenging target because expression and purification is challenging, as is extensively noted in the literature. We present an expression and purification procedure in *E. coli* to isolate active TMPRSS2 protease domain. Additionally, while inhibitors of TMPRSS2 exist, there are no selective chemical probes and thus its broader biological roles remain poorly characterized. By integration of biochemical and computational methodologies, we are able to identify novel reversible TMPRSS2 inhibitors.

1.9 References

- (1) Frye, S. V. The Art of the Chemical Probe. *Nat. Chem. Biol.* **2010**, *6* (3), 159–161. <https://doi.org/10.1038/nchembio.296>.
- (2) Bunnage, M. E.; Chekler, E. L. P.; Jones, L. H. Target Validation Using Chemical Probes. *Nat. Chem. Biol.* **2013**, *9* (4), 195–199. <https://doi.org/10.1038/nchembio.1197>.
- (3) Garbaccio, R. M.; Parmee, E. R. The Impact of Chemical Probes in Drug Discovery: A Pharmaceutical Industry Perspective. *Cell Chem. Biol.* **2016**, *23* (1), 10–17. <https://doi.org/10.1016/j.chembiol.2015.11.011>.
- (4) Nero, T. L.; Morton, C. J.; Holien, J. K.; Wielens, J.; Parker, M. W. Oncogenic Protein Interfaces: Small Molecules, Big Challenges. *Nat. Rev. Cancer* **2014**, *14* (4), 248–262. <https://doi.org/10.1038/nrc3690>.
- (5) Ran, X.; Gestwicki, J. E. Inhibitors of Protein–Protein Interactions (PPIs): An Analysis of Scaffold Choices and Buried Surface Area. *Curr. Opin. Chem. Biol.* **2018**, *44*, 75–86. <https://doi.org/10.1016/j.cbpa.2018.06.004>.
- (6) Barrios-Rodiles, M.; Brown, K. R.; Ozdamar, B.; Bose, R.; Liu, Z.; Donovan, R. S.; Shinjo, F.; Liu, Y.; Dembowy, J.; Taylor, I. W.; Luga, V.; Przulj, N.; Robinson, M.; Suzuki, H.; Hayashizaki, Y.; Jurisica, I.; Wrana, J. L. High-Throughput Mapping of a Dynamic Signaling Network in Mammalian Cells. *Science* **2005**, *307* (5715), 1621–1625. <https://doi.org/10.1126/science.1105776>.
- (7) Kholodenko, B. N. Cell-Signalling Dynamics in Time and Space. *Nat. Rev. Mol. Cell Biol.* **2006**, *7* (3), 165–176. <https://doi.org/10.1038/nrm1838>.
- (8) Yeger-Lotem, E.; Sattath, S.; Kashtan, N.; Itzkovitz, S.; Milo, R.; Pinter, R. Y.; Alon, U.; Margalit, H. Network Motifs in Integrated Cellular Networks of Transcription–Regulation and Protein–Protein Interaction. *Proc. Natl. Acad. Sci.* **2004**, *101* (16), 5934–5939. <https://doi.org/10.1073/pnas.0306752101>.
- (9) Stein, A.; Pache, R. A.; Bernadó, P.; Pons, M.; Aloy, P. Dynamic Interactions of Proteins in Complex Networks: A More Structured View. *FEBS J.* **2009**, 5390–5405. [https://doi.org/10.1111/j.1742-4658.2009.07251.x@10.1002/\(ISSN\)1742-4658\(CAT\)FreeReviewContent\(VI\)Reviews0809](https://doi.org/10.1111/j.1742-4658.2009.07251.x@10.1002/(ISSN)1742-4658(CAT)FreeReviewContent(VI)Reviews0809).
- (10) Heo, M.; Maslov, S.; Shakhnovich, E. Topology of Protein Interaction Network Shapes Protein Abundances and Strengths of Their Functional and Nonspecific Interactions. *Proc. Natl. Acad. Sci.* **2011**, *108* (10), 4258–4263. <https://doi.org/10.1073/pnas.1009392108>.
- (11) Jones, S.; Thornton, J. M. Principles of Protein-Protein Interactions. *Proc. Natl. Acad. Sci. U. S. A.* **1996**, *93* (1), 13–20. <https://doi.org/10.1073/pnas.93.1.13>.
- (12) Khan, S. H.; Ahmad, F.; Ahmad, N.; Flynn, D. C.; Kumar, R. Protein-Protein Interactions: Principles, Techniques, and Their Potential Role in New Drug Development. *J. Biomol. Struct. Dyn.* **2011**, *28* (6), 929–938. <https://doi.org/10.1080/07391102.2011.10508619>.
- (13) Ku, B.; Liang, C.; Jung, J. U.; Oh, B.-H. Evidence That Inhibition of BAX Activation by BCL-2 Involves Its Tight and Preferential Interaction with the BH3 Domain of BAX. *Cell Res.* **2011**, *21* (4), 627–641. <https://doi.org/10.1038/cr.2010.149>.
- (14) Manero, F.; Gautier, F.; Gallenne, T.; Cauquil, N.; Grée, D.; Cartron, P.-F.; Geneste, O.; Grée, R.; Vallette, F. M.; Juin, P. The Small Organic Compound HA14-1 Prevents Bcl-2

- Interaction with Bax to Sensitize Malignant Glioma Cells to Induction of Cell Death. *Cancer Res.* **2006**, *66* (5), 2757–2764. <https://doi.org/10.1158/0008-5472.CAN-05-2097>.
- (15) Samuel, S.; Tumilasci, V. F.; Oliere, S.; Li  n-Anh Nguy  n, T.; Shamy, A.; Bell, J.; Hiscott, J. VSV Oncolysis in Combination With the BCL-2 Inhibitor Obatoclax Overcomes Apoptosis Resistance in Chronic Lymphocytic Leukemia. *Mol. Ther.* **2010**, *18* (12), 2094–2103. <https://doi.org/10.1038/mt.2010.188>.
 - (16) Demarest, S. J.; Martinez-Yamout, M.; Chung, J.; Chen, H.; Xu, W.; Dyson, H. J.; Evans, R. M.; Wright, P. E. Mutual Synergistic Folding in Recruitment of CBP/P300 by P160 Nuclear Receptor Coactivators. *Nature* **2002**, *415* (6871), 549–553. <https://doi.org/10.1038/415549a>.
 - (17) Demarest, S. J.; Deechongkit, S.; Dyson, H. J.; Evans, R. M.; Wright, P. E. Packing, Specificity, and Mutability at the Binding Interface between the P160 Coactivator and CREB-Binding Protein. *Protein Sci. Publ. Protein Soc.* **2004**, *13* (1), 203–210. <https://doi.org/10.1110/ps.03366504>.
 - (18) Feng, H.; Zhou, B.-R.; Bai, Y. Binding Affinity and Function of the Extremely Disordered Protein Complex Containing Human Linker Histone H1.0 and Its Chaperone ProT  . *Biochemistry* **2018**, *57* (48), 6645–6648. <https://doi.org/10.1021/acs.biochem.8b01075>.
 - (19) Borgia, A.; Borgia, M. B.; Bugge, K.; Kissling, V. M.; Heidarsson, P. O.; Fernandes, C. B.; Sottini, A.; Soranno, A.; Buholzer, K. J.; Nettels, D.; Kragelund, B. B.; Best, R. B.; Schuler, B. Extreme Disorder in an Ultrahigh-Affinity Protein Complex. *Nature* **2018**, *555* (7694), 61–66. <https://doi.org/10.1038/nature25762>.
 - (20) Wang, W.; Wang, D. Extreme Fuzziness: Direct Interactions between Two IDPs. *Biomolecules* **2019**, *9* (3). <https://doi.org/10.3390/biom9030081>.
 - (21) Mittag, T.; Kay, L. E.; Forman-Kay, J. D. Protein Dynamics and Conformational Disorder in Molecular Recognition. *J. Mol. Recognit. JMR* **2010**, *23* (2), 105–116. <https://doi.org/10.1002/jmr.961>.
 - (22) 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**, *291* (13), 6714–6722. <https://doi.org/10.1074/jbc.R115.692020>.
 - (23) Breen, M. E.; Mapp, A. K. Modulating the Masters: Chemical Tools to Dissect CBP and P300 Function. *Curr. Opin. Chem. Biol.* **2018**, *45*, 195–203. <https://doi.org/10.1016/j.cbpa.2018.06.005>.
 - (24) Mapp, A. K.; Pricer, R.; Sturlis, S. Targeting Transcription Is No Longer a Quixotic Quest. *Nat. Chem. Biol.* **2015**, *11* (12), 891–894. <https://doi.org/10.1038/nchembio.1962>.
 - (25) Euskirchen, G. M.; Auerbach, R. K.; Davidov, E.; Gianoulis, T. A.; Zhong, G.; Rozowsky, J.; Bhardwaj, N.; Gerstein, M. B.; Snyder, M. Diverse Roles and Interactions of the SWI/SNF Chromatin Remodeling Complex Revealed Using Global Approaches. *PLoS Genet.* **2011**, *7* (3). <https://doi.org/10.1371/journal.pgen.1002008>.
 - (26) Walter, S.; Buchner, J. Molecular Chaperones—Cellular Machines for Protein Folding. *Angew. Chem. Int. Ed.* **2002**, *41* (7), 1098–1113. [https://doi.org/10.1002/1521-3773\(20020402\)41:7<1098::AID-ANIE1098>3.0.CO;2-9](https://doi.org/10.1002/1521-3773(20020402)41:7<1098::AID-ANIE1098>3.0.CO;2-9).
 - (27) Batada, N. N.; Hurst, L. D.; Tyers, M. Evolutionary and Physiological Importance of Hub Proteins. *PLoS Comput. Biol.* **2006**, *2* (7). <https://doi.org/10.1371/journal.pcbi.0020088>.

- (28) Muller, P.; Ruckova, E.; Halada, P.; Coates, P. J.; Hrstka, R.; Lane, D. P.; Vojtesek, B. C-Terminal Phosphorylation of Hsp70 and Hsp90 Regulates Alternate Binding to Co-Chaperones CHIP and HOP to Determine Cellular Protein Folding/Degradation Balances. *Oncogene* **2013**, *32* (25), 3101–3110. <https://doi.org/10.1038/onc.2012.314>.
- (29) Fuxreiter, M.; Tompa, P.; Simon, I.; Uversky, V. N.; Hansen, J. C.; Asturias, F. J. Malleable Machines Take Shape in Eukaryotic Transcriptional Regulation. *Nat. Chem. Biol.* **2008**, *4* (12), 728. <https://doi.org/10.1038/nchembio.127>.
- (30) Gavin, A.-C.; Superti-Furga, G. Protein Complexes and Proteome Organization from Yeast to Man. *Curr. Opin. Chem. Biol.* **2003**, *7* (1), 21–27. [https://doi.org/10.1016/S1367-5931\(02\)00007-8](https://doi.org/10.1016/S1367-5931(02)00007-8).
- (31) Kasahara, K.; Terazawa, H.; Takahashi, T.; Higo, J. Studies on Molecular Dynamics of Intrinsically Disordered Proteins and Their Fuzzy Complexes: A Mini-Review. *Comput. Struct. Biotechnol. J.* **2019**, *17*, 712–720. <https://doi.org/10.1016/j.csbj.2019.06.009>.
- (32) Marsh, J. A.; Teichmann, S. A. Structure, Dynamics, Assembly, and Evolution of Protein Complexes. *Annu. Rev. Biochem.* **2015**, *84* (1), 551–575. <https://doi.org/10.1146/annurev-biochem-060614-034142>.
- (33) Thompson, A. D.; Dugan, A.; Gestwicki, J. E.; Mapp, A. K. Fine-Tuning Multiprotein Complexes Using Small Molecules. *ACS Chem. Biol.* **2012**, *7* (8), 1311–1320. <https://doi.org/10.1021/cb300255p>.
- (34) Fernández-Fernández, M. R.; Valpuesta, J. M. Hsp70 Chaperone: A Master Player in Protein Homeostasis. *Fl000Research* **2018**, *7*. <https://doi.org/10.12688/fl000research.15528.1>.
- (35) A. Assimon, V.; T. Gillies, A.; N. Rauch, J.; E. Gestwicki, J. Hsp70 Protein Complexes as Drug Targets. *Curr. Pharm. Des.* **2013**, *19* (3), 404–417. <https://doi.org/10.2174/138161213804143699>.
- (36) Cossins, B. P.; Lawson, A. D. G. Small Molecule Targeting of Protein–Protein Interactions through Allosteric Modulation of Dynamics. *Molecules* **2015**, *20* (9), 16435–16445. <https://doi.org/10.3390/molecules200916435>.
- (37) Smith, M. C.; Gestwicki, J. E. Features of Protein–Protein Interactions That Translate into Potent Inhibitors: Topology, Surface Area and Affinity. *Expert Rev. Mol. Med.* **2012**, *14*. <https://doi.org/10.1017/erm.2012.10>.
- (38) Arkin, M. R.; Tang, Y.; Wells, J. A. Small-Molecule Inhibitors of Protein-Protein Interactions: Progressing toward the Reality. *Chem. Biol.* **2014**, *21* (9), 1102–1114. <https://doi.org/10.1016/j.chembiol.2014.09.001>.
- (39) Cesa, L. C.; Mapp, A. K.; Gestwicki, J. E. Direct and Propagated Effects of Small Molecules on Protein–Protein Interaction Networks. *Front. Bioeng. Biotechnol.* **2015**, *3*. <https://doi.org/10.3389/fbioe.2015.00119>.
- (40) Semmler, L.; Reiter-Brennan, C.; Klein, A. BRCA1 and Breast Cancer: A Review of the Underlying Mechanisms Resulting in the Tissue-Specific Tumorigenesis in Mutation Carriers. *J. Breast Cancer* **2019**, *22* (1), 1–14. <https://doi.org/10.4048/jbc.2019.22.e6>.
- (41) Takaoka, M.; Miki, Y. BRCA1 Gene: Function and Deficiency. *Int. J. Clin. Oncol.* **2018**, *23* (1), 36–44. <https://doi.org/10.1007/s10147-017-1182-2>.
- (42) Lee, T. I.; Young, R. A. Transcriptional Regulation and Its Misregulation in Disease. *Cell* **2013**, *152* (6), 1237–1251. <https://doi.org/10.1016/j.cell.2013.02.014>.

- (43) Wang, F.; Marshall, C. B.; Ikura, M. Transcriptional/Epigenetic Regulator CBP/P300 in Tumorigenesis: Structural and Functional Versatility in Target Recognition. *Cell. Mol. Life Sci.* **2013**, *70* (21), 3989–4008. <https://doi.org/10.1007/s00018-012-1254-4>.
- (44) Gill, G.; Ptashne, M. Negative Effect of the Transcriptional Activator GAL4. *Nature* **1988**, *334* (6184), 721–724. <https://doi.org/10.1038/334721a0>.
- (45) Wightman, R.; Bell, R.; Reece, R. J. Localization and Interaction of the Proteins Constituting the GAL Genetic Switch in *Saccharomyces Cerevisiae*. *Eukaryot. Cell* **2008**, *7* (12), 2061–2068. <https://doi.org/10.1128/EC.00261-08>.
- (46) Zaika, A.; Marchenko, N.; Moll, U. M. Cytoplasmically “Sequestered” Wild Type P53 Protein Is Resistant to Mdm2-Mediated Degradation. *J. Biol. Chem.* **1999**, *274* (39), 27474–27480. <https://doi.org/10.1074/jbc.274.39.27474>.
- (47) Kalabova, D.; Filandr, F.; Alblova, M.; Petrvalska, O.; Horvath, M.; Man, P.; Obsil, T.; Obsilova, V. 14-3-3 Protein Binding Blocks the Dimerization Interface of Caspase-2. *FEBS J.* **2020**. <https://doi.org/10.1111/febs.15215>.
- (48) Silhan, J.; Vacha, P.; Strnadova, P.; Vecer, J.; Herman, P.; Sulc, M.; Teisinger, J.; Obsilova, V.; Obsil, T. 14-3-3 Protein Masks the DNA Binding Interface of Forkhead Transcription Factor FOXO4. *J. Biol. Chem.* **2009**, *284* (29), 19349–19360. <https://doi.org/10.1074/jbc.M109.002725>.
- (49) Wang, Y.; Shen, J.; Arenzana, N.; Tirasophon, W.; Kaufman, R. J.; Prywes, R. Activation of ATF6 and an ATF6 DNA Binding Site by the Endoplasmic Reticulum Stress Response. *J. Biol. Chem.* **2000**, *275* (35), 27013–27020.
- (50) Hahn, S. Phase Separation, Protein Disorder, and Enhancer Function. *Cell* **2018**, *175* (7), 1723–1725. <https://doi.org/10.1016/j.cell.2018.11.034>.
- (51) Boija, A.; Klein, I. A.; Sabari, B. R.; Dall’Agnese, A.; Coffey, E. L.; Zamudio, A. V.; Li, C. H.; Shrinivas, K.; Manteiga, J. C.; Hannett, N. M.; Abraham, B. J.; Afeyan, L. K.; Guo, Y. E.; Rimel, J. K.; Fant, C. B.; Schuijers, J.; Lee, T. I.; Taatjes, D. J.; Young, R. A. Transcription Factors Activate Genes through the Phase-Separation Capacity of Their Activation Domains. *Cell* **2018**, *175* (7), 1842–1855.e16. <https://doi.org/10.1016/j.cell.2018.10.042>.
- (52) Lu, Y.; Wu, T.; Gutman, O.; Lu, H.; Zhou, Q.; Henis, Y. I.; Luo, K. Phase Separation of TAZ Compartmentalizes the Transcription Machinery to Promote Gene Expression. *Nat. Cell Biol.* **2020**, *22* (4), 453–464. <https://doi.org/10.1038/s41556-020-0485-0>.
- (53) Trojanowski, J.; Frank, L.; Rademacher, A.; Grigaitis, P.; Rippe, K. Transcription Activation Is Enhanced by Multivalent Interactions Independent of Liquid-Liquid Phase Separation. *bioRxiv* **2021**, 2021.01.27.428421. <https://doi.org/10.1101/2021.01.27.428421>.
- (54) Yang, W.; Hong, Y. H.; Shen, X.-Q.; Frankowski, C.; Camp, H. S.; Leff, T. Regulation of Transcription by AMP-Activated Protein Kinase PHOSPHORYLATION OF P300 BLOCKS ITS INTERACTION WITH NUCLEAR RECEPTORS. *J. Biol. Chem.* **2001**, *276* (42), 38341–38344. <https://doi.org/10.1074/jbc.C100316200>.
- (55) Duan, G.; Walther, D. The Roles of Post-Translational Modifications in the Context of Protein Interaction Networks. *PLoS Comput. Biol.* **2015**, *11* (2). <https://doi.org/10.1371/journal.pcbi.1004049>.
- (56) Nussinov, R.; Tsai, C.-J.; Xin, F.; Radivojac, P. Allosteric Post-Translational Modification Codes. *Trends Biochem. Sci.* **2012**, *37* (10), 447–455. <https://doi.org/10.1016/j.tibs.2012.07.001>.

- (57) Walsh, S. T. R. A Biosensor Study Indicating That Entropy, Electrostatics, and Receptor Glycosylation Drive the Binding Interaction between Interleukin-7 and Its Receptor. *Biochemistry* **2010**, *49* (40), 8766–8778. <https://doi.org/10.1021/bi101050h>.
- (58) Chan, P.; Han, X.; Zheng, B.; DeRan, M.; Yu, J.; Jarugumilli, G. K.; Deng, H.; Pan, D.; Luo, X.; Wu, X. Autopalmitoylation of TEAD Proteins Regulates Transcriptional Output of Hippo Pathway. *Nat. Chem. Biol.* **2016**, *12* (4), 282–289. <https://doi.org/10.1038/nchembio.2036>.
- (59) Stetz, G.; Tse, A.; Verkhivker, G. M. Dissecting Structure-Encoded Determinants of Allosteric Cross-Talk between Post-Translational Modification Sites in the Hsp90 Chaperones. *Sci. Rep.* **2018**, *8* (1), 6899. <https://doi.org/10.1038/s41598-018-25329-4>.
- (60) Swain, J. F.; Gierasch, L. M. The Changing Landscape of Protein Allostery. *Curr. Opin. Struct. Biol.* **2006**, *16* (1), 102–108. <https://doi.org/10.1016/j.sbi.2006.01.003>.
- (61) Tsai, C.-J.; Sol, A. del; Nussinov, R. Protein Allostery , Signal Transmission and Dynamics: A Classification Scheme of Allosteric Mechanisms. *Mol. Biosyst.* **2009**, *5* (3), 207–216. <https://doi.org/10.1039/B819720B>.
- (62) Abrusán, G.; Marsh, J. A. Ligand-Binding-Site Structure Shapes Allosteric Signal Transduction and the Evolution of Allostery in Protein Complexes. *Mol. Biol. Evol.* **2019**, *36* (8), 1711–1727. <https://doi.org/10.1093/molbev/msz093>.
- (63) Thal, D. M.; Glukhova, A.; Sexton, P. M.; Christopoulos, A. Structural Insights into G-Protein-Coupled Receptor Allostery. *Nature* **2018**, *559* (7712), 45–53. <https://doi.org/10.1038/s41586-018-0259-z>.
- (64) Peng, H.; Prokop, J.; Karar, J.; Park, K.; Cao, L.; Harbour, J. W.; Bowcock, A. M.; Malkowicz, S. B.; Cheung, M.; Testa, J. R.; Rauscher, F. J. Familial and Somatic BAP1 Mutations Inactivate ASXL1/2-Mediated Allosteric Regulation of BAP1 Deubiquitinase by Targeting Multiple Independent Domains. *Cancer Res.* **2018**, *78* (5), 1200–1213. <https://doi.org/10.1158/0008-5472.CAN-17-2876>.
- (65) Papaleo, E.; Saladino, G.; Lambrughi, M.; Lindorff-Larsen, K.; Gervasio, F. L.; Nussinov, R. The Role of Protein Loops and Linkers in Conformational Dynamics and Allostery. *Chem. Rev.* **2016**, *116* (11), 6391–6423. <https://doi.org/10.1021/acs.chemrev.5b00623>.
- (66) Henkel, T.; Zabel, U.; van Zee, K.; Müller, J. M.; Fanning, E.; Baeuerle, P. A. Intramolecular Masking of the Nuclear Location Signal and Dimerization Domain in the Precursor for the P50 NF-Kappa B Subunit. *Cell* **1992**, *68* (6), 1121–1133. [https://doi.org/10.1016/0092-8674\(92\)90083-o](https://doi.org/10.1016/0092-8674(92)90083-o).
- (67) Furnham, N.; Sillitoe, I.; Holliday, G. L.; Cuff, A. L.; Laskowski, R. A.; Orengo, C. A.; Thornton, J. M. Exploring the Evolution of Novel Enzyme Functions within Structurally Defined Protein Superfamilies. *PLoS Comput. Biol.* **2012**, *8* (3), e1002403. <https://doi.org/10.1371/journal.pcbi.1002403>.
- (68) Lu, S.; Huang, W.; Zhang, J. Recent Computational Advances in the Identification of Allosteric Sites in Proteins. *Drug Discov. Today* **2014**, *19* (10), 1595–1600. <https://doi.org/10.1016/j.drudis.2014.07.012>.
- (69) Nussinov, R.; Tsai, C.-J.; Csermely, P. Allo-Network Drugs: Harnessing Allostery in Cellular Networks. *Trends Pharmacol. Sci.* **2011**, *32* (12), 686–693. <https://doi.org/10.1016/j.tips.2011.08.004>.

- (70) Huang, W.; Lu, S.; Huang, Z.; Liu, X.; Mou, L.; Luo, Y.; Zhao, Y.; Liu, Y.; Chen, Z.; Hou, T.; Zhang, J. Allosite: A Method for Predicting Allosteric Sites. *Bioinformatics* **2013**, *29* (18), 2357–2359. <https://doi.org/10.1093/bioinformatics/btt399>.
- (71) Panjkovich, A.; Daura, X. Assessing the Structural Conservation of Protein Pockets to Study Functional and Allosteric Sites: Implications for Drug Discovery. *BMC Struct. Biol.* **2010**, *10* (1), 9. <https://doi.org/10.1186/1472-6807-10-9>.
- (72) Fang, Z.; Grütter, C.; Rauh, D. Strategies for the Selective Regulation of Kinases with Allosteric Modulators: Exploiting Exclusive Structural Features. *ACS Chem. Biol.* **2013**, *8* (1), 58–70. <https://doi.org/10.1021/cb300663j>.
- (73) Filippakopoulos, P.; Knapp, S. Targeting Bromodomains: Epigenetic Readers of Lysine Acetylation. *Nat. Rev. Drug Discov.* **2014**, *13* (5), 337–356. <https://doi.org/10.1038/nrd4286>.
- (74) Wootten, D.; Christopoulos, A.; Sexton, P. M. Emerging Paradigms in GPCR Allostery: Implications for Drug Discovery. *Nat. Rev. Drug Discov.* **2013**, *12* (8), 630–644. <https://doi.org/10.1038/nrd4052>.
- (75) Gao, Z.-G.; Jacobson, K. A. Allosteric Modulation and Functional Selectivity of G Protein-Coupled Receptors. *Drug Discov. Today Technol.* **2013**, *10* (2), e237–e243. <https://doi.org/10.1016/j.ddtec.2012.08.004>.
- (76) Leach, K.; Sexton, P. M.; Christopoulos, A. Allosteric GPCR Modulators: Taking Advantage of Permissive Receptor Pharmacology. *Trends Pharmacol. Sci.* **2007**, *28* (8), 382–389. <https://doi.org/10.1016/j.tips.2007.06.004>.
- (77) Beshore, D. C.; N. Di Marco, C.; Chang, R. K.; Greshock, T. J.; Ma, L.; Wittmann, M.; Seager, M. A.; Koeplinger, K. A.; Thompson, C. D.; Fuerst, J.; Hartman, G. D.; Bilodeau, M. T.; Ray, W. J.; Kuduk, S. D. MK-7622: A First-in-Class M1 Positive Allosteric Modulator Development Candidate. *ACS Med. Chem. Lett.* **2018**, *9* (7), 652–656. <https://doi.org/10.1021/acsmchemlett.8b00095>.
- (78) Karim, R. M.; Chan, A.; Zhu, J.-Y.; Schönbrunn, E. Structural Basis of Inhibitor Selectivity in the BRD7/9 Subfamily of Bromodomains. *J. Med. Chem.* **2020**. <https://doi.org/10.1021/acs.jmedchem.9b01980>.
- (79) Olp, M. D.; Sprague, D. J.; Goetz, C. J.; Kathman, S. G.; Wynia-Smith, S. L.; Shishodia, S.; Summers, S. B.; Xu, Z.; Statsyuk, A. V.; Smith, B. C. Covalent-Fragment Screening of BRD4 Identifies a Ligandable Site Orthogonal to the Acetyl-Lysine Binding Sites. *ACS Chem. Biol.* **2020**. <https://doi.org/10.1021/acschembio.0c00058>.
- (80) Bum-Erdene, K.; Zhou, D.; Gonzalez-Gutierrez, G.; Ghosayel, M. K.; Si, Y.; Xu, D.; Shannon, H. E.; Bailey, B. J.; Corson, T. W.; Pollok, K. E.; Wells, C. D.; Meroueh, S. O. Small-Molecule Covalent Modification of Conserved Cysteine Leads to Allosteric Inhibition of the TEAD·Yap Protein-Protein Interaction. *Cell Chem. Biol.* **2019**, *26* (3), 378–389.e13. <https://doi.org/10.1016/j.chembiol.2018.11.010>.
- (81) Li, Q.; Sun, Y.; Jarugumilli, G. K.; Liu, S.; Dang, K.; Cotton, J. L.; Xiol, J.; Chan, P. Y.; DeRan, M.; Ma, L.; Li, R.; Zhu, L. J.; Li, J. H.; Leiter, A. B.; Ip, Y. T.; Camargo, F. D.; Luo, X.; Johnson, R. L.; Wu, X.; Mao, J. Lats1/2 Sustain Intestinal Stem Cells and Wnt Activation through TEAD-Dependent and Independent Transcription. *Cell Stem Cell* **2020**, *26* (5), 675–692.e8. <https://doi.org/10.1016/j.stem.2020.03.002>.
- (82) Quambusch, L.; Landel, I.; Depta, L.; Weisner, J.; Uhlenbrock, N.; Müller, M. P.; Glanemann, F.; Althoff, K.; Siveke, J. T.; Rauh, D. Covalent-Allosteric Inhibitors to

- Achieve Akt Isoform-Selectivity. *Angew. Chem.* **2019**, *131* (52), 18999–19005. <https://doi.org/10.1002/ange.201909857>.
- (83) Henderson, A. R.; Henley, M. J.; Foster, N. J.; Peiffer, A. L.; Beyersdorf, M. S.; Stanford, K. D.; Sturlis, S. M.; Linhares, B. M.; Hill, Z. B.; Wells, J. A.; Cierpicki, T.; Brooks, C. L.; Fierke, C. A.; Mapp, A. K. Conservation of Coactivator Engagement Mechanism Enables Small-Molecule Allosteric Modulators. *Proc. Natl. Acad. Sci.* **2018**, *115* (36), 8960–8965. <https://doi.org/10.1073/pnas.1806202115>.
- (84) Zhang, J.; Xiong, B.; Zhen, X.; Zhang, A. Dopamine D1 Receptor Ligands: Where Are We Now and Where Are We Going. *Med. Res. Rev.* **2009**, *29* (2), 272–294. <https://doi.org/10.1002/med.20130>.
- (85) Hao, J.; Beck, J. P.; Schaus, J. M.; Krushinski, J. H.; Chen, Q.; Beadle, C. D.; Vidal, P.; Reinhard, M. R.; Dressman, B. A.; Massey, S. M.; Boulet, S. L.; Cohen, M. P.; Watson, B. M.; Tupper, D.; Gardinier, K. M.; Myers, J.; Johansson, A. M.; Richardson, J.; Richards, D. S.; Hembre, E. J.; Remick, D. M.; Coates, D. A.; Bhardwaj, R. M.; Diseroad, B. A.; Bender, D.; Stephenson, G.; Wolfangel, C. D.; Diaz, N.; Getman, B. G.; Wang, X.; Heinz, B. A.; Cramer, J. W.; Zhou, X.; Maren, D. L.; Falcone, J. F.; Wright, R. A.; Mitchell, S. N.; Carter, G.; Yang, C. R.; Bruns, R. F.; Svensson, K. A. Synthesis and Pharmacological Characterization of 2-(2,6-Dichlorophenyl)-1-((1S,3R)-5-(3-Hydroxy-3-Methylbutyl)-3-(Hydroxymethyl)-1-Methyl-3,4-Dihydroisoquinolin-2(1H)-Yl)Ethan-1-One (LY3154207), a Potent, Subtype Selective, and Orally Available Positive Allosteric Modulator of the Human Dopamine D1 Receptor. *J. Med. Chem.* **2019**, *62* (19), 8711–8732. <https://doi.org/10.1021/acs.jmedchem.9b01234>.
- (86) Gräb, J.; Berg, T. The Selectivity of Fosfosal for STAT5b over STAT5a Is Mediated by Arg566 in the Linker Domain. *ChemBioChem* **2020**, *n/a* (n/a). <https://doi.org/10.1002/cbic.202000111>.
- (87) Elumalai, N.; Berg, A.; Natarajan, K.; Scharow, A.; Berg, T. Nanomolar Inhibitors of the Transcription Factor STAT5b with High Selectivity over STAT5a. *Angew. Chem. Int. Ed.* **2015**, *54* (16), 4758–4763. <https://doi.org/10.1002/anie.201410672>.
- (88) Gräb, J.; Berg, A.; Blechschmidt, L.; Klüver, B.; Rubner, S.; Fu, D. Y.; Meiler, J.; Gräber, M.; Berg, T. The STAT5b Linker Domain Mediates the Selectivity of Catechol Bisphosphates for STAT5b over STAT5a. *ACS Chem. Biol.* **2019**, *14* (4), 796–805. <https://doi.org/10.1021/acscchembio.9b00137>.
- (89) Ishii, K.; Zhou, M.; Uchiyama, S. Native Mass Spectrometry for Understanding Dynamic Protein Complex. *Biochim. Biophys. Acta BBA - Gen. Subj.* **2018**, *1862* (2), 275–286. <https://doi.org/10.1016/j.bbagen.2017.09.019>.
- (90) Huang, C.; Kalodimos, C. G. Structures of Large Protein Complexes Determined by Nuclear Magnetic Resonance Spectroscopy. *Annu. Rev. Biophys.* **2017**, *46* (1), 317–336. <https://doi.org/10.1146/annurev-biophys-070816-033701>.
- (91) Merk, A.; Bartesaghi, A.; Banerjee, S.; Falconieri, V.; Rao, P.; Davis, M. I.; Pragani, R.; Boxer, M. B.; Earl, L. A.; Milne, J. L. S.; Subramaniam, S. Breaking Cryo-EM Resolution Barriers to Facilitate Drug Discovery. *Cell* **2016**, *165* (7), 1698–1707. <https://doi.org/10.1016/j.cell.2016.05.040>.
- (92) Schmidt, C.; Urlaub, H. Combining Cryo-Electron Microscopy (Cryo-EM) and Cross-Linking Mass Spectrometry (CX-MS) for Structural Elucidation of Large Protein Assemblies. *Curr. Opin. Struct. Biol.* **2017**, *46*, 157–168. <https://doi.org/10.1016/j.sbi.2017.10.005>.

- (93) El Khattabi, L.; Zhao, H.; Kalchschmidt, J.; Young, N.; Jung, S.; Van Blerkom, P.; Kieffer-Kwon, P.; Kieffer-Kwon, K.-R.; Park, S.; Wang, X.; Krebs, J.; Tripathi, S.; Sakabe, N.; Sobreira, D. R.; Huang, S.-C.; Rao, S. S. P.; Pruett, N.; Chauss, D.; Sadler, E.; Lopez, A.; Nóbrega, M. A.; Aiden, E. L.; Asturias, F. J.; Casellas, R. A Pliable Mediator Acts as a Functional Rather Than an Architectural Bridge between Promoters and Enhancers. *Cell* **2019**, *178* (5), 1145-1158.e20. <https://doi.org/10.1016/j.cell.2019.07.011>.
- (94) Abdella, R.; Talyzina, A.; Chen, S.; Inouye, C. J.; Tjian, R.; He, Y. Structure of the Human Mediator-Bound Transcription Preinitiation Complex. *Science* **2021**. <https://doi.org/10.1126/science.abg3074>.
- (95) Kaledhonkar, S.; Fu, Z.; Caban, K.; Li, W.; Chen, B.; Sun, M.; Gonzalez, R. L.; Frank, J. Late Steps in Bacterial Translation Initiation Visualized Using Time-Resolved Cryo-EM. *Nature* **2019**, *570* (7761), 400–404. <https://doi.org/10.1038/s41586-019-1249-5>.
- (96) Cesa, L. C.; Patury, S.; Komiyama, T.; Ahmad, A.; Zuiderweg, E. R. P.; Gestwicki, J. E. Inhibitors of Difficult Protein-Protein Interactions Identified by High-Throughput Screening of Multiprotein Complexes. *ACS Chem. Biol.* **2013**, *8* (9), 1988–1997. <https://doi.org/10.1021/cb400356m>.
- (97) Taylor, I. R.; Dunyak, B. M.; Komiyama, T.; Shao, H.; Ran, X.; Assimon, V. A.; Kalyanaraman, C.; Rauch, J. N.; Jacobson, M. P.; Zuiderweg, E. R. P.; Gestwicki, J. E. High-Throughput Screen for Inhibitors of Protein–Protein Interactions in a Reconstituted Heat Shock Protein 70 (Hsp70) Complex. *J. Biol. Chem.* **2018**, *293* (11), 4014–4025. <https://doi.org/10.1074/jbc.RA117.001575>.
- (98) Patwardhan, C. A.; Alfa, E.; Lu, S.; Chadli, A. Progesterone Receptor Chaperone Complex–Based High-Throughput Screening Assay: Identification of Capsaicin as an Inhibitor of the Hsp90 Machine. *J. Biomol. Screen.* **2014**. <https://doi.org/10.1177/1087057114549147>.
- (99) Monroy, C. A.; Mackie, D. I.; Roman, D. L. A High Throughput Screen for RGS Proteins Using Steady State Monitoring of Free Phosphate Formation. *PLOS ONE* **2013**, *8* (4), e62247. <https://doi.org/10.1371/journal.pone.0062247>.
- (100) Niu, S.; Ruotolo, B. T. Collisional Unfolding of Multiprotein Complexes Reveals Cooperative Stabilization upon Ligand Binding. *Protein Sci.* **2015**, *24* (8), 1272–1281. <https://doi.org/10.1002/pro.2699>.
- (101) Rabuck-Gibbons, J. N.; Keating, J. E.; Ruotolo, B. T. Collision Induced Unfolding and Dissociation Differentiates ATP-Competitive from Allosteric Protein Tyrosine Kinase Inhibitors. *Int. J. Mass Spectrom.* **2018**, *427*, 151–156. <https://doi.org/10.1016/j.ijms.2017.12.002>.
- (102) Dixit, S. M.; Polasky, D. A.; Ruotolo, B. T. Collision Induced Unfolding of Isolated Proteins in the Gas Phase: Past, Present, and Future. *Curr. Opin. Chem. Biol.* **2018**, *42*, 93–100. <https://doi.org/10.1016/j.cbpa.2017.11.010>.
- (103) Lea, W. A.; Simeonov, A. Fluorescence Polarization Assays in Small Molecule Screening. *Expert Opin. Drug Discov.* **2011**, *6* (1), 17–32. <https://doi.org/10.1517/17460441.2011.537322>.
- (104) Majmudar, C. Y.; Højfeldt, J. W.; Arevang, C. J.; Pomerantz, W. C.; Gagnon, J. K.; Schultz, P. J.; Cesa, L. C.; Doss, C. H.; Rowe, S. P.; Vásquez, V.; Tamayo-Castillo, G.; Cierpicki, T.; Brooks, C. L.; Sherman, D. H.; Mapp, A. K. Sekikaic Acid and Lobaric

- Acid Target a Dynamic Interface of the Coactivator CBP/P300. *Angew. Chem. Int. Ed.* **2012**, *51* (45), 11258–11262. <https://doi.org/10.1002/anie.201206815>.
- (105) Chan, A. I.; McGregor, L. M.; Liu, D. R. Novel Selection Methods for DNA-Encoded Chemical Libraries. *Curr. Opin. Chem. Biol.* **2015**, *26*, 55–61. <https://doi.org/10.1016/j.cbpa.2015.02.010>.
- (106) Jr, R. A. G.; Dumelin, C. E.; Keefe, A. D. DNA-Encoded Chemistry: Enabling the Deeper Sampling of Chemical Space. *Nat. Rev. Drug Discov.* **2016**, *16* (2), nrd.2016.213. <https://doi.org/10.1038/nrd.2016.213>.
- (107) Salamon, H.; Klika Škopić, M.; Jung, K.; Bugain, O.; Brunschweiler, A. Chemical Biology Probes from Advanced DNA-Encoded Libraries. *ACS Chem. Biol.* **2016**, *11* (2), 296–307. <https://doi.org/10.1021/acscchembio.5b00981>.
- (108) Gong, Z.; Hu, G.; Li, Q.; Liu, Z.; Wang, F.; Zhang, X.; Xiong, J.; Li, P.; Xu, Y.; Ma, R.; Chen, S.; Li, J. Compound Libraries: Recent Advances and Their Applications in Drug Discovery. *Curr. Drug Discov. Technol.* **2017**, *14* (4), 216–228. <https://doi.org/10.2174/1570163814666170425155154>.
- (109) Jin, X.; Lee, K.; Kim, N. H.; Kim, H. S.; Yook, J. I.; Choi, J.; No, K. T. Natural Products Used as a Chemical Library for Protein–Protein Interaction Targeted Drug Discovery. *J. Mol. Graph. Model.* **2018**, *79*, 46–58. <https://doi.org/10.1016/j.jmgm.2017.10.015>.
- (110) Moret, N.; Clark, N. A.; Hafner, M.; Wang, Y.; Lounkine, E.; Medvedovic, M.; Wang, J.; Gray, N.; Jenkins, J.; Sorger, P. K. Cheminformatics Tools for Analyzing and Designing Optimized Small-Molecule Collections and Libraries. *Cell Chem. Biol.* **2019**, *26* (5), 765–777.e3. <https://doi.org/10.1016/j.chembiol.2019.02.018>.
- (111) Doak, B. C.; Norton, R. S.; Scanlon, M. J. The Ways and Means of Fragment-Based Drug Design. *Pharmacol. Ther.* **2016**, *167*, 28–37. <https://doi.org/10.1016/j.pharmthera.2016.07.003>.
- (112) Kirsch, P.; Hartman, A. M.; Hirsch, A. K. H.; Empting, M. Concepts and Core Principles of Fragment-Based Drug Design. *Molecules* **2019**, *24* (23), 4309. <https://doi.org/10.3390/molecules24234309>.
- (113) Backus, K. M.; Correia, B. E.; Lum, K. M.; Forli, S.; Horning, B. D.; González-Páez, G. E.; Chatterjee, S.; Lanning, B. R.; Teijaro, J. R.; Olson, A. J.; Wolan, D. W.; Cravatt, B. F. Proteome-Wide Covalent Ligand Discovery in Native Biological Systems. *Nature* **2016**, *534* (7608), 570–574. <https://doi.org/10.1038/nature18002>.
- (114) Parker, C. G.; Galmozzi, A.; Wang, Y.; Correia, B. E.; Sasaki, K.; Joslyn, C. M.; Kim, A. S.; Cavallaro, C. L.; Lawrence, R. M.; Johnson, S. R.; Narvaiza, I.; Saez, E.; Cravatt, B. F. Ligand and Target Discovery by Fragment-Based Screening in Human Cells. *Cell* **2017**, *168* (3), 527–541.e29. <https://doi.org/10.1016/j.cell.2016.12.029>.
- (115) 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* **2013**, *503* (7477), 548–551. <https://doi.org/10.1038/nature12796>.
- (116) Sadowsky, J. D.; Burlingame, M. A.; Wolan, D. W.; McClendon, C. L.; Jacobson, M. P.; Wells, J. A. Turning a Protein Kinase on or off from a Single Allosteric Site via Disulfide Trapping. *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108* (15), 6056–6061. <https://doi.org/10.1073/pnas.1102376108>.
- (117) Wang, N.; Majmudar, C. Y.; Pomerantz, W. C.; Gagnon, J. K.; Sadowsky, J. D.; Meagher, J. L.; Johnson, T. K.; Stuckey, J. A.; Brooks, C. L.; Wells, J. A.; Mapp, A. K.

- Ordering a Dynamic Protein via a Small-Molecule Stabilizer. *J. Am. Chem. Soc.* **2013**, *135* (9), 3363–3366. <https://doi.org/10.1021/ja3122334>.
- (118) Sijbesma, E.; Hallenbeck, K. K.; Leysen, S.; de Vink, P. J.; Skóra, L.; Jahnke, W.; Brunsveld, L.; Arkin, M. R.; Ottmann, C. Site-Directed Fragment-Based Screening for the Discovery of Protein-Protein Interaction Stabilizers. *J. Am. Chem. Soc.* **2019**, *141* (8), 3524–3531. <https://doi.org/10.1021/jacs.8b11658>.
- (119) Arntson, K. E.; Pomerantz, W. C. K. Protein-Observed Fluorine NMR: A Bioorthogonal Approach for Small Molecule Discovery. *J. Med. Chem.* **2016**, *59* (11), 5158–5171. <https://doi.org/10.1021/acs.jmedchem.5b01447>.
- (120) Dalvit, C.; Vulpetti, A. Ligand-Based Fluorine NMR Screening: Principles and Applications in Drug Discovery Projects. *J. Med. Chem.* **2019**, *62* (5), 2218–2244. <https://doi.org/10.1021/acs.jmedchem.8b01210>.
- (121) Furukawa, A.; Konuma, T.; Yanaka, S.; Sugase, K. Quantitative Analysis of Protein–Ligand Interactions by NMR. *Prog. Nucl. Magn. Reson. Spectrosc.* **2016**, *96*, 47–57. <https://doi.org/10.1016/j.pnmrs.2016.02.002>.
- (122) Gupta, A. K.; Wang, X.; Pagba, C. V.; Prakash, P.; Sarkar-Banerjee, S.; Putkey, J.; Gorfe, A. A. Multi-Target, Ensemble-Based Virtual Screening Yields Novel Allosteric KRAS Inhibitors at High Success Rate. *Chem. Biol. Drug Des.* **2019**, *94* (2), 1441–1456. <https://doi.org/10.1111/cbdd.13519>.
- (123) De Vivo, M.; Masetti, M.; Bottegoni, G.; Cavalli, A. Role of Molecular Dynamics and Related Methods in Drug Discovery. *J. Med. Chem.* **2016**, *59* (9), 4035–4061. <https://doi.org/10.1021/acs.jmedchem.5b01684>.
- (124) Greener, J. G.; Sternberg, M. J. Structure-Based Prediction of Protein Allostery. *Curr. Opin. Struct. Biol.* **2018**, *50*, 1–8. <https://doi.org/10.1016/j.sbi.2017.10.002>.
- (125) Ma, X.; Meng, H.; Lai, L. Motions of Allosteric and Orthosteric Ligand-Binding Sites in Proteins Are Highly Correlated. *J. Chem. Inf. Model.* **2016**, *56* (9), 1725–1733. <https://doi.org/10.1021/acs.jcim.6b00039>.
- (126) A. Senisterra, G.; Patrick J. Finerty, J. High Throughput Methods of Assessing Protein Stability and Aggregation. *Mol. Biosyst.* **2009**, *5* (3), 217–223. <https://doi.org/10.1039/B814377C>.
- (127) Gao, K.; Oerlemans, R.; Groves, M. R. Theory and Applications of Differential Scanning Fluorimetry in Early-Stage Drug Discovery. *Biophys. Rev.* **2020**, *12* (1), 85–104. <https://doi.org/10.1007/s12551-020-00619-2>.
- (128) Niesen, F. H.; Berglund, H.; Vedadi, M. The Use of Differential Scanning Fluorimetry to Detect Ligand Interactions That Promote Protein Stability. *Nat. Protoc.* **2007**, *2* (9), 2212–2221. <https://doi.org/10.1038/nprot.2007.321>.
- (129) Attarha, S.; Reithmeier, A.; Busker, S.; Desroses, M.; Page, B. D. G. Validating STAT Protein-Inhibitor Interactions Using Biochemical and Cellular Thermal Shift Assays. *ACS Chem. Biol.* **2020**. <https://doi.org/10.1021/acschembio.0c00046>.
- (130) Desroses, M.; Busker, S.; Astorga-Wells, J.; Attarha, S.; Kolosenko, I.; Zubarev, R. A.; Helleday, T.; Grandér, D.; Page, B. D. G. STAT3 Differential Scanning Fluorimetry and Differential Scanning Light Scattering Assays: Addressing a Missing Link in the Characterization of STAT3 Inhibitor Interactions. *J. Pharm. Biomed. Anal.* **2018**, *160*, 80–88. <https://doi.org/10.1016/j.jpba.2018.07.018>.

- (131) DeSantis, K.; Reed, A.; Rahhal, R.; Reinking, J. Use of Differential Scanning Fluorimetry as a High-Throughput Assay to Identify Nuclear Receptor Ligands: *Nucl. Recept. Signal.* **2012**. <https://doi.org/10.1621/nrs.10002>.
- (132) Mac Sweeney, A.; Chambovey, A.; Wicki, M.; Müller, M.; Artico, N.; Lange, R.; Bijelic, A.; Breibeck, J.; Rompel, A. The Crystallization Additive Hexatungstotellurate Promotes the Crystallization of the HSP70 Nucleotide Binding Domain into Two Different Crystal Forms. *PLoS ONE* **2018**, *13* (6). <https://doi.org/10.1371/journal.pone.0199639>.
- (133) Shao, H.; Oltion, K.; Wu, T.; E. Gestwicki, J. Differential Scanning Fluorimetry (DSF) Screen to Identify Inhibitors of Hsp60 Protein–Protein Interactions. *Org. Biomol. Chem.* **2020**. <https://doi.org/10.1039/D0OB00928H>.
- (134) Sun, C.; Li, Y.; Yates, E. A.; Fernig, D. G. SimpleDSFviewer: A Tool to Analyze and View Differential Scanning Fluorimetry Data for Characterizing Protein Thermal Stability and Interactions. *Protein Sci.* **2020**, *29* (1), 19–27. <https://doi.org/10.1002/pro.3703>.
- (135) Wu, T.; Yu, J.; Gale-Day, Z.; Woo, A.; Suresh, A.; Hornsby, M.; Gestwicki, J. E. Three Essential Resources to Improve Differential Scanning Fluorimetry (DSF) Experiments. *bioRxiv* **2020**, 2020.03.22.002543. <https://doi.org/10.1101/2020.03.22.002543>.
- (136) Sorenson, A. E.; Schaeffer, P. M. High-Throughput Differential Scanning Fluorimetry of GFP-Tagged Proteins. In *Targeting Enzymes for Pharmaceutical Development: Methods and Protocols*; Labrou, N. E., Ed.; Methods in Molecular Biology; Springer US: New York, NY, 2020; pp 69–85. https://doi.org/10.1007/978-1-0716-0163-1_5.
- (137) Molina, D. M.; Jafari, R.; Ignatushchenko, M.; Seki, T.; Larsson, E. A.; Dan, C.; Sreekumar, L.; Cao, Y.; Nordlund, P. Monitoring Drug Target Engagement in Cells and Tissues Using the Cellular Thermal Shift Assay. *Science* **2013**, *341* (6141), 84–87. <https://doi.org/10.1126/science.1233606>.
- (138) Henderson, M. J.; Holbert, M. A.; Simeonov, A.; Kallal, L. A. High-Throughput Cellular Thermal Shift Assays in Research and Drug Discovery. *SLAS Discov. Adv. Sci. Drug Discov.* **2020**, *25* (2), 137–147. <https://doi.org/10.1177/2472555219877183>.
- (139) Shaw, J.; Dale, I.; Hemsley, P.; Leach, L.; Dekki, N.; Orme, J. P.; Talbot, V.; Narvaez, A. J.; Bista, M.; Martinez Molina, D.; Dabrowski, M.; Main, M. J.; Gianni, D. Positioning High-Throughput CETSA in Early Drug Discovery through Screening against B-Raf and PARP1. *Slas Discov.* **2019**, *24* (2), 121–132. <https://doi.org/10.1177/2472555218813332>.
- (140) Nagasawa, I.; Muroi, M.; Kawatani, M.; Ohishi, T.; Ohba, S.; Kawada, M.; Osada, H. Identification of a Small Compound Targeting PKM2-Regulated Signaling Using 2D Gel Electrophoresis-Based Proteome-Wide CETSA. *Cell Chem. Biol.* **2020**, *27* (2), 186–196.e4. <https://doi.org/10.1016/j.chembiol.2019.11.010>.
- (141) Jessani, N.; Cravatt, B. F. The Development and Application of Methods for Activity-Based Protein Profiling. *Curr. Opin. Chem. Biol.* **2004**, *8* (1), 54–59. <https://doi.org/10.1016/j.cbpa.2003.11.004>.
- (142) Pai, M. Y.; Lomenick, B.; Hwang, H.; Schiestl, R.; McBride, W.; Loo, J. A.; Huang, J. Drug Affinity Responsive Target Stability (DARTS) for Small Molecule Target Identification. *Methods Mol. Biol. Clifton NJ* **2015**, *1263*, 287–298. https://doi.org/10.1007/978-1-4939-2269-7_22.
- (143) Savitski, M. M.; Reinhard, F. B. M.; Franken, H.; Werner, T.; Savitski, M. F.; Eberhard, D.; Molina, D. M.; Jafari, R.; Dovega, R. B.; Klaeger, S.; Kuster, B.; Nordlund, P.;

- Bantscheff, M.; Drewes, G. Tracking Cancer Drugs in Living Cells by Thermal Profiling of the Proteome. *Science* **2014**, *346* (6205). <https://doi.org/10.1126/science.1255784>.
- (144) Shrimp, J. H.; Grose, C.; Widmeyer, S. R. T.; Thorpe, A. L.; Jadhav, A.; Meier, J. L. Chemical Control of a CRISPR-Cas9 Acetyltransferase. *ACS Chem. Biol.* **2018**, *13* (2), 455–460. <https://doi.org/10.1021/acscchembio.7b00883>.
- (145) Zucconi, B. E.; Luef, B.; Xu, W.; Henry, R. A.; Nodelman, I. M.; Bowman, G. D.; Andrews, A. J.; Cole, P. A. Modulation of P300/CBP Acetylation of Nucleosomes by Bromodomain Ligand I-CBP112. *Biochemistry* **2016**, *55* (27), 3727–3734. <https://doi.org/10.1021/acs.biochem.6b00480>.
- (146) Huang, W.; Peng, Y.; Kiselar, J.; Zhao, X.; Albaqami, A.; Mendez, D.; Chen, Y.; Chakravarthy, S.; Gupta, S.; Ralston, C.; Kao, H.-Y.; Chance, M. R.; Yang, S. Multidomain Architecture of Estrogen Receptor Reveals Interfacial Cross-Talk between Its DNA-Binding and Ligand-Binding Domains. *Nat. Commun.* **2018**, *9*. <https://doi.org/10.1038/s41467-018-06034-2>.
- (147) Bondeson, D. P.; Smith, B. E.; Burslem, G. M.; Buhimschi, A. D.; Hines, J.; Jaime-Figueroa, S.; Wang, J.; Hamman, B. D.; Ishchenko, A.; Crews, C. M. Lessons in PROTAC Design from Selective Degradation with a Promiscuous Warhead. *Cell Chem. Biol.* **2018**, *25* (1), 78-87.e5. <https://doi.org/10.1016/j.chembiol.2017.09.010>.
- (148) Gadd, M. S.; Testa, A.; Lucas, X.; Chan, K.-H.; Chen, W.; Lamont, D. J.; Zengerle, M.; Ciulli, A. Structural Basis of PROTAC Cooperative Recognition for Selective Protein Degradation. *Nat. Chem. Biol.* **2017**, *13* (5), 514–521. <https://doi.org/10.1038/nchembio.2329>.

CHAPTER 2

Targeting the Dynamic Coactivator Med25 with a Selective Small Molecule Probe[†]

2.1 Abstract

Here we report the results of a high throughput screen that identified the natural product norstictic acid as a highly selective, potent (low μM) allosteric inhibitor of Med25 AcID activator interactions. Further, we elucidated a unique mechanism of action for the small molecule, showing that it preferentially modifies lysine residues within a dynamic loop on one face of the protein affecting binding at both faces. Structural models in combination with biochemical data indicate that the molecule likely functions via an orthosteric/allosteric mechanism. Norstictic acid also engages with cellular Med25, and treatment of mammalian cells with norstictic acid downregulates genes known to be regulated by Med25, recapitulating a Med25 KO phenotype. Taken together, the discovery of norstictic acid confirms that functionally modulating AcID with small molecules by targeting dynamic loop regions of the protein provides selective allosteric modulators that can affect cellular function of Med25.

2.2 Introduction

Transcriptional coactivators play an integral role in the regulation of gene expression, serving as hub proteins for transcriptional machinery assembly through interactions with transcriptional activators.^{1–9} Coactivators use conformationally plastic activator binding domains

[†] The majority of this chapter is reproduced from Garlick J.M., Sturlis S.M., Bruno P.A., Yates J.A., Peiffer A.L., Goo L., Bao L.W., De Salle S.N., Tamayo-Castillo G., Brooks C.L. III, Merajver S.D., Mapp A.K. Norstictic acid is a selective allosteric transcriptional regulator (DOI: 10.1101/2021.03.26.437253). The research described in this chapter was a collaborative effort. Dr. Steve Sturlis and Dr. Paul Bruno conceived and developed the HTS, including Med25 expression, ran the HTS and validated the initial hits as covalent binders, identifying the orthophenolic aldehyde as key functional group for binding and function. Dr. Steve Sturlis made, characterized, and tested many of the Lys to Arg mutants. He was also consulted on the design and implementation of experiments. Julie Garlick carried out full selectivity assessment of NA; expanded Lys to Arg mutant study to exclude all possible Lys as sites of modification and tested mutational effects on ETV5 binding and inhibition by NA; developed and implemented CETSA and co-IP studies in HeLa lysates; developed and implemented studies of NA in VARI068 WT and KO cell lines. Amanda Peiffer conducted MD simulations to construct structural model of NA binding and function. Yeujin Liu carried out synergy experiments of NA in MB-MDA-231 cells with lapatinib. Dr. Joel Yates and Laura Goo generated Med25 CRISPR KO cell lines. Sam De Salle generated K422R mutant.

(ABDs) to interact with a variety of DNA-bound transcriptional activators. This interaction leads to recruitment of additional proteins, such as chromatin remodeling enzymes and general transcription factors, and ultimately RNA Pol II (Figure 2.1). Alterations in the network of coactivator-activator protein-protein interactions (PPIs) contribute to the onset and perpetuation of numerous diseases, leading to significant interest in synthetic probes for mechanistic studies and therapeutic applications.^{10–16} This is especially true for coactivators whose function is highly context-dependent, required only for a subset of genes or only at particular times or locations in the life cycle of an organism.¹⁷ Thus, synthetic modulation of such coactivators would also be context-dependent, providing an advantageous layer of specificity. The coactivator Med25 falls into this category; homozygous deletion of Med25, for example, is nonlethal, impacting approximately 900 genes.¹⁸ Med25 is a subunit of the Mediator complex. Mediator acts as a functional bridge between DNA bound transcriptional factors and RNA Polymerase II. The ABD of coactivators, such as Med25, within the tail of the complex, interact directly with the transcriptional activation domain (TAD) of activator proteins. As a substoichiometric component of the Mediator complex, it is not required for basal transcription. Several lines of evidence indicate that dysregulation of the PPI network of Med25 and the ETV/PEA3 transcriptional activators contributes to oncogenesis as well as metastatic phenotypes in certain breast and prostate cancers, heightening a need for Med25-selective inhibitors.^{19–22}

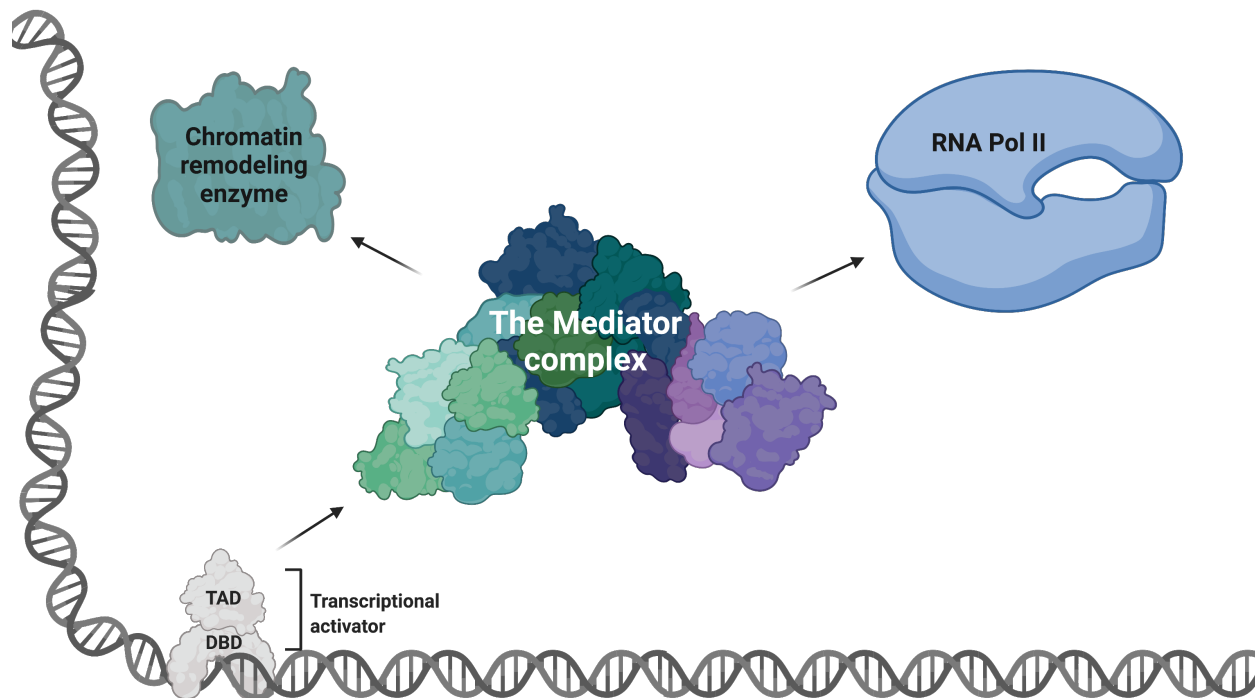


Figure 2.1. Transcriptional activation is mediated by coactivator proteins such as those found in the tail of the Mediator complex. A highly regulated network of protein-protein interactions (PPIs) occurs at gene promoters between transcription factors and the transcriptional machinery. Specifically, interaction between the transcriptional activation domain (TAD) of a DNA bound activators and coactivators are fundamental in the process of activated gene expression. The Mediator Complex contains multiple coactivator proteins and acts as a functional bridge between transcriptional activators and the additional machinery, such as chromatin remodeling enzymes and general transcription factors, that ultimately led to recruitment of the RNA Polymerase II preinitiation complex. Figure created using BioRender.

The domain of Med25 that interacts with activators is the Activator-Interacting Domain (AcID).^{23–25} This domain contains a 7-stranded beta barrel core flanked with dynamic loops and three alpha helices (Figure 2.2). Like other ABDs within coactivators, AcID does not contain defined binding pockets but rather uses highly dynamic, largely hydrophobic interfaces to interact with transcription factors to regulate gene expression.¹¹ Specifically, Med25 contains two distinct binding faces it uses to interact with different TADs (Figure 2.2).^{19,25,26} This interaction has been well studied particularly within the context of AcID•VP16 and AcID•ETV5.^{19,23,27–29} These qualities make AcID, as well as other coactivators, challenging to target selectively. Recent efforts by our lab and others have been elucidating successful approaches to target these classically “undruggable” proteins^{20,30–33}. The Mapp lab recently reported that dynamic substructures, such as helices and loops, within activator binding domains are very important for formation of

activator-coactivator complexes and are the driving component of allosteric networks connecting binding sites.²⁰ For example, Tethering of a small molecule to a cysteine residue within a loop at the base of Med25 AcID, visible in Figure 2.2 (right) on the H1 face highlighted in blue, enhances complex formation with transcriptional activators.²⁰ In another unrelated example, a class of small molecule inhibitors of the SH2 domain of STAT transcription factors were found to bind at an adjacent dynamic linker rather than within the SH2 domain itself.³² The differences in this linker region between different STAT proteins, even closely related homologs such as STAT5b and STAT5a, could be exploited to influence inhibitor selectivity.³⁴ Thus, we hypothesize that targeting the dynamic substructures in coactivators proteins can provide an effective strategy for modulation of coactivator protein-protein interactions. Further, because such substructures are more likely to access conformations with topologically unique binding surfaces, one might anticipate that small-molecule modulators of such sites would exhibit enhanced selectivity compared to purely orthosteric ligands.^{12,35–39}

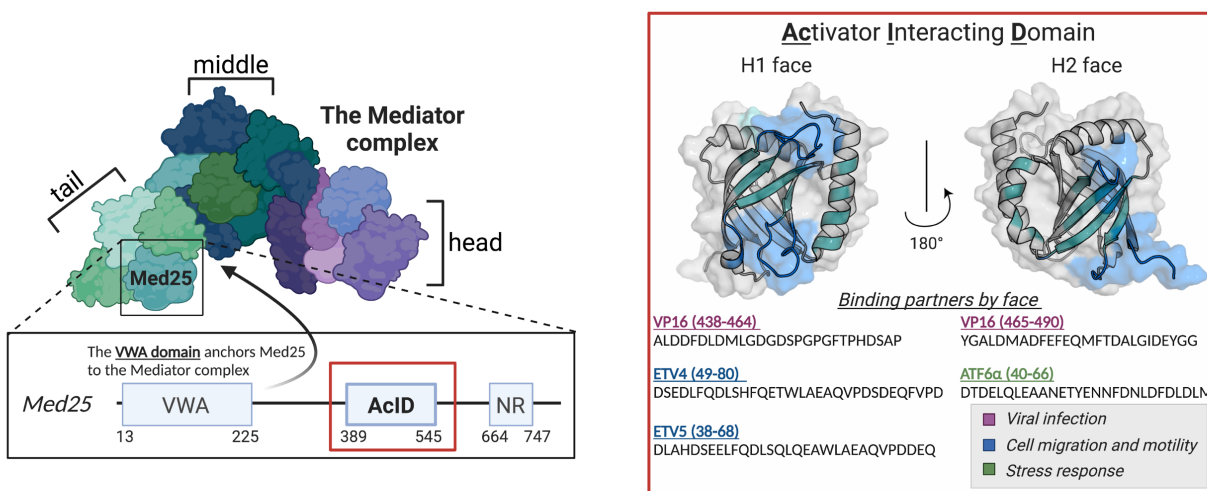


Figure 2.2. Coactivator and hub protein Med25 is a subunit of the Mediator coactivator complex. Left: The Mediator complex is comprised of ~30 proteins classified into multiple sections. The variable tail portion is known to contain transcriptional coactivators, including Med25. Med25 is made up of two main structured domains, the VWA domain which anchors the protein to Mediator, and the Activator-Interacting Domain (AcID). Right: AcID forms a PPI network with transcriptional activators using two binding surfaces (H1, H2) and in doing so regulates key cellular processes often associated with disease. PDB 2XNF used to generate figure. Figure created using BioRender.

This chapter focuses on the identification of norstictic acid as a potent and selective Med25 inhibitor. The mechanism of action of NA is elucidated, binding and covalent modification of a

flexible loop, further suggesting that targeting dynamic substructures leads to allosteric, selective, protein modulation. Finally, NA is validated in as a useful probe of Med25 biological function.

2.3 Results & Discussion

High throughput screening of Med25 leads to identification of covalent depsidone inhibitors

To identify inhibitors of Med25 AcID, we utilized a high-throughput fluorescence polarization (FP) assay interrogating a complex of AcID and fluorescein-tagged VP16(465-490) (Figure 2.3). As previously reported, this VP16 sequence contains the minimal binding sequence for interaction with AcID ($K_D = 0.60 \pm 0.06 \mu\text{M}$) and interacts with the H1 and H2 binding surfaces.^{20,25} Several commercially available libraries (MS Spectrum 2000, Focused Collections, and BioFocus NCC libraries) with a combined total of 4046 compounds were screened using this format (Figure 2.3, see Methods for additional details). The primary screening campaign had an average Z' score of 0.87, indicating an excellent assay, and a 1.6% hit rate. Following the primary screen, hits were filtered and compounds with known chemically reactive properties as well as those compounds that demonstrated native fluorescence greater than ten percent of the fluorescence produced by the tracer were removed. Compounds with activity greater than 3 standard deviations relative to the negative control (DMSO) were subjected to dose-response assessment with freshly purchased material, as well as secondary selectivity assays. From this, the natural lichen products norstictic acid (NA) and psoromic acid (PA) emerged as the best inhibitors, with apparent IC_{50} values of $2.3 \pm 0.1 \mu\text{M}$ and $3.9 \pm 0.3 \mu\text{M}$, respectively (Figure 2.3).

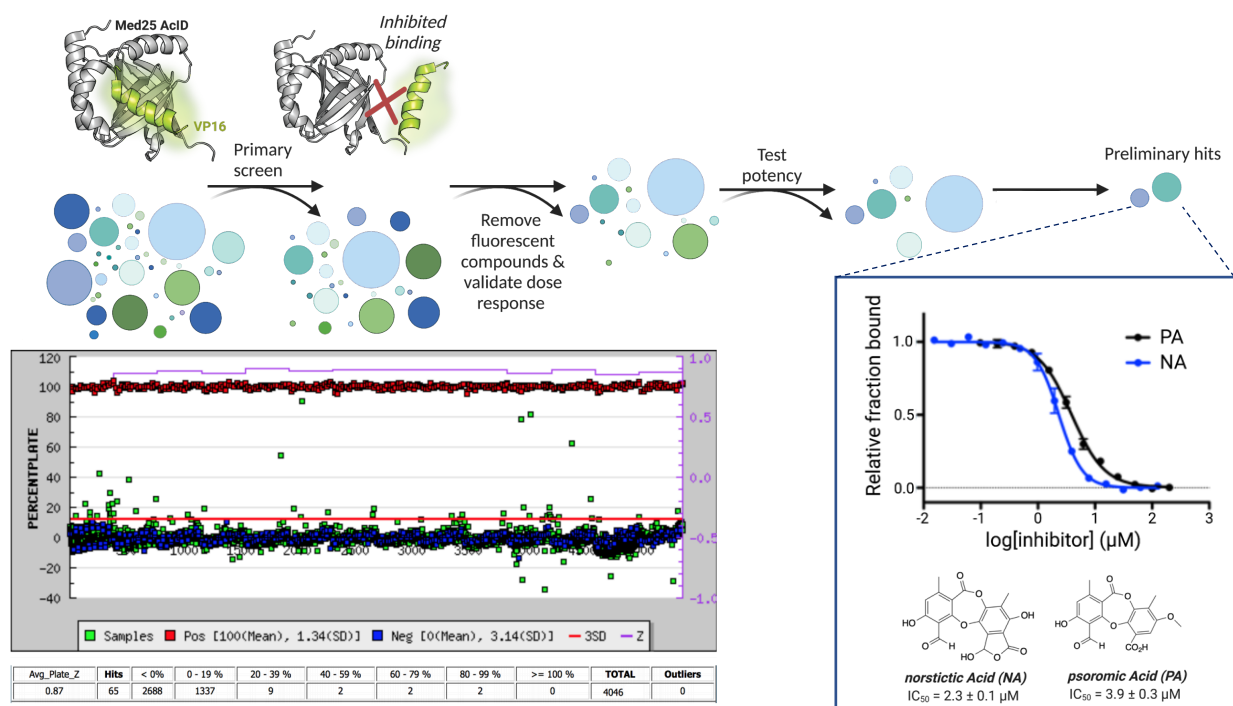


Figure 2.3. Norstictic acid is discovered as the result of a high throughput screen for inhibitors of the AcID•VP16(465-490) PPI. Top: Schematic of HTS fluorescence polarization based screen to identify inhibitors of the Med25 AcID PPI network. Bottom left: Campaign view of HTS screening results. 65 hits were identified from a 2688 compound screen. Bottom right: after confirmatory screening, two compounds emerged as preliminary hits, with low micromolar IC₅₀ values, norstictic acid and psoromic acid. Data obtained by Dr. Steven Sturlis. Figure made using BioRender.

Both NA and PA are natural products in the depsidone family containing an orthophenolic aldehyde moiety. The presence of a reactive aldehyde functionality suggested a potential covalent mechanism of action via imine formation with lysine side chains. Consistent with this hypothesis, analysis of NA-treated Med25 AcID using mass spectrometry showed the presence of concentration-dependent covalent adduct(s) (Figure 2.4). Treatment with the reducing agent NaBH₄ led to incorporation of H₂ into the adduct, indicating initial formation of a Schiff base followed by reduction. Data from a time-course experiment revealed that at 5 minutes, significant inhibition is observed, with full activity after 30 minutes (Figure 2.4).

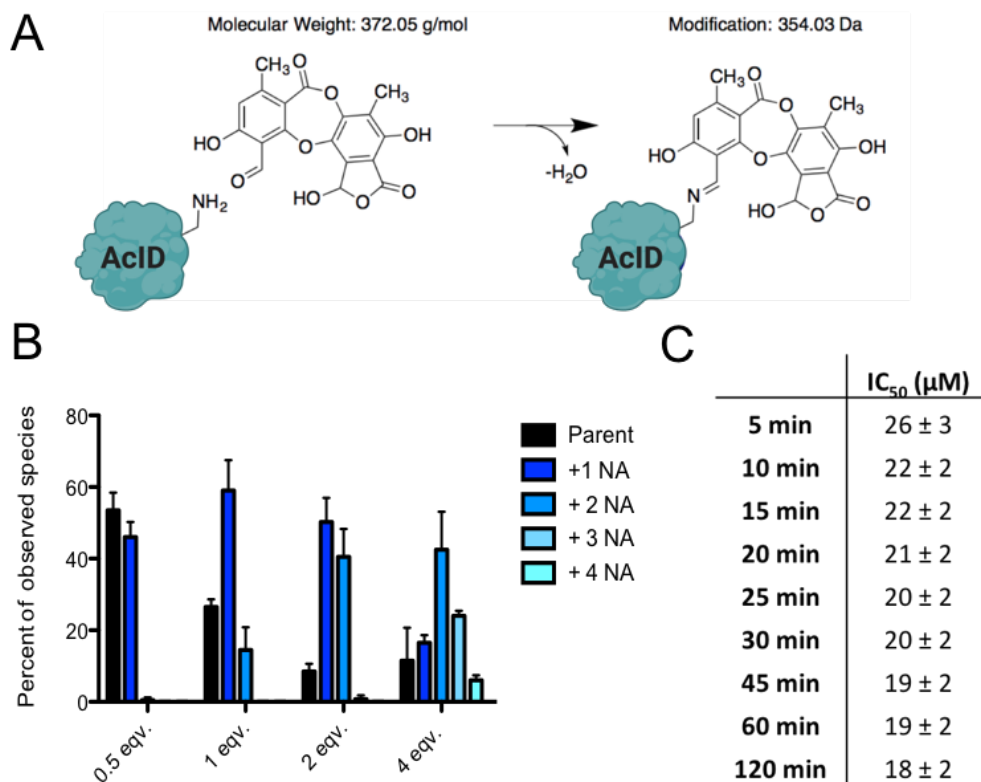


Figure 2.4. Norstictic acid is identified as a covalent inhibitor. A) Schematic of covalent bond formation with AcID via reaction of the orthophenolic aldehyde of NA with a primary amine group from a lysine residue, resulting in an imine. B) Concentration dependent NA adduct formation is observed with AcID (20 μM) using mass spectrometry. Data was obtained after incubation of Med25 with NA for 30 minutes. X-axis labels represent the equivalents of NA used compared to protein concentration. Experiments were conducted in technical triplicate and the average value is shown with error bars representing the standard deviation of the mean. Experiment conducted with Dr. Steven Sturlis. C) The apparent IC_{50} of norstictic acid was determined for inhibition of the VP16(465-490)•AcID interaction at various time points to determine how quickly inhibition was achieved. Experiment conducted by Dr. Steven Sturlis.

An examination of related structures indicates that the orthophenolic aldehyde is necessary, but not sufficient for either interaction with Med25 AcID or inhibitory activity. Stictic acid, in which the phenol is masked as a methyl ether, inhibits Med25 interactions poorly ($\text{IC}_{50} > 250 \mu\text{M}$). Additionally, salicylaldehyde efficiently labels Med25 AcID, but does not impact binding of activators (Figure 2.5). These data suggest that noncovalent interactions play essential roles in the inhibitor function of NA.

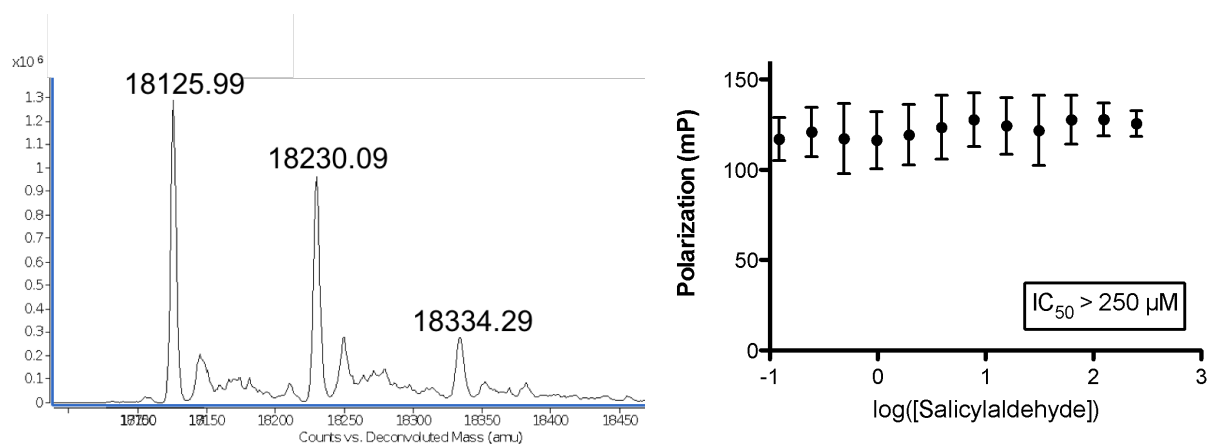


Figure 2.5. Salicylaldehyde covalently modifies, but does not inhibit, Med25 AcID. Mass spectrometry shows dosing with 50X salicylaldehyde leads to formation of +104 Med25 adduct (left). Salicylaldehyde does not inhibit the Med25•VP16(465-490) interaction using fluorescence polarization (right). Curves represent the mean values of three independent experiments with vertical error bars representing the standard deviation of the fraction of tracer bound at the indicated AcID concentration. Data obtained by Dr. Steven Sturlis.

Evaluating the in vitro selectivity of norstictic acid

ABDs in coactivators are typically hydrophobic and conformationally dynamic to allow for interaction with many distinct transcriptional activation domains.^{9,28,40–42} The implication for inhibitor discovery is that molecules discovered as inhibitors of one ABD will most often interact with other off target ABDs. It has been suggested that small molecules that take advantage of dynamic binding interfaces within coactivator proteins could allosterically modulate the target protein while maintain selectivity.^{20,30,43} We have also found that peptides engaging the beta-barrel surface are not very specific for one or the other faces of Med25 (or other coactivators), while interaction with loops and helices confers more selectivity— for example, VP16(465-490) interacts with both binding surfaces and engages largely with the beta-barrel surfaces. In contrast, the ETV/PEA3 activators form specific contacts with dynamic loops flanking the H1 beta-barrel surface and are highly selective for that binding site.²⁰ We hypothesized that interaction with the dynamic substructures within AcID would confer additional inhibitor selectivity. Screening, via competition FP, of a small panel of coactivators comprised of various common motifs showed that NA is indeed selective for Med25 AcID (Figure 2.6). The demonstrated selectivity of NA is particularly significant because selective targeting of coactivator proteins has remained a challenge to date. Notably, NA inhibits Med25 PPIs at both binding surfaces, including those formed with transcriptional activators ETV5 (H1 binding surface) and ATF6 α (H2 binding surface).

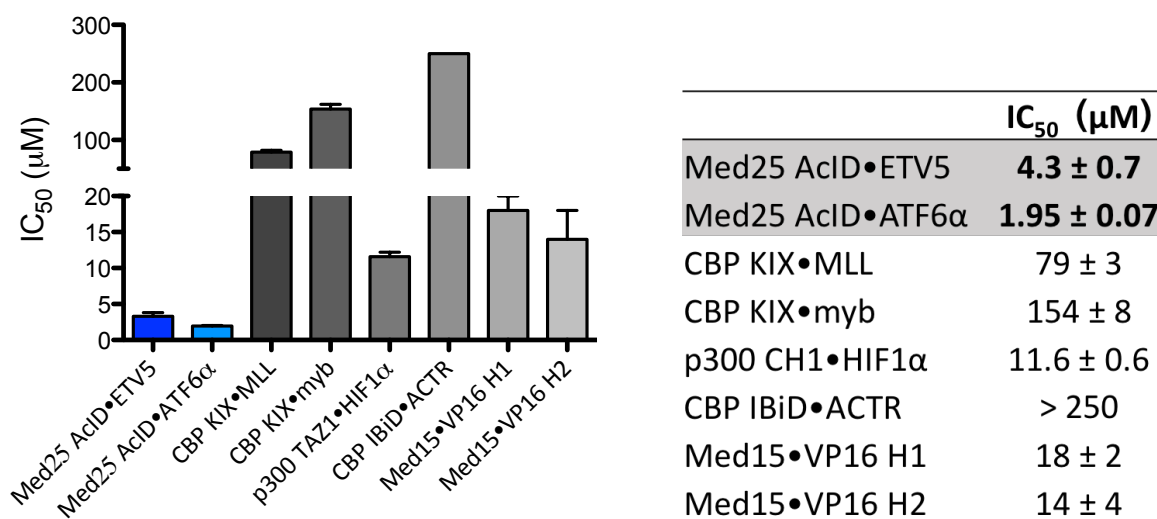


Figure 2.6. Norstictic acid is selective for AcID PPIs. NA was tested against a panel of transcriptional PPIs using competition FP. Left: Graphical view. Right: Apparent IC₅₀ values obtained after a 30 minute incubation of compound with protein. Experiments were performed in technical triplicate, And IC₅₀ values are the average with error shows as the standard deviation of the mean.

Selectivity with regards to lysine modification can also be observed with NA. For example, CBP KIX and AcID contain a similar number of lysine residues, 8 and 11 respectively; however the data clearly demonstrates that NA covalently modifies AcID at a much faster rate (Figure 2.7). After incubation for 30 minutes using 1 equivalent NA, 75% of AcID has been covalently modified whereas for KIX less than 5% has been modified. Even at 4 equivalents NA after 30 minutes only slightly more than 50% of KIX protein is labelled. Similar behavior is observed when dosing a mixture of both proteins with NA. Even when there is 4-fold excess KIX protein compared to AcID, upon adding NA and incubating for 30 minutes almost exclusively AcID protein is covalently modified. Thus, both binding and proper positioning of the lysine residue is key for covalent adduct formation to occur, and may play a role in the selectivity of the compound.

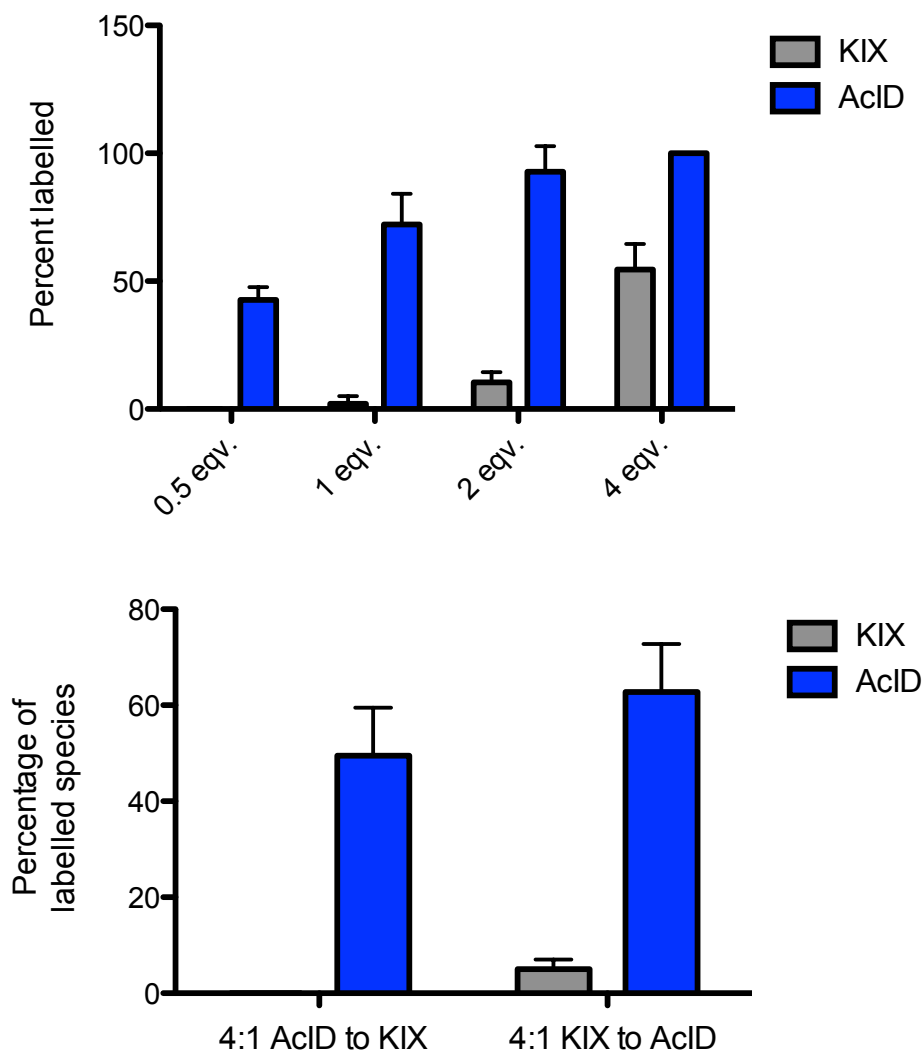


Figure 2.7. Comparison of norstictic acid reactivity between AcID and KIX. KIX and Med25 were separately dosed with varying concentrations of norstictic acid (0.5-4 equivalents) and subjected to LC-MS analysis. The percent of labelled species (any +NA modification) was calculated and plotted. This experiment was conducted in technical triplicate and the average value is shown with error bars representing the standard deviation of the mean. (top). A mixture of Med25 and KIX was dosed with NA and the percent of labelled species (any +NA modification) observed for each protein was calculated. This experiment was conducted in technical duplicate and the average value is shown with error bars representing the standard deviation of the mean (bottom).

Elucidation of NA binding site and inhibitory mechanism

Several lines of evidence suggested that the engagement site of NA is a lysine-rich dynamic loop that borders the H2 binding site. There are 11 lysine residues within Med25 AcID, 6 of which are found on dynamic loop regions flanking the two known activator binding surfaces (Figure 2.8).

By mutating these residues to arginine, we are able to remove the ability of NA to covalently modify at that particular site while maintaining the same chemical environment (i.e. keeping the positive charge). As expected, the lysine to arginine mutations had minimal effects on binding with the cognate transcriptional activator binding partners (Figure 2.8).

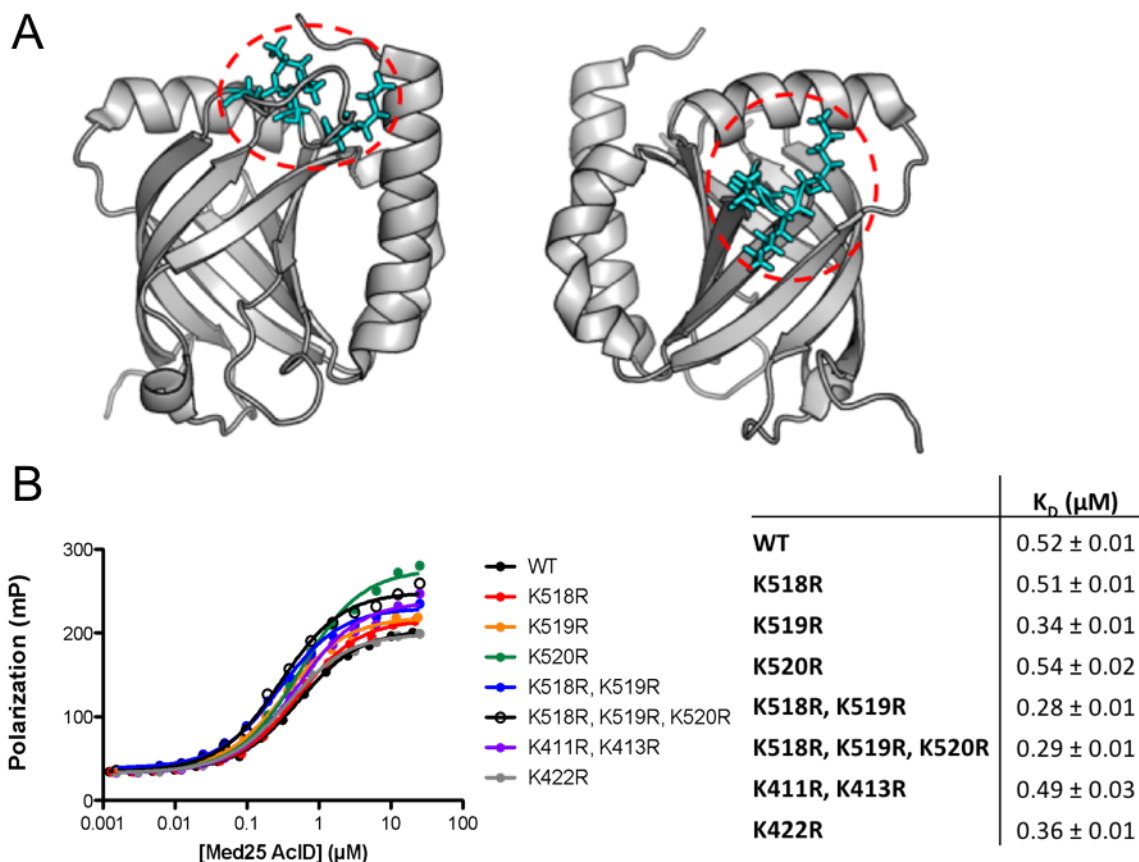


Figure 2.8. Mutation of dynamic loop lysine residues to arginine has minimal impact on ETV5 binding. A) There are two dynamic loops containing 3 lysine residues each located on AcID. The residues are shown in light blue. PDB: 2XNF B) Direct binding fluorescence polarization experiments with FITC labelled ETV5 and AcID show that mutation of residue(s) to arginine has minimal effect on K_D . Experiments were conducted in technical triplicate with the average shown and error represented as the standard deviation of the mean.

Similarly, mutations within the H1 binding surface had minimal impact on both NA binding, determined by mass spectrometric analysis, and inhibition in an *in vitro* binding assay. In contrast, mutation of K519, located within a lysine-rich dynamic loop that flanks the H2 face, had a profound effect on NA binding and inhibition (Figure 2.9). Mutation of the additional lysine residues within the H2 loop to arginine, further impacts NA function, suggesting K518 and K520

may also be important. Thus, we can conclude that the molecule inhibits via binding at the H2 lysine loop, with K519 the primary site of modification.

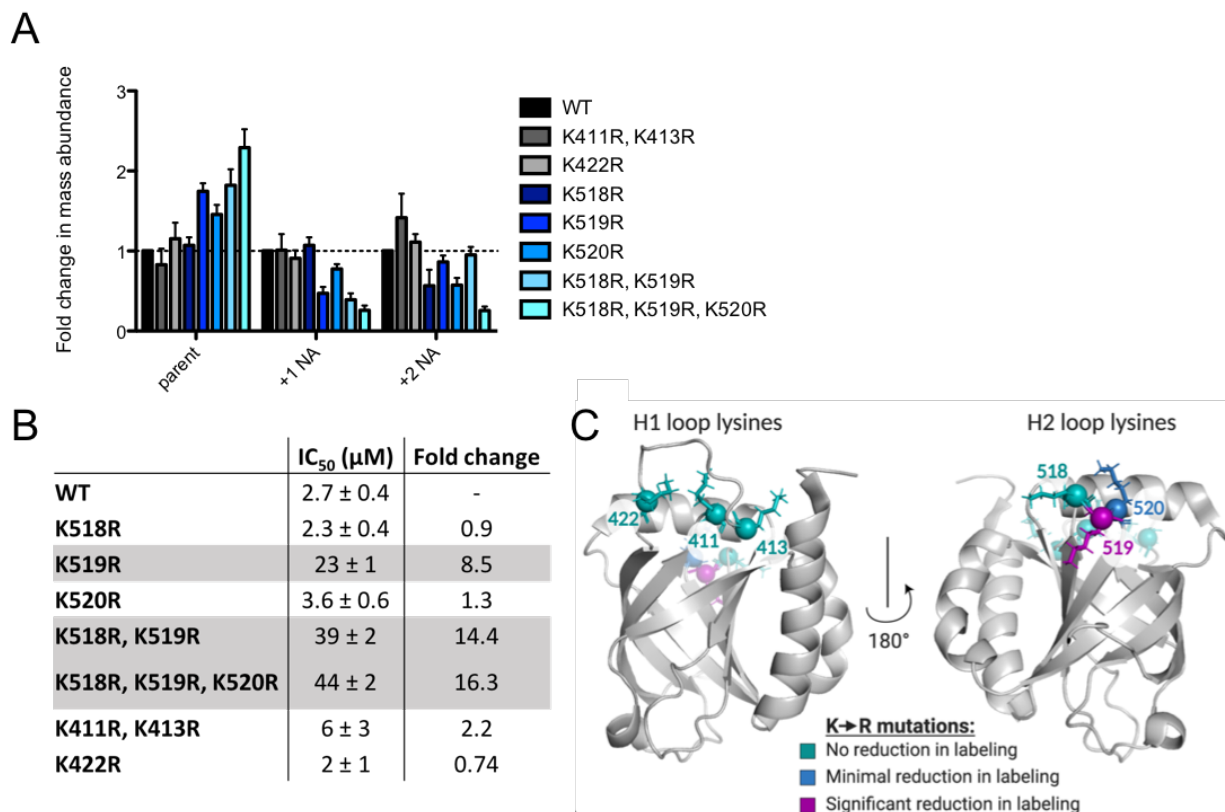


Figure 2.9. Mutation of lysine residues in H2 loop to arginine impacts NA binding and inhibition. A) Mass spec data obtained after incubation of AcID protein with NA for 30 minutes. From the deconvoluted data, the fold change in mass abundance of each potential species (parent, +1 NA, +2 NA) was calculated for the mutant compared to the WT. Data obtained in technical duplicate with error bars representing the standard deviation of the mean. B) Competition FP experiments with NA against FITC labelled ETV5 and AcID. Mutants highlighted in gray show a significant decrease in NA inhibitory ability (increase in IC₅₀). C) The structure of AcID, summarizing importance of each lysine residue analyzed for NA binding and inhibition determined from mass spec and competition FP experiments.

To develop a structural model of NA binding and function, molecular dynamics simulations of the covalent NA-Med25 AcID complex in which NA is covalently linked to K519 were carried out, and the results compared to the case of unbound Med25 AcID. As illustrated in Figure 2.10, minimal restructuring in the lysine loop adjacent to the H2 binding interface is observed. However, helix α 1 shows significant conformational changes, resulting in partial unfolding. More surprising, the only detectable dynamical changes in NA binding occur on the H1 face, with residues in the two loops on that face showing up to 50% reduction in root-mean-square fluctuations (RMSF).

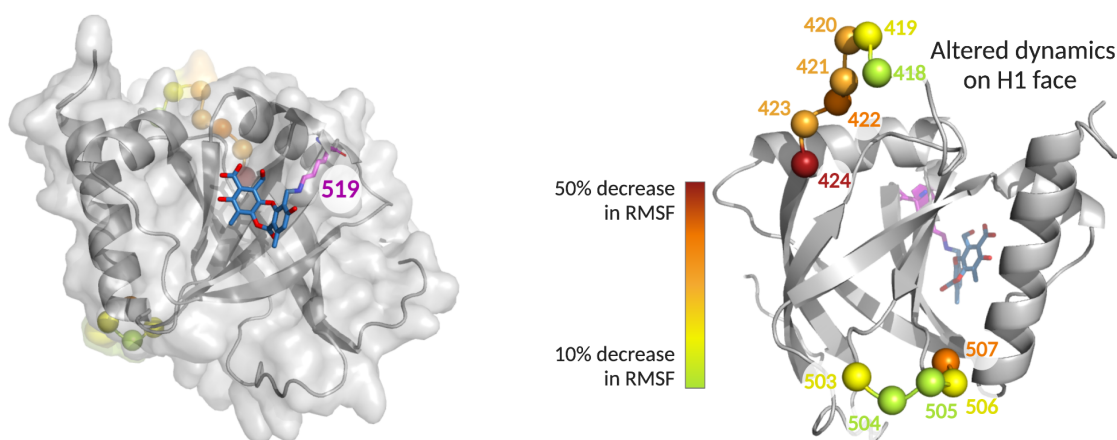


Figure 2.10. Molecular dynamics of AcID with NA modification at K519. Centroid structure of the most populated cluster from molecular dynamics simulations, where norstictic acid binds to the H2 face of Med25 and covalently links to K519. (Right) The residues that showed the greatest reduction in fluctuations (RMSF) upon activator binding all occur on dynamic substructures on the H1 face. Experiments conducted by Amanda Peiffer.

Taken together, the data indicate that NA serves as both an orthosteric inhibitor of H2-binding transcription factors (e.g. ATF6 α) and an allosteric inhibitor of H1 binding transcriptional activators (e.g. ETV5).

Engagement of norstictic acid with Med25 in cells

Next, the engagement of full-length Med25 by NA was tested, along with the resulting impact of PPI formation and function in mammalian cells. Thermal shift assays using freshly prepared HeLa nuclear extracts demonstrate that NA stabilizes endogenous Med25 protein, indicating engagement with the AcID motif in the context of full-length protein (Figure 2.11). Quantification of band density using ImageJ software shows that Med25 is generally stabilized at all tested concentrations, with only slight thermal denaturation occurring at the highest temperature in the 25 μ M condition. To determine an approximate ΔT_m of Med25 + NA from thermal shift assays, extended temperature points were tested (Figure 2.11). Ultimately, it was calculated that an approximate increase in T_m of 13.3°C occurs from NA binding to endogenous Med25.

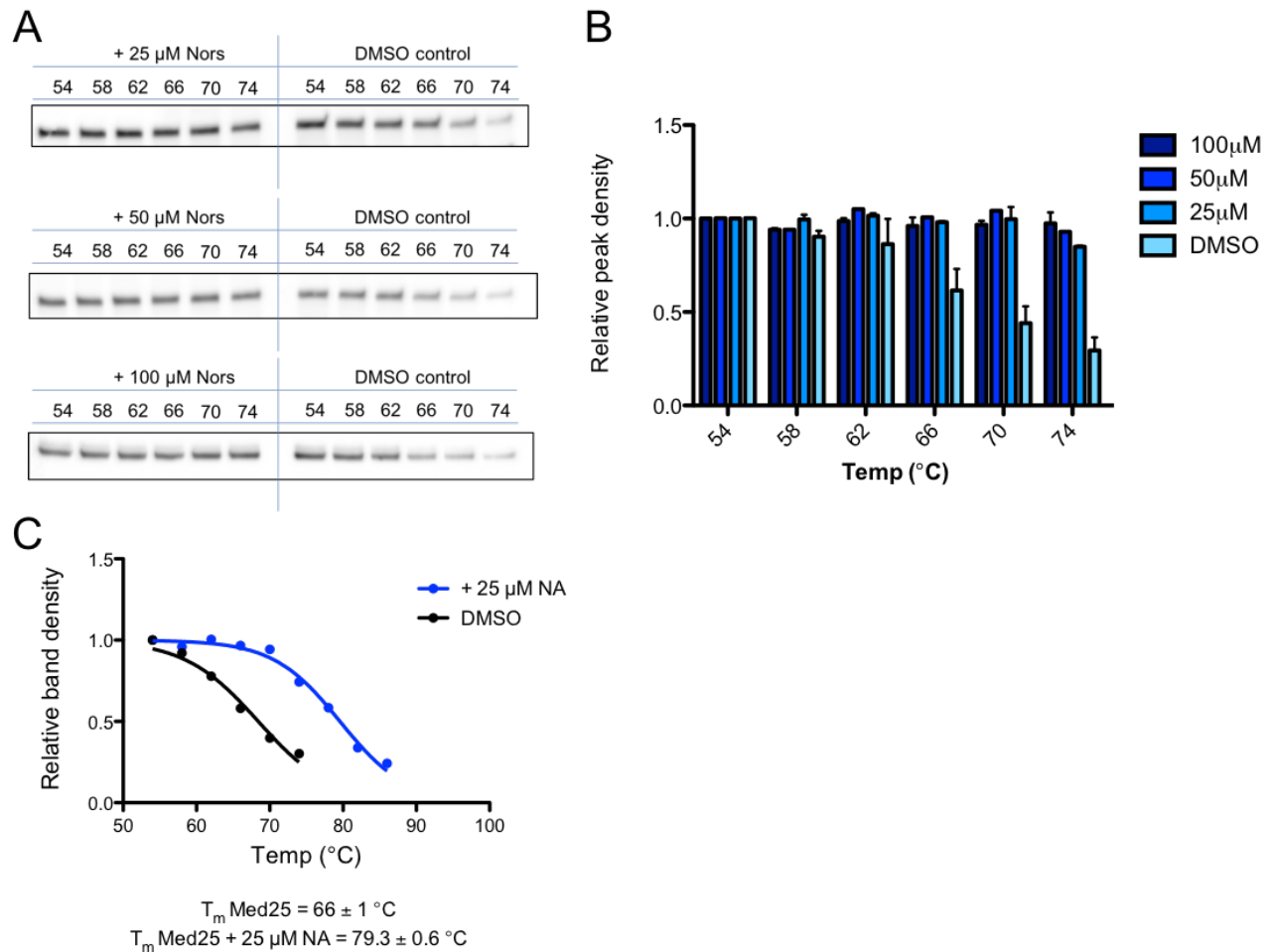


Figure 2.11. Thermal shift assays profiling NA engagement with endogenous Med25. A) Western blot for anti-Med25 in HeLa nuclear extracts heated at various temperature either with or without dosing with NA. B) Quantification of relative band densities from Western blot. NA dosed conditions are the average of biological duplicate. DMSO data is the average of six biological replicates. Error bars represent the standard deviation of the mean. C) NA stabilization of endogenous Med25 tested at extended temperatures (up to 86 $^{\circ}$ C). Band density was then plotted against temperature to determine T_m .

The ability of NA to block PPIs formed between endogenous Med25 and cognate activators was then assessed. ETV5 is a member of the ETV/PEA3 subfamily of transcriptional activators, comprised of ETV5, ETV1, and ETV4. This transcriptional activator trio has nearly identical domains that utilize a PPI with the H1 surface of Med25 for function, and the PPIs are dysregulated in cancer through overexpression of one or both of the binding partners.^{19,44,45} As shown in Figure 2.12, NA treatment of HeLa cells blocks formation of the Med25-ETV complex, consistent with the *in vitro* binding data of Figure 2D.

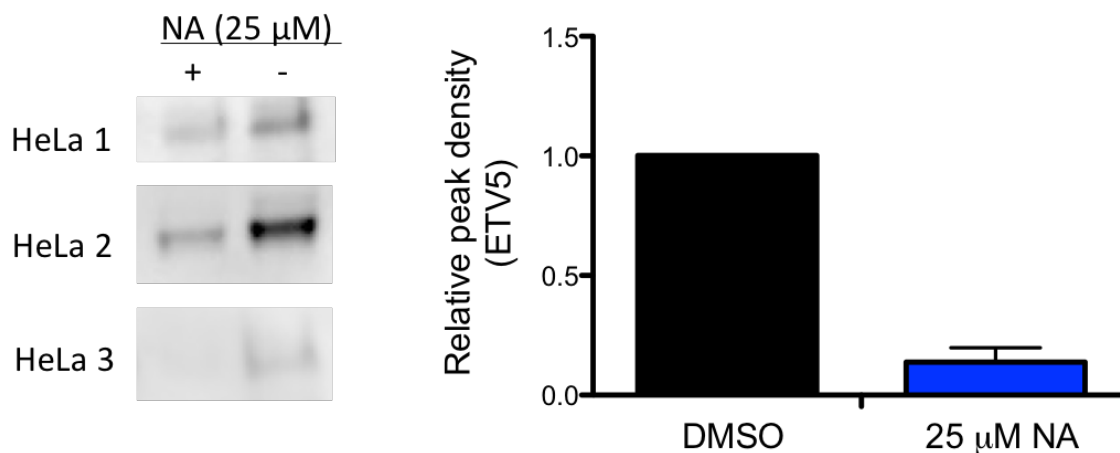


Figure 2.12. Co-immunoprecipitation of ETV5 with Med25 is inhibited by NA in HeLa cells. Western blot against ETV5 in HeLa cells dosed with 25 μ M NA or DMSO (right). Relative band density of anti-ETV5 calculated using ImageJ, normalized to anti-Med25 signal. Values represent the average of biological triplicate experiments (right).

The Med25•ETV/PEA3 PPIs regulate proliferation, invasion, and migration pathways and in at least a subset of cancers, as demonstrated as part of a tyrosine kinase receptor-driven RAS-RAF-MEK-MAPK circuit.^{46,47} Thus, if NA blocks Med25•ETV/PEA3 PPIs, positive synergy with a Her2 inhibitor would be anticipated in some systems. To test this, the combination of NA and lapatinib was tested for synergy by the isobologram method (Figure 2.13) in MDA-MB-231 breast cancer cells, an established model.¹⁹ As can be seen, strong synergy was observed, consistent with NA engagement of Med25 blocking PPIs formed with ETV/PEA3 activators.

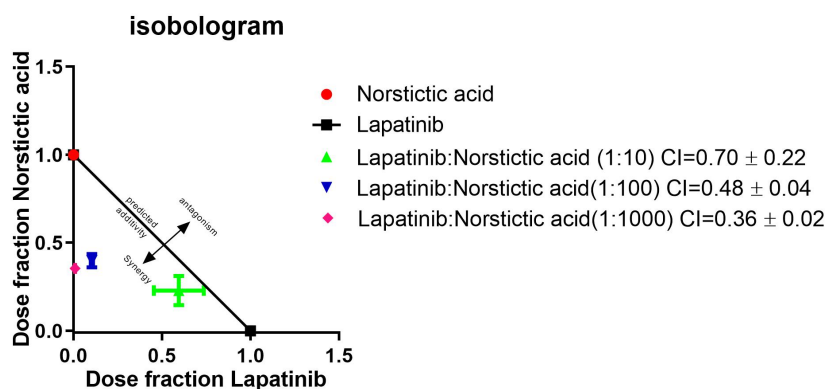


Figure 2.13. Isobologram showing synergy between NA and Lapatinib. NA shows positive synergy with an on-pathway kinase inhibitor, lapatinib. IC₅₀ values of fixed dose ratios of NA and lapatinib were measured in MDA-MB-231 cells after 2 days of dosing and plotted on an isobologram. Experiment conducted by Yejun Liu.

This model was further tested in the patient-derived early passage triple negative breast cancer cell line VARI068 with robust EGFR expression.^{48,49} VARI068 cells exhibit approximately 2-fold upregulation of Med25 relative to normal-like non-tumorigenic MCF10A cells. Treatment of VARI068 cells with NA blocks the Med25•ETV5 complex (Figure 2.14). To further explore the effects of Med25 regulation on cancer, CRISPR-Cas9-mediated knockout of Med25 in VARI068 cells leads to downregulation of Med25•ETV/PEA3-regulated MMP2 (~5 fold) and NA treatment leads to substantial down-regulation of MMP2 relative to vehicle (Figure 2.14). Taken together, these data are consistent with a mechanism in which NA engages Med25 in cells and alters its PPI network with downstream effects on tumor phenotype.

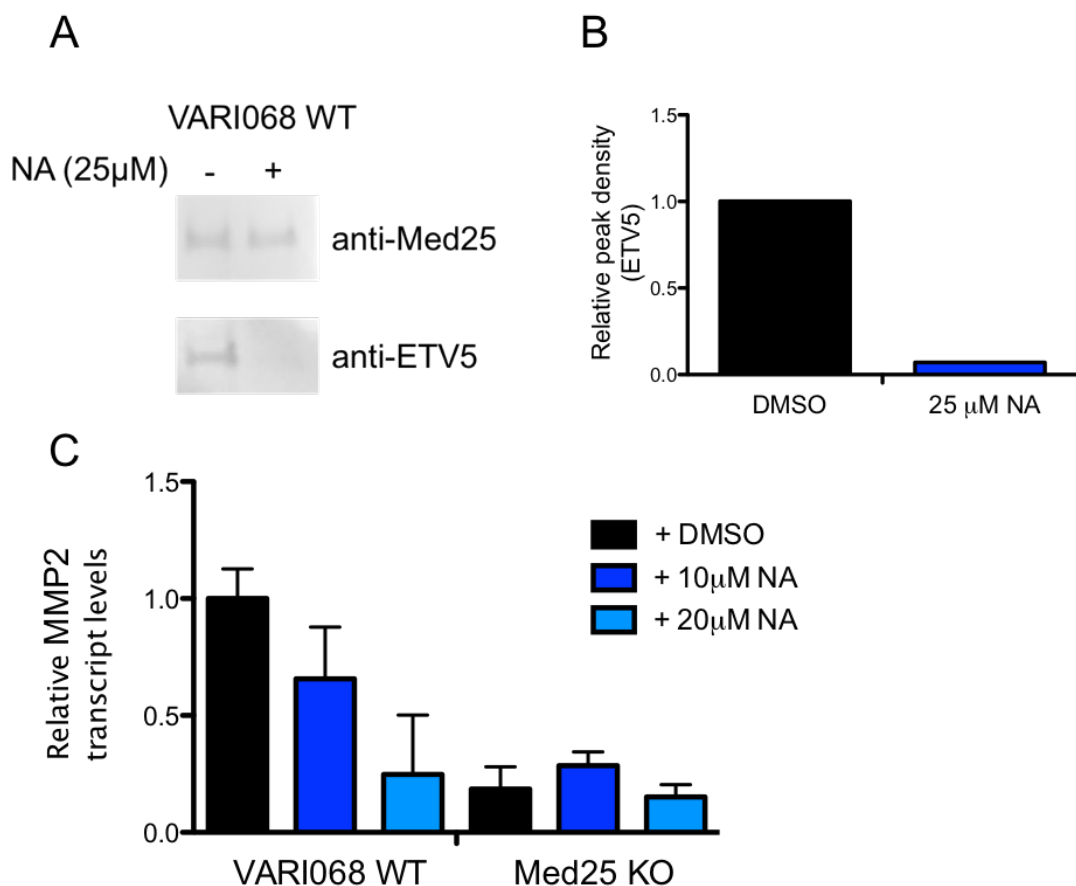


Figure 2.14. NA inhibits Med25 ETV5 interaction in triple negative breast cancer cell line VARI068. Co-immunoprecipitation of ETV5 with Med25 dosed with either NA or DMSO (top left). Relative band density of anti-ETV5 calculated using ImageJ, normalized to anti-Med25 signal (top right). Analysis of MMP2 transcript levels by qPCR indicates that NA treatment decreases MMP2 levels to that of a KO variant of the cell line. MMP2 transcript levels are normalized to the reference gene RPL19. Results shown are the average of technical triplicate experiments, conducted in biological duplicate with error bars representing the standard deviation of the mean (bottom).

In vitro inhibition data indicated that NA would also be a good inhibitor of Med25 interaction with ATF6. Thus, cellular experiments looking at NA's ability to regulate Med25•ATF6 were also conducted. Importantly, for the Med25•ATF6 to occur in cells, the unfolded protein response must be activated. This is accomplished by treating cells with the compound thapsigargin to induce ER stress and activate this pathway prior to dosing with inhibitor (in this case NA).^{26,50} Co-immunoprecipitation experiments looking at Med25•ATF6 after treatment with thapsigargin and dosing with NA or DMSO revealed that NA inhibited Med25•ATF6 (Figure 2.15). To further explore the use of NA as a Med25•ATF6 probe, qPCR experiments looking at NA's effects on HSPA5 expression were conducted. HSPA5 expression has been shown to be regulated by the Med25•ATF6 interaction.²⁶ In VARI068 cells, Med25 KO decreases stimulated HSPA5 expression by about 3.75-fold. Dosing WT cells with NA led to a slightly less than 2-fold decrease in HSPA5 expression (Figure 2.15). Thapsigargin induces the unfolded protein response and expression of HSPA5 through multiple pathways,^{51,52} which could explain why dosing with NA has less of an effect on transcriptional output of the gene target HSPA5 than MMP-2, which we hypothesize is primarily regulated via Med25•ETV5.

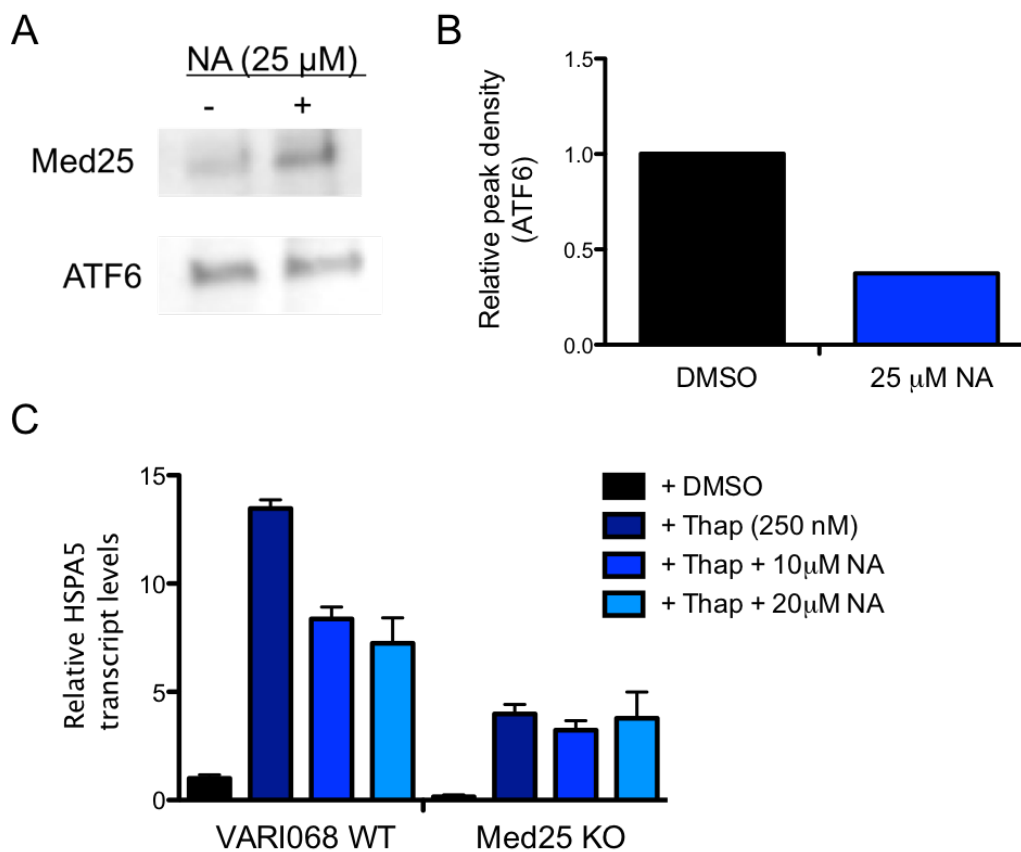


Figure 2.15. NA inhibits Med25 ATF6 interaction mammalian cells line. A) Co-immunoprecipitation of ATF6 with Med25 dosed with either NA or DMSO. B) Relative band density of anti-ATF6 calculated using ImageJ, normalized to anti-Med25 signal. Analysis of HSPA5 transcript levels by qPCR indicates that NA treatment decreases HSPA5 levels in VARI068. C) HSPA5 transcript levels are normalized to the reference gene RPL19. Results shown are the average of technical triplicate experiments with error bars representing the standard deviation of the mean.

2.4 Conclusions

The demonstration of NA as a selective allosteric inhibitor of Med25 function validates the importance of dynamic loops in coactivators in molecular recognition and their utility as targets. We also provide evidence that modulation of these protein-protein interactions would provide a useful tool to study cancer at the bench. Our work suggests that Med25 is a potentially viable clinical target, thus justifying the search for a drug-like compound for testing the clinical utility of the strategy. We anticipate that NA will be a useful probe compound for dissecting the Med25•ETV/PEA3 axis in cancers with Med25 dysregulation. Further, the strategy of targeting dynamic substructures within coactivators should be generalizable beyond Med25. This work, in line with the information presented in Chapter 1, continues to support the hypothesis that targeting

dynamic loops rather than primary binding surfaces (allostery vs orthostery) affords compounds with better selectivity.

2.5 Materials & Methods

Protein expression and purification

The Med25 expression plasmid, referred to as pAcID(394-543)-His6, was generously provided by Prof. Patrick Cramer.²⁷ pAcID(394-543)-His6 mutants, with the exception of K420R, were prepared using site directed mutagenesis with the primers listed below. Plasmid sequence identity was confirmed via standard Sanger sequencing methods at the University of Michigan DNA Sequencing Core and analyzed using SeqMan Pro from the Lasergene DNASTAR software suite.

Primers used for SDM:

pAcID-(K518R)-His6

F Pr. TCATGCTCCTGTACTCGTCCAGGAAGAAGATCTTCATGGGCCTCATCCC

R Pr. GGGATGAGGCCCATGAAGATCTTCTTCCTGGACGAGTACAGGAGCATGA

pAcID-(K519R)-His6

F Pr. TCATGCTCCTGTACTCGTCCAAGAGGAAGATCTTCATGGGCCTCATCCC

R Pr. GGGATGAGGCCCATGAAGATCTTCCTCTTGGACGAGTACAGGAGCATGA

pAcID-(K520R)-His6

F Pr. TCATGCTCCTGTACTCGTCCAAGAAGAGGATCTTCATGGGCCTCATCCC

R Pr. GGGATGAGGCCCATGAAGATCCTCTTCTTGGACGAGTACAGGAGCATGA

pAcID-(KK518RR)-His6

F Pr. TCATGCTCCTGTACTCGTCCAGGAGGAAGATCTTCATGGGCCTCATCCC

R Pr. GGGATGAGGCCCATGAAGATCTTCCTCCTGGACGAGTACAGGAGCATGA

pAcID-(KKK518RRR)-His6

F Pr. TCATGCTCCTGTACTCGTCCAGGAGGAGGATCTTCATGGGCCTCATCCC

R Pr. GGGATGAGGCCCATGAAGATCCTCCTCCTGGACGAGTACAGGAGCATGA

pAcID-(K411R/K413R)-His6

F Pr. GGGGTCCTGGAGTGGCAAGAGAGACCCAGACCTGCCTCAGTGGATGCCAAC

R Pr. GTTGGCATCCACTGAGGCAGGTCTGGGTCTCTCTTGCCACTCCAGGACCCC

Mutant K420R was made using PCR-driven overlap extension.⁵³ The flanking primers A and D were designed to extend into the pET21b vector. Primers B and C contain the mutated codon. The primer designs are as follows:

Primer A: 5'-GAA GGA GAT ATA CAT ATG TCA GTC TCC AAT AAG-3'

Primer B: 5'-GAT GCC AAC ACC CGA CTG ACG CGG TCA CTG-3'

Primer C: 5' CAG TGA CCG CGT CAG TCG GGT GTT GGC ATC-3'

Primer D: 5'-GTG GTG GTG CTC GAG GTT GGT GAT GAC-3'

WT Med25 and mutants were expressed and purified from heat-shock competent Rosetta pLysS cells (Novagen), in Terrific Broth (TB) containing 0.1 mg/mL ampicillin and 0.034 mg/mL chloramphenicol, using previously described conditions.^{20,27} Cells were grown at 37 °C to an optical density (OD_{600nm}) of 0.8. Temperature was reduced to 18°C and protein expression was induced upon addition of IPTG to a final concentration of 0.5 mM. Post-induction, cells were incubated 16 hours at 18°C. Cells were pelleted via centrifugation at 6000xg for 20 mins at 4°C. Cell pellets were stored at -80°C prior to purification. The harvested pellet was thawed on ice and resuspended in 20 mL of lysis buffer (50 mM phosphate, 300 mM sodium chloride, 10 mM imidazole, pH 6.8). Cells were then lysed by sonication on ice and cellular lysates were cleared by centrifugation at 9500 rpm for 20 min at 4°C. The supernatant lysate was then added to 750 μ L Ni-NTA beads (Qiagen) and incubated for 1 hour at 4°C. The resin was pelleted by centrifugation at 2500 rpm for 2 min at 4°C and washed with wash buffer (50 mM phosphate, 300 mM sodium chloride, 30 mM imidazole, pH 6.8) a total of five times. Protein was then eluted with 2 mL of elution buffer (50 mM phosphate, 300 mM sodium chloride, 400 mM imidazole, pH 6.8) a total of three times. Eluent was then pooled and purified by cation exchange FPLC (Source 15S, GE Healthcare) using a gradient of Buffer B (50 mM phosphate, 100 mM NaCl, 1 mM DTT, pH 6.8) in Buffer A (50 mM phosphate, 1 mM DTT). The FPLC purified protein was then dialyzed into storage buffer (10 mM phosphate, 50 mM NaCl, 10% v/v glycerol, 0.001% v/v NP-40, pH 6.8) overnight, concentrated, aliquoted, and stored at -80°C. Final protein was greater than 90% pure as determined by Coomassie stained polyacrylamide gel. Protein concentration was determined by UV-Vis spectroscopy using an extinction coefficient, $\epsilon = 22,460 \text{ M}^{-1} \text{ cm}^{-1}$.

CBP KIX (586-672) was expressed in BL21 DE3 *E. coli* as previously described.³⁰ Cells were grown to an optical density (OD 600nm) of 0.8 (37°C, 250 rpm), induced with 0.25 mM isopropyl b-d-1-thiogalactopyranoside (IPTG) for 16 hours at 20°C, harvested by centrifugation (20 min, 6500xg) and stored at -80°C. Cell pellets were lysed via sonication in lysis buffer (10 mM phosphate, 300 mM NaCl, 10 mM imidazole, pH 7.2) containing 2-mercaptoethanol and cOmplete protease inhibitor (Roche, 11873580001). The Hisx6 tagged protein was affinity purified using immobilized metal ion affinity chromatography (IMAC) on a HisTrap HP Ni Sepharose column (GE Healthcare). Elution was conducted using an imidazole gradient of 10 mM to 600 mM imidazole. An additional round of purification was completed using ion-exchange chromatography on a Source S column (GE Healthcare) in phosphate buffer (50 mM, pH 7.2) by eluting with a NaCl gradient from 0 to 1M. Purified protein was buffer-exchanged by dialysis (overnight, 4°C) into 10 mM phosphate, 100 mM NaCl, 10% glycerol, pH 6.8. Purity was determined by Coomassie stained polyacrylamide gel. Protein concentration was determined by UV-Vis spectroscopy using an extinction coefficient, $\epsilon = 12,950 \text{ M}^{-1} \text{ cm}^{-1}$. Purified protein samples (>90% pure) were aliquoted and stored at -80°C.

Med15 (1-345) was expressed and purified by as previously described.⁴³

The expression plasmid for p300 TAZ1(324-423) was generously provided by Prof. Paramjit Arora.⁵⁴ Protein was expressed in BL21 DE3 *E. coli*. Cells were grown in LB containing 0.1 mg/mL ampicillin and 1 mM ZnCl₂ to an optical density (OD 600nm) of 0.8 (37°C, 250 rpm), cooled to 22°C and induced with 100 μ M IPTC for 5 hours. Cells were harvested by centrifugation (20 min, 6500xg) and stored at -80°C. Cell pellets were lysed by sonication in lysis buffer (50 mM Tris, 150 mM NaCl, pH 6.3) containing cOmplete protease inhibitor (Roche, 11873580001). The GST tagged protein was affinity purified using a GSTrap column (GE Healthcare). After initial binding of the protein to the column, elution was conducted using a buffer containing 10 mM reduced glutathione. An additional round of purification was completed using ion-exchange chromatography on a Source S column (GE Healthcare) in phosphate buffer (50 mM, 1 mM DTT, pH 7.2) by eluting with a NaCl gradient from 0 to 1M. Purified protein was buffer-exchanged by dialysis (overnight, 4°C) into 10 mM Tris, 100 mM NaH₂PO₄, 10% glycerol, 1mM DTT, 100 μ M

ZnCl₂, pH 6.8. Purity was determined by Coomassie stained polyacrylamide gel. Protein concentration was determined by UV-Vis spectroscopy using an extinction coefficient, $\epsilon = 49,110 \text{ M}^{-1} \text{ cm}^{-1}$. Purified protein samples (>90% pure) were aliquoted and stored at -80°C.

Synthesis of transcriptional activation domain peptides

The peptides listed below were prepared following standard Fmoc solid-phase synthesis methods on a Liberty Blue Microwave Synthesizer (CEM). Fmoc deprotections were completed by suspending the resin in 20% piperidine (ChemImpex) in DMF supplemented with 0.2 M Oxyma Pure (CEM) and irradiating under variable power to maintain a temperature of 90°C for 60 seconds. Coupling reactions were completed by combining the amino acid (5 eq relative to resin; CEM, ChemImpex, and NovaBiochem), diisopropylcarbodiimide (7 eq, ChemImpex), and Oxyma Pure (5 eq) in DMF and irradiating under variable power to maintain a temperature of 90°C for 4 minutes. The resin was rinsed four times with an excess of DMF between all deprotection and coupling steps. N-terminal addition of fluoresceine isothiocyanate (FITC) residue was conducted by adding 1.2 eq in 5% diisopropylethylamine in dimethyl formamide for 18 hours at RT. Peptides were deprotected and cleaved from the resin for 4 hours in 90% trifluoroacetic acid (TFA), 5% thioanisole, 3% ethanedithiol (EDT) and 2% anisole. Crude peptides were filtered to remove resin, dried under nitrogen stream, and precipitated from cold ether. Peptide suspensions were transferred to a 15 mL falcon tube, centrifuged at 4000 g for 5 minutes at 4°C, and ether decanted. Crude peptides were resuspended in 20-40% acetonitrile, purified via HPLC on an Agilent 1260 HPLC using a semi-prep C18 column (Phenomenex). Pure fractions were collected and lyophilized to afford pure peptides. Final purity was determined via analytical HPLC and identify was confirmed using mass spectrometry. Analytical spectra were obtained using an analytical C18 column (Phenomenex) on an Agilent 1260 HPLC. Mass spectra were obtained using an Agilent 6230 LC/TOF and an Agilent 6545 LC/Q-TOF. Analytical spectra of peptides can be found in Appendix A.

VP16 (465-490) - Fluorescein isothiocyanate and β -alanine were coupled to the N-terminus to produce FITC- β A-YGALDMADFEFEQMFTDALGIDEYGG. A gradient of 10-40% acetonitrile over 30 min was used for HPLC purification.

ETV5 (38-68) - Fluorescein isothiocyanate and β -alanine were coupled to the N-terminus to produce FITC- β A-ALDMADFEFEQMFTDALG. A gradient of 10-40% acetonitrile over 30 min was used for HPLC purification

ATF6 α (40-66) - Fluorescein isothiocyanate and β -alanine were coupled to the N-terminus to produce FITC- β A-DTDELQLEAANETYENNFDNLDFDLDM. A gradient of 10-40% acetonitrile over 30 min was used for HPLC purification

MLL (840-858) – Fluorescein isothiocyanate and β -alanine were coupled to the N-terminus to produce FITC- β A- DCGNILPSDIMDFVLKNTP. A gradient of 10-40% acetonitrile over 30 min was used for HPLC purification

myb (291-316) - Fluorescein isothiocyanate and β -alanine were coupled to the N-terminus to produce FITC- β A-KEKRIKELELLLMSTENELKGQQVLP. A gradient of 10-40% acetonitrile over 30 min was used for HPLC purification

IBiD (2063–2111) – The N-terminus was acetylated to produce Ac-SPSALQDLLRTLKSPSSPQQQQQVLNILKSNPQLMAAFIKQRTAKYVAN. A gradient of 10-50% acetonitrile over 40 min was used for HPLC purification.

ACTR (1041-1088) – Fluorescein isothiocyanate and β -alanine were coupled to the N-terminus to produce FITC- β A-PSNLEGQSDERALLDQLHTLLSNTDATGLEEIDRALGIPELVNQGQAL. A gradient of 10-50% acetonitrile over 40 min was used for HPLC purification.

HIF1 α (786-826) - Fluorescein isothiocyanate and β -alanine were coupled to the N-terminus to produce FITC- β A-SMDESGLPQLTSYDCEVNAPIQGSRNLLQGEELLRALDQVN. A gradient of 10-30% acetonitrile over 60 min was used for HPLC purification.

Direct binding and competition experiments

Direct binding and competition experiments were performed using fluorescence polarization as previously described.²⁰ Low volume, non-binding, black 384-well plates (Corning) were used and fluorescence polarization was measured using a PHERAStar plate reader with polarized excitation at 485 nm and emission intensity measured through a parallel and perpendicularly polarized 535 nm filters.

Data was analyzed using GraphPad Prism. For direct binding experiments, a binding isotherm accounting for ligand depletion (assuming a 1:1 binding model) was fit to the obtained polarization values as a function of protein concentration to obtain the equilibrium dissociation constant, K_d :

$$y = c + (b - c) \times \frac{(K_d + a + x) - \sqrt{(K_d + a + x)^2 - 4ax}}{2a}$$

“a” and “x” are the total concentrations of fluorescent peptide and protein, respectively, “y” is the observed anisotropy at a given protein 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 representing the standard deviation the mean. For competition experiments, curves were fit with a non-linear regression using the “log(inhibitor) vs response – variable slope” equation, allowing calculation of IC_{50} .

AcID•VP16 High-throughput screening

Assays were performed in a final volume of 20 μ L in low volume, non-binding, black 384-well plates (Corning) and read by plate reader (Pherastar) with excitation at 485 nm and emission intensity measured through parallel and perpendicularly polarized 535 nm filters. Optimization of fluorescence polarization assay for high throughput was conducted by testing stability of the VP16(465-490)•AcID interaction (K_d) over time, with combinations of DMSO and NP-40 (for more details see the dissertation of Dr. Steven Sturilis: <http://hdl.handle.net/2027.42/120661>). The assay shows little variance in affinity over time, up to 20 hours as well as tolerance to DMSO (5%) and NP-40 (0.001%). 4046 compounds were tested from the MS Spectrum 2000, Focused

Collections, and BioFocus NCC libraries, which include known bioactive molecules, secondary metabolites, natural products, and FDA approved drugs. 200 nL of each compound in DMSO was first plated, followed by addition of 10 μ L FITC-VP16(465-490). The compounds were then tested for fluorescence quenching before 10 μ L of Med25 AcID protein was added. Plates were incubated for thirty minutes at room temperature and read by plate reader, as described above with gain settings determined based on a well from columns 23-24 (tracer only). Final concentration of AcID protein was 850 nM, final concentration of FITC-VP16(465-490) was 20 nM, and compounds were assayed at a concentration of 20 μ M with a final DMSO concentration of 1% v/v. Data was published to and analyzed using MScreen (<http://mscreen.lsi.umich.edu>). Conducted by Steven Sturlis; published at <http://hdl.handle.net/2027.42/120661>.

Mass spectrometry analysis of covalent adducts

Protein (Med25 WT and mutants) was diluted to a concentration of 20 μ M using storage buffer (10 mM phosphate, 50 mM NaCl, 10% v/v glycerol, 0.001% v/v NP-40, pH 6.8). Norstictic acid was added to the protein to give a final concentration of 20 μ M (1 equivalent, unless otherwise noted). Samples were incubated for 30 minutes at room temperature with gentle mixing on an orbital shaker. Analysis was conducted by mass spectrometry using an Agilent QToF LC/MS equipped with a Poroshell 300SB C8 reverse-phased column with a gradient of 5-100% acetonitrile with 0.1% formic acid in water with 0.1% formic acid over five minutes. Analysis of data was completed using the Agilent Qualitative Analysis Program with background subtraction and deconvolution settings for an intact protein of 10,000- 30,000 Da.

Molecular dynamics simulations

Modeling was performed using the NMR structure of Med25 AcID (PDB 2xnf). The norstictic acid was parameterized using CGENFF, which was then covalently linked to Med25 K519 through a PATCH that was created in CHARMM, with the molecules oriented out in space to allow for full, unbiased exploration around the protein before binding. The system was solvated using TIP3P water molecules as well as 100 mM NaCl using the MMTSB toolset so that the linked complex was in a cubic box with a minimum cutoff distance being 10 Å from the box edges. Simulations were unbiased molecular dynamics simulations using the CHARMM36 and CGENFF force fields for 100 ns of sampling at 298 K after allowing for 2ns of equilibration of the system.

The simulation was run in the NVT ensemble using the Langevin dynamics algorithm with a friction coefficient of 5 ps^{-1} . The SHAKE algorithm was used to fix bond lengths during simulations. PME and vswitch were used for nonbonded interactions with a 12 Å cutoff. All molecular dynamics simulations were performed using GPUs through the CHARMM compatible OpenMM interface. Five independent trials of simulations were performed for each molecule. Conducted by Amanda Peiffer.

Cell Culture

HeLa cells were purchased from ATCC (HTB-55). VARI068 cells were derived from a patient-derived xenograft orthotopically implanted in NSG mice first and subsequent passaged through nude mice. HeLa cells were grown in DMEM (Gibco, 11965-092) supplemented with 10% FBS (R&D, S11150). VARI068 cells were grown in DMEM (Gibco, 11965-092) supplemented with 10% FBS (R&D, S11150), and 1X Antibiotic-Antimycotic. Both cell lines were grown at 37C with 5% CO₂.

Thermal shift assays

HeLa cells were grown, counted, and harvested using standard protocols (300,000 cells per temperature). Pelleted cells were resuspended in 1 mL cold PBS and transferred to a 1.5 mL Eppendorf tube on ice. Cells were again pelleted by centrifugation at 1500 G for 3 minutes at 4°C. Nuclear extracts were generated using a NE-PER kit (Thermo Scientific, 78833) and manufacturer's protocol. After isolating the nuclear extract, a buffer exchange into PBS was conducted using a Zeba Spin Desalting Column 7K MWCO (Thermo Scientific, 89882). Prepared nuclear extracts were split into 2 epitubes. Norethisterone acid (dissolved in DMSO) was added to one tube and an equivalent volume of DMSO was added to the other. Final concentration of DMSO was 0.1% v/v. Dosed nuclear extracts were incubated at room temperature for 30 minutes. After incubation, samples were aliquoted into thin-walled PCR tubes (15 µl per tube, the equivalent of 300,000 cells per tube).

A Labnet Multigene OPTIMAX PCR was used to heat each sample for 3 minutes. Six temperature points were tested, 54, 58, 62, 66, 70, and 74C. Contents were transferred to epitubes and centrifuged at 17000 g for 1 minute at 4°C to remove precipitated proteins. Contents of each epitube were carefully transferred to a clean epitube, leaving precipitated protein behind. LDS

loading dye was added and samples were boiled for 5 minutes at 95°C. 10µL of each sample was loaded onto a 4-20% mini-PROTEAN TGX gel (BioRad, 4561096) gel was run at 180V for 45 minutes. Protein was transferred from gels to PVDF membrane using a BioRaD TransferBox Turbo following the standard protocols. Membrane was blocked for 1 hour at room temperature using Super Block (Thermo Scientific, 37515). Med25 antibody (Novus biologicals, NBP2-55868) was added to membrane (1:1000 dilution in Super Block) and incubated overnight at 4°C with gentle shaking. After removal of primary antibody and three washes with PBST, Secondary antibody (Santa Cruz SC-2004, 1:20,000 in Super Block) was added to membrane and incubated at RT for 1hr with shaking. After removal of secondary antibody with three washes with PBST, HRP substrate (Thermo Scientific, 34095) was added and after 1 minute Western blot was visualized using Chemiluminescence on an Azure Biosystems c600 imager. Analysis was conducted using ImageJ software.

Co-Immunoprecipitation experiments

Med25 antibody (Santa Cruz, sc-393759) was chemically crosslinked to Dynabeads Protein G (Invitrogen, 10004D) using Bis(sulfosuccinimidyl) suberate (BS³). Briefly, 20 µL Dynabeads Protein G were washed with 250 µL PBST 3 times. Med25 Santa Cruz antibody (24 µL, 4.8ug) in 400 µL PBST was added to beads and incubated on with rotation at 4°C for 1hr. Antibody coupled beads were washed twice with conjugation buffer (20mM sodium phosphate, 150 mM NaCl, pH = 7.5), resuspended in 250 µL of 5mM BS³ (Thermo Scientific, 21580), and incubated at RT with rotation. After 30 minutes, 12.5 µL quenching buffer (1M Tris HCl, pH 7.5) was added and beads were incubated an additional 15 min with rotation. Crosslinked beads were washed three times with PBST and immediately used or stored at 4°C in PBST and used within 24 hours.

Mammalian cells were harvested as described in 1-9. Nuclear extracts were generated from mammalian cells using a NE-PER kit (Thermo Scientific, 78833). Extracts were buffer exchanged into PBS using Zeba Spin Desalting Column 7K MWCO (Thermo Scientific, 89882). Extracts were dosed with either Norstictic acid (dissolved in DMSO) or an equivalent volume of DMSO (0.1% v/v). After washing crosslinked beads with PBS, nuclear extracts were added and incubated for 3 hours at 4°C with rotation. Flowthrough was collected and saved, and beads were washed 3X with PBST. LDS sample buffer (Invitrogen, NP0007; 2X final concentration) was added to beads

and incubated at 95°C for 10 minutes to elute immunoprecipitated proteins. Samples were run on 4-20% mini-PROTEAN TGX gels (BioRad, 4561096). Transfer and blotting was conducted using standard protocols (see cellular thermal shift assay protocol). ETV5 antibody (Proteintech, 13011-1-AP) diluted 1:2000 in SuperBlock for use.

Quantitative polymerase chain reaction experiments

For endogenous gene expression analysis, HeLa cells were seeded into a 24-well plate (1×10^5 cells/well) and allowed to adhere overnight. Media was removed and replaced with OptiMem media containing vehicle (DMSO) or compound delivered (in DMSO, 0.5% v/v) at the indicated concentrations. After 6 h, the media was removed and total RNA was isolated using RNeasy Plus RNA isolation kits (Qiagen) according to manufacturer's instructions. Each RNA sample was used to synthesize cDNA using iScript cDNA synthesis kits (Bio-Rad). RT-qPCR reactions were carried out in triplicate on an Applied Biosystems StepPlusOne instrument using SYBR green master mix and primers for:

human RPL19 Forward. 5':ATGTATCACAGCCTGTACCTG:3'

human RPL19 Reverse. 5':TTCTTGGTCTCTCTTCCTCCTTG:3'

MMP-2 Forward. 5':CATTCCAGGCATCTGCGATGAG:3'

MMP-2 Reverse. 5':AGCGAGTGGATGCCGCCTTTAA:3'

RT-qPCR analysis was carried out using the comparative CT Method ($\Delta\Delta$ CT Method) to estimate MMP-2 mRNA levels relative to the reference RPL19 mRNA levels. Experiments were conducted in biological duplicate, technical triplicate.

Synergy experiments with lapatinib (performed by Yejun Liu)

Cell culture: MDA-MB-231 cells were purchased on ATCC website. Cells were seed at 3000 cells per well in 96 well plates and cultured in DMEM(Gibco) with 10% FBS and no antibiotics. After 24 hours, cells were adhered on the plate. Medium was changed from 10% FBS to 1% FBS and at the same time, added appropriate amount of compound (Keep DMSO at 1%). After 24 hours, old medium was removed and new 1% FBS DMEM medium and compounds were

added to each well. The day after this treatment, cell viability was measured by Cell Proliferation Kit (MTT) from Roche following the manufacturer's instructions.

Calculation of synergy: All calculations were performed in Graphpad Prism and Microsoft Excel 365. Isobolograms were generated based on the dose fractions calculated by the IC_{50} s of either Norstictic acid, Laptinb alone or combination of them in different ratios. In this case, dose fraction is defined as the IC_{50} of one component (Norstictic acid or Laptinb) in a combination divided by the IC_{50} of that component in isolation required to exert the same effect. The dose fractions of Norstictic acid and Laptinb represent the x/y coordinates on the isobologram. Combination index (CI) is the sum of two dose fractions. For single agent, CI is 1 and for combination, synergy effect is present when $CI < 1$.

Dose fraction of Compound A for combination AB = $IC_{50}(A \text{ in AB}) / IC_{50}(A \text{ in isolation})$

Dose fraction of Compound B for combination AB = $IC_{50}(B \text{ in AB}) / IC_{50}(B \text{ in isolation})$

CI (combination index) = dose fraction A + dose fraction B.

2.5 References

- (1) Krishnamurthy, M.; Dugan, A.; Nwokoye, A.; Fung, Y.-H.; Lancia, J. K.; Majmudar, C. Y.; Mapp, A. K. Caught in the Act: Covalent Cross-Linking Captures Activator–Coactivator Interactions in Vivo. *ACS Chem. Biol.* **2011**, *6* (12), 1321–1326. <https://doi.org/10.1021/cb200308e>.
- (2) Sabari, B. R.; Dall’Agnese, A.; Boija, A.; Klein, I. A.; Coffey, E. L.; Shrinivas, K.; Abraham, B. J.; Hannett, N. M.; Zamudio, A. V.; Manteiga, J. C.; Li, C. H.; Guo, Y. E.; Day, D. S.; Schuijers, J.; Vasile, E.; Malik, S.; Hnisz, D.; Lee, T. I.; Cisse, I. I.; Roeder, R. G.; Sharp, P. A.; Chakraborty, A. K.; Young, R. A. Coactivator Condensation at Super-Enhancers Links Phase Separation and Gene Control. *Science* **2018**, *361* (6400). <https://doi.org/10.1126/science.aar3958>.
- (3) Yang, F.; Vought, B. W.; Satterlee, J. S.; Walker, A. K.; Jim Sun, Z.-Y.; Watts, J. L.; DeBeaumont, R.; Saito, R. M.; Hyberts, S. G.; Yang, S.; Macol, C.; Iyer, L.; Tjian, R.; van den Heuvel, S.; Hart, A. C.; Wagner, G.; Näär, A. M. An ARC/Mediator Subunit Required for SREBP Control of Cholesterol and Lipid Homeostasis. *Nature* **2006**, *442* (7103), 700–704. <https://doi.org/10.1038/nature04942>.
- (4) Larivière, L.; Seizl, M.; Cramer, P. A Structural Perspective on Mediator Function. *Curr. Opin. Cell Biol.* **2012**, *24* (3), 305–313. <https://doi.org/10.1016/j.ceb.2012.01.007>.
- (5) Thakur, J. K.; Yadav, A.; Yadav, G. Molecular Recognition by the KIX Domain and Its Role in Gene Regulation. *Nucleic Acids Res.* **2014**, *42* (4), 2112–2125. <https://doi.org/10.1093/nar/gkt1147>.
- (6) Wang, H.; Dienemann, C.; Stützer, A.; Urlaub, H.; Cheung, A. C. M.; Cramer, P. Structure of the Transcription Coactivator SAGA. *Nature* **2020**, *577* (7792), 717–720. <https://doi.org/10.1038/s41586-020-1933-5>.
- (7) Kurokawa, R.; Kalafus, D.; Ogliastro, M.-H.; Kioussi, C.; Xu, L.; Torchia, J.; Rosenfeld, M. G.; Glass, C. K. Differential Use of CREB Binding Protein-Coactivator Complexes. *Science* **1998**, *279* (5351), 700–703. <https://doi.org/10.1126/science.279.5351.700>.
- (8) York, B.; O’Malley, B. W. Steroid Receptor Coactivator (SRC) Family: Masters of Systems Biology*. *J. Biol. Chem.* **2010**, *285* (50), 38743–38750. <https://doi.org/10.1074/jbc.R110.193367>.
- (9) Näär, A. M.; Lemon, B. D.; Tjian, R. Transcriptional Coactivator Complexes. *Annu. Rev. Biochem.* **2001**, *70* (1), 475–501. <https://doi.org/10.1146/annurev.biochem.70.1.475>.
- (10) Bushweller, J. H. Targeting Transcription Factors in Cancer — from Undruggable to Reality. *Nat. Rev. Cancer* **2019**, *19* (11), 611–624. <https://doi.org/10.1038/s41568-019-0196-7>.
- (11) Mapp, A. K.; Pricer, R.; Sturlis, S. Targeting Transcription Is No Longer a Quixotic Quest. *Nat. Chem. Biol.* **2015**, *11* (12), 891–894. <https://doi.org/10.1038/nchembio.1962>.
- (12) Garlick, J. M.; Mapp, A. K. Selective Modulation of Dynamic Protein Complexes. *Cell Chem. Biol.* **2020**, *27* (8), 986–997. <https://doi.org/10.1016/j.chembiol.2020.07.019>.
- (13) Breen, M. E.; Mapp, A. K. Modulating the Masters: Chemical Tools to Dissect CBP and P300 Function. *Curr. Opin. Chem. Biol.* **2018**, *45*, 195–203. <https://doi.org/10.1016/j.cbpa.2018.06.005>.
- (14) Tien, J. C.-Y.; Xu, J. Steroid Receptor Coactivator-3 as a Potential Molecular Target for Cancer Therapy. *Expert Opin. Ther. Targets* **2012**, *16* (11), 1085–1096. <https://doi.org/10.1517/14728222.2012.718330>.

- (15) Nierenberg, A. A.; Ghaznavi, S. A.; Sande Mathias, I.; Ellard, K. K.; Janos, J. A.; Sylvia, L. G. Peroxisome Proliferator-Activated Receptor Gamma Coactivator-1 Alpha as a Novel Target for Bipolar Disorder and Other Neuropsychiatric Disorders. *Biol. Psychiatry* **2018**, *83* (9), 761–769. <https://doi.org/10.1016/j.biopsych.2017.12.014>.
- (16) Yu, Y.; Su, X.; Qin, Q.; Hou, Y.; Zhang, X.; Zhang, H.; Jia, M.; Chen, Y. Yes-Associated Protein and Transcriptional Coactivator with PDZ-Binding Motif as New Targets in Cardiovascular Diseases. *Pharmacol. Res.* **2020**, *159*, 105009. <https://doi.org/10.1016/j.phrs.2020.105009>.
- (17) Pricer, R.; Gestwicki, J. E.; Mapp, A. K. From Fuzzy to Function: The New Frontier of Protein-Protein Interactions. *Acc. Chem. Res.* **2017**, *50* (3), 584–589. <https://doi.org/10.1021/acs.accounts.6b00565>.
- (18) El Khattabi, L.; Zhao, H.; Kalchschmidt, J.; Young, N.; Jung, S.; Van Blerkom, P.; Kieffer-Kwon, P.; Kieffer-Kwon, K.-R.; Park, S.; Wang, X.; Krebs, J.; Tripathi, S.; Sakabe, N.; Sobreira, D. R.; Huang, S.-C.; Rao, S. S. P.; Pruett, N.; Chauss, D.; Sadler, E.; Lopez, A.; Nóbrega, M. A.; Aiden, E. L.; Asturias, F. J.; Casellas, R. A Pliable Mediator Acts as a Functional Rather Than an Architectural Bridge between Promoters and Enhancers. *Cell* **2019**, *178* (5), 1145–1158.e20. <https://doi.org/10.1016/j.cell.2019.07.011>.
- (19) Verger, A.; Baert, J.-L.; Verreman, K.; Dewitte, F.; Ferreira, E.; Lens, Z.; de Launoit, Y.; Villeret, V.; Monté, D. The Mediator Complex Subunit MED25 Is Targeted by the N-Terminal Transactivation Domain of the PEA3 Group Members. *Nucleic Acids Res.* **2013**, *41* (9), 4847–4859. <https://doi.org/10.1093/nar/gkt199>.
- (20) Henderson, A. R.; Henley, M. J.; Foster, N. J.; Peiffer, A. L.; Beyersdorf, M. S.; Stanford, K. D.; Sturlis, S. M.; Linhares, B. M.; Hill, Z. B.; Wells, J. A.; Cierpicki, T.; Brooks, C. L.; Fierke, C. A.; Mapp, A. K. Conservation of Coactivator Engagement Mechanism Enables Small-Molecule Allosteric Modulators. *Proc. Natl. Acad. Sci.* **2018**, *115* (36), 8960–8965. <https://doi.org/10.1073/pnas.1806202115>.
- (21) Pellecchia, A.; Pescucci, C.; De Lorenzo, E.; Luceri, C.; Passaro, N.; Sica, M.; Notaro, R.; De Angioletti, M. Overexpression of ETV4 Is Oncogenic in Prostate Cells through Promotion of Both Cell Proliferation and Epithelial to Mesenchymal Transition. *Oncogenesis* **2012**, *1* (7), e20. <https://doi.org/10.1038/oncsis.2012.20>.
- (22) Landrieu, I.; Verger, A.; Baert, J.-L.; Rucktooa, P.; Cantrelle, F.-X.; Dewitte, F.; Ferreira, E.; Lens, Z.; Villeret, V.; Monté, D. Characterization of ERM Transactivation Domain Binding to the ACID/PTOV Domain of the Mediator Subunit MED25. *Nucleic Acids Res.* **2015**, *43* (14), 7110–7121. <https://doi.org/10.1093/nar/gkv650>.
- (23) Mittler, G.; Stühler, T.; Santolin, L.; Uhlmann, T.; Kremmer, E.; Lottspeich, F.; Berti, L.; Meisterernst, M. A Novel Docking Site on Mediator Is Critical for Activation by VP16 in Mammalian Cells. *EMBO J.* **2003**, *22* (24), 6494–6504. <https://doi.org/10.1093/emboj/cdg619>.
- (24) Yang, F.; DeBeaumont, R.; Zhou, S.; Näär, A. M. The Activator-Recruited Cofactor/Mediator Coactivator Subunit ARC92 Is a Functionally Important Target of the VP16 Transcriptional Activator. *Proc. Natl. Acad. Sci.* **2004**, *101* (8), 2339–2344. <https://doi.org/10.1073/pnas.0308676100>.
- (25) Vojnic, E.; Mourão, A.; Seizl, M.; Simon, B.; Wenzek, L.; Larivière, L.; Baumli, S.; Baumgart, K.; Meisterernst, M.; Sattler, M.; Cramer, P. Structure and VP16 Binding of the Mediator Med25 Activator Interaction Domain. *Nat. Struct. Mol. Biol.* **2011**, *18* (4), nsmb.1997. <https://doi.org/10.1038/nsmb.1997>.

- (26) Sela, D.; Conkright, J. J.; Chen, L.; Gilmore, J.; Washburn, M. P.; Florens, L.; Conaway, R. C.; Conaway, J. W. Role for Human Mediator Subunit MED25 in Recruitment of Mediator to Promoters by Endoplasmic Reticulum Stress-Responsive Transcription Factor ATF6 α . *J. Biol. Chem.* **2013**, *288* (36), 26179–26187. <https://doi.org/10.1074/jbc.M113.496968>.
- (27) Vojnic, E.; Mourão, A.; Seizl, M.; Simon, B.; Wenzek, L.; Larivière, L.; Baumli, S.; Baumgart, K.; Meisterernst, M.; Sattler, M.; Cramer, P. Structure and VP16 Binding of the Mediator Med25 Activator Interaction Domain. *Nat. Struct. Mol. Biol.* **2011**, *18* (4), 404–409. <https://doi.org/10.1038/nsmb.1997>.
- (28) Henley, M. J.; Linhares, B. M.; Morgan, B. S.; Cierpicki, T.; Fierke, C. A.; Mapp, A. K. Unexpected Specificity within Dynamic Transcriptional Protein–Protein Complexes. *Proc. Natl. Acad. Sci.* **2020**, *117* (44), 27346–27353. <https://doi.org/10.1073/pnas.2013244117>.
- (29) Landrieu, I.; Verger, A.; Baert, J.-L.; Rucktooa, P.; Cantrelle, F.-X.; Dewitte, F.; Ferreira, E.; Lens, Z.; Villeret, V.; Monté, D. Characterization of ERM Transactivation Domain Binding to the ACID/PTOV Domain of the Mediator Subunit MED25. *Nucleic Acids Res.* **2015**, *43* (14), 7110–7121. <https://doi.org/10.1093/nar/gkv650>.
- (30) Wang, N.; Majmudar, C. Y.; Pomerantz, W. C.; Gagnon, J. K.; Sadowsky, J. D.; Meagher, J. L.; Johnson, T. K.; Stuckey, J. A.; Brooks, C. L.; Wells, J. A.; Mapp, A. K. Ordering a Dynamic Protein via a Small-Molecule Stabilizer. *J. Am. Chem. Soc.* **2013**, *135* (9), 3363–3366. <https://doi.org/10.1021/ja3122334>.
- (31) Lao, B. B.; Drew, K.; Guarracino, D. A.; Brewer, T. F.; Heindel, D. W.; Bonneau, R.; Arora, P. S. Rational Design of Topographical Helix Mimics as Potent Inhibitors of Protein–Protein Interactions. *J. Am. Chem. Soc.* **2014**, *136* (22), 7877–7888. <https://doi.org/10.1021/ja502310r>.
- (32) Gräb, J.; Berg, A.; Blechschmidt, L.; Klüver, B.; Rubner, S.; Fu, D. Y.; Meiler, J.; Gräber, M.; Berg, T. The STAT5b Linker Domain Mediates the Selectivity of Catechol Bisphosphates for STAT5b over STAT5a. *ACS Chem. Biol.* **2019**, *14* (4), 796–805. <https://doi.org/10.1021/acschembio.9b00137>.
- (33) Elumalai, N.; Berg, A.; Natarajan, K.; Scharow, A.; Berg, T. Nanomolar Inhibitors of the Transcription Factor STAT5b with High Selectivity over STAT5a. *Angew. Chem. Int. Ed.* **2015**, *54* (16), 4758–4763. <https://doi.org/10.1002/anie.201410672>.
- (34) Gräb, J.; Berg, T. The Selectivity of Fosfosal for STAT5b over STAT5a Is Mediated by Arg566 in the Linker Domain. *ChemBioChem* **2020**, *n/a* (n/a). <https://doi.org/10.1002/cbic.202000111>.
- (35) Chakraborty, S.; Inukai, T.; Fang, L.; Golkowski, M.; Maly, D. J. Targeting Dynamic ATP-Binding Site Features Allows Discrimination between Highly Homologous Protein Kinases. *ACS Chem. Biol.* **2019**, *14* (6), 1249–1259. <https://doi.org/10.1021/acschembio.9b00214>.
- (36) Lake, E. W.; Muretta, J. M.; Thompson, A. R.; Rasmussen, D. M.; Majumdar, A.; Faber, E. B.; Ruff, E. F.; Thomas, D. D.; Levinson, N. M. Quantitative Conformational Profiling of Kinase Inhibitors Reveals Origins of Selectivity for Aurora Kinase Activation States. *Proc. Natl. Acad. Sci. U. S. A.* **2018**, *115* (51), E11894–E11903. <https://doi.org/10.1073/pnas.1811158115>.
- (37) Sandhu, M.; Touma, A. M.; Dysthe, M.; Sadler, F.; Sivaramakrishnan, S.; Vaidehi, N. Conformational Plasticity of the Intracellular Cavity of GPCR–G-Protein Complexes

- Leads to G-Protein Promiscuity and Selectivity. *Proc. Natl. Acad. Sci.* **2019**, *116* (24), 11956–11965. <https://doi.org/10.1073/pnas.1820944116>.
- (38) Hollingsworth, S. A.; Kelly, B.; Valant, C.; Michaelis, J. A.; Mastromihalis, O.; Thompson, G.; Venkatakrishnan, A. J.; Hertig, S.; Scammells, P. J.; Sexton, P. M.; Felder, C. C.; Christopoulos, A.; Dror, R. O. Cryptic Pocket Formation Underlies Allosteric Modulator Selectivity at Muscarinic GPCRs. *Nat. Commun.* **2019**, *10* (1), 1–9. <https://doi.org/10.1038/s41467-019-11062-7>.
 - (39) Fang, Z.; Grütter, C.; Rauh, D. Strategies for the Selective Regulation of Kinases with Allosteric Modulators: Exploiting Exclusive Structural Features. *ACS Chem. Biol.* **2013**, *8* (1), 58–70. <https://doi.org/10.1021/cb300663j>.
 - (40) Mittag, T.; Kay, L. E.; Forman-Kay, J. D. Protein Dynamics and Conformational Disorder in Molecular Recognition. *J. Mol. Recognit. JMR* **2010**, *23* (2), 105–116. <https://doi.org/10.1002/jmr.961>.
 - (41) Thakur, J. K.; Yadav, A.; Yadav, G. Molecular Recognition by the KIX Domain and Its Role in Gene Regulation. *Nucleic Acids Res.* **2014**, *42* (4), 2112–2125. <https://doi.org/10.1093/nar/gkt1147>.
 - (42) Taatjes, D. J.; Näär, A. M.; Andel, F.; Nogales, E.; Tjian, R. Structure, Function, and Activator-Induced Conformations of the CRSP Coactivator. *Science* **2002**, *295* (5557), 1058–1062. <https://doi.org/10.1126/science.1065249>.
 - (43) Majmudar, C. Y.; Højfeldt, J. W.; Arevang, C. J.; Pomerantz, W. C.; Gagnon, J. K.; Schultz, P. J.; Cesa, L. C.; Doss, C. H.; Rowe, S. P.; Vásquez, V.; Tamayo-Castillo, G.; Cierpicki, T.; Brooks, C. L.; Sherman, D. H.; Mapp, A. K. Sekikaic Acid and Lobaric Acid Target a Dynamic Interface of the Coactivator CBP/P300. *Angew. Chem. Int. Ed.* **2012**, *51* (45), 11258–11262. <https://doi.org/10.1002/anie.201206815>.
 - (44) Helgeson, B. E.; Tomlins, S. A.; Shah, N.; Laxman, B.; Cao, Q.; Prensner, J. R.; Cao, X.; Singla, N.; Montie, J. E.; Varambally, S.; Mehra, R.; Chinnaiyan, A. M. Characterization of TMPRSS2:ETV5 and SLC45A3:ETV5 Gene Fusions in Prostate Cancer. *Cancer Res.* **2008**, *68* (1), 73–80. <https://doi.org/10.1158/0008-5472.CAN-07-5352>.
 - (45) Tomlins, S. A.; Laxman, B.; Dhanasekaran, S. M.; Helgeson, B. E.; Cao, X.; Morris, D. S.; Menon, A.; Jing, X.; Cao, Q.; Han, B.; Yu, J.; Wang, L.; Montie, J. E.; Rubin, M. A.; Pienta, K. J.; Roulston, D.; Shah, R. B.; Varambally, S.; Mehra, R.; Chinnaiyan, A. M. Distinct Classes of Chromosomal Rearrangements Create Oncogenic ETS Gene Fusions in Prostate Cancer. *Nature* **2007**, *448* (7153), 595–599. <https://doi.org/10.1038/nature06024>.
 - (46) Kar, A.; Gutierrez-Hartmann, A. Molecular Mechanisms of ETS Transcription Factor-Mediated Tumorigenesis. *Crit. Rev. Biochem. Mol. Biol.* **2013**, *48* (6), 522–543. <https://doi.org/10.3109/10409238.2013.838202>.
 - (47) Shepherd, T. G.; Kockeritz, L.; Szrajber, M. R.; Muller, W. J.; Hassell, J. A. The Pea3 Subfamily Ets Genes Are Required for HER2/Neu-Mediated Mammary Oncogenesis. *Curr. Biol. CB* **2001**, *11* (22), 1739–1748. [https://doi.org/10.1016/s0960-9822\(01\)00536-x](https://doi.org/10.1016/s0960-9822(01)00536-x).
 - (48) Aw Yong, K. M.; Ulintz, P. J.; Caceres, S.; Cheng, X.; Bao, L.; Wu, Z.; Jiagge, E. M.; Merajver, S. D. Heterogeneity at the Invasion Front of Triple Negative Breast Cancer Cells. *Sci. Rep.* **2020**, *10* (1), 5781. <https://doi.org/10.1038/s41598-020-62516-8>.
 - (49) Liu, M.; Liu, Y.; Deng, L.; Wang, D.; He, X.; Zhou, L.; Wicha, M. S.; Bai, F.; Liu, S. Transcriptional Profiles of Different States of Cancer Stem Cells in Triple-Negative Breast Cancer. *Mol. Cancer* **2018**, *17* (1), 65. <https://doi.org/10.1186/s12943-018-0809-x>.

- (50) Abdullahi, A.; Stanojcic, M.; Parousis, A.; Patsouris, D.; Jeschke, M. G. Modeling Acute ER Stress in Vivo and in Vitro. *Shock Augusta Ga* **2017**, *47* (4), 506–513. <https://doi.org/10.1097/SHK.0000000000000759>.
- (51) Hetz, C.; Zhang, K.; Kaufman, R. J. Mechanisms, Regulation and Functions of the Unfolded Protein Response. *Nat. Rev. Mol. Cell Biol.* **2020**, *21* (8), 421–438. <https://doi.org/10.1038/s41580-020-0250-z>.
- (52) Wu, L.; Huang, X.; Kuang, Y.; Xing, Z.; Deng, X.; Luo, Z. Thapsigargin Induces Apoptosis in Adrenocortical Carcinoma by Activating Endoplasmic Reticulum Stress and the JNK Signaling Pathway: An in Vitro and in Vivo Study. *Drug Des. Devel. Ther.* **2019**, *13*, 2787–2798. <https://doi.org/10.2147/DDDT.S209947>.
- (53) Heckman, K. L.; Pease, L. R. Gene Splicing and Mutagenesis by PCR-Driven Overlap Extension. *Nat. Protoc.* **2007**, *2* (4), 924–932. <https://doi.org/10.1038/nprot.2007.132>.
- (54) Block, K. M.; Wang, H.; Szabó, L. Z.; Polaske, N. W.; Henchey, L. K.; Dubey, R.; Kushal, S.; László, C. F.; Makhoul, J.; Song, Z.; Meuillet, E. J.; Olenyuk, B. Z. Direct Inhibition of Hypoxia-Inducible Transcription Factor Complex with Designed Dimeric Epidithiodiketopiperazine. *J. Am. Chem. Soc.* **2009**, *131* (50), 18078–18088. <https://doi.org/10.1021/ja807601b>.

CHAPTER 3

High-Throughput Method to Identify Selective Modulators of Activator•Coactivator Interactions

3.1 Abstract

The protein-protein interactions (PPIs) that regulate transcription are often dysregulated in disease and therefore make attractive pharmacological targets. However, a major challenge facing the discovery of small-molecule modulators of transcriptional coactivator•transcription factor complexes has been selectivity. High-throughput screening (HTS) methods have most often relied upon binding inhibition against the activator•coactivator complex of interest alone. This approach is prone to discovery of compounds that act as nonspecific mimetics of the transcriptional activator and thus interact with multiple coactivator proteins, limiting their utility as biological probes. An iterative approach involving several rounds of selectivity screening against multiple PPIs has led to more potent and selective inhibitors. Still, this approach considers selectivity after primary screening and initial triage has been conducted, risking exclusion of slightly less potent but selective scaffolds. To address this, a fluorescence polarization-based assay that simultaneously monitors multiple activator•coactivator interactions has been developed by simultaneously evaluating different protein complexes using peptide tracers tagged with spectroscopically distinct fluorophores. This method enables assessment of both selectivity and potency of candidate inhibitors in a single screen. To demonstrate the utility of the method, a duplex assay containing the coactivators Med25 AcID and CBP KIX has been optimized and a pilot screen has been conducted. The pilot screen was able to categorize active compounds as Med25 AcID-selective, CBP KIX-selective, or dual inhibitors. Representative compounds from each subset were further evaluated in secondary screening. We observed that for compounds retaining activity in secondary screening, selectivity profiles from the primary screen are replicated. Ultimately, this approach led to the identification of a highly selective compound that displays *in vitro* inhibition of the interaction between Med25 AcID and the transcriptional activator ETV5.

3.2 Introduction

As highlighted in the previous chapter, the interactions between activators and coactivators are fundamental in the process of activated gene expression.¹⁻⁴ Further elucidation of this network of PPIs is crucial for insights into the compositionally and conformationally plastic complexes that govern transcriptional regulation.⁵ Additionally, dysregulation of the protein-protein interactions involved in transcription is characteristic of many human diseases including cancer.⁴ Thus, transcriptionally relevant PPI's are potentially attractive therapeutic targets and there is great interest in expanding the suite of chemical tools used to study these systems.⁶⁻¹⁰

A significant hurdle for studying these interactions has been the task of discovering chemical modulators due to the dynamic nature of activator and coactivator proteins. Coactivators use conformationally plastic activator binding domains (ABDs) to interact with a variety of activators.¹¹⁻¹³ For example, the KIX domain of the coactivator CBP regulates transcription via interaction with more than 15 distinct transcription factors using only two binding interfaces.¹⁴ Further, ABDs adopt distinct complex conformations upon interaction with different binding partners.^{11,15,16} This has been highlighted in work analyzing the unique structural changes that occur when KIX interacts with different TADs or small molecules ligands.^{14,17-21} Additionally, transcriptional activators interact with many different proteins using only short recognition motifs within their transactivation domains (TADs).^{22,23} Most activators are intrinsically disordered and only adopt secondary structural elements upon binding to a coactivator while others retain conformational heterogeneity even after binding, forming “fuzzy” complexes (Figure 3.1).^{13,24} Considering the structural and functional complexity of the proteins involved in these PPIs, the challenge then becomes discovering molecules that selectively target one of these interactions over others.

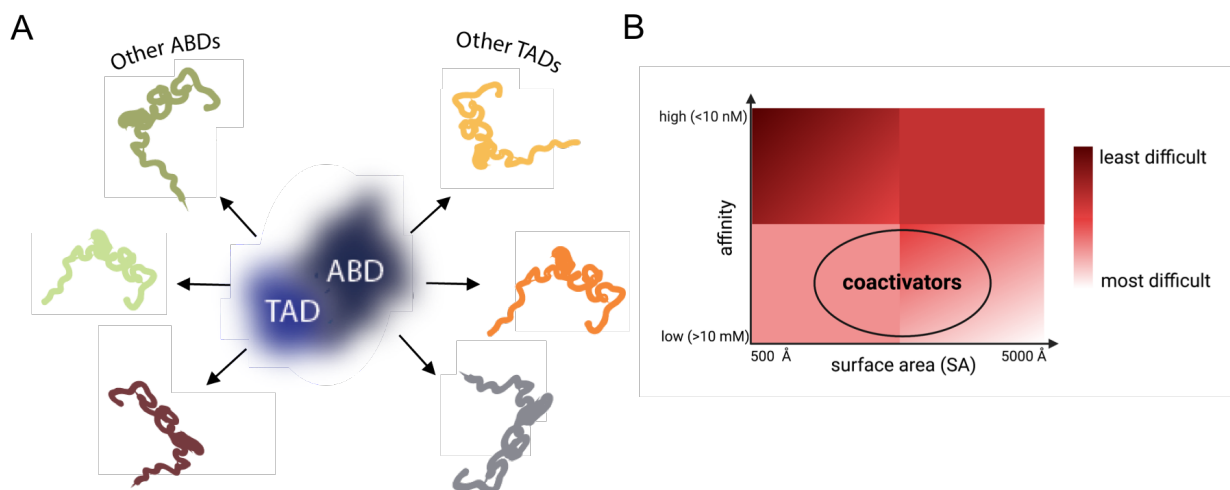


Figure 3.1. Coactivator•activator interactions are classically a challenge to target selectively. A) ABDs and TADs form transient, conformationally heterogeneous, complexes (often referred to as “fuzzy”). Both members of this complex also interact with a separate network of proteins, including other TADs or ABDs, to regulate alternate functions. B) Coactivator activator interactions are classified among the most challenging to target with inhibitors due to the attributes of the PPI interface - lower affinity occurring over a large surface area.

The previous chapter demonstrated that an iterative *in vitro* screening platform can be successfully applied to determine potent inhibitors of difficult to target ABDs. In addition, two natural product inhibitors of CBP/KIX interactions, lobaric acid and sekikaic acid, were discovered using this approach.²⁵ However, a significant limitation of the iterative workflow is performing separate screens to determine potency and selectivity. Depending on library size and screening capabilities, it is possible that promising selective scaffolds of moderate potency are filtered out early on in the screening process and never pursued. In addition, many high potency hits identified in a HTS for ABD inhibitors mimic the structure of the bound TAD and thus compete for binding to the ABD. Because TAD peptides often interact with multiple binding partners, these amphipathic mimetics are highly nonspecific for the profiled interaction and can affect many other PPIs.²⁶ For example, this can be seen with the isoxazolidine mimetics, utilized due to their chemical similarity to natural transcriptional activators.²⁷ This class of molecules was found to interact with more than 10 different coactivators,^{28,29} underscoring the unique challenge of selectivity when screening for coactivator•activator modulators. Thus, the current HTS strategy has room for improvement, as the maximum amount of information is not currently obtained during initial triage of a large compound library.

The work outlined in this chapter focuses on the development and implementation of a HTS platform for PPIs that improves the ability to detect selective hits from the primary screening level. This is particularly useful for dynamic coactivator•transcription factor complexes because of the outlined challenges with regards to inhibitor selectivity. A fluorescence polarization (FP) based assay that simultaneously monitors multiple activator•coactivator interactions is presented. With this method, each protein complex can be individually evaluated using peptide tracers labelled with spectroscopically distinct fluorophores. This method enables characterization of both selectivity and potency of candidate inhibitors in a single screen.

3.2 Results & Discussion

Development of duplex fluorescence polarization assay with Med25 AcID and CBP KIX

HTS affords excellent opportunities to identify chemical scaffolds that can target unique conformations of activator-coactivator interactions. A common primary screening methodology selected for HTS protein-protein interactions is fluorescence polarization (FP). FP is an attractive HTS technique due to its low cost, rapid readout, and amenability to a high throughput format.^{30,31} With an activator-coactivator system, FP reports on the binding of a synthetic activator TAD peptide to the ABD motif of a coactivator. This approach was used in Chapter 2 to identify norstictic acid as a potent and selective Med25 inhibitor. As some TAD•ABD interactions can be highly specific, we hypothesized it would be possible to multiplex an FP based assay, incorporating multiple TAD•ABD pairs, with the fluorescent tracers labeled with spectroscopically distinct fluorophores. This would allow for information on specificity and potency of compounds to be determined in a single primary screen. In addition, multiple PPIs can be screened against a library of compounds simultaneously.

The KIX domain of the coactivator CBP and the AcID domain of Med25 were chosen as candidate ABDs to utilize in the development of a duplex FP assay due to their structural and biological relevance (Figure 3.2). KIX is a prototypical three-helix bundle, a motif associated with many ABDs.^{14,32,33} AcID, as described in the previous chapter, is structurally unique, containing a central seven-stranded beta barrel surrounded by three alpha helices.³⁴ Additionally, both AcID and KIX are implicated in interactions that influence transcription in diseases such as diabetes and cancer.^{35–39}


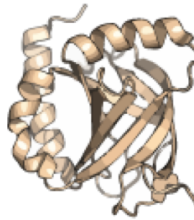
 <p>PDB: 4I9O</p>	<p>CBP KIX</p> <ul style="list-style-type: none"> • Prototypical 3-helix bundle motif • Interaction with MLL is associated with leukemia 	<p>Med25 AcID</p> <ul style="list-style-type: none"> • Unique structure (central beta barrel) • Interaction with ETV5 linked to metastasis 	 <p>PDB: 2XNF</p>
----------------------------------------------------------------------------------------------------	-------------------------------------------------------------------------------------------------------------------------------------------------------------------------	---------------------------------------------------------------------------------------------------------------------------------------------------------------------------	------------------------------------------------------------------------------------------------------

Figure 3.2. Overview of first generation duplex screen proteins Med25 AcID and CBP KIX. Structures of the two activator binding domains are shown with the corresponding PDB codes. Qualities of these proteins that made them attractive candidates for the first-generation duplex assay are listed.³⁹⁻⁴²

The format of the duplex screen expands upon the FP technique we and others have used for screening transcriptional PPIs. A typical assay profiles the direct interaction of a single ABD and a fluorescently-labeled peptide derived from the transcriptional activation domain (TAD) of an activator as the tracer. The duplex FP format incorporates two different ABD•TAD pairs into one screen, each labeled with a distinct fluorophore. Based upon their distinct spectral properties, fluorescein (FITC) and Texas Red (TR) were selected for tracer labelling.⁴³ For development of an AcID•ETV5/KIX•MLL duplex assay, a FITC-labelled peptide of the ETV5 TAD (residues 38-68) and a Texas-Red labelled peptide of the MLL TAD (residues 840-858) were synthesized as the tracers (Figure 3.3).

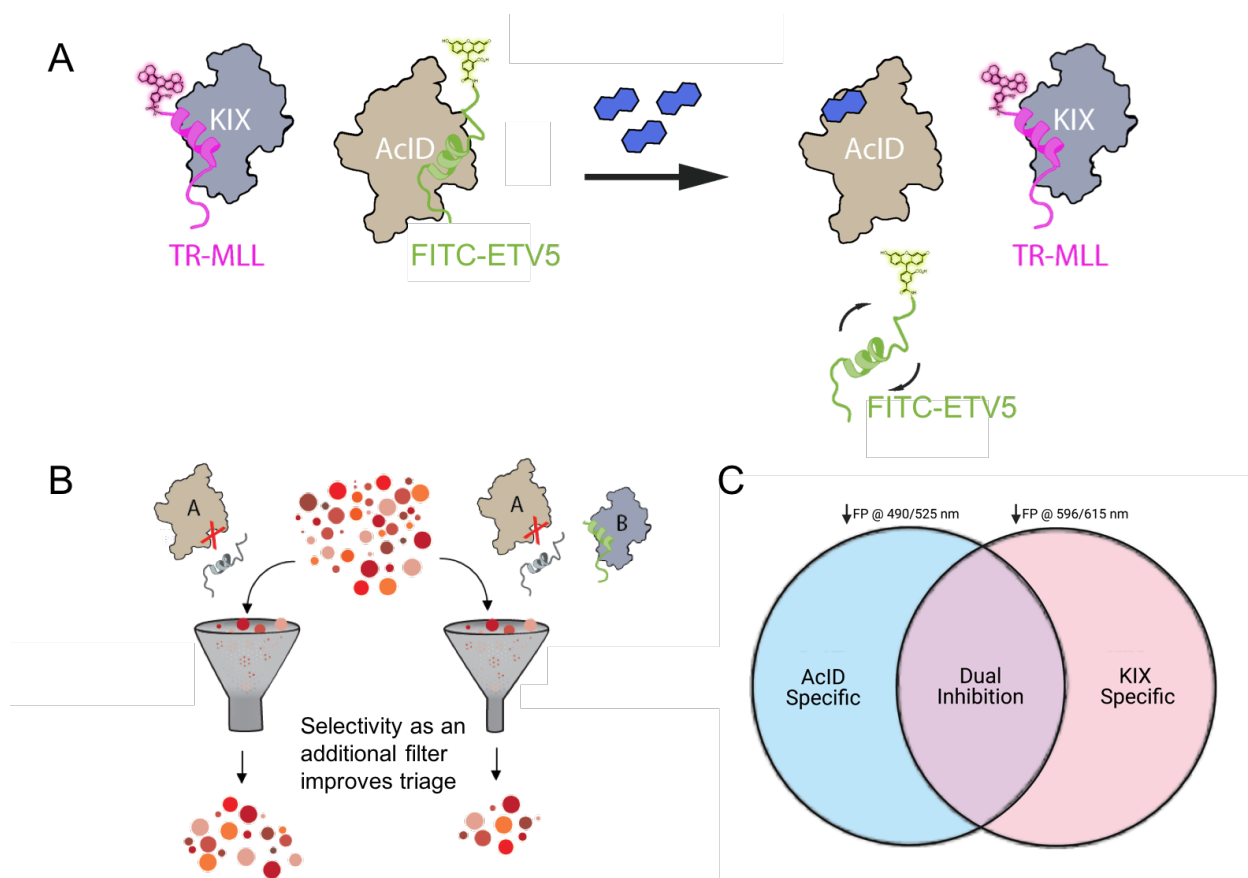


Figure 3.3. Description of first generation duplex screen fluorescence polarization assay. A) Overview of assay concept with KIX•TR-MLL and AcID•FITC-ETV5. In this example, low fluorescence polarization in the Fluorescein channel and high fluorescence polarization in the Texas-Red channel would indicate an AcID selective inhibitor. B) Advantages to using the duplex screening approach. More data at the primary screening level allows for more thorough refinement of hits. C) This method allows for rapid classification of ligand selectivity.

For the proposed screen to accurately report on activity of a small molecule, it must be demonstrated that any observed decrease in polarization is due to direct inhibition of the activator peptide-coactivator protein interaction. Direct binding experiments of FITC-ETV5 and TR-MLL tracers were performed in isolation, with KIX or AcID, and in duplex, with both KIX and AcID. As shown in Figure 3.4, the dissociation constants of the tracers were unaffected in the duplex format. Additionally, each tracer shows at least 10-fold selectivity for the cognate ABD. The Z factors for AcID•FITC-ETV5 and KIX•TR-MLL in the duplex screen were determined to be 0.9 and 0.65, respectively, indicating an excellent assay for HTS.⁴⁴

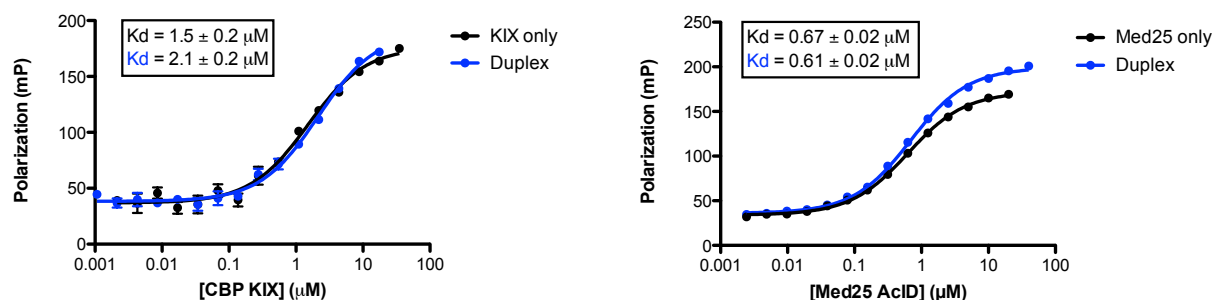
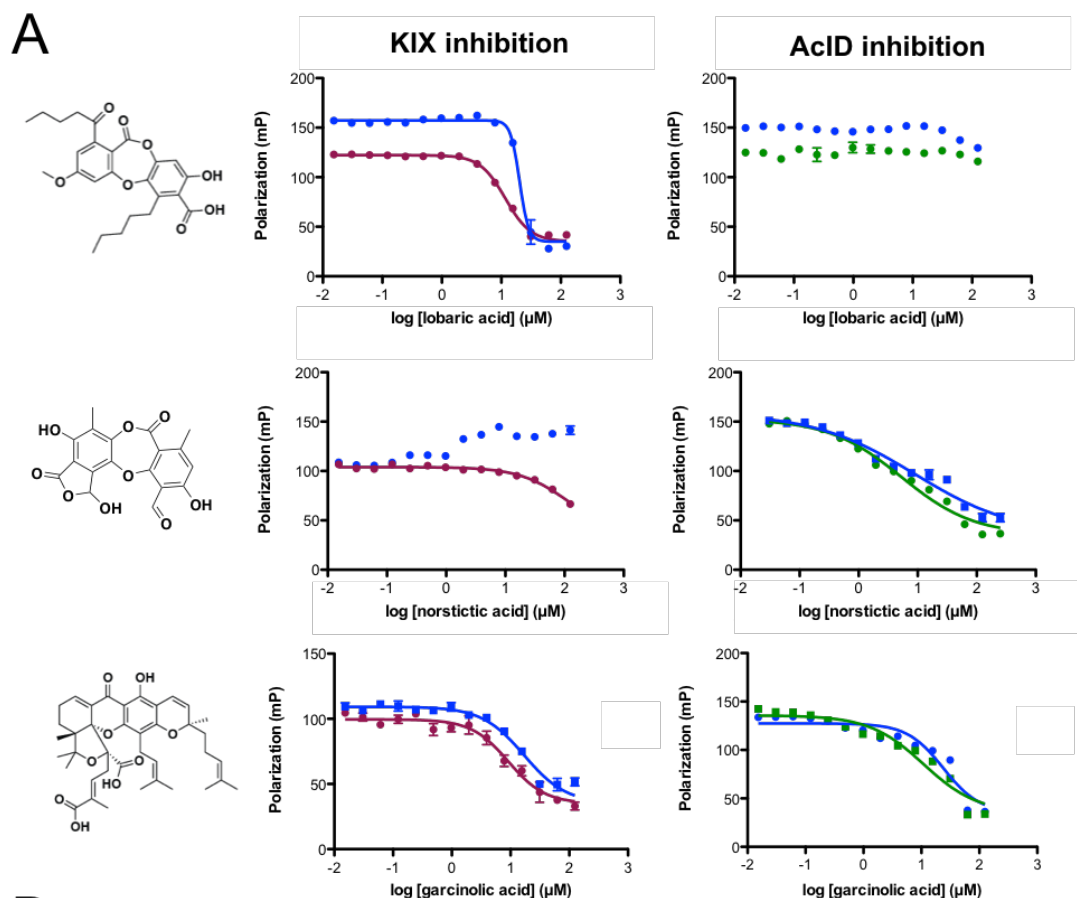


Figure 3.4. Direct binding experiments under duplex conditions accurately represent interaction of the cognate ABD•TAD pairs. Left: TR-MLL. Right: FITC-ETV5. Fluorescence polarization data for each concentration point was obtained in technical triplicate, with the average plotted and the error shown as the standard deviation of the mean.

To further assess the dynamic range of the assay, as well as to probe its ability to distinguish the binding profiles of small molecule ligands, inhibitors previously identified in the Mapp laboratory were examined with this method (Figure 3.5). Lobaric acid is a CBP/p300 KIX-selective inhibitor reported in 2012.²⁵ Norstictic acid is a Med25 AcID-selective inhibitor (presented in Ch. 2) and garcinolic acid is an inhibitor with good activity for both AcID and KIX (unpublished work from Dr. Meg Breen). Competition experiments were performed by adding a mixture of the pre-formed AcID•FITC-ETV5 and KIX•TR-MLL complexes at 50% bound tracer to serially diluted small molecule solution aliquots in 384-well plates and incubating at RT for 30 min. Fluorescence polarization measurements were obtained using both the fluorescein and Texas Red channels on a Pherastar plate reader. The inhibition curves obtained accurately reflect the known activity of the molecules for their specific targets. These measurements were performed in triplicate on different days and produced an average percent coefficient of variability (%CV) of 8.6%, consistent with a reproducible assay.⁴⁵



B

	CBP KIX	Duplex (CBP KIX)	Med25 AcID	Duplex (Med25 AcID)
lobaric acid	11.7 ± 0.4	20.2 ± 0.7	N/A	N/A
norstictic acid	114 ± 8	N/A	5.3 ± 0.7	9 ± 2
garcinolic acid	9 ± 1	17 ± 2	11 ± 1	24 ± 3

Figure 3.5. Duplex assay can accurately characterize the selectivity of known inhibitors lobaric acid, norstictic acid, and garcinolic acid. A) Competition fluorescence polarization was conducted for each inhibitor with Med25 AcID, CBP KIX, or under duplex screen conditions. Plots of KIX inhibition are shown in the first column (red – in isolation, blue – under duplex conditions). Plots of AcID inhibition are shown on the second column (green – in isolation, blue – under duplex conditions). Assays were conducted in technical triplicate, and the average polarization value was plotted with the error shown as the standard deviation of the mean. B) IC_{50} values (shown in μM) calculated from the inhibition curves shown in A.

Pilot screen of small library using duplex assay

With the duplex assay validated against known inhibitors of both the KIX•MLL and AcID•ETV5 interactions, a pilot screen of 2,846 compounds (NIH Clinical Collection and MicroSource 2400) was completed with the Center for Chemical Genomics to test the suitability

of the duplex format for high-throughput screening. A mixture of AcID (0.6 μ M final concentration) and KIX (2 μ M final concentration) was plated in 384-well format and treated with compound (20 μ M final concentration). The fluorophore-labeled tracers were added (20 nM final concentration, leading to 50% bound complex) and the plates incubated for 30 minutes at room temperature before fluorescence polarization in the fluorescein and Texas Red channels were measured using a PheraStar plate reader. Percent inhibition for each compound was determined using wells treated with DMSO as a negative control and wells containing fluorophore-tagged peptide alone as a positive control. Each plate contained 32 negative control wells, containing the protein-tracer complex mixture dosed with DMSO, and 32 positive control wells, containing tracer only. The assay demonstrated excellent performance with Z factors of 0.88 and 0.64 for the AcID•FITC-ETV5 and KIX•TR-MLL measurements, respectively.⁴⁴ The initial activity threshold was set at a polarization value greater than 3 S.D. below the average signal of that of the negative controls, leading to the identification of 26 molecules as hits against the AcID•ETV5 interaction (0.91% hit-rate) and 94 compounds as hits against the KIX•MLL interaction (3.3% hit-rate). A comparison of the individual data sets revealed that 6 compounds inhibited both interactions above the 3 S.D. cutoff, leading to their preliminary classification as dual inhibitors.

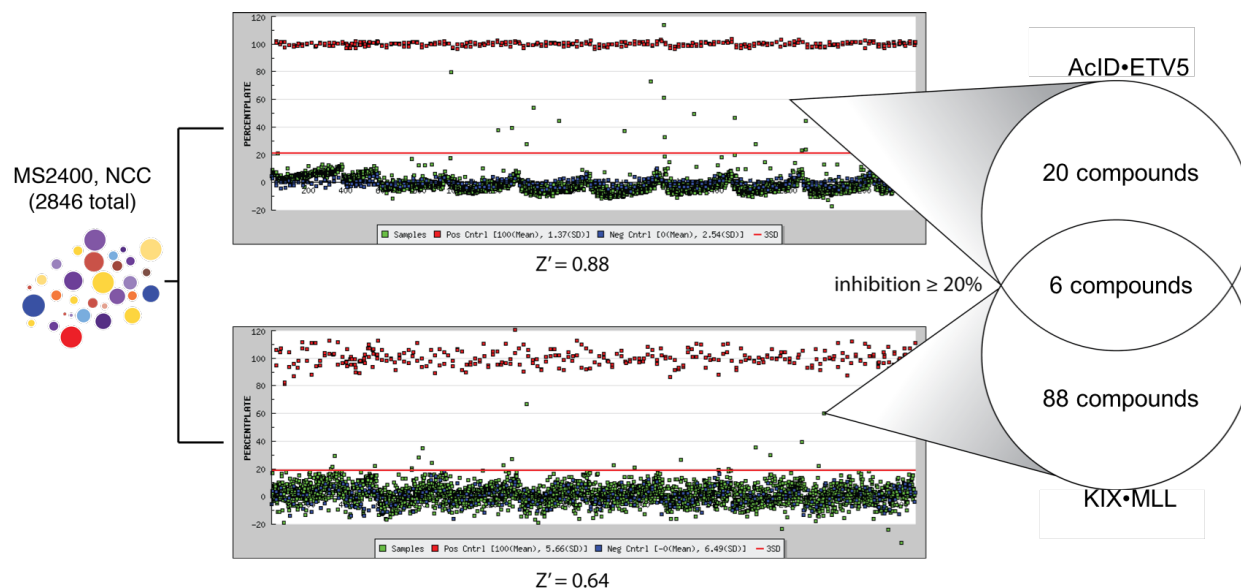


Figure 3.6. Duplex assay can accurately characterize the selectivity of known inhibitors lobaric acid, norstictic acid, and garcinolic acid. Campaign view of primary screen conducted with duplex assay; percent inhibition for each data point. Red dots represent positive controls (100 % inhibition, tracer only), blue dots represent negative controls (0% inhibition, tracer + protein), and green dots represent tested compounds. The red line represents the 3SD, 20% inhibition, threshold used to determine hits. Data for Med25 AcID is shown on the top, CBP KIX is shown on the bottom. 94 compounds were identified as hits and immediately sorted for selectivity based on this data.

Confirmatory screening using dose-response (8 different concentrations, from 120 μM to 3.3 μM) in duplicate was conducted with the 114 hit compounds. Additionally, at this step known fluorophores such as calcein and fluorescein that appeared as hits were excluded. As an additional measure to check for artifacts, the order of addition was switched for the assay components (tracers first, protein second). The top three active compounds from each category were selected for further profiling via purchase of fresh material (Figure 3.7).

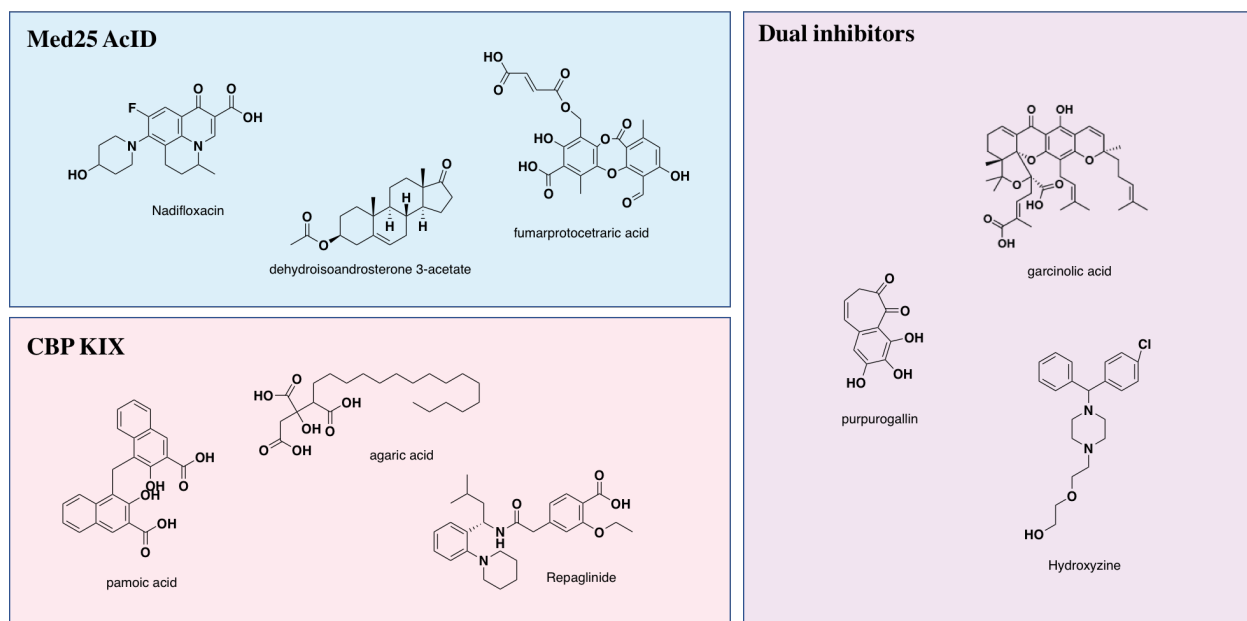


Figure 3.7. Lead compounds selected for follow up screening. Structures, names, and selectivity profiles are shown. AcID selective molecules = blue panel, KIX selective molecules = pink panel, inhibitors of both proteins = purple panel.

Upon purchase, these compounds were tested in competition FP assays with the individual duplex screen ABD•TAD pairs to obtain full inhibition curves and IC_{50} values (Figure 3.8). At this stage, the fluorophore pairs were swapped to test if the fluorophore was not having an effect on the inhibition data as FITC is negatively charged whereas Texas Red is neutral. Previous work has shown that FITC does increase the binding affinity of the TAD peptide for the ABD, and fluorescein shows up as a Med25 binder in high throughput screens. Thus, it is possible that would add some bias to the screening method.

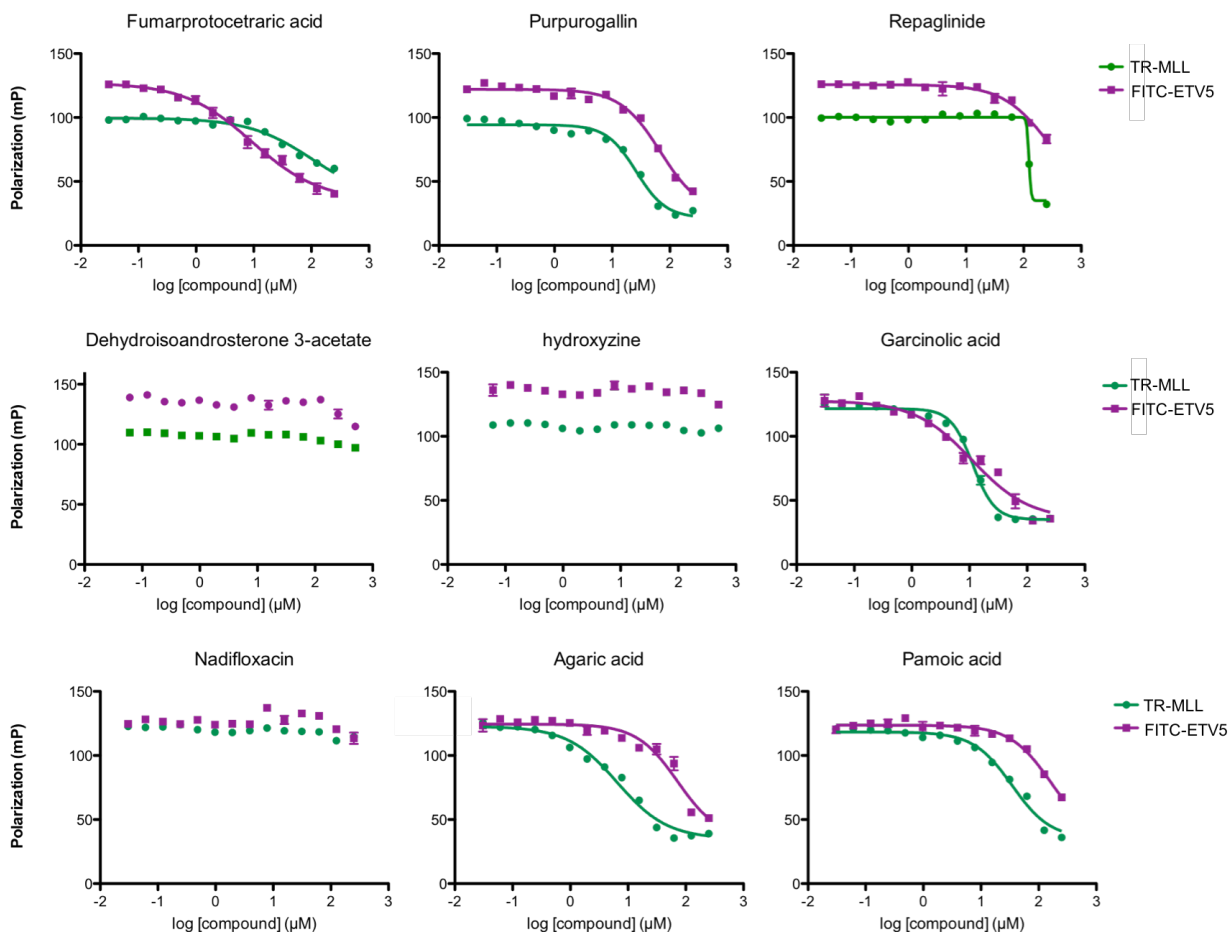
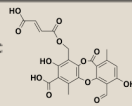
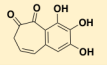
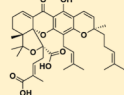
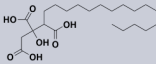
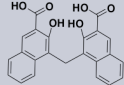


Figure 3.8. Inhibition curves of representative duplex screen hits. Competition FP was conducted for each compound using AcID•TR-ETV5 and KIX•FITC-MLL in isolation. The average polarization from assays conducted in technical triplicate is plotted with error bars representing the standard deviation of the mean. Calculated IC_{50} values for compounds showing activity are shown in Table 3.1.

Only five compounds still displayed activity: fumarprotocetraric acid, garcinolic acid, purpurogallin, pamoic acid, and agaric acid (Table 3.1). The most potent inhibitor of CBP KIX was garcinolic acid, which was not surprising because it had already been identified in a KIX targeting screen of the same library conducted by a previous lab member. However, this screen also classified garcinolic acid as a dual inhibitor, which was confirmed in secondary screening. Agaric acid, with a similar IC_{50} value to garcinolic acid, was determined to be the most potent *and* selective inhibitor of CBP KIX, showing greater than 10-fold selectivity for KIX over AcID. Fumarprotocetraric acid (FPA) stood out as the best Med25 AcID inhibitor, with a low μM IC_{50} as well as greater than 10-fold selectivity for AcID over KIX. As shown in Table 1, the five

compounds retaining activity maintained the same selectivity profiles from the primary screen, illustrating the reliability of the selectivity assessment obtained from the duplex assay.

Table 3.1. Summary of KIX and AcID inhibition by duplex screen lead molecules

Compound	Duplex selectivity	IC50 KIX (μM)	IC50 AcID (μM)	Selectivity for target
Fumarprotocetraric acid 	AcID	105 ± 8	8.3 ± 0.9	>10 fold
Purpurogallin 	Dual	26 ± 2	60 ± 5	~2 fold for CBP KIX
Garcinolic acid 	Dual	5.4 ± 0.2	6 ± 1	Dual
Agaric acid 	KIX	6.5 ± 0.5	72 ± 6	>10 fold
Pamoic acid 	KIX	35 ± 2	160 ± 10	>4 fold

Additional selectivity profiling of lead compounds from pilot screen

To further analyze selectivity, the five lead compounds were tested against additional coactivators, p300 CH1 and CBP IBiD (Figure 3.9). IBiD is known to be intrinsically disordered and unstructured until interaction with an activator protein occurs, when it becomes structured into a prototypical three helix bundle.^{46–48} CH1, also known as TAZ1, is a three-helix bundle, however it is unique in the fact that it requires specific coordination of three zinc ions for proper folding and function.⁴⁹ The sequences of TADs that interact with both of these coactivators are published, and thus these coactivators are readily adaptable to FP assay.^{46,50–53} The CH1•HIF1α interaction plays a key role in transcribing proteins that promote cancer cell survival under the hypoxic condition of tumor environment, thus targeting this interaction could be an effective anti-cancer strategy.^{50,54,55} IBiD is the most conformationally dynamic of the CBP/p300 ABDs, only adopting its three-helix bundle structure upon ligand binding.^{46–48} While IBiD has been implicated in cancer progression, interacting with p53 as well as the activator for thyroid hormone and retinoid receptors (ACTR) domain of the p160 nuclear-receptor co-activator, no small molecule inhibitors have been reported.^{47,53,56} Thus these are interesting targets to test with regards to selectivity for multiple reasons, first due to both their structural similarities as three helix bundles, and differences, one as an intrinsically disordered protein and CH1 as a zinc coordinating protein. Additionally, since they are both domains within the master coactivator CBP/p300, like KIX, they would be likely present and able to compete for small molecule binding in a cellular

environment.^{56,57}

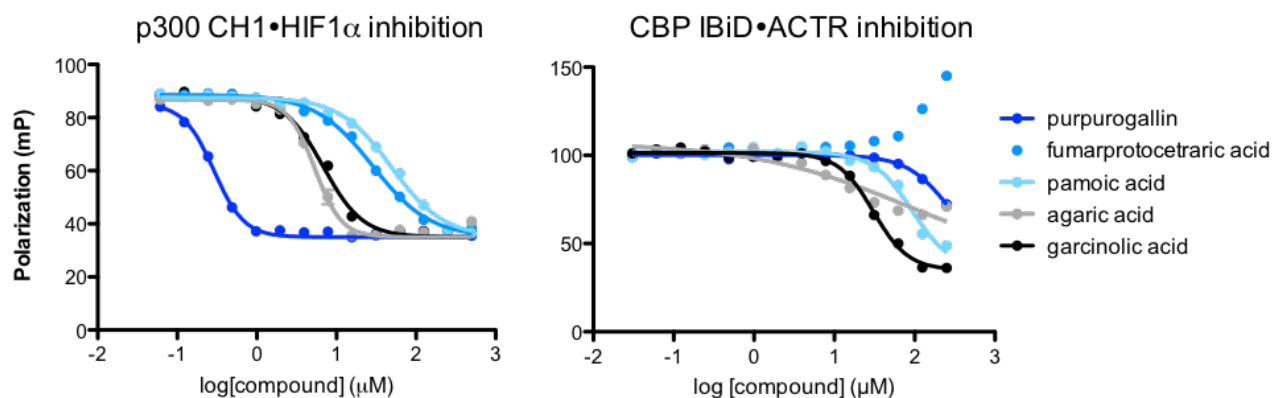


Figure 3.9. Competition FP looking at inhibition of CH1 and IBD by duplex screen lead compounds.

Assays were conducted in technical triplicate and the average for each data point is shown with error bars representing the standard deviation of the mean.

Most of the compounds were able to inhibit the CH1•HIF1 α interaction. Notably, purpurogallin had a much lower IC₅₀ for this interaction than the other compounds tested, at $0.29 \pm 0.02 \mu\text{M}$. Purpurogallin also inhibited this interaction much better than any other interaction, with its next best IC₅₀ at $26 \pm 2 \mu\text{M}$. Purpurogallin is a known metal chelator, and part of the polyphenol family of natural products, which have been well characterized for their antioxidant properties.^{58–60} Therefore, it is likely that this compound inhibits CH1 not by direct PPI inhibition, but by removal and sequestering of the zinc ions necessary for proper folding and TAD binding. This inhibitory mechanism has been thoroughly characterized in the context of other CH1•HIF1 α inhibitors.^{61–63} Garcinolic acid demonstrated single digit μM IC₅₀ values against all interactions tested, further confirming it as a nonspecific inhibitor of multiple coactivators. Pamoic acid, while a more modest inhibitor of KIX at $35 \pm 2 \mu\text{M}$, had some selectivity for its target (>2 fold). Out of all compounds tested, FPA stood out as the most useful hit, with a low μM IC₅₀ value for Med25•ETV5 and also the greatest selectivity for its target (Figure 3.10).

	IC ₅₀ (μM)	Approx. Fold selectivity
Med25 AcID•ETV5	4.9 ± 0.5	-
Med25 AcID•ATF6α	11.4 ± 0.9	2
CBP KIX•MLL	104 ± 6	20
CBP KIX•myb	> 250	> 250
p300 CH1•HIF1α	28.9 ± 0.9	5.5
CBP IBiD•ACTR	> 250	> 250

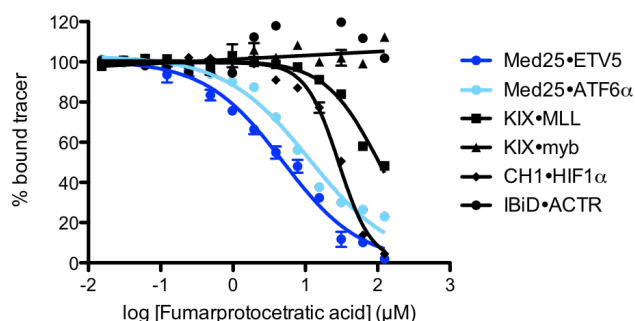


Figure 3.10. Selectivity profiling of fumarprotocetraric acid. Assays were conducted in technical triplicate and the average for each data point is shown with error bars representing the standard deviation of the mean. Left: calculated IC₅₀ values for each interaction and fold selectivity based on AcID•ETV5 inhibition. Right: Inhibition curves obtained from competition FP experiments and used to calculate IC₅₀ values

FPA is a member of the depsidone family of natural products and like NA, discussed in the previous chapter, contains an orthophenolic aldehyde, suggesting the ability to covalently modify protein targets. However, NA inhibits AcID•ATF6α better than AcID•ETV5 while FPA showed modest facial selectivity at the H1 face, inhibiting AcID•ETV5 2-fold better than AcID•ATF6α. Thus, while these molecules have high structural similarity, they may bind to unique sites on AcID or, alternatively, bind at similar sites yet have distinct effects on conformational tuning. The main difference between FPA and NA is the presence of a fumarate functionality on FPA (Figure 3.11). Further interrogation of the binding mechanism of FPA will be necessary to fully understand inhibition. However, this preliminary data suggests that fine tuning of coactivator targeting inhibitors could provide even selectivity at the individual PPI level for those proteins with multiple binding interfaces.

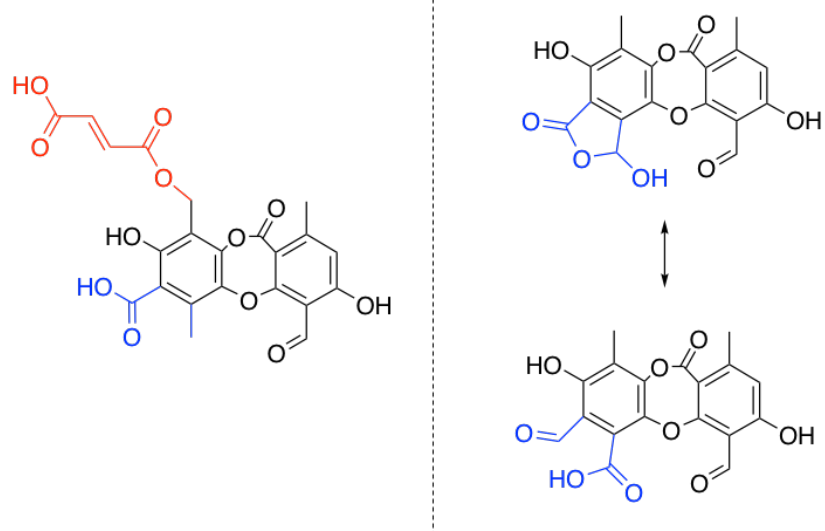


Figure 3.11. Comparison of FPA(left) and NA(right). The non-orthophenolic aldehyde containing benzene ring of FPA is differently substituted than the corresponding ring of NA. Most notable is the presence of the fumarate moiety (red), although there are additional subtle changes (blue). Potentially significant, NA contains two aldehydes, while FPA only contains one.

Optimization of duplex assay for IBiD

With the successes of this approach using AcID•ETV5 and KIX•MLL, the duplex approach was tested with another ABD of interest, IBiD. Direct binding experiments of TR-ETV5 and FITC-ACTR tracers were performed in isolation, with IBiD or AcID, and in duplex, with both IBiD and AcID. As shown in Figure 3.12, the dissociation constants of the tracers were unaffected in the duplex format. Additionally, each tracer shows high selectivity for the cognate ABD (>10 fold).

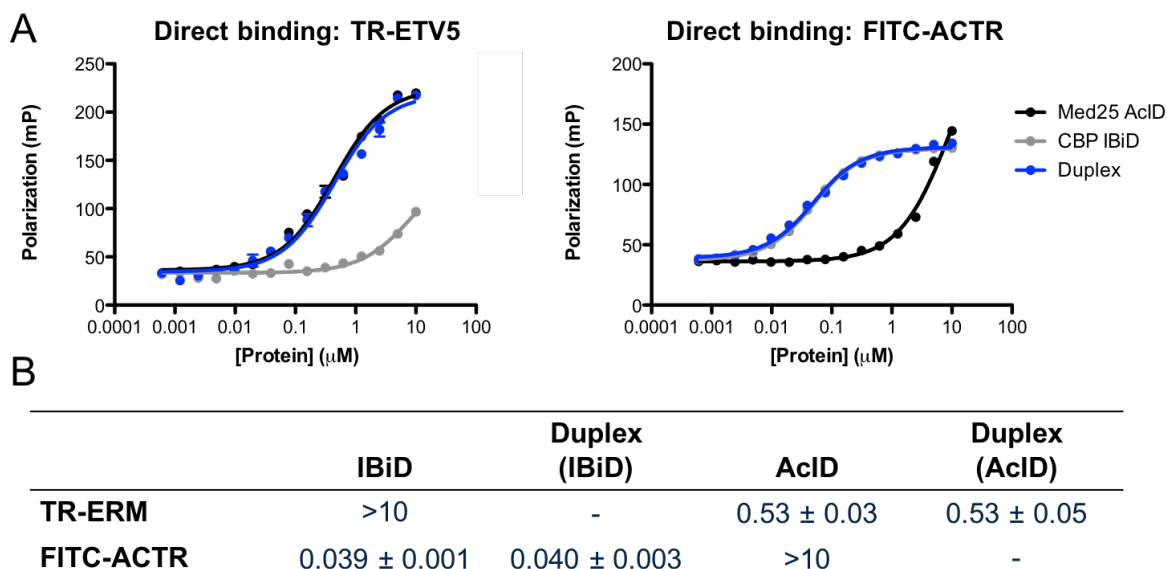


Figure 3.12. Development of duplex assay utilizing Med25 AcID and CBP IBiD. A) Direct binding fluorescence polarization data was obtained in technical triplicate, with the average plotted and error bars representing the standard deviation of the mean. B) Calculated binding constants (K_d), shown in μM , for each peptide tracer under the various conditions tested. The K_d for IBiD•TR-ETV5 and AcID•FITC-ACTR could not be determined because it is greater than $\frac{1}{2}$ the maximum protein concentration used in this assay.

Conditions for competition experiments were optimized for signal-to-noise, with best results obtained with AcID•TR-ETV5 at 50% bound and IBiD•FITC-ACTR at 80% bound. With these conditions, the Z' factors are 0.86 for IBiD and 0.68 for AcID. Thus, the duplex screen platform appears to be a generalizable strategy for coactivator protein inhibitor discovery.

3.4 Conclusions

Taken together, the results shown here demonstrate that the duplex FP format is a viable HTS platform for activator•coactivator interactions. Further, considering selectivity throughout the screening process streamlines discovery of selective small molecule modulators. Demonstrated here, using this approach with Med25 AcID and CBP KIX, screening of a small pilot library resulted in identification of a novel Med25 inhibitor, fumarprotocetraric acid, that demonstrated enhanced selectivity for AcID•ETV5 over a panel of other related coactivators and interactions. Additionally, while this approach was validated with Med25 AcID and CBP KIX, preliminary data suggests that this would be applicable to any set of coactivators, if there is demonstrated specificity of the selected TAD tracers.

3.5 Materials & Methods

Protein expression and purification

Proteins were expressed and purified as described in Chapter 2.

Synthesis of transcriptional activation domain peptides

The synthesis of peptides FITC-MLL(840-858), FITC-ETV5(38-68), FITC- ATF6 α (40-66), FITC-myb(291-316), FITC-HIF1 α (786-826), FITC-ACTR(1041-1088), and Ac-IBiD(2063-2111) were described in the previous chapter.

To generate Texas-Red labelled peptide tracers of ETV5(38-68) and MLL(840-858), first the peptides sequences were synthesized following standard Fmoc solid-phase synthesis methods on a Liberty Blue Microwave Synthesizer (CEM), with the addition of an N-terminal β -alanine (as described in depth in Chapter 2). Because Texas-Red sulfonyl chloride was used and hydrolysis

of the sulfonyl-chloride group can easily occur leading to poor yields, only dry reagents were used. Briefly, peptides still on resin were washed 5x with dry dichloromethane (DCM). For coupling, a slight excess (1.2 equivalents) of Texas-Red was dissolved in 1.9 mL dry DCM with 100 μ M diisopropylethylamine and added to a foiled reaction vessel containing the peptide on resin. Coupling occurred for 18 hours on a rotator at RT. Peptides were deprotected and cleaved from the resin for 3 hours in 95% trifluoroacetic acid, 2.5% ethanedithiol, 2.5% water. Crude peptides were filtered to remove resin, dried under nitrogen stream, and precipitated from cold ether. Peptide suspensions were transferred to a 15 mL falcon tube, centrifuged at 4000 g for 5 minutes at 4°C, and ether decanted. Crude peptides were resuspended in 40% acetonitrile, purified via HPLC on an Agilent 1260 HPLC using a semi-prep C18 column (Phenomenex). A gradient of 15-65% acetonitrile over 50 minutes was used, collecting fractions based on absorbance at 590 nm. Pure fractions were collected and lyophilized to afford pure peptides. Final purity was determined via analytical HPLC and identity was confirmed using mass spectrometry. Analytical spectra were obtained using an analytical C18 column (Phenomenex) on an Agilent 1260 HPLC. Mass spectra were obtained using an Agilent 6545 LC/Q-TOF. Analytical spectra of peptides can be found in Appendix A.

Fluorescence polarization assay protocol – direct binding and competition.

Fluorescence polarization assays were conducted as described in Chapter 2.

Duplex HTS protocol

The pilot screen of 2,846 compounds was completed in the Center for Chemical Genomics. AcID (0.6 μ M final concentration) and KIX (2 μ M final concentration) were plated in 384-well format and treated with 20 μ M (final concentration) of each library compound. The fluorophore-labeled tracers were added (20 nM final concentration; leading to 50% bound complex) and the plates incubated for 30 min at room temperature before fluorescence polarization measurements in the fluorescein and Texas Red channels were made using a PheraStar plate reader. Percent inhibition for each compound was determined using wells treated with DMSO as a negative control and wells containing fluorophore-tagged peptide alone as a positive control. Each plate contained 32 negative control wells, containing the protein-tracer complex mixture dosed with DMSO, and 32 positive control wells, containing tracer only. The assay demonstrated excellent performance

with Z' factors of 0.88 and 0.64 for the AcID•FITC-ETV5 and KIX•TR-MLL measurements, respectively. Using an initial activity threshold a polarization value ≥ 3 S.D. below the average signal of that of the negative controls, 26 molecules were identified as hits against the AcID•ETV5 interaction (0.91% hit-rate) and 94 compounds were identified as hits against the KIX•MLL interaction (3.3% hit-rate). A comparison of the individual data sets revealed that 6 compounds inhibited both interactions at the 3 S.D. cutoff, leading to their preliminary classification as dual inhibitors.

For preliminary development of the IBiD and AcID duplex screen, the final concentration of protein per well found to be optimal for the assay was 100 nM and 530 nM respectively. To determine the Z-Score, protein was plated followed by addition of TR-ETV5 and FITC-ACTR tracers (final concentration 20 nM; representing 50% bound complex for AcID•ETV5 and 80% bound complex for IBiD•ACTR). Protein and tracer were incubated for 30 min at room temperature before fluorescence polarization measurements in the fluorescein and Texas Red channels were made using a PheraStar plate reader. 48 negative control wells and 48 positive control wells were tested to determine the Z factor.

3.6 References

- (1) Ptashne, M.; Gann, A. Transcriptional Activation by Recruitment. *Nature* **1997**, *386* (6625), 569–577. <https://doi.org/10.1038/386569a0>.
- (2) Tan, K.; Shlomi, T.; Feizi, H.; Ideker, T.; Sharan, R. Transcriptional Regulation of Protein Complexes within and across Species. *Proc. Natl. Acad. Sci.* **2007**, *104* (4), 1283–1288. <https://doi.org/10.1073/pnas.0606914104>.
- (3) Tang, H.; Sun, X.; Reinberg, D.; Ebright, R. H. Protein-Protein Interactions in Eukaryotic Transcription Initiation: Structure of the Preinitiation Complex. *Proc. Natl. Acad. Sci. U. S. A.* **1996**, *93* (3), 1119–1124.
- (4) Lee, T. I.; Young, R. A. Transcriptional Regulation and Its Misregulation in Disease. *Cell* **2013**, *152* (6), 1237–1251. <https://doi.org/10.1016/j.cell.2013.02.014>.
- (5) Fuxreiter, M.; Tompa, P.; Simon, I.; Uversky, V. N.; Hansen, J. C.; Asturias, F. J. Malleable Machines Take Shape in Eukaryotic Transcriptional Regulation. *Nat. Chem. Biol.* **2008**, *4* (12), 728. <https://doi.org/10.1038/nchembio.127>.
- (6) Mapp, A. K.; Pricer, R.; Sturlis, S. Targeting Transcription Is No Longer a Quixotic Quest. *Nat. Chem. Biol.* **2015**, *11* (12), 891–894. <https://doi.org/10.1038/nchembio.1962>.
- (7) Cesa, L. C.; Mapp, A. K.; Gestwicki, J. E. Direct and Propagated Effects of Small Molecules on Protein–Protein Interaction Networks. *Front. Bioeng. Biotechnol.* **2015**, *3*. <https://doi.org/10.3389/fbioe.2015.00119>.
- (8) Scott, D. E.; Bayly, A. R.; Abell, C.; Skidmore, J. Small Molecules, Big Targets: Drug Discovery Faces the Protein–Protein Interaction Challenge. *Nat. Rev. Drug Discov.* **2016**, *15* (8), nrd.2016.29. <https://doi.org/10.1038/nrd.2016.29>.
- (9) Arkin, M. R.; Tang, Y.; Wells, J. A. Small-Molecule Inhibitors of Protein-Protein Interactions: Progressing toward the Reality. *Chem. Biol.* **2014**, *21* (9), 1102–1114. <https://doi.org/10.1016/j.chembiol.2014.09.001>.
- (10) Arkin, M. R.; Randal, M.; DeLano, W. L.; Hyde, J.; Luong, T. N.; Oslob, J. D.; Raphael, D. R.; Taylor, L.; Wang, J.; McDowell, R. S.; Wells, J. A.; Braisted, A. C. Binding of Small Molecules to an Adaptive Protein–Protein Interface. *Proc. Natl. Acad. Sci.* **2003**, *100* (4), 1603–1608. <https://doi.org/10.1073/pnas.252756299>.
- (11) Scholes, N. S.; Weinzierl, R. O. J. Molecular Dynamics of “Fuzzy” Transcriptional Activator-Coactivator Interactions. *PLOS Comput. Biol.* **2016**, *12* (5), e1004935. <https://doi.org/10.1371/journal.pcbi.1004935>.
- (12) Liu, W.-L.; Coleman, R. A.; Ma, E.; Grob, P.; Yang, J. L.; Zhang, Y.; Dailey, G.; Nogales, E.; Tjian, R. Structures of Three Distinct Activator–TFIID Complexes. *Genes Dev.* **2009**, *23* (13), 1510–1521. <https://doi.org/10.1101/gad.1790709>.
- (13) Staby, L.; O’Shea, C.; Willemoës, M.; Theisen, F.; Kragelund, B. B.; Skriver, K. Eukaryotic Transcription Factors: Paradigms of Protein Intrinsic Disorder. *Biochem. J.* **2017**, *474* (15), 2509–2532. <https://doi.org/10.1042/BCJ20160631>.
- (14) Thakur, J. K.; Yadav, A.; Yadav, G. Molecular Recognition by the KIX Domain and Its Role in Gene Regulation. *Nucleic Acids Res.* **2014**, *42* (4), 2112–2125. <https://doi.org/10.1093/nar/gkt1147>.
- (15) Taatjes, D. J.; Näär, A. M.; Andel, F.; Nogales, E.; Tjian, R. Structure, Function, and Activator-Induced Conformations of the CRSP Coactivator. *Science* **2002**, *295* (5557), 1058–1062. <https://doi.org/10.1126/science.1065249>.

- (16) Ebert, M.-O.; Bae, S.-H.; Dyson, H. J.; Wright, P. E. NMR Relaxation Study of the Complex Formed Between CBP and the Activation Domain of the Nuclear Hormone Receptor Coactivator ACTR[†]. *Biochemistry* **2008**, *47* (5), 1299–1308. <https://doi.org/10.1021/bi701767j>.
- (17) 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. Publ. Protein Soc.* **2014**, *23* (7), 962–969. <https://doi.org/10.1002/pro.2480>.
- (18) 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.* **2004**, *337* (3), 521–534. <https://doi.org/10.1016/j.jmb.2004.01.038>.
- (19) Pomerantz, W. C.; Wang, N.; Lipinski, A. K.; Wang, R.; Cierpicki, T.; Mapp, A. K. Profiling the Dynamic Interfaces of Fluorinated Transcription Complexes for Ligand Discovery and Characterization. *ACS Chem. Biol.* **2012**, *7* (8), 1345–1350. <https://doi.org/10.1021/cb3002733>.
- (20) 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. U. S. A.* **2014**, *111* (33), 12061–12066. <https://doi.org/10.1073/pnas.1406033111>.
- (21) Rabuck-Gibbons, J. N.; Lodge, J. M.; Mapp, A. K.; Ruotolo, B. T. Collision Induced Unfolding Reveals Unique Fingerprints for Remote Protein Interaction Sites in the KIX Regulation Domain. *J. Am. Soc. Mass Spectrom.* **2019**, *30* (1), 94–102. <https://doi.org/10.1007/s13361-018-2043-6>.
- (22) Van Roey, K.; Uyar, B.; Weatheritt, R. J.; Dinkel, H.; Seiler, M.; Budd, A.; Gibson, T. J.; Davey, N. E. Short Linear Motifs: Ubiquitous and Functionally Diverse Protein Interaction Modules Directing Cell Regulation. *Chem. Rev.* **2014**, *114* (13), 6733–6778. <https://doi.org/10.1021/cr400585q>.
- (23) Neduva, V.; Linding, R.; Su-Angrand, I.; Stark, A.; Masi, F. de; Gibson, T. J.; Lewis, J.; Serrano, L.; Russell, R. B. Systematic Discovery of New Recognition Peptides Mediating Protein Interaction Networks. *PLOS Biol.* **2005**, *3* (12), e405. <https://doi.org/10.1371/journal.pbio.0030405>.
- (24) Mittag, T.; Kay, L. E.; Forman-Kay, J. D. Protein Dynamics and Conformational Disorder in Molecular Recognition. *J. Mol. Recognit. JMR* **2010**, *23* (2), 105–116. <https://doi.org/10.1002/jmr.961>.
- (25) Majmudar, C. Y.; Højfeldt, J. W.; Arevang, C. J.; Pomerantz, W. C.; Gagnon, J. K.; Schultz, P. J.; Cesa, L. C.; Doss, C. H.; Rowe, S. P.; Vásquez, V.; Tamayo-Castillo, G.; Cierpicki, T.; Brooks, C. L.; Sherman, D. H.; Mapp, A. K. Sekikaic Acid and Lobaric Acid Target a Dynamic Interface of the Coactivator CBP/P300. *Angew. Chem. Int. Ed.* **2012**, *51* (45), 11258–11262. <https://doi.org/10.1002/anie.201206815>.
- (26) Buhrlage, S. J.; Bates, C. A.; Rowe, S. P.; Minter, A. R.; Brennan, B. B.; Majmudar, C. Y.; Wemmer, D. E.; Al-Hashimi, H.; Mapp, A. K. Amphipathic Small Molecules Mimic the Binding Mode and Function of Endogenous Transcription Factors. *ACS Chem. Biol.* **2009**, *4* (5), 335–344. <https://doi.org/10.1021/cb900028j>.
- (27) Minter, A. R.; Brennan, B. B.; Mapp, A. K. A Small Molecule Transcriptional Activation Domain. *J. Am. Chem. Soc.* **2004**, *126* (34), 10504–10505. <https://doi.org/10.1021/ja0473889>.

- (28) Casey, R. J.; Desaulniers, J.-P.; Hojfeldt, J. W.; Mapp, A. K. Expanding the Repertoire of Small Molecule Transcriptional Activation Domains. *Bioorg. Med. Chem.* **2009**, *17* (3), 1034–1043. <https://doi.org/10.1016/j.bmc.2008.02.045>.
- (29) Bates, C. A.; Pomerantz, W. C.; Mapp, A. K. Transcriptional Tools: Small Molecules for Modulating CBP KIX-Dependent Transcriptional Activators. *Biopolymers* **2011**, *95* (1), 17–23. <https://doi.org/10.1002/bip.21548>.
- (30) Lea, W. A.; Simeonov, A. Fluorescence Polarization Assays in Small Molecule Screening. *Expert Opin. Drug Discov.* **2011**, *6* (1), 17–32. <https://doi.org/10.1517/17460441.2011.537322>.
- (31) Janzen, W. P. Screening Technologies for Small Molecule Discovery: The State of the Art. *Chem. Biol.* **2014**, *21* (9), 1162–1170. <https://doi.org/10.1016/j.chembiol.2014.07.015>.
- (32) Horng, J.-C.; Tracz, S. M.; Lumb, K. J.; Raleigh, D. P. Slow Folding of a Three-Helix Protein via a Compact Intermediate. *Biochemistry* **2005**, *44* (2), 627–634. <https://doi.org/10.1021/bi048852p>.
- (33) Ortiz, A. R.; Kolinski, A.; Skolnick, J. Tertiary Structure Prediction of the KIX Domain of CBP Using Monte Carlo Simulations Driven by Restraints Derived from Multiple Sequence Alignments. *Proteins* **1998**, *30* (3), 287–294.
- (34) Bontems, F.; Verger, A.; Dewitte, F.; Lens, Z.; Baert, J.-L.; Ferreira, E.; Launoit, Y. de; Sizun, C.; Guittet, E.; Villeret, V.; Monté, D. NMR Structure of the Human Mediator MED25 ACID Domain. *J. Struct. Biol.* **2011**, *174* (1), 245–251. <https://doi.org/10.1016/j.jsb.2010.10.011>.
- (35) Goodman, R. H.; Smolik, S. CBP/P300 in Cell Growth, Transformation, and Development. *Genes Dev.* **2000**, *14* (13), 1553–1577. <https://doi.org/10.1101/gad.14.13.1553>.
- (36) 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.* **2002**, *277* (45), 43168–43174. <https://doi.org/10.1074/jbc.M207660200>.
- (37) 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.* **2005**, *30* (5), 250–255. <https://doi.org/10.1016/j.tibs.2005.03.002>.
- (38) Lee, H.-K.; Park, U.-H.; Kim, E.-J.; Um, S.-J. MED25 Is Distinct from TRAP220/MED1 in Cooperating with CBP for Retinoid Receptor Activation. *EMBO J.* **2007**, *26* (15), 3545–3557. <https://doi.org/10.1038/sj.emboj.7601797>.
- (39) Verger, A.; Baert, J.-L.; Verreman, K.; Dewitte, F.; Ferreira, E.; Lens, Z.; de Launoit, Y.; Villeret, V.; Monté, D. The Mediator Complex Subunit MED25 Is Targeted by the N-Terminal Transactivation Domain of the PEA3 Group Members. *Nucleic Acids Res.* **2013**, *41* (9), 4847–4859. <https://doi.org/10.1093/nar/gkt199>.
- (40) Llauradó, M.; Abal, M.; Castellví, J.; Cabrera, S.; Gil-Moreno, A.; Pérez-Benavente, A.; Colás, E.; Doll, A.; Dolcet, X.; Matias-Guiu, X.; Vazquez-Levin, M.; Reventós, J.; Ruiz, A. ETV5 Transcription Factor Is Overexpressed in Ovarian Cancer and Regulates Cell Adhesion in Ovarian Cancer Cells. *Int. J. Cancer* **2012**, *130* (7), 1532–1543. <https://doi.org/10.1002/ijc.26148>.

- (41) Currie, S. L.; Doane, J. J.; Evans, K. S.; Bhachech, N.; Madison, B. J.; Lau, D. K. W.; McIntosh, L. P.; Skalicky, J. J.; Clark, K. A.; Graves, B. J. ETV4 and AP1 Transcription Factors Form Multivalent Interactions with Three Sites on the MED25 Activator-Interacting Domain. *J. Mol. Biol.* **2017**, *429* (20), 2975–2995. <https://doi.org/10.1016/j.jmb.2017.06.024>.
- (42) Slany, R. K. The Molecular Biology of Mixed Lineage Leukemia. *Haematologica* **2009**, *94* (7), 984–993. <https://doi.org/10.3324/haematol.2008.002436>.
- (43) Goding, J. W. 12 - Immunofluorescence. In *Monoclonal Antibodies (Third Edition)*; Academic Press: London, 1996; pp 352–399. <https://doi.org/10.1016/B978-012287023-1/50060-2>.
- (44) Zhang, null; Chung, null; Oldenburg, null. A Simple Statistical Parameter for Use in Evaluation and Validation of High Throughput Screening Assays. *J. Biomol. Screen.* **1999**, *4* (2), 67–73. <https://doi.org/10.1177/108705719900400206>.
- (45) Reed, G. F.; Lynn, F.; Meade, B. D. Use of Coefficient of Variation in Assessing Variability of Quantitative Assays. *Clin. Diagn. Lab. Immunol.* **2002**, *9* (6), 1235–1239. <https://doi.org/10.1128/CDLI.9.6.1235-1239.2002>.
- (46) Demarest, S. J.; Martinez-Yamout, M.; Chung, J.; Chen, H.; Xu, W.; Dyson, H. J.; Evans, R. M.; Wright, P. E. Mutual Synergistic Folding in Recruitment of CBP/P300 by P160 Nuclear Receptor Coactivators. *Nature* **2002**, *415* (6871), 549–553. <https://doi.org/10.1038/415549a>.
- (47) Demarest, S. J.; Deechongkit, S.; Dyson, H. J.; Evans, R. M.; Wright, P. E. Packing, Specificity, and Mutability at the Binding Interface between the P160 Coactivator and CREB-Binding Protein. *Protein Sci. Publ. Protein Soc.* **2004**, *13* (1), 203–210. <https://doi.org/10.1110/ps.03366504>.
- (48) Lin, C. H.; Hare, B. J.; Wagner, G.; Harrison, S. C.; Maniatis, T.; Fraenkel, E. A Small Domain of CBP/P300 Binds Diverse Proteins: Solution Structure and Functional Studies. *Mol. Cell* **2001**, *8* (3), 581–590. [https://doi.org/10.1016/S1097-2765\(01\)00333-1](https://doi.org/10.1016/S1097-2765(01)00333-1).
- (49) 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* **2005**, *44* (2), 490–497. <https://doi.org/10.1021/bi048161t>.
- (50) Dames, S. A.; Martinez-Yamout, M.; De Guzman, R. N.; Dyson, H. J.; Wright, P. E. Structural Basis for Hif-1 Alpha /CBP Recognition in the Cellular Hypoxic Response. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99* (8), 5271–5276. <https://doi.org/10.1073/pnas.082121399>.
- (51) Chu, W.-T.; Chu, X.; Wang, J. Investigations of the Underlying Mechanisms of HIF-1 α and CITED2 Binding to TAZ1. *Proc. Natl. Acad. Sci.* **2020**, *117* (11), 5595–5603. <https://doi.org/10.1073/pnas.1915333117>.
- (52) De Guzman, R. N.; Martinez-Yamout, M. A.; Dyson, H. J.; Wright, P. E. Interaction of the TAZ1 Domain of the CREB-Binding Protein with the Activation Domain of CITED2: REGULATION BY COMPETITION BETWEEN INTRINSICALLY UNSTRUCTURED LIGANDS FOR NON-IDENTICAL BINDING SITES*. *J. Biol. Chem.* **2004**, *279* (4), 3042–3049. <https://doi.org/10.1074/jbc.M310348200>.
- (53) Finlan, L.; Hupp, T. R. The N-Terminal Interferon-Binding Domain (IBiD) Homology Domain of P300 Binds to Peptides with Homology to the P53 Transactivation Domain. *J. Biol. Chem.* **2004**, *279* (47), 49395–49405. <https://doi.org/10.1074/jbc.M405974200>.

- (54) Blagosklonny, M. V. Hypoxia-Inducible Factor: Achilles' Heel of Antiangiogenic Cancer Therapy (Review). *Int. J. Oncol.* **2001**, *19* (2), 257–262. <https://doi.org/10.3892/ijo.19.2.257>.
- (55) Semenza, G. L. HIF-1 and Human Disease: One Highly Involved Factor. *Genes Dev.* **2000**, *14* (16), 1983–1991. <https://doi.org/10.1101/gad.14.16.1983>.
- (56) 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. U. S. A.* **2007**, *104* (17), 7009–7014. <https://doi.org/10.1073/pnas.0702010104>.
- (57) Breen, M. E.; Mapp, A. K. Modulating the Masters: Chemical Tools to Dissect CBP and P300 Function. *Curr. Opin. Chem. Biol.* **2018**, *45*, 195–203. <https://doi.org/10.1016/j.cbpa.2018.06.005>.
- (58) O'Coinceanainn, M.; Astill, C.; Baderschneider, B. Coordination of Aluminium with Purpurogallin and Theaflavin Digallate. *J. Inorg. Biochem.* **2003**, *96* (4), 463–468. [https://doi.org/10.1016/S0162-0134\(03\)00248-4](https://doi.org/10.1016/S0162-0134(03)00248-4).
- (59) Mahmoud, M. E.; Hafez, O. F.; Osman, M. M.; Yakout, A. A.; Alrefaay, A. Hybrid Inorganic/Organic Alumina Adsorbents-Functionalized-Purpurogallin for Removal and Preconcentration of Cr(III), Fe(III), Cu(II), Cd(II) and Pb(II) from Underground Water. *J. Hazard. Mater.* **2010**, *176* (1), 906–912. <https://doi.org/10.1016/j.jhazmat.2009.11.122>.
- (60) Hynes, M. J.; Ó Coinceanainn, M. The Kinetics and Mechanisms of the Reaction of Iron(III) with Gallic Acid, Gallic Acid Methyl Ester and Catechin. *J. Inorg. Biochem.* **2001**, *85* (2), 131–142. [https://doi.org/10.1016/S0162-0134\(01\)00205-7](https://doi.org/10.1016/S0162-0134(01)00205-7).
- (61) Cook, K. M.; Hilton, S. T.; Mecinovic, J.; Motherwell, W. B.; Figg, W. D.; Schofield, C. J. Epidithiodiketopiperazines Block the Interaction between Hypoxia-Inducible Factor-1 α (HIF-1 α) and P300 by a Zinc Ejection Mechanism. *J. Biol. Chem.* **2009**, *284* (39), 26831–26838. <https://doi.org/10.1074/jbc.M109.009498>.
- (62) Jayatunga, M. K. P.; Thompson, S.; McKee, T. C.; Chan, M. C.; Reece, K. M.; Hardy, A. P.; Sekirnik, R.; Seden, P. T.; Cook, K. M.; McMahon, J. B.; Figg, W. D.; Schofield, C. J.; Hamilton, A. D. Inhibition of the HIF1 α -P300 Interaction by Quinone- and Indandione-Mediated Ejection of Structural Zn(II). *Eur. J. Med. Chem.* **2015**, *94*, 509–516. <https://doi.org/10.1016/j.ejmech.2014.06.006>.
- (63) Viziteu, E.; Grandmougin, C.; Goldschmidt, H.; Seckinger, A.; Hose, D.; Klein, B.; Moreaux, J. Chetomin, Targeting HIF-1 α /P300 Complex, Exhibits Antitumour Activity in Multiple Myeloma. *Br. J. Cancer* **2016**, *114* (5), 519–523. <https://doi.org/10.1038/bjc.2016.20>.

CHAPTER 4

Targeting the Transmembrane Serine Protease TMPRSS2 to Prevent SARS-CoV-2 Infection[‡]

4.1 Abstract

The COVID-19 pandemic has highlighted the need for new antiviral targets, as many of the currently approved drugs have proven ineffective against mitigating SARS-CoV-2 infections. The host transmembrane serine protease TMPRSS2 is a highly promising antiviral target, as it plays a direct role in priming the spike protein before viral entry occurs. Further, unlike other targets such as ACE2, TMPRSS2 has no known biological role but is implicated in infectivity of many additional viruses. There are few published TMPRSS2 inhibitors, and those that exist are not ideal biological probe compounds or drugs. Thus, TMPRSS2 is an attractive therapeutic target and there is a need for novel inhibitors. In this chapter, an *in vitro* biochemical platform with which TMPRSS2 protease can be studied and novel inhibitors can be identified is developed. Ultimately, this approach is used in combination with *in silico* screening to identify a set of TMPRSS2 inhibitors from already approved drugs. Characterization of lead compounds and known inhibitors is conducted, providing insight into their TMPRSS2 inhibitory mechanism. Additional preliminary screening data of multiple libraries is presented.

4.2 Introduction

The emergence of COVID-19 in late 2019 and the rapid transmission of the disease around the globe has prompted an urgent need for effective treatments.¹ As with many coronaviruses,

[‡] The majority of this chapter is reproduced from Peiffer A.L.*, Garlick J.M.*, Wu Y., Soellner M.B., Brooks C.L. III, Mapp A.K. TMPRSS2 inhibitor discovery facilitated through an *in silico* and biochemical screening platform (DOI: 10.1101/2021.03.22.436465). The research described in this chapter was a collaborative effort. Julie Garlick and Amanda Peiffer conceived the project, targeting TMPRSS2 as an anti-SARS-CoV-2 strategy. Julie Garlick led experimental approach on the biochemical side. Amanda Peiffer led experimental approach on the computational side. Julie Garlick and Amanda Peiffer developed expression and purification protocols for TMPRSS2 protease. Julie Garlick conducted all assay development, mass spec experiments, and screening. Amanda Peiffer constructed the homology model as well as interpreted computational results, such as analysis of docked structures. Yujin Wu developed the flexible docking protocol and scoring functions, and used them to conduct *in silico* screening. Tessa Epstein curated the guanidinium screening library.

infection with SARS-CoV-2 requires host cell cooperation; the spike (S) protein located on the viral coat requires priming by TMPRSS2, a human transmembrane serine protease, for viral entry via the receptor ACE2 (Figure 4.1).²⁻⁵ While many groups are focusing on blocking the interactions between ACE2 and the S protein, ACE2 also plays an important role in healthy cell function by counterbalancing ACE to lower and maintain healthy blood pressure.⁶ Alternatively, there is little known about the biological function of TMPRSS2, with data suggesting it is likely functionally redundant.^{7,8} TMPRSS2^{-/-} knockout mice have little phenotypic differences compared to wild-type animals, yet conferred resistance to viral infections, suggesting that the protein is not essential.⁹ Inhibiting transcription of TMPRSS2 via BET inhibitors leads to decreased infectivity of SARS-CoV-2 in human lung cells, further suggesting the viability of TMPRSS2 inhibition as an antiviral strategy.¹⁰ Besides its role in viral infection, the other documented functions of TMPRSS2 are related to prostate cancer.¹¹⁻¹³ TMPRSS2 expression is regulated in an androgen dependent manner, and has been shown to be overexpressed in prostate cancer, and potentially linked to cancer progression and metastasis.^{11,12,14} However, it is unclear if TMPRSS2 overexpression is functionally significant, or its overexpression is just the result of the real issue, misregulation of the androgen receptor. In addition, translocations of the ETS transcription factor genes to the TMPRSS2 promotor is known to occur. This leads to over expression of the ETS transcription factors, now regulated by the androgen receptor.^{15,16} Thus at this point TMPRSS2 is only known to have detrimental functions in humans, aiding in viral infection and cancer metastasis. Additionally, as a human protein target rather than a viral protein target, TMPRSS2-targeting therapeutics should be less susceptible to drug-resistance due to viral mutation. Taken together, this information indicates that TMPRSS2 is a desirable drug target for preventing SARS-CoV-2 infection.

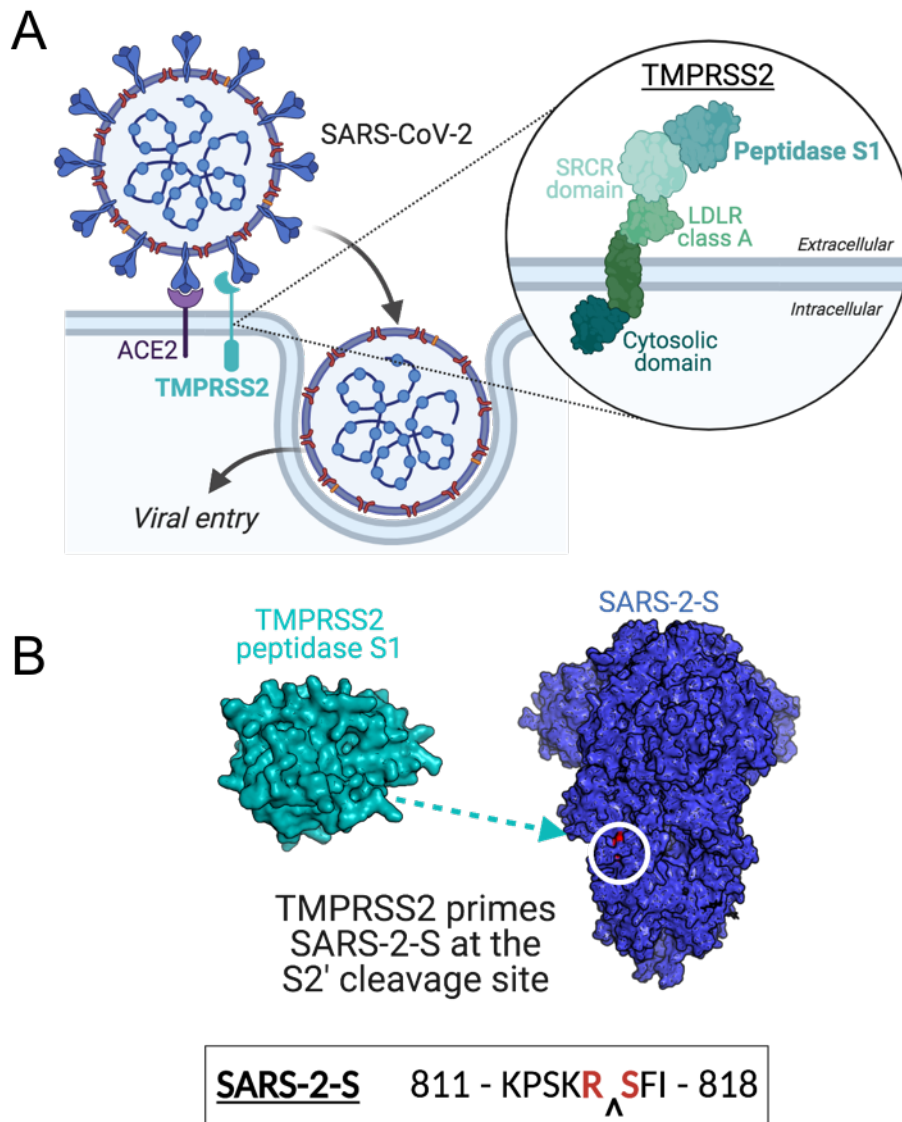


Figure 4.1. TMPRSS2 aids in SARS-CoV-2 viral infection. A) TMPRSS2 primes the viral S protein (SARS-2-S), which promotes membrane fusion and ultimately viral entry. TMPRSS2, part of the type II transmembrane serine protease family and hepsin/TMPRSS subfamily, is anchored at the cell membrane.⁴ The protein is mostly extracellular, with a small intracellular cytosolic domain. The extracellular portion of the protein is composed of a LDLR class A domain, an SRCR domain, and finally the peptidase S1 domain required for catalytic activity.¹⁷ B) The peptidase S1 domain of TMPRSS2 cleaves SARS-2-S at the S2' cleavage site.^{18,19} Figure created using BioRender.

To date, there are few reported TMPRSS2 inhibitors. Camostat, a compound initially discovered as a Matriptase 2 inhibitor, also inhibits TMPRSS2.^{4,20} Nafamostat and gabexate have also been reported to inhibit TMPRSS2. However, camostat, nafamostat, and gabexate lack TMPRSS2 specificity and inhibit a wide range of serine proteases. Additionally, camostat is

metabolized in mammals to structures with poorly defined activity.^{21–23} Further, these compounds have only recently been explored for clinical applications with SARS-CoV-2, and the underlying inhibitory mechanism of action is still being investigated.^{20,22} It has been reported that each of these compounds form a covalent bond with the active site serine of serine proteases via the central ester, also a site of metabolic breakdown (Figure 4.2).^{24,25} Additionally, molecular modeling studies on TMPRSS2 supports covalent bond formation with both camostat and nafamostat, as well as the camostat metabolite FOY 251.²⁶ Thus, TMPRSS2 inhibitors with less reactive architectures are highly desirable.

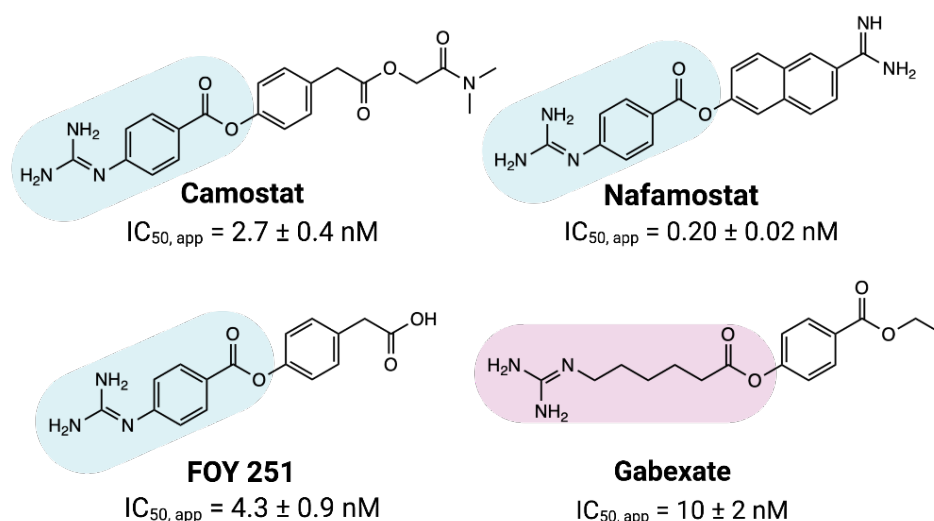


Figure 4.2. The molecules that have commonly been cited as TMPRSS2 inhibitors. All four inhibitors contain a reactive ester, which can form a covalent bond with the activated catalytic serine. Highlighted are the adducts that would form. Camostat, nafamostat, and FOY 251 all form the same adduct (blue). Apparent IC_{50} values for the four inhibitors against TMPRSS2 after 30 minutes incubation time are shown.

TMPRSS2 is part of the type II transmembrane serine protease family and hepsin/TMPRSS subfamily. The protein, anchored at the cell membrane, is mostly extracellular, with a small intracellular cytosolic domain. The extracellular portion of the protein is composed of a LDLR class A domain, an SRCR domain, and finally the peptidase S1 domain, which is required for catalytic activity (Figure 4.1).^{4,17} Thus, the most obvious site for TMPRSS2 inhibition is the active site of the peptidase itself, and indeed this is where prior efforts have been focused.^{11,27–29} The classic triad of His/Asp/Ser is present (H296, D345, S441) in the active site, and mutating any of these residues abrogates activity *in vitro*.¹⁷ Determining the endogenous substrates and function of TMPRSS2 has been challenging, with even complete knockout in mice producing no measurable

effects histopathologically and phenotypically.⁷ A few studies exist looking at the substrate scope of TMPRSS2, for example matriptase and PAR2 have been shown to undergo TMPRSS2 mediated activation in prostate cancer proliferation and migration pathways, and positional scanning data is available for P1-P4 substrate peptides.^{12,30,31} These data indicate that TMPRSS2 strongly prefers Arg as the P1 residue.¹¹ In addition, results from substrate-based inhibitors suggests analogues with hydrophobic D-configuration P3 residues are the most successful for TMPRSS2 inhibition.²⁸ However, further studies will be necessary to define chemotypes that are particularly potent and selective for TMPRSS2.

The most significant roadblock in TMPRSS2 inhibitor discovery is the well-documented challenge of obtaining catalytically active enzyme, full-length or the peptidase domain;^{33,35,40} even commercial sources of TMPRSS2 exhibit low to no activity, making it difficult to carry out extensive screening campaigns.⁴³ This chapter begins with design of a novel expression construct which eliminates the need for multiple rounds of purification. In addition, re-folding conditions were optimized, leading to isolation of sufficient TMPRSS2 peptidase domain for a full screening and follow-up campaign for up to 5000 compounds with a single 1-L expression.⁴⁸ This allowed for biochemical characterization of known TMPRSS2 inhibitors as well as identification of novel active-site inhibitors, including two new TMPRSS2 inhibitor architectures, using a hybrid *in silico/in vitro* screening platform.

4.3 Results & Discussion

Recombinant expression of TMPRSS2 Peptidase S1 domain in E. coli

Like other serine proteases, TMPRSS2 is natively expressed as a zymogen; activation occurs via autocatalysis of the peptide bond between Arg-255 and Ile-256, leaving the *N*-terminal portion of Ile-256 to undergo a conformational change to stabilize the active state.^{32–34} Careful consideration was taken when designing the gene fragment for recombinant expression and purification of the TMPRSS2 peptidase domain (residues 256–492). *C*-terminal affinity tags appear to disrupt catalytic activity; thus, an *N*-terminal affinity tag is required but must be removed, leaving residue Ile-256 with a free *N*-terminus. Previous approaches to expression and purification in bacteria have utilized an orthogonal protease to allow for cleavage directly *N*-terminal to Ile-256, such as the TAGzyme system.²⁸ We hypothesized that utilizing a construct with an extended *N*-terminal portion would be enough to promote enzyme autocatalysis, removing any *N*-terminal

affinity tag as well as yielding the necessary free isoleucine without the need for multiple rounds of purification and introduction of an orthogonal cleavage site. Thus, a gene fragment for the catalytic domain of TMPRSS2 was constructed (see Methods), called TMPRSS2(247–492), with an N-terminal 6xHis tag for purification (Figure 4.3).

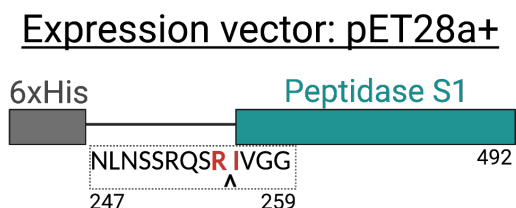


Figure 4.3. Proposed construct for TMPRSS2(256–492) peptidase expression in *E. Coli*. The proposed amino acid sequence was inserted into pET28a+ vector with a kanamycin resistance cassette. Autocleavage of the peptide bond between R255 and I256 (highlighted in red and denoted with an arrow) should allow for removal of the 6xHis purification tag once the protein is activated.

Overexpression of the TMPRSS2(247–492) construct in *E. coli* led to protein aggregation in insoluble inclusion bodies, which enabled simple separation from the remaining cell lysate. Solubilizing the inclusion bodies required denaturing the aggregates utilizing 8 M urea. The N-terminal 6xHis tag was used to remove the denatured, unfolded protein from remaining impurities by batch binding with Ni-NTA resin. Purified, denatured TMPRSS2 was then subjected to refolding by rapid dilution in 1:100 refolding buffer. Development of a modified refolding procedure using a syringe pump for slow and controlled dilution proved to be instrumental in producing active protein, likely due to providing a more optimal environment for the three internal disulfide bonds to correctly form. Concentration (10-fold) and subsequent dialysis into a 50 mM Tris 500 mM NaCl 0.01% NP-40 pH 8 buffer led to activation of the enzymatic activity demonstrated by the autocatalytic cleavage of the 6xHis tag (Figure 4.4).

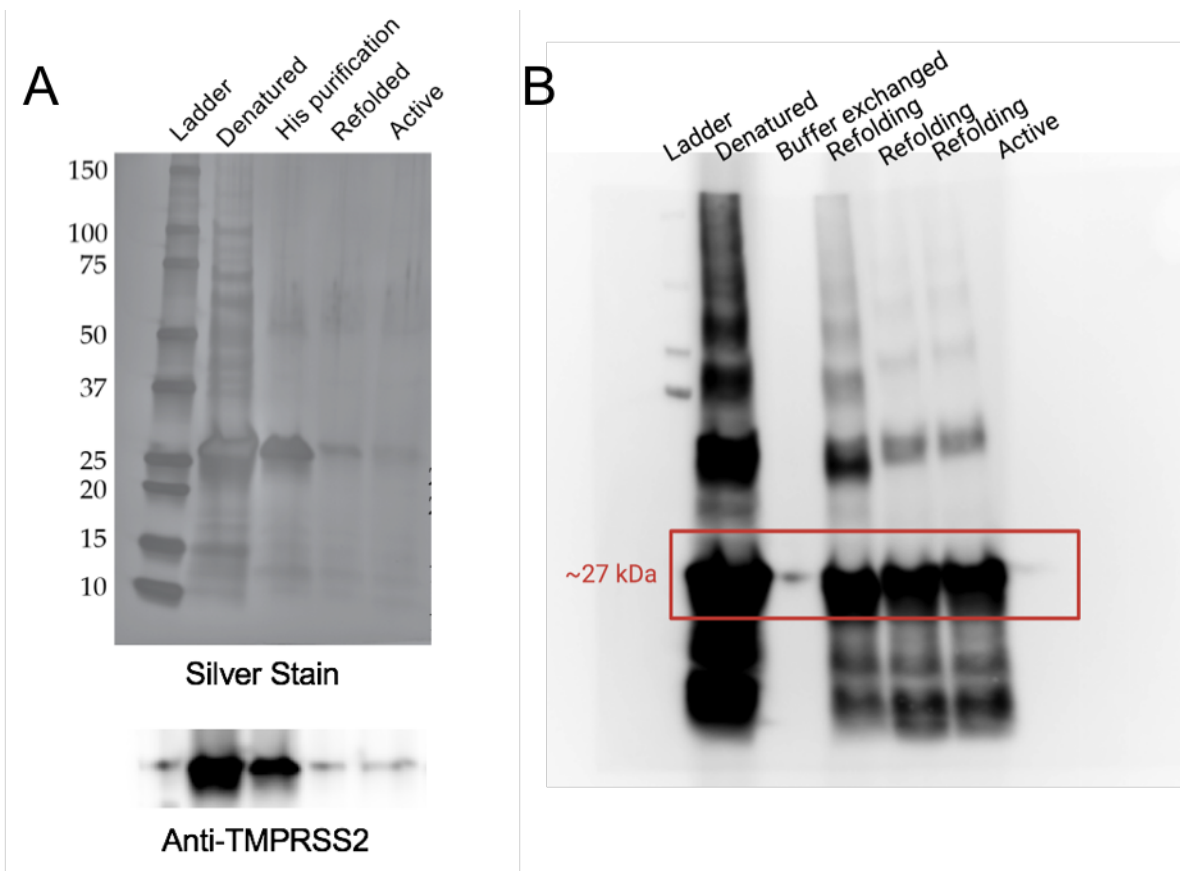


Figure 4.4. Expression and purification of TMPRSS2 peptidase domain. A) Protein is expressed and purified by expression in inclusion bodies (labeled denatured), purified by batch binding to Ni-NTA resin, refolded, and dialyzed into assay buffer. Silver staining shows purity and protein levels at each step (TMPRSS2, ~26kDa). Western blot with TMPRSS2 antibody raised against the protease domain confirms the identity of the 26 kDa protein observed with silver staining as TMPRSS2. B) Western blot for 6xHis tag at various stages of TMPRSS2 purification. Significant signal is seen after denaturation and at various time points during refolding. A small amount of His-tagged protein is observed immediately after exchange of concentrated refolding solution into activation buffer, but ultimately no signal is observed, indicating autocleavage and removal of 6xHis tag, in the final active product. Experiments conducted with Amanda Peiffer.

Activity of TMPRSS2 peptidase domain

The canonical trypsin substrate Boc-QAR-AMC has been reported as a TMPRSS2 substrate, which we used to confirm activity of the purified protein (assay overview in Figure 4.7).^{11,21} To determine approximate concentration of active protein, we measured TMPRSS2 activity while titrating in the covalent protease inhibitor FPR-chloromethylketone (also known as PPACK).³⁵⁻³⁷ This inhibitor reacts with the histidine in the catalytic triad of serine proteases to irreversibly alkylate and inactivate the enzyme, in a 1:1 complex (Figure 4.5). We determined the K_M to be $5.1 \pm 0.4 \mu M$.

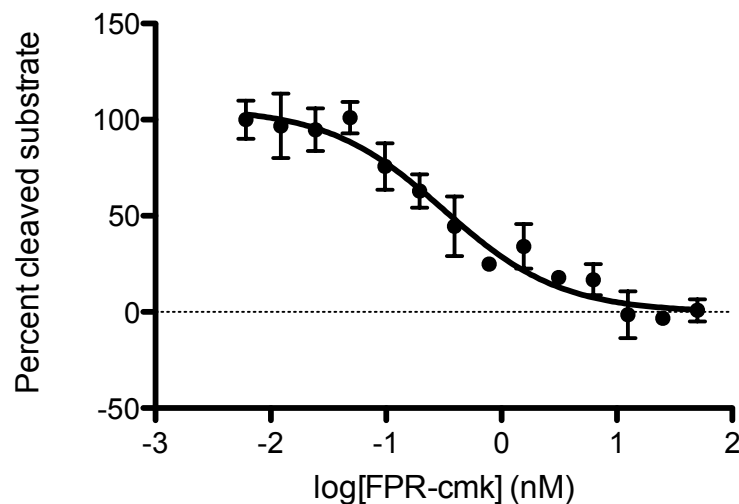


Figure 4.5. Inhibition curve of TMPRSS2 activity obtained with increasing [FPR-cmk]. The IC_{50} obtained here is 0.32 ± 0.09 nM, allowing us to approximate 100% inhibition at about 0.64 nM. Thus, assuming 1:1 complex formation of protein and inhibitor, we can determine the protein concentration of this sample to be about 0.64 nM. Data points represent the average of technical triplicate data points with error bars indicating the standard deviation of the mean.

with TMPRSS2, which is comparable to the K_M for trypsin with this substrate, 7.8 ± 0.7 μ M (Figure 4.6). Experimental conditions for high throughput screening (HTS) in 384-well plates were optimized using 2.5 μ M substrate and 0.5 nM TMPRSS2. The concentration of substrate in this assay was set below the K_M to enable identification of competitive inhibitors. After incubation at room temperature for 30 min, endpoint fluorescence was determined (ex: 355 nm, em: 450 nm). At 30 min, less than 20% of the substrate was cleaved as determined by comparing fluorescence to that of 2.5 μ M free AMC (i.e. 100% substrate conversion), enabling inhibition to be monitored while activity is still within the linear range. The S:B and Z factor were calculated to be 10.6 and

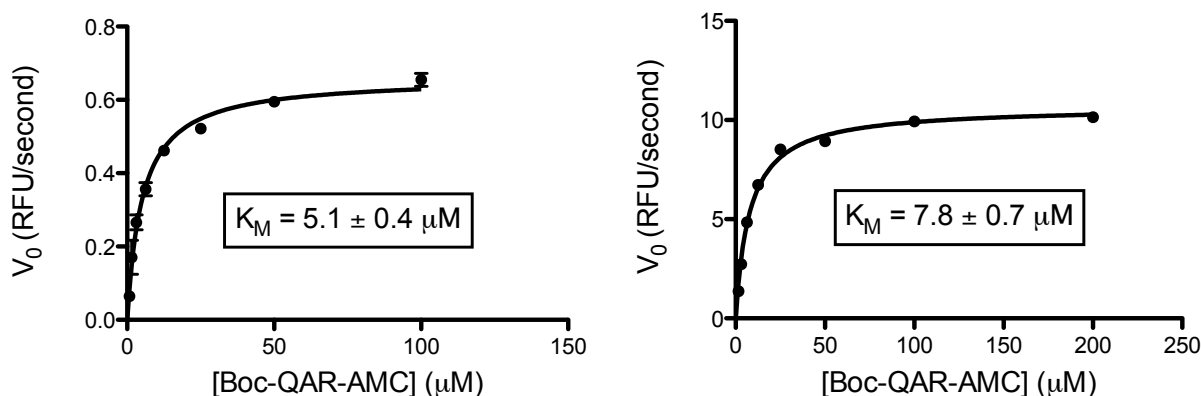


Figure 4.6. Boc-QAR-AMC as a substrate for TMPRSS2 and Trypsin. The K_M of Boc-Gln-Ala-Arg-AMC is determined for TMPRSS2 (left) and Trypsin (right) by plotting initial velocity, calculated as the slope of the kinetic trace at less than 10 percent substrate cleavage, vs substrate concentration. Data was fit to the Michaelis-Menten equation using GraphPad Prism. Individual data points are kinetic experiments conducted in duplicate with error bars as the standard deviation of the mean.

0.79, respectively, indicating an excellent HTS assay (Figure 4.7). This biochemical assay was then utilized to test the apparent IC_{50} of known inhibitors camostat, nafamostat, and gabexate, where the compounds were incubated with TMPRSS2 for 30 min before substrate was added (Figure 4.7). IC_{50} values for these compounds agree with previously reported values obtained using similar biochemical approaches.²¹ Notably, the IC_{50} values for these compounds are very low, approaching the lower limit of detection based on the protein concentration in the assay.

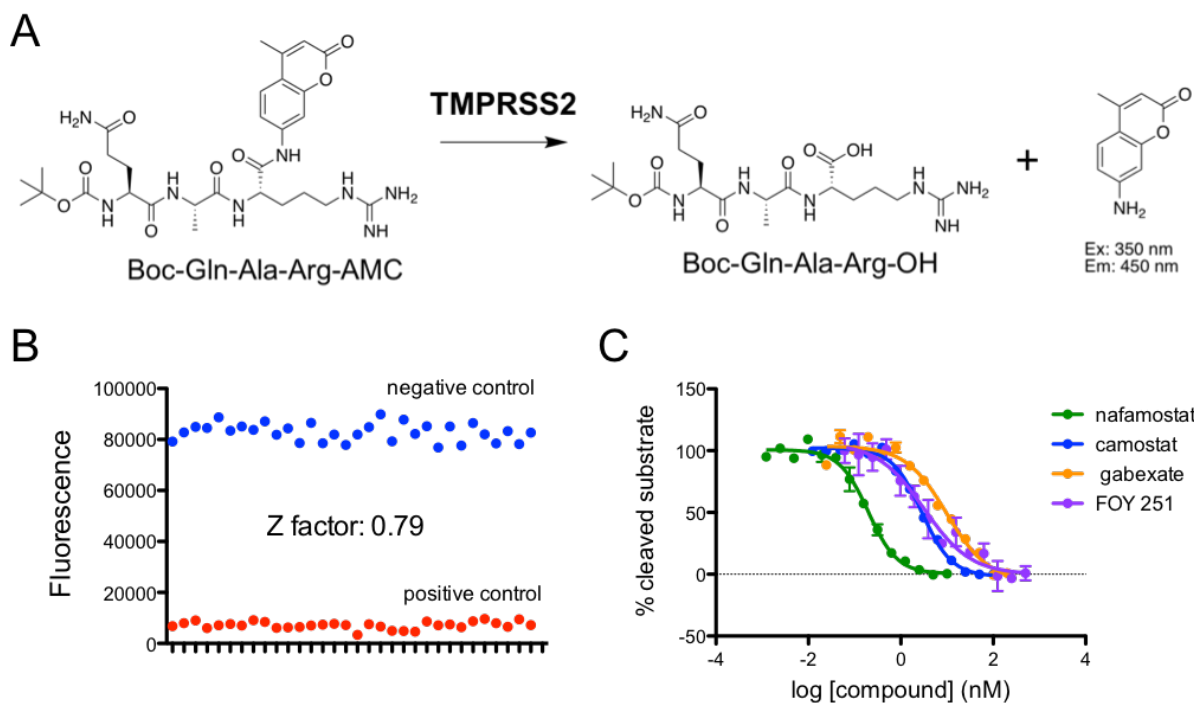


Figure 4.7. Development of biochemical TMPRSS2 activity assay. A) Schematic of the biochemical assay used to monitor TMPRSS2 activity. Cleavage of the peptide bond between Arg and AMC in the substrate results in release of the AMC fluorophore, for which fluorescence can be monitored. B) TMPRSS2 activity assay was adapted to 384 well format with the Z factor indicating an excellent assay. Negative controls correspond to protein + substrate, while positive controls correspond to substrate only. C) Inhibition of TMPRSS2 by compounds camostat, nafamostat, FOY 251, and gabexate, with 30 minutes preincubation of protein and inhibitor. Data points represent the average of technical triplicate with error bars indicating the standard deviation of the mean. IC₅₀ values are listed in Figure 5.2.

Analysis of covalent TMPRSS2 inhibitors

While it has not been confirmed experimentally, others have suggested that the inhibitors camostat and nafamostat covalently modify TMPRSS2 with a similar mechanism to other proteases.²⁶ This involves initial binding, acylation, and ultimately hydrolysis. While it would be ideal to validate formation of a covalent adduct with TMPRSS2 using mass spectrometry, the purification of the protease domain yields low concentrations of active protein that is very sensitive to buffer conditions, making mass spectrometric analysis intractable. However, we demonstrate covalent adduct formation via mass spectrometry using trypsin (Figure 4.8). With equimolar protein and compound, an increase in mass of 161 Da is observed for camostat, nafamostat, and FOY 251. Gabexate requires 10X compound compared to protein for adduct formation to be observed, which corresponds to a mass increase of 155 Da. Additionally, a time-

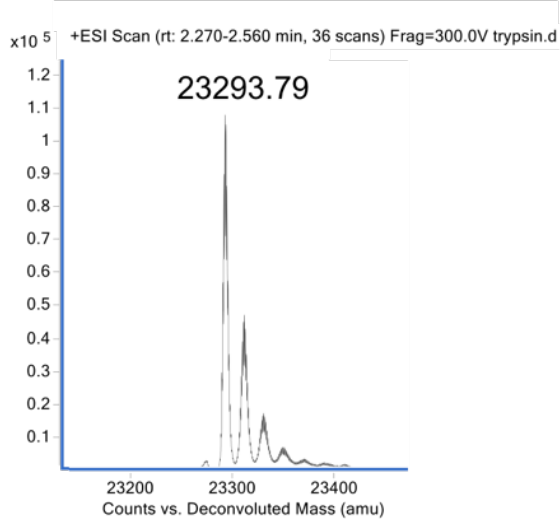
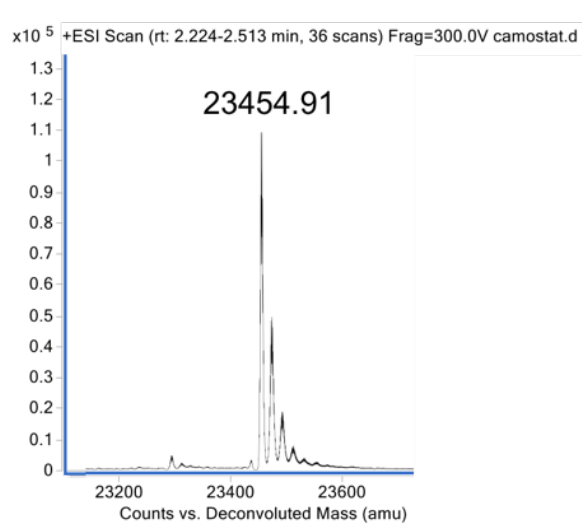
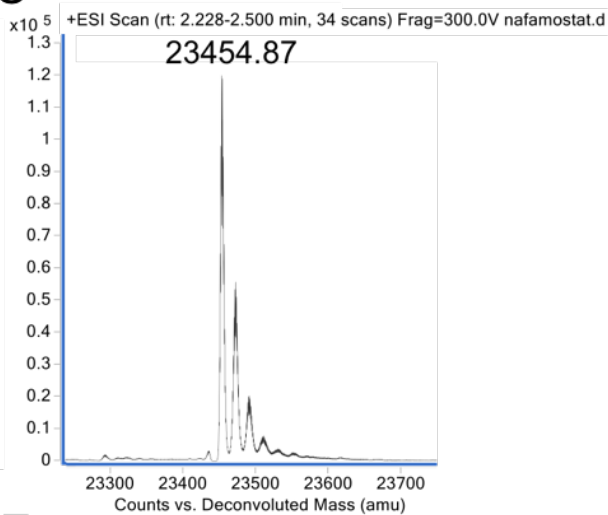
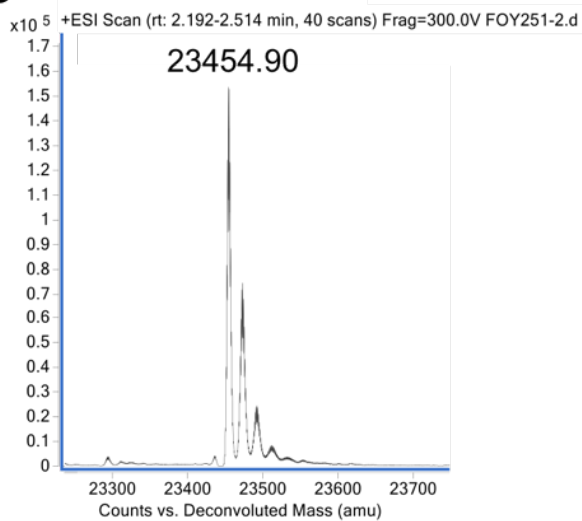
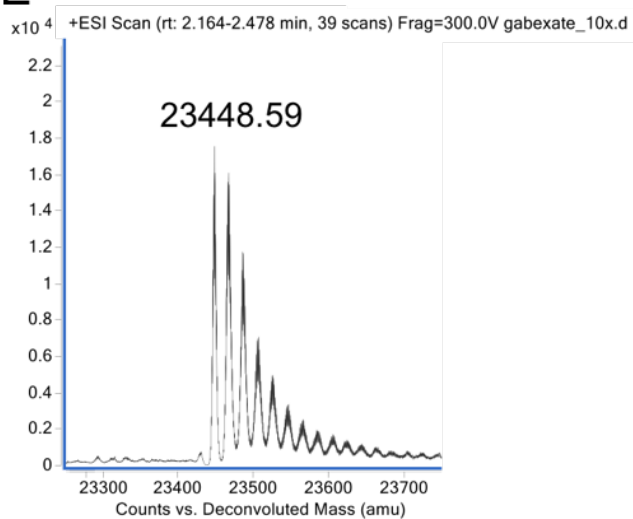
A**B****C****D****E**

Figure 4.8. Camostat, nafamostat, FOY 251, and gabexate covalently modify Trypsin. A-D) Deconvoluted LC-MS data of trypsin (A), incubated with equimolar camostat (B) nafamostat (C), or FOY 251 (D). After 30 minutes incubation, a mass increase of about 161 Da is observed, indicating covalent adduct formation. E) Trypsin incubated with 10 equivalents gabexate. After 30 minutes incubation, a mass increase of about 155 Da is observed, indicating covalent adduct formation. No mass change was observed after 30 minutes incubation with equimolar inhibitor

dependent decrease in apparent IC_{50} is observed with camostat in biochemical activity assays for both trypsin and TMPRSS2 (Figure 4.9). This shift in IC_{50} vs. time fits to an equation for one phase decay, with the IC_{50} values approaching the limiting concentration of the protein used in the assay, suggesting that covalent bond formation is occurring at the active site (Figure 5A). Thus, while these compounds exhibit low nM IC_{50} values in vitro, this is likely due to the covalent mechanism of action.

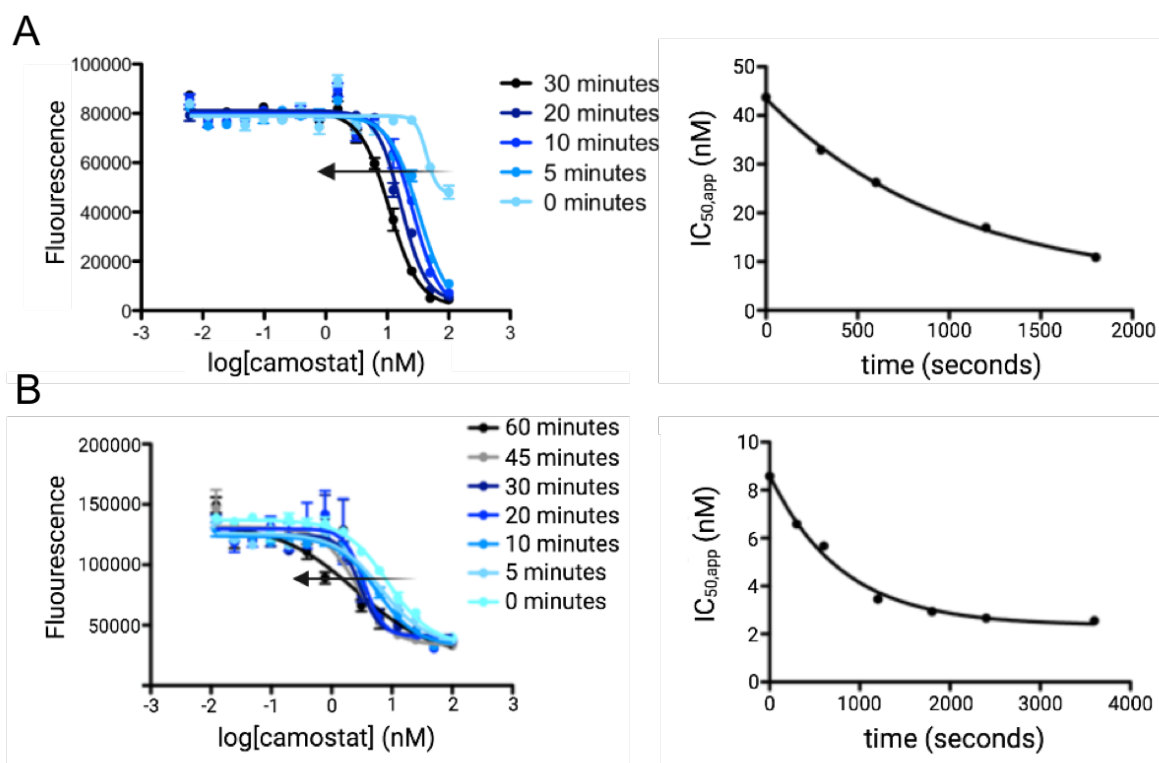


Figure 4.9. Decreasing IC_{50} values for camostat with increased incubation time for trypsin (A) and TMPRSS2 (B) is consistent with covalent inhibition. Left: IC_{50} curves of camostat obtained at after varying incubation times with protein. Values plotted are the average of triplicate data points with the error represented as the standard deviation of the mean. Right: Plot of camostat IC_{50} vs incubation time, fit to an equation for one phase decay.

It is important to note that in a cellular context camostat has a very short half-life, <1 min.³⁸ Rapid hydrolysis to (4-(4-guanidinobenzoyl-oxy)phenylacetic acid), also known as the protease

inhibitor FOY 251, occurs both *in vitro* in serum as well as *in vivo*.^{39–41} We observed similar potencies of FOY 251 and camostat in our biochemical assay, with apparent $IC_{50} = 4.3 \pm 0.9$ nM and 2.7 ± 0.4 nM, respectively (Figure 4.2). The half-life of FOY 251 is longer than camostat, though it is metabolized to 4-guanidinobenzoic acid, which shows minimal TMPRSS2 inhibition.³⁹ Thus, it would be advantageous to identify novel inhibitors that have alternative molecular scaffolds.

Combined in silico and biochemical screening leads to identification of novel inhibitors

Virtual screening methods have greatly improved over the past two decades, leading many drug discovery campaigns by filtering out thousands/millions of molecules before testing them *in vitro*.^{42–46} However, such studies require a structural model in which to dock compounds into the active site. Because no crystallographic or NMR-based models exist for TMPRSS2, we developed a homology model for the active soluble domain starting from prediction using the SWISS-MODEL web-interface (Figure 4.10).^{47–51} This structure was built based on sequence homology of hepsin (PDB 5CE1). The structure showed high homology with the TMPRSS2 peptidase domain (34% sequence similarity with 70% sequence coverage) and also contained the bound ligand 2-[6-(1-hydroxycyclohexyl)pyridin-2-yl]-1H-indole-5-carboximidamide, which served as one of the templates for pharmacophore-based docking of putative ligands. Full descriptions of homology model development and virtual screening methodology are available in Appendix C.

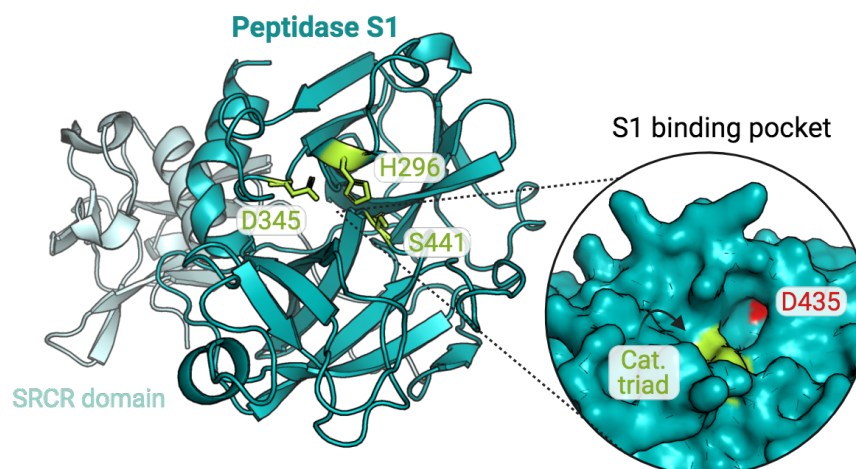


Figure 4.10. TMPRSS2 homology model. The catalytic triad is highlighted in lime green. D435, key for substrate recognition, is shown in red. Experiment conducted by Amanda Peiffer. Figure generated using BioRender.

A total of 134,109 molecules were collected from multiple databases (ZINC, SweetLead, and the Center for Chemical Genomics) and subjected to initial screening, leading to 4,307 Level 1 candidates which were ranked by a novel scoring function developed by Yujin Wu. The virtual screen was successful on many fronts. MD-based flexible docking identified residues near the active site that are conformationally dynamic to accommodate different ligands. Specifically, residues Gln-438 and Lys-342 show the greatest conformational change upon ligand binding, suggesting that they participate in stabilizing bound compounds. The known inhibitors camostat, nafamostat, and gabexate ranked in the top 5 compounds, and all adopted poses that demonstrate how the catalytic serine residue positions itself to ultimately react with the inhibitors while the guanidinium functionality forms a salt-bridge interaction with the active site Asp-435. Upon substrate binding in the active site of a serine protease, the active site serine is deprotonated by the histidine in the catalytic triad, turning it into a reactive nucleophile. This is key for the mechanism of covalent inhibition by compounds such as camostat. Thus, we performed a subsequent docking experiment using the three molecules with these charge changes (Figure 4.11). For all three molecules, the deprotonated serine allows for more advantageous position in the active site for nucleophilic attack the carbonyl carbon to occur. For instance, the distance between the serine oxygen and camostat carbonyl carbon decreased from 4.9 Å to 3.5 Å, and the distance for gabexate

decreased from 5.9 Å to 3.3 Å. While nafamostat appears to be further away from the reactive carbon (4.4 Å to 4.8 Å), the molecule flips so that the carbonyl is positioned for reactivity.

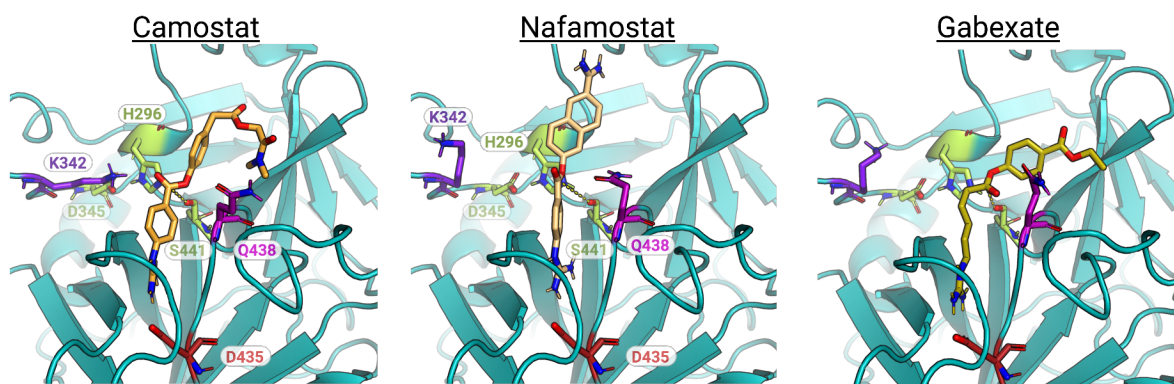


Figure 4.11. Docked poses of camostat, nafamostat, and gabexate in the TMPRSS2 active site with the proton transfer from Ser-441 to His-296 that occurs after substrate binding. Shown in red is D435, which resides at the bottom of the S1 binding pocket, and lime green corresponds to the catalytic triad (His-296, Asp-345, Ser-441). Dashed lines shown here indicate the distance between the catalytic serine oxygen and the carbonyl carbon of each inhibitor (camostat = 3.5 Å, nafamostat = 4.8 Å, gabexate = 3.5 Å). Experiments done by Yujin Wu and Amanda Peiffer.

Several clinically approved drugs emerged as top ranked compounds in the virtual screen, and these were selected for initial testing in the biochemical assay and validation of the platform. Like the covalent inhibitors, pentamidine, propamidine, and debrisoquine all contain a guanidinium moiety and docked into the active site of TMPRSS2 with the positive charge pointing towards Asp-435 (Figure 4.12). Biochemically, all three molecules inhibited TMPRSS2 activity with varied potencies, with debrisoquine being the least potent (Figure 4.12). The docked poses of pentamidine and propamidine show both compounds are positioned to block the active site residues, whereas debrisoquine does not fully span the catalytic triad, which likely correlates to the differences in potency. Pentamidine and propamidine are of similar size to camostat and nafamostat, typical of small molecule inhibitors (>350 MW), while debrisoquine is quite small, at

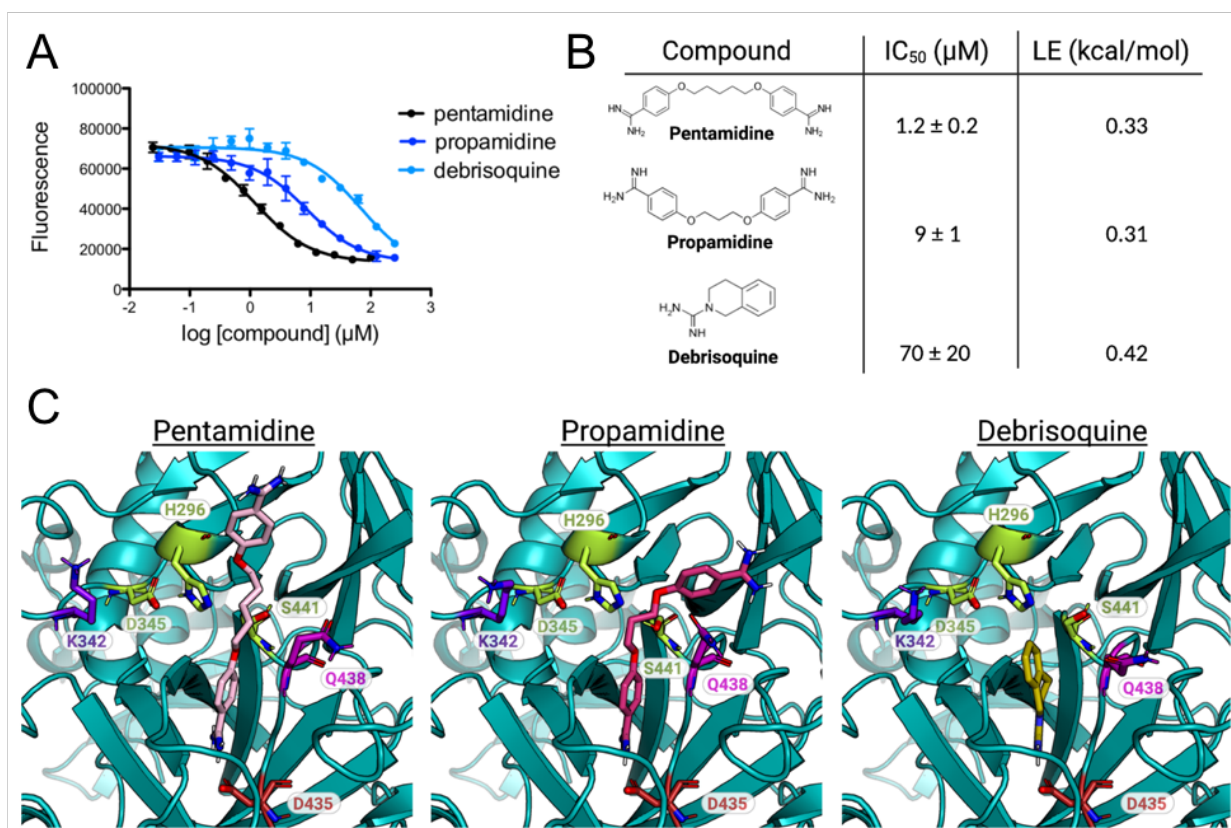


Figure 4.12. Profiling of clinically approved drugs identified as top hits from computational screening. A) Inhibition of TMPRSS2 in biochemical activity assay. Raw inhibition values used to obtain IC₅₀ curves. Data is the average of duplicate experiments conducted in technical triplicate. B) Structures of hits, calculated IC₅₀ values, and ligand efficiencies (LE). C) Docking results for the three drugs identified as hits both in the virtual screen and the *in vitro* assay. All three molecules fit into the active site (Asp-435 at the bottom of the pocket shown in red). Pentamidine and propamidine obstruct access to the catalytic triad (shown in lime green), whereas the fragment debrisoquine only partially reaches these residues. Experiments conducted with Yujin Wu and Amanda Peiffer.

175.2 MW, classifying it as a fragment rather than a small molecule. However, debrisoquine has the greatest calculated ligand efficiency (LE) at 0.42 compared to pentamidine and propamidine, which are 0.33 and 0.31 respectively. Thus, for its small size debrisoquine binds well to TMPRSS2. Furthermore, a LE = 0.42 is suggestive of an excellent starting point for fragment expansion efforts. It is worth noting that each of these molecules also inhibit trypsin activity (Figure 4.13). However, the fragment debrisoquine has modest selectivity for TMPRSS2 over trypsin (~3 fold). As observed with the covalent inhibitors (camostat, nafamostat, and gabexate; Figure 4.11), the docking studies with the noncovalent compounds showed significant structural rearrangements for the two non-catalytic residues, Lys-342 and Gln-438 (Figure 4.12). While Gln-438 is conserved across the serine proteases tested here as well as those used to construct the

homology model (i.e. TMPRSS2, trypsin, hepsin; lysine residue in human plasma kallikrein), Lys-342 only appears in TMPRSS2. In fact, the entire loop where this residue resides greatly differs in length and conformation among the four proteases. Thus, derivatization of a fragment like debrisoquine towards Lys-342 may confer additional selectivity across similar serine proteases.

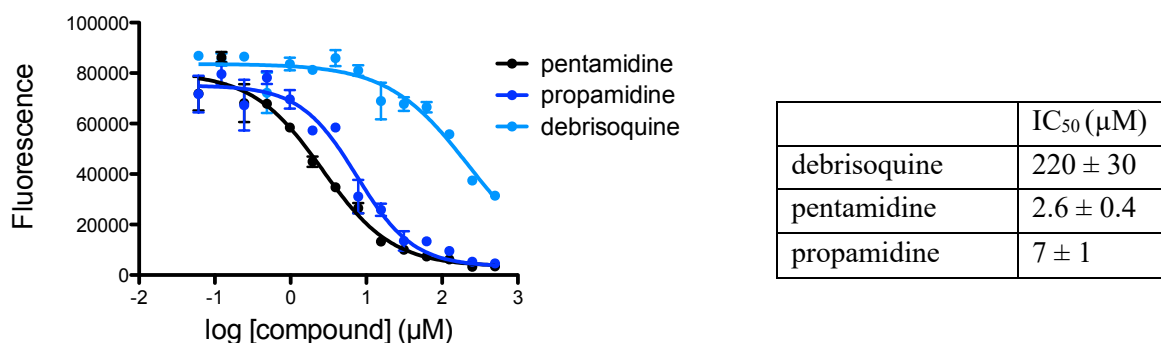


Figure 4.13. Inhibition of Trypsin by pentamidine, propamidine, and debrisoquine. Inhibitors were preincubated with protein for 30 minutes before addition of substrate. Data points represent the average of technical triplicate with error bars indicating the standard deviation (left). Calculated IC₅₀ values (right).

Curated libraries based on guanidinium functionality

As previously discussed, a commonality of the TMPRSS2 inhibitors identified thus far is the presence of one or more guanidinium groups. Thus, an attractive approach to TMPRSS2 inhibitor identification would be to create a TMPRSS2-directed library, comprised of compounds all containing this moiety. Prof. Alison Narayan's lab, through utilizing enzymes in the biocatalytic pathway of saxitoxin synthesis, has the ability to generate compounds that fall into this category.^{52,53} An initial screen was conducted of a small library of compounds made by Narayan lab members using the enzyme Saxitoxin G (library collated by Tessa Epstein). When tested at 10 μM, two compounds showed >15% inhibition of TMPRSS2 activity, indicating that these would be good starting points to further derivatize for increased potency (Figure 4.14).

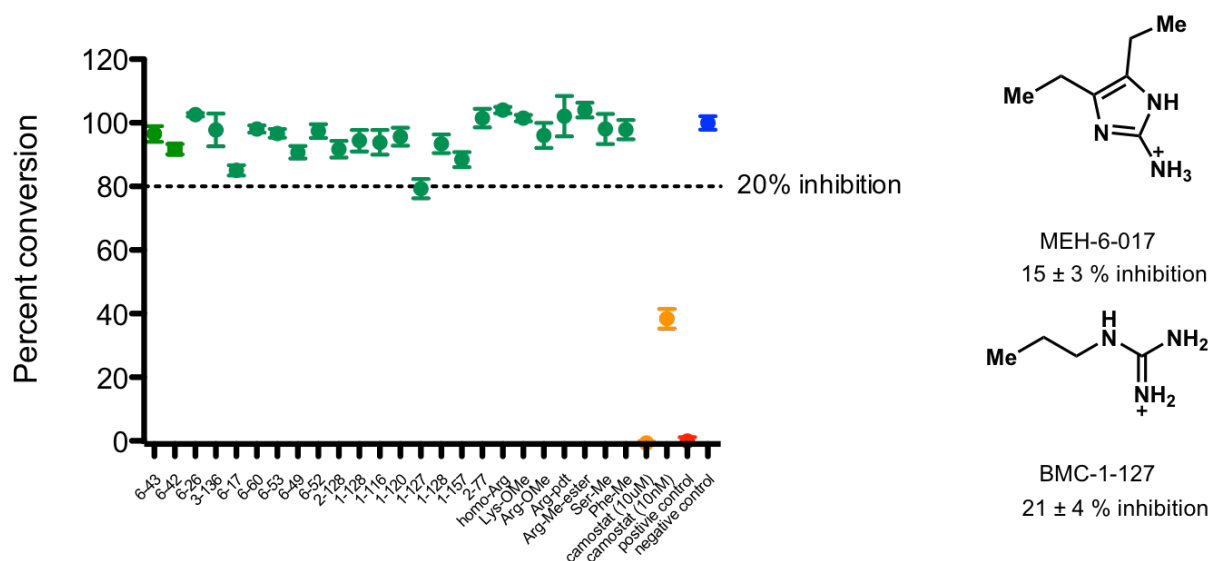


Figure 4.14. SxtG fragment screening results. Fragments were preincubated with protein for 30 minutes before addition of substrate. Single concentration points (10 μ M) of each fragment were tested in triplicate and are represented in green. Camostat, in orange, was tested on the same plate as a control at both 10 μ M and 10 nM, both in triplicate. The positive control, substrate only, representing 0% substrate conversion, was tested in 6 replicate wells. The negative control, substrate and protein, representing 100% substrate conversion, was tested in 12 replicate wells. Data points represent the average with error bars indicating the standard deviation. The two lead fragments with >15% inhibition are shown on the right.

Covalent inhibitors of TMPRSS2

As shown with camostat and related compounds, covalent inhibition can lead to extremely potent small molecule inhibitors. Additionally, covalent chemical probes have many advantages, such as allowing for proteome-wide target identification and increasing specificity among related enzymes.^{54,55} Serine proteases utilize an activated serine residue to cleave substrate amide bonds via a covalent acyl-enzyme intermediate.^{56–58} Thus, this same reaction can be exploited to lead to inhibition of the active site serine via formation of a covalent bond. Importantly, the active site serine is not “activated” until substrate binding occurs, leading to rearrangement of the catalytic triad and subsequent deprotonation of the active site serine by histidine.³⁴ Based on this mechanism, even with a covalent warhead a small molecule must bind at the active site for bond formation to be possible, which may be advantageous for selectivity. Serine proteases in general have been thoroughly studied in the literature, and thus potential chemical warheads that can react with the active site serine are well documented.^{55,59–63} A library of serine focused covalent fragments, 1280 compounds, was purchased from enamine for screening against TMPRSS2.

Functional groups within this library included sulfonyl fluorides, epoxides, β -Propiolactone; γ -butyrolactone; β -lactams, boronic acids, piperidines and piperazines.

For high-throughput screening, the compounds were plated at 25 μ M concentration followed by addition of TMPRSS2 protein. Compound and protein were incubated for 1 hour at room temperature to allow for covalent bond formation. After 1 hour, Boc-QAR-AMC substrate was added and endpoint fluorescence was read after 30min, at which 16% substrate conversion had occurred in the negative control. The Z factor was calculated to be 0.7, indicating an excellent assay. A cutoff of 3 SD above the mean of the negative control (zero percent inhibition), was used to determine hits and with this criteria 45 hit compounds were identified (Figure 4.15, full screening data available through MScreen). Of these 45 compounds, 38 displayed greater than 20% inhibition. A subset of structures from the top 45 compounds are shown in Figure 4.15 The hits contained a variety of warheads, with sulfonyl fluorides being the most common (25), with aldehydes (8) and boronic acids (5) also appearing multiple times. Further chemoinformatic analysis will be necessary to determine if these functionalities are enriched based on the library composition. Additionally, follow up screening for both time and concentration dependence as well as selectivity will be necessary to determine the best scaffolds for further derivatization.

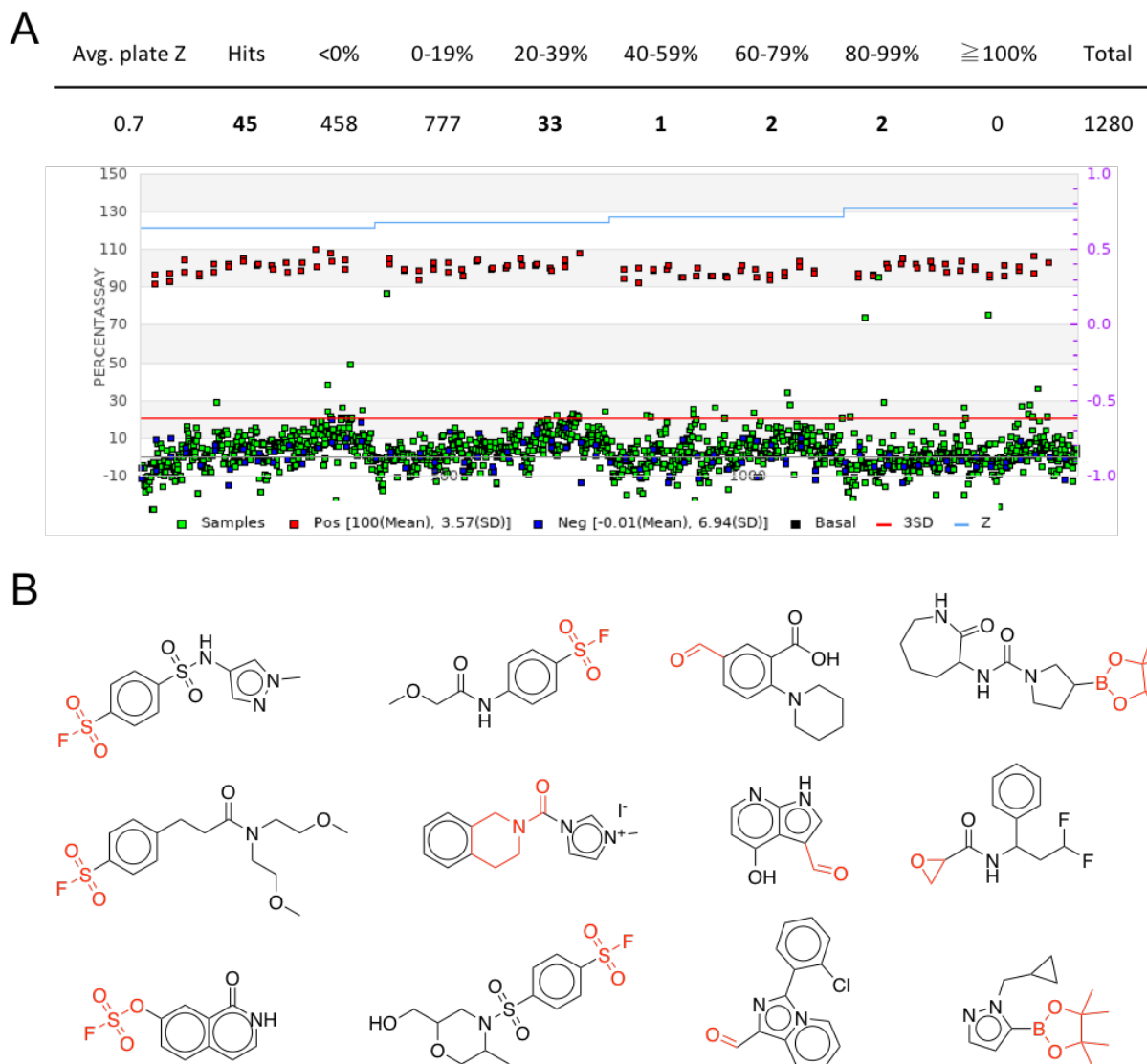


Figure 4.15. Enamine library screening results. A) Campaign view of HTS. Green dots represent compounds, red dots represent the positive controls (substrate only), and blue dots represent the negative controls (protein and substrate). The red line indicated the 3SD cutoff used to determine hits. Figure generated on MScreen. B) Representative structures from the top 45 hits. Compounds were selected to emphasize chemical diversity, and covalent functionality is highlighted in red.

4.4 Conclusions

With new strains of SARS-CoV-2 emerging, and the significance of TMPRSS2 in viral entry for multiple coronaviruses, it is pivotal that we uncover novel strategies to inhibit TMPRSS2 protease activity.^{64,65} However, TMPRSS2 provides obstacles in multiple areas of the discovery pipeline. For instance, the lack of an experimentally solved structure makes docking studies a challenge, relying instead on the use of a homology model developed based on other serine

protease domains. Biochemically, the protease domain of TMPRSS2 has proven difficult to purify and refold recombinantly in *E. coli*, which has stalled many high-throughput screening efforts. This chapter has demonstrated that combined computational and experimental techniques can be used to identify new TMPRSS2 protease inhibitors. Having identified promising scaffolds with high ligand efficiencies, future work will be dedicated towards improving potency and selectivity of these inhibitors. While we have developed an effective expression and purification protocol for the TMPRSS2 peptidase domain, it remains a challenge to obtain high yields of active protein. Thus, the combined virtual and biochemical screening approach presented here is attractive because it enables an initial triage through large compound libraries before testing a smaller number of molecules, more likely to be functionally relevant, in an assay. Current efforts are directed toward testing more hits prioritized from the computational screen for biochemical activity. The results of the two primary fragment screens presented here are currently being tested for concentration response and in secondary screening to identify additional lead scaffolds.

4.5 Materials & Methods

Mass spectrometric analysis of trypsin

Trypsin (Sigma, T9201), 10 μ M, was incubated with compound, 10 μ M, for 30 minutes at room temperature in assay buffer (50 mM Tris, 500 mM NaCl, 0.001% NP-40, pH 7.5) before being subjected to analysis by mass spectrometry using an Agilent QToF LC-MS equipped with a Poroshell 300SB C8 reverse phase column. A 5-100% gradient of acetonitrile with 0.1% formic acid in water to 0.1% formic acid over five minutes was used. Raw data was deconvoluted (intact protein of 20,000-25,000 Da) using BioConfirm software with background subtraction.

Vector design

Plasmids were constructed by Twist Bioscience by inserting the gene for protease domain of TMPRSS2, specifically amino acids 247-492, into the pET28a+ vector using the NdeI_XhoI restriction enzyme cut-sites.

Protein expression

The pET28a+ plasmids containing the TMPRSS2 genes were transformed into BL21(DE3) and plated on LB agar with kanamycin. The bacteria were grown in small 5 mL LB (+ kanamycin)

cultures overnight at 37 °C. The 5 mL starters were used to inoculate 1 L LB (+ kanamycin) cultures, which were grown to OD = 0.8 at 37 °C with shaking at 250 rpm. Expression was induced using 1 mM Isopropyl β -D-1-thiogalactopyranoside (IPTG), which we let grow for 5 hours. The cells were then spun down at 9,500 x g for 15 min. The pellets were collected, flash frozen and stored at -80 °C.

Chemical lysis and denaturing

Before lysing, the cell pellet was first fully thawed until it reached room temperature. Chemical lysis was performed by resuspending the pellet using B-PER reagent (Fisher, PI78243) with lysozyme (Fisher, 90082) and DNase I (Fisher, 90083) following manufacturer's protocols. The cells were then spun at 15,000 x g for 5 minutes, and the cell lysate was collected and saved. The insoluble portion, which contained inclusion bodies of TMPRSS2, was resuspended / washed using lysis buffer containing detergent (50 mM Tris HCl, 150 mM NaCl, 1% Triton X-100, pH 7.5) and then spun at 15,000 x g for 5 minutes. After removing the supernatant, the pellet was washed once more with a lysis buffer that did not contain detergent (50 mM Tris HCl, 150 mM NaCl, pH 7.5) and then spun at 15,000 x g for 5 minutes.

The pellet was resuspended and the inclusion bodies were denatured by adding 20 mL denaturing buffer (8 M urea, 10 mM Tris, 100 mM sodium phosphate, pH 8.0) plus reducing agent (1:1000 BME). Denaturing occurred at room temperature with rotation for at least 30 minutes. The concentration of protein was determined via nanodrop, and additional denaturing buffer was added to reduce the concentration to below 1mg/mL. Denaturing occurred at room temperature on a rotator for at least 30 minutes before being spun down and decanted (20,000 x g, 15 minutes).

Batch binding

Ni-NTA agarose (Qiagen, 30210) was prepared by washing 3 times with binding buffer (8 M urea, 10 mM Tris, 100 mM sodium phosphate, pH 8.0). Denatured protein was added to Ni-NTA resin (750 μ L) and incubated at 4 °C on a rotator for 1.5 hours. Resin was pelleted by centrifugation at 2500 x g and flowthrough was removed. Resin was washed 3 times with wash buffer (8 M urea, 10 mM Tris, 100 mM sodium phosphate, 20 mM imidazole, pH 6.5), followed by addition of elution buffer (8 M urea, 10 mM Tris, 100 mM sodium phosphate, 500 mM

Imidazole, pH 6.5). Eluting was performed on a rotator at 4 °C for 30 minutes. The resin was again pelleted by centrifugation at 2500 x g, and the sample was collected.

Refolding

The denatured sample was diluted 1:100 into refolding buffer (50 mM Tris, 0.5 M arginine, 20 mM CaCl₂, 1 mM EDTA, 100 mM NaCl, 0.01% NP-40, 0.05 mM GSSG, 0.5 mM GSH, pH 7.5) at room temperature using a syringe pump (flow rate 1 mL/min) while allowing the solution to gently stir. The refolding protein was left at 4 °C for 3 days with gentle stirring.

The sample was concentrated 10-fold using Amicon Stirred Cells with 10 kDa Ultrafugation disks (Millipore, UFC801024). Once concentrated, the sample was dialyzed overnight into assay buffer (50 mM Tris, 500 mM NaCl, 0.001% NP-40, pH 7.5) at 4 °C.

Protein gel and silver staining

LDS loading dye was added to protein samples and samples were boiled for 5 minutes at 95 °C. 10uL of each sample was loaded onto a 4-20% mini-PROTEAN TGX gel (BioRad, 4561096) and run at 180V for 45 minutes. Total protein was visualized using a Pierce Silver Staining Kit (Thermo, PI24612) following manufacturer's protocols.

Western Blot

After running gel as described above, protein was transferred to PVDF membrane using a BioRaD TransferBox Turbo following the standard protocols. Membrane was blocked for 1 hour at room temperature using Super Block (Thermo Scientific, 37515). TMPRSS2 antibody (Novus biologicals, NBP1-20984) was added to membrane (1:1000 dilution in Super Block) and incubated overnight at 4C with gentle shaking. After removal of primary antibody and three washes with TBST, HRP conjugated secondary antibody (abcam, ab6741, 1:20,000 in Super Block) was added to membrane and incubated at RT for 1hr with shaking. After removal of secondary antibody with three washes with TBST, HRP substrate (Thermo Scientific, 34095) was added and after 1 minute Western blot was visualized using Chemiluminescence on an Azure Biosystems c600 imager.

Kinetic assays

Assays were conducted on a Molecular Devices Spectramax Spectrophotometer using 96-well plates (Fisher, 12-565-501). Protein was first plated, followed by addition of substrate, Boc-QAR-AMC (Bachem, 4017019.0005) at concentrations to give the indicated final concentration in a 100 μ L volume. After addition of substrate, fluorescence was immediately read (Ex: 380, Em: 460nm), taking measurements every 30 seconds for 20 minutes. Active protein was quantified by titrating in the known active site protease inhibitor FPR-CMK (Haematologic Technologies). To determine the K_M the initial fluorescence data, at less than 10% substrate conversion, was fit to a linear equation and the slope was determined, V_0 . V_0 was plotted vs substrate concentration and the data was fit to the Michaelis Menten equation using GraphPad Prism.

Fluorescence endpoint assays for IC_{50} determination

Assays were conducted in 384 well black plates (Costar, 4514) using an Envision plate reader, ex. filter 350 nm and em. filter 450 nm. The compounds were first plated (10 μ L, at various concentrations) followed by addition of TMPRSS2 protein (8 μ L, 0.5 nM final concentration). After 30 minute incubation (unless otherwise specified) at room temperature, substrate (2 μ L, 2.5 μ M final concentration) was added. At 30 minutes, corresponding to less than 20% substrate cleavage as measured by comparing fluorescence of the negative control to free AMC (Millipore, 257370), fluorescence was read. Wells containing no TMPRSS2 protein (substrate only) served as positive controls. Wells containing no inhibitors (TMPRSS2 and substrate only) served as negative controls. Fluorescence readout was plotted against the log of inhibitor concentration and fit to log(inhibitor) vs response - variable slope equation in GraphPad Prism. Fluorescence endpoint assays with trypsin were conducted utilizing 1 nM protein and 5 μ M substrate

References

- (1) Zhu, N.; Zhang, D.; Wang, W.; Li, X.; Yang, B.; Song, J.; Zhao, X.; Huang, B.; Shi, W.; Lu, R.; Niu, P.; Zhan, F.; Ma, X.; Wang, D.; Xu, W.; Wu, G.; Gao, G. F.; Tan, W. A Novel Coronavirus from Patients with Pneumonia in China, 2019. *N. Engl. J. Med.* **2020**, *382* (8), 727–733. <https://doi.org/10.1056/NEJMoa2001017>.
- (2) Letko, M.; Marzi, A.; Munster, V. Functional Assessment of Cell Entry and Receptor Usage for SARS-CoV-2 and Other Lineage B Betacoronaviruses. *Nat. Microbiol.* **2020**, *5* (4), 562–569. <https://doi.org/10.1038/s41564-020-0688-y>.
- (3) Shang, J.; Wan, Y.; Luo, C.; Ye, G.; Geng, Q.; Auerbach, A.; Li, F. Cell Entry Mechanisms of SARS-CoV-2. *Proc. Natl. Acad. Sci.* **2020**, *117* (21), 11727–11734. <https://doi.org/10.1073/pnas.2003138117>.
- (4) Hoffmann, M.; Kleine-Weber, H.; Schroeder, S.; Krüger, N.; Herrler, T.; Erichsen, S.; Schiergens, T. S.; Herrler, G.; Wu, N.-H.; Nitsche, A.; Müller, M. A.; Drosten, C.; Pöhlmann, S. SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor. *Cell* **2020**. <https://doi.org/10.1016/j.cell.2020.02.052>.
- (5) V'kovski, P.; Kratzel, A.; Steiner, S.; Stalder, H.; Thiel, V. Coronavirus Biology and Replication: Implications for SARS-CoV-2. *Nat. Rev. Microbiol.* **2021**, *19* (3), 155–170. <https://doi.org/10.1038/s41579-020-00468-6>.
- (6) Turner, A. J. ACE2 Cell Biology, Regulation, and Physiological Functions. *Prot. Arm Renin Angiotensin Syst. RAS* **2015**, 185–189. <https://doi.org/10.1016/B978-0-12-801364-9.00025-0>.
- (7) Kim, T. S.; Heinlein, C.; Hackman, R. C.; Nelson, P. S. Phenotypic Analysis of Mice Lacking the Tmprss2-Encoded Protease. *Mol. Cell. Biol.* **2006**, *26* (3), 965–975. <https://doi.org/10.1128/MCB.26.3.965-975.2006>.
- (8) Stopsack, K. H.; Mucci, L. A.; Antonarakis, E. S.; Nelson, P. S.; Kantoff, P. W. TMPRSS2 and COVID-19: Serendipity or Opportunity for Intervention? *Cancer Discov.* **2020**, *10* (6), 779–782. <https://doi.org/10.1158/2159-8290.CD-20-0451>.
- (9) Hatesuer, B.; Bertram, S.; Mehnert, N.; Bahgat, M. M.; Nelson, P. S.; Pöhlman, S.; Schughart, K. Tmprss2 Is Essential for Influenza H1N1 Virus Pathogenesis in Mice. *PLOS Pathog.* **2013**, *9* (12), e1003774. <https://doi.org/10.1371/journal.ppat.1003774>.
- (10) Qiao, Y.; Wang, X.-M.; Mannan, R.; Pitchiaya, S.; Zhang, Y.; Wotring, J. W.; Xiao, L.; Robinson, D. R.; Wu, Y.-M.; Tien, J. C.-Y.; Cao, X.; Simko, S. A.; Apel, I. J.; Bawa, P.; Kregel, S.; Narayanan, S. P.; Raskind, G.; Ellison, S. J.; Parolia, A.; Zelenka-Wang, S.; McMurry, L.; Su, F.; Wang, R.; Cheng, Y.; Delekta, A. D.; Mei, Z.; Pretto, C. D.; Wang, S.; Mehra, R.; Sexton, J. Z.; Chinnaiyan, A. M. Targeting Transcriptional Regulation of SARS-CoV-2 Entry Factors ACE2 and TMPRSS2. *Proc. Natl. Acad. Sci.* **2021**, *118* (1). <https://doi.org/10.1073/pnas.2021450118>.
- (11) Lucas, J. M.; Heinlein, C.; Kim, T.; Hernandez, S. A.; Malik, M. S.; True, L. D.; Morrissey, C.; Corey, E.; Montgomery, B.; Mostaghel, E.; Clegg, N.; Coleman, I.; Brown, C. M.; Schneider, E. L.; Craik, C.; Simon, J. A.; Bedalov, A.; Nelson, P. S. The Androgen-Regulated Protease TMPRSS2 Activates a Proteolytic Cascade Involving Components of the Tumor Microenvironment and Promotes Prostate Cancer Metastasis. *Cancer Discov.* **2014**, *4* (11), 1310–1325. <https://doi.org/10.1158/2159-8290.CD-13-1010>.
- (12) Lucas, J. M.; True, L.; Hawley, S.; Matsumura, M.; Morrissey, C.; Vessella, R.; Nelson, P. S. The Androgen-Regulated Type II Serine Protease TMPRSS2 Is Differentially

- Expressed and Mislocalized in Prostate Adenocarcinoma. *J. Pathol.* **2008**, *215* (2), 118–125. <https://doi.org/10.1002/path.2330>.
- (13) Chen, Y.-W.; Lee, M.-S.; Lucht, A.; Chou, F.-P.; Huang, W.; Havighurst, T. C.; Kim, K.; Wang, J.-K.; Antalis, T. M.; Johnson, M. D.; Lin, C.-Y. TMPRSS2, a Serine Protease Expressed in the Prostate on the Apical Surface of Luminal Epithelial Cells and Released into Semen in Prostatasomes, Is Misregulated in Prostate Cancer Cells. *Am. J. Pathol.* **2010**, *176* (6), 2986–2996. <https://doi.org/10.2353/ajpath.2010.090665>.
 - (14) Ko, C.-J.; Hsu, T.-W.; Wu, S.-R.; Lan, S.-W.; Hsiao, T.-F.; Lin, H.-Y.; Lin, H.-H.; Tu, H.-F.; Lee, C.-F.; Huang, C.-C.; Chen, M.-J. M.; Hsiao, P.-W.; Huang, H.-P.; Lee, M.-S. Inhibition of TMPRSS2 by HAI-2 Reduces Prostate Cancer Cell Invasion and Metastasis. *Oncogene* **2020**, *39* (37), 5950–5963. <https://doi.org/10.1038/s41388-020-01413-w>.
 - (15) Helgeson, B. E.; Tomlins, S. A.; Shah, N.; Laxman, B.; Cao, Q.; Prensner, J. R.; Cao, X.; Singla, N.; Montie, J. E.; Varambally, S.; Mehra, R.; Chinnaiyan, A. M. Characterization of TMPRSS2:ETV5 and SLC45A3:ETV5 Gene Fusions in Prostate Cancer. *Cancer Res.* **2008**, *68* (1), 73–80. <https://doi.org/10.1158/0008-5472.CAN-07-5352>.
 - (16) St. John, J.; Powell, K.; Conley-LaComb, M. K.; Chinni, S. R. TMPRSS2-ERG Fusion Gene Expression in Prostate Tumor Cells and Its Clinical and Biological Significance in Prostate Cancer Progression. *J. Cancer Sci. Ther.* **2012**, *4* (4), 94–101. <https://doi.org/10.4172/1948-5956.1000119>.
 - (17) Paoloni-Giacobino, A.; Chen, H.; Peitsch, M. C.; Rossier, C.; Antonarakis, S. E. Cloning of the TMPRSS2 Gene, Which Encodes a Novel Serine Protease with Transmembrane, LDLRA, and SRCR Domains and Maps to 21q22.3. *Genomics* **1997**, *44* (3), 309–320. <https://doi.org/10.1006/geno.1997.4845>.
 - (18) Xia, X. Domains and Functions of Spike Protein in SARS-Cov-2 in the Context of Vaccine Design. *Viruses* **2021**, *13* (1), 109. <https://doi.org/10.3390/v13010109>.
 - (19) Hoffmann, M.; Kleine-Weber, H.; Pöhlmann, S. A Multibasic Cleavage Site in the Spike Protein of SARS-CoV-2 Is Essential for Infection of Human Lung Cells. *Mol. Cell* **2020**, *78* (4), 779–784.e5. <https://doi.org/10.1016/j.molcel.2020.04.022>.
 - (20) Breining, P.; Frølund, A. L.; Højen, J. F.; Gunst, J. D.; Staerke, N. B.; Saedder, E.; Cases-Thomas, M.; Little, P.; Nielsen, L. P.; Søgaard, O. S.; Kjolby, M. Camostat Mesylate against SARS-CoV-2 and COVID-19—Rationale, Dosing and Safety. *Basic Clin. Pharmacol. Toxicol.* **2021**, *128* (2), 204–212. <https://doi.org/10.1111/bcpt.13533>.
 - (21) Shrimp, J. H.; Kales, S. C.; Sanderson, P. E.; Simeonov, A.; Shen, M.; Hall, M. D. An Enzymatic TMPRSS2 Assay for Assessment of Clinical Candidates and Discovery of Inhibitors as Potential Treatment of COVID-19. *ACS Pharmacol. Transl. Sci.* **2020**, *3* (5), 997–1007. <https://doi.org/10.1021/acsptsci.0c00106>.
 - (22) Hofmann-Winkler, H.; Moerer, O.; Alt-Epping, S.; Bräuer, A.; Büttner, B.; Müller, M.; Fricke, T.; Grundmann, J.; Harnisch, L.-O.; Heise, D.; Kernchen, A.; Pressler, M.; Stephani, C.; Tampe, B.; Kaul, A.; Gärtner, S.; Kramer, S.; Pöhlmann, S.; Winkler, M. S. Camostat Mesylate May Reduce Severity of Coronavirus Disease 2019 Sepsis: A First Observation. *Crit. Care Explor.* **2020**, *2* (11), e0284. <https://doi.org/10.1097/CCE.0000000000000284>.
 - (23) Wang, D.; Li, Z.; Liu, Y. An Overview of the Safety, Clinical Application and Antiviral Research of the COVID-19 Therapeutics. *J. Infect. Public Health* **2020**, *13* (10), 1405–1414. <https://doi.org/10.1016/j.jiph.2020.07.004>.

- (24) Spraggon, G.; Hornsby, M.; Shipway, A.; Tully, D. C.; Bursulaya, B.; Danahay, H.; Harris, J. L.; Lesley, S. A. Active Site Conformational Changes of Prostatic Trypsin Provide a New Mechanism of Protease Regulation by Divalent Cations. *Protein Sci. Publ. Protein Soc.* **2009**, *18* (5), 1081–1094. <https://doi.org/10.1002/pro.118>.
- (25) Sun, W.; Zhang, X.; Cummings, M. D.; Albarazani, K.; Wu, J.; Wang, M.; Alexander, R.; Zhu, B.; Zhang, Y.; Leonard, J.; Lanter, J.; Lenhard, J. Targeting Enteropeptidase with Reversible Covalent Inhibitors to Achieve Metabolic Benefits. *J. Pharmacol. Exp. Ther.* **2020**. <https://doi.org/10.1124/jpet.120.000219>.
- (26) Hempel, T.; Raich, L.; Olsson, S.; P. Azouz, N.; M. Klingler, A.; Hoffmann, M.; Pöhlmann, S.; E. Rothenberg, M.; Noé, F. Molecular Mechanism of Inhibiting the SARS-CoV-2 Cell Entry Facilitator TMPRSS2 with Camostat and Nafamostat. *Chem. Sci.* **2021**, *12* (3), 983–992. <https://doi.org/10.1039/D0SC05064D>.
- (27) Hu, X.; Shrimp, J. H.; Guo, H.; Xu, M.; Chen, C. Z.; Zhu, W.; Zakharov, A. V.; Jain, S.; Shinn, P.; Simeonov, A.; Hall, M. D.; Shen, M. Discovery of TMPRSS2 Inhibitors from Virtual Screening as a Potential Treatment of COVID-19. *ACS Pharmacol. Transl. Sci.* **2021**. <https://doi.org/10.1021/acspsci.0c00221>.
- (28) Meyer, D.; Sielaff, F.; Hammami, M.; Böttcher-Friebertshäuser, E.; Garten, W.; Steinmetzer, T. Identification of the First Synthetic Inhibitors of the Type II Transmembrane Serine Protease TMPRSS2 Suitable for Inhibition of Influenza Virus Activation. *Biochem. J.* **2013**, *452* (2), 331–343. <https://doi.org/10.1042/BJ20130101>.
- (29) Idris, M. O.; Yekeen, A. A.; Alakanse, O. S.; Durojaye, O. A. Computer-Aided Screening for Potential TMPRSS2 Inhibitors: A Combination of Pharmacophore Modeling, Molecular Docking and Molecular Dynamics Simulation Approaches. *J. Biomol. Struct. Dyn.* **2020**, *0* (0), 1–19. <https://doi.org/10.1080/07391102.2020.1792346>.
- (30) Ko, C.-J.; Huang, C.-C.; Lin, H.-Y.; Juan, C.-P.; Lan, S.-W.; Shyu, H.-Y.; Wu, S.-R.; Hsiao, P.-W.; Huang, H.-P.; Shun, C.-T.; Lee, M.-S. Androgen-Induced TMPRSS2 Activates Matrilysin and Promotes Extracellular Matrix Degradation, Prostate Cancer Cell Invasion, Tumor Growth, and Metastasis. *Cancer Res.* **2015**, *75* (14), 2949–2960. <https://doi.org/10.1158/0008-5472.CAN-14-3297>.
- (31) WILSON, S.; GREER, B.; HOOPER, J.; ZIJLSTRA, A.; WALKER, B.; QUIGLEY, J.; HAWTHORNE, S. The Membrane-Anchored Serine Protease, TMPRSS2, Activates PAR-2 in Prostate Cancer Cells. *Biochem. J.* **2005**, *388* (3), 967–972. <https://doi.org/10.1042/BJ20041066>.
- (32) Afar, D. E. H.; Vivanco, I.; Hubert, R. S.; Kuo, J.; Chen, E.; Saffran, D. C.; Raitano, A. B.; Jakobovits, A. Catalytic Cleavage of the Androgen-Regulated TMPRSS2 Protease Results in Its Secretion by Prostate and Prostate Cancer Epithelia. *Cancer Res.* **2001**, *61* (4), 1686–1692.
- (33) Hedstrom, L.; Lin, T.-Y.; Fast, W. Hydrophobic Interactions Control Zymogen Activation in the Trypsin Family of Serine Proteases. *Biochemistry* **1996**, *35* (14), 4515–4523. <https://doi.org/10.1021/bi951928k>.
- (34) Hedstrom, L. Serine Protease Mechanism and Specificity. *Chem. Rev.* **2002**, *102* (12), 4501–4524. <https://doi.org/10.1021/cr000033x>.
- (35) Salvagnini, C.; Gharbi, S.; Boxus, T.; Marchand-Brynaert, J. Synthesis and Evaluation of a Small Library of Graftable Thrombin Inhibitors Derived from (L)-Arginine. *Eur. J. Med. Chem.* **2007**, *42* (1), 37–53. <https://doi.org/10.1016/j.ejmech.2006.07.010>.

- (36) Maitz, M. F.; Sperling, C.; Werner, C. Immobilization of the Irreversible Thrombin Inhibitor D-Phe-Pro-Arg-Chloromethylketone: A Concept for Hemocompatible Surfaces? *J. Biomed. Mater. Res. A* **2010**, *94A* (3), 905–912. <https://doi.org/10.1002/jbm.a.32780>.
- (37) Powers, J. C.; Asgian, J. L.; Ekici, Ö. D.; James, K. E. Irreversible Inhibitors of Serine, Cysteine, and Threonine Proteases. *Chem. Rev.* **2002**, *102* (12), 4639–4750. <https://doi.org/10.1021/cr010182v>.
- (38) Midgley, I.; Hood, A. J.; Proctor, P.; Chasseaud, L. F.; Irons, S. R.; Cheng, K. N.; Brindley, C. J.; Bonn, R. Metabolic Fate of ¹⁴C-Camostat Mesylate in Man, Rat and Dog after Intravenous Administration. *Xenobiotica Fate Foreign Compd. Biol. Syst.* **1994**, *24* (1), 79–92. <https://doi.org/10.3109/00498259409043223>.
- (39) Hoffmann, M.; Hofmann-Winkler, H.; Smith, J. C.; Krüger, N.; Arora, P.; Sørensen, L. K.; Søgaard, O. S.; Hasselstrøm, J. B.; Winkler, M.; Hempel, T.; Raich, L.; Olsson, S.; Danov, O.; Jonigk, D.; Yamazoe, T.; Yamatsuta, K.; Mizuno, H.; Ludwig, S.; Noé, F.; Kjolby, M.; Braun, A.; Sheltzer, J. M.; Pöhlmann, S. Camostat Mesylate Inhibits SARS-CoV-2 Activation by TMPRSS2-Related Proteases and Its Metabolite GBPA Exerts Antiviral Activity. *EBioMedicine* **2021**, 103255. <https://doi.org/10.1016/j.ebiom.2021.103255>.
- (40) Beckh, K.; Göke, B.; Müller, R.; Arnold, R. Elimination of the Low-Molecular Weight Proteinase Inhibitor Camostate (FOY 305) and Its Degradation Products by the Rat Liver. *Res. Exp. Med. (Berl.)* **1987**, *187* (6), 401–406. <https://doi.org/10.1007/BF01852177>.
- (41) Beckh, K.; Weidenbach, H.; Weidenbach, F.; Müller, R.; Adler, G. Hepatic and Pancreatic Metabolism and Biliary Excretion of the Protease Inhibitor Camostat Mesilate. *Int. J. Pancreatol. Off. J. Int. Assoc. Pancreatol.* **1991**, *10* (3–4), 197–205. <https://doi.org/10.1007/BF02924157>.
- (42) Jorgensen, W. L. Efficient Drug Lead Discovery and Optimization. *Acc. Chem. Res.* **2009**, *42* (6), 724–733. <https://doi.org/10.1021/ar800236t>.
- (43) Ghanakota, P.; Carlson, H. A. Driving Structure-Based Drug Discovery through Cosolvent Molecular Dynamics. *J. Med. Chem.* **2016**, *59* (23), 10383–10399. <https://doi.org/10.1021/acs.jmedchem.6b00399>.
- (44) Irwin, J. J.; Shoichet, B. K. Docking Screens for Novel Ligands Conferring New Biology. *J. Med. Chem.* **2016**, *59* (9), 4103–4120. <https://doi.org/10.1021/acs.jmedchem.5b02008>.
- (45) da Silva Rocha, S. F. L.; Olanda, C. G.; Fokoue, H. H.; Sant’Anna, C. M. R. Virtual Screening Techniques in Drug Discovery: Review and Recent Applications. *Curr. Top. Med. Chem.* **2019**, *19* (19), 1751–1767. <https://doi.org/10.2174/1568026619666190816101948>.
- (46) Lin, X.; Li, X.; Lin, X. A Review on Applications of Computational Methods in Drug Screening and Design. *Molecules* **2020**, *25* (6), 1375. <https://doi.org/10.3390/molecules25061375>.
- (47) Waterhouse, A.; Bertoni, M.; Bienert, S.; Studer, G.; Tauriello, G.; Gumienny, R.; Heer, F. T.; de Beer, T. A. P.; Rempfer, C.; Bordoli, L.; Lepore, R.; Schwede, T. SWISS-MODEL: Homology Modelling of Protein Structures and Complexes. *Nucleic Acids Res.* **2018**, *46* (W1), W296–W303. <https://doi.org/10.1093/nar/gky427>.
- (48) Bienert, S.; Waterhouse, A.; de Beer, T. A. P.; Tauriello, G.; Studer, G.; Bordoli, L.; Schwede, T. The SWISS-MODEL Repository—New Features and Functionality. *Nucleic Acids Res.* **2017**, *45* (Database issue), D313–D319. <https://doi.org/10.1093/nar/gkw1132>.

- (49) Guex, N.; Peitsch, M. C.; Schwede, T. Automated Comparative Protein Structure Modeling with SWISS-MODEL and Swiss-PdbViewer: A Historical Perspective. *Electrophoresis* **2009**, *30 Suppl 1*, S162-173. <https://doi.org/10.1002/elps.200900140>.
- (50) Studer, G.; Rempfer, C.; Waterhouse, A. M.; Gumienny, R.; Haas, J.; Schwede, T. QMEANDisCo—Distance Constraints Applied on Model Quality Estimation. *Bioinformatics* **2020**, *36* (6), 1765–1771. <https://doi.org/10.1093/bioinformatics/btz828>.
- (51) Bertoni, M.; Kiefer, F.; Biasini, M.; Bordoli, L.; Schwede, T. Modeling Protein Quaternary Structure of Homo- and Hetero-Oligomers beyond Binary Interactions by Homology. *Sci. Rep.* **2017**, *7* (1), 10480. <https://doi.org/10.1038/s41598-017-09654-8>.
- (52) Chun, S. W.; Hinze, M. E.; Skiba, M. A.; Narayan, A. R. H. Chemistry of a Unique Polyketide-like Synthase. *J. Am. Chem. Soc.* **2018**, *140* (7), 2430–2433. <https://doi.org/10.1021/jacs.7b13297>.
- (53) Lukowski, A. L.; Denomme, N.; Hinze, M. E.; Hall, S.; Isom, L. L.; Narayan, A. R. H. Biocatalytic Detoxification of Paralytic Shellfish Toxins. *ACS Chem. Biol.* **2019**, *14* (5), 941–948. <https://doi.org/10.1021/acscchembio.9b00123>.
- (54) Backus, K. M.; Correia, B. E.; Lum, K. M.; Forli, S.; Horning, B. D.; González-Páez, G. E.; Chatterjee, S.; Lanning, B. R.; Teijaro, J. R.; Olson, A. J.; Wolan, D. W.; Cravatt, B. F. Proteome-Wide Covalent Ligand Discovery in Native Biological Systems. *Nature* **2016**, *534* (7608), 570–574. <https://doi.org/10.1038/nature18002>.
- (55) Sutanto, F.; Konstantinidou, M.; Dömling, A. Covalent Inhibitors: A Rational Approach to Drug Discovery. *RSC Med. Chem.* **2020**, *11* (8), 876–884. <https://doi.org/10.1039/D0MD00154F>.
- (56) Breau, E. J.; Bender, M. L. Direct Spectrophotometric Observation of an Acyl-Enzyme Intermediate in Elastase Catalysis. *Biochem. Biophys. Res. Commun.* **1976**, *70* (1), 235–240. [https://doi.org/10.1016/0006-291X\(76\)91133-5](https://doi.org/10.1016/0006-291X(76)91133-5).
- (57) Steitz, T. A.; Shulman, R. G. Crystallographic and NMR Studies of the Serine Proteases. *Annu. Rev. Biophys. Bioeng.* **1982**, *11*, 419–444. <https://doi.org/10.1146/annurev.bb.11.060182.002223>.
- (58) Kraut, J. Serine Proteases: Structure and Mechanism of Catalysis. *Annu. Rev. Biochem.* **1977**, *46* (1), 331–358. <https://doi.org/10.1146/annurev.bi.46.070177.001555>.
- (59) Zhu, L.; George, S.; Schmidt, M. F.; Al-Gharabli, S. I.; Rademann, J.; Hilgenfeld, R. Peptide Aldehyde Inhibitors Challenge the Substrate Specificity of the SARS-Coronavirus Main Protease. *Antiviral Res.* **2011**, *92* (2), 204–212. <https://doi.org/10.1016/j.antiviral.2011.08.001>.
- (60) Narayanan, A.; H. Jones, L. Sulfonyl Fluorides as Privileged Warheads in Chemical Biology. *Chem. Sci.* **2015**, *6* (5), 2650–2659. <https://doi.org/10.1039/C5SC00408J>.
- (61) Kalgutkar, A. S.; Dalvie, D. K. Drug Discovery for a New Generation of Covalent Drugs. *Expert Opin. Drug Discov.* **2012**, *7* (7), 561–581. <https://doi.org/10.1517/17460441.2012.688744>.
- (62) Liu, Y.; Patricelli, M. P.; Cravatt, B. F. Activity-Based Protein Profiling: The Serine Hydrolases. *Proc. Natl. Acad. Sci. U. S. A.* **1999**, *96* (26), 14694–14699. <https://doi.org/10.1073/pnas.96.26.14694>.
- (63) Diaz, D. B.; Yudin, A. K. The Versatility of Boron in Biological Target Engagement. *Nat. Chem.* **2017**, *9* (8), 731–742. <https://doi.org/10.1038/nchem.2814>.
- (64) Chen, R. E.; Zhang, X.; Case, J. B.; Winkler, E. S.; Liu, Y.; VanBlargan, L. A.; Liu, J.; Errico, J. M.; Xie, X.; Suryadevara, N.; Gilchuk, P.; Zost, S. J.; Tahan, S.; Droit, L.;

- Turner, J. S.; Kim, W.; Schmitz, A. J.; Thapa, M.; Wang, D.; Boon, A. C. M.; Presti, R. M.; O'Halloran, J. A.; Kim, A. H. J.; Deepak, P.; Pinto, D.; Fremont, D. H.; Crowe, J. E.; Corti, D.; Virgin, H. W.; Ellebedy, A. H.; Shi, P.-Y.; Diamond, M. S. Resistance of SARS-CoV-2 Variants to Neutralization by Monoclonal and Serum-Derived Polyclonal Antibodies. *Nat. Med.* **2021**, 1–10. <https://doi.org/10.1038/s41591-021-01294-w>.
- (65) Garcia-Beltran, W. F.; Lam, E. C.; St. Denis, K.; Nitido, A. D.; Garcia, Z. H.; Hauser, B. M.; Feldman, J.; Pavlovic, M. N.; Gregory, D. J.; Poznansky, M. C.; Sigal, A.; Schmidt, A. G.; Iafrate, A. J.; Naranbhai, V.; Balazs, A. B. Multiple SARS-CoV-2 Variants Escape Neutralization by Vaccine-Induced Humoral Immunity. *Cell* **2021**. <https://doi.org/10.1016/j.cell.2021.03.013>.

CHAPTER 5

Conclusions and Future Directions

5.1 Conclusions

The work presented in this dissertation focuses on selective inhibitor discovery for dynamic protein targets. While in the past these targets have been considered “undruggable,” Chapter One includes some of the advancements that have led to identification of successful chemical modulators, bringing the emergence of general principles for drug/probe discovery for dynamic proteins. One such principle is that dynamic proteins often operate through allosteric mechanisms, providing opportunities to discover or develop allosterically active small molecules. For this reason, screening methodologies that are geared toward identification of allosteric inhibitors, rather than relying on serendipitous identification, would help streamline selective inhibitor identification.

In Chapter Two we set out to target the transcriptional coactivator Med25 with the hypothesis that due to the unique structure of its activator binding domain, only found in one other protein to date,¹ Med25 AcID could be selectively targeted with a small molecule. In addition, known binding partners of Med25 are implicated in various disease related pathways, such as viral infection,^{2,3} the unfolded protein response,^{4,5} and metastasis,^{6–8} making it a potentially viable therapeutic target. The natural product norstictic acid (NA), was discovered as a potent and selective inhibitor of AcID PPIs. Utilizing a combination of mutational analysis, mass spectrometry, and inhibitory studies, we discovered that NA covalently modifies lysine residues within a dynamic loop, leading to inhibition in a mixed orthosteric/allosteric mode, and concluded that this mechanism of action contributes to its unique selectivity. Finally, we showed that NA engages with cellular Med25, inhibiting endogenous PPIs as well as decreasing mRNA levels of the Med25 regulated gene MMP-2. The demonstration of NA as a selective allosteric inhibitor of Med25 function complements the discussion in Chapter 1, suggesting that targeting dynamic substructures rather than primary binding surfaces (allostery vs orthostery) affords compounds with better selectivity, and this strategy should be generalizable beyond Med25. Additionally, this work supports the hypothesis that dynamic substructures are key attributes for molecular

recognition in coactivator proteins, complimenting other recent publications coming from the Mapp lab.^{9–12} Our work suggests that Med25 is a potential therapeutic target in metastatic breast cancer, justifying the search for drug-like compounds and further exploration of the phenotypic effects of Med25 PPI inhibition.

In Chapter Three, a duplex fluorescence polarization based assay was developed as a primary screening method for transcriptional PPI inhibitors. We hypothesized that by looking at multiple interactions simultaneously, selective scaffolds could be prioritized immediately in the inhibitor discovery process. A pilot screen conducted with AcID-ETV5 and KIX-MLL illustrated the utility of this method for rapidly classifying both the potency and selectivity of chemical scaffolds for a given target. Additionally, this led to identification of novel inhibitors for both interactions, including fumarprotocetraric acid (FPA), which showed facial selectivity for Med25 PPI inhibition, with more potent inhibition of AcID•ETV5 than AcID•ATF6 α . These results demonstrate that the duplex FP format is a useful high-throughput screening platform for activator•coactivator interactions, allowing selectivity to be assessed during the first step of the inhibitor discovery process. This approach should be applicable to any set of coactivators, if there is demonstrated specificity of the selected TAD tracers. Further, considering selectivity throughout the screening process streamlines discovery of selective small molecule modulators.

In Chapter Four, we hypothesized that targeting the human serine protease TMPRSS2 would be an effective strategy to prevent viral infection. Studies with TMPRSS2 suggest its primary function in healthy cells is redundant,^{13–15} however it aids in the priming of multiple viral coat proteins, including SARS-CoV-2, to facilitate receptor binding and cellular entry.^{16–19} We recognized that there was a lack of selective TMPRSS2 probes in the literature and hypothesized that we would be able to find inhibitors of protease activity using high-throughput screening. An expression and purification method was developed using *E. coli* and produced active protein for use in an optimized biochemical assay. Characterization of the known TMPRSS2 inhibitors camostat, nafamostat, and gabexate confirmed the hypothesis presented by us and others^{20–23} that these molecules have such low IC₅₀ values due to their covalent mechanism of action, which occurs with multiple other serine proteases.^{24,25} Multiple curated libraries were screened, including a subset of molecules identified from in silico docking method using a homology model of TMPRSS2, demonstrating that combining computational and experimental techniques is an effective strategy to identify new inhibitors. Further, for dynamic proteins that prove challenge to express

and purify in high yields, a similar approach could streamline inhibitor discovery because it enables initial triage through large compound libraries before testing a smaller number of molecules, requiring less protein than a typical HTS. Ultimately, we identified promising scaffolds with good affinity and high ligand efficiencies, providing starting points for development of potent and selective antivirals.

5.2 Future Directions

The conclusions from this dissertation have provided key insights towards chemically targeting dynamic proteins and presents them as tractable therapeutic targets. Future efforts will build upon these results and methods presented, leading to an improved model of selectivity, frameworks of molecular recognition, and a better understanding of the targets presented, specifically Med25 and TMPRSS2, in a physiological context.

A few general considerations will be key for future work. First, the *in vitro* studies presented here were conducted with simplified protein constructs, single domains. As allosteric mechanisms are likely significant for selectivity, increasing the complexity of the proteins constructs used in these studies, for example, utilizing full length Med25 and multi-domain constructs of TMPRSS2, will make the *in vitro* models more closely resemble the protein in a native context and improve inhibitor discovery. Second, continued advancement of the screening approaches presented here will help develop enhanced models of selectivity. And finally, utilizing the chemical probes identified in this thesis and beyond to better understand how dynamic targets function in a native context will enhance future inhibitor discovery efforts.

Targeting Med25 with NA and beyond

While NA demonstrated excellent selectivity towards AcID *in vitro* and the ability to engage with Med25 in a cellular context, global selectivity of the molecule remains unknown. To better understand the utility of NA as a Med25 probe and the extent of its selectivity, it would be beneficial to conduct whole cell proteomics. Since NA is a covalent molecule, this affords a unique opportunity where specific signs of protein-NA adduct formation could be identified using MS/MS. Proteomics has successfully led to specificity mapping with other covalent inhibitors and other bioactive small molecules.^{26–28} Alternatively, there are other means to profile selectivity such as drug affinity responsive target stability (DARTS)^{29,30} or thermal proteomic profiling (TPP)^{31,32}

that do not rely on covalent adduct identification. CETSA could also be used to look at thermal stability of hypothetical off targets.^{33,34} We have already observed that NA is able to modify CBP *in vitro* (both the KIX and TAZ1 domain), thus there is the potential that this could occur in cells and a simple experiment would be to look at the effect of NA on endogenous CBP thermal stability. Also, analysis of NA's effect on typical benchmark proteins, such as GAPDH, Tubulin or Actin using CETSA could provide insights into the extent of NA selectivity. If NA global selectivity can be verified, it would also be beneficial to look at how proteins and total RNA levels are effected by NA, that would pinpoint additional gene and protein targets regulated by Med25 AcID. A combination of RNAseq analysis and proteomics would allow this information to be gleaned.

Ultimately, there is still a lack of knowledge with regards to Med25 biology. For example, when analyzing Med25 using Western blot, two potential species are observed: one with a lower MW (~70 kDa) which is expected of full length Med25, and one that is a higher MW (~90 kDa). When looking at cytoplasmic components vs nuclear extracts, the lower MW form is observed primarily in the cytoplasm, whereas the higher MW form is the main form in the nucleus (Figure 5.1).

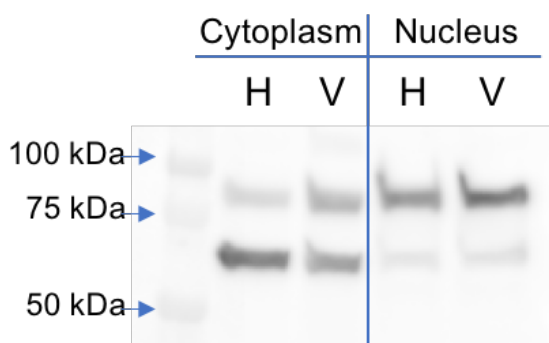


Figure 5.1. Western blot comparing Med25 localization in HeLa (H) and VARI068 (V). Two bands occur when blotting for Med25, one that appears between the 75 and 100 kDa marker on the ladder and one that appears directly below the 75 kDa marker. In both cell lines the higher MW form of Med25 is highly abundant in the nucleus, while the lower MW form is almost exclusively in the cytoplasm.

At this point it is unclear if these two proteins are both Med25. However, as a coactivator Med25 must be located in the nucleus, thus the 92 kDa band likely corresponds to functional Med25. Indeed, when Med25 was first identified via a pulldown experiment with the VP16 TAD, it was called Arc92 because the observed protein was around 92 kDa.^{35,36} Thus, there is a discrepancy between the posited size of Med25 based on the amino acid sequence and that observed *in cellulo*. I hypothesize that this size difference could indicate a PTM occurs (or multiple PTMs) on Med25,

potentially linked to its function as a coactivator. For example, a small ubiquitin like modifier (SUMO), typically 12-14kDa, is close to the size of the observed MW difference.³⁷ An online SUMOylation predictor, based on protein amino acid sequence, identified multiple sites within Med25 of potential modification.³⁸ Interestingly, the two most probable sites are within the AcID domain (Figure 5.2). Additionally, the tool PhosphoSite reports multiple potential PTMs within Med25 from published proteomic datasets (Figure 5.2).³⁹

A

SUMOylation prediction	
(Score ≥ 0.5 for SUMOylation site, Score < 0.5 for non-SUMOylation site)	
Position	Confidence Score
440	0.732692
551	0.652669

B

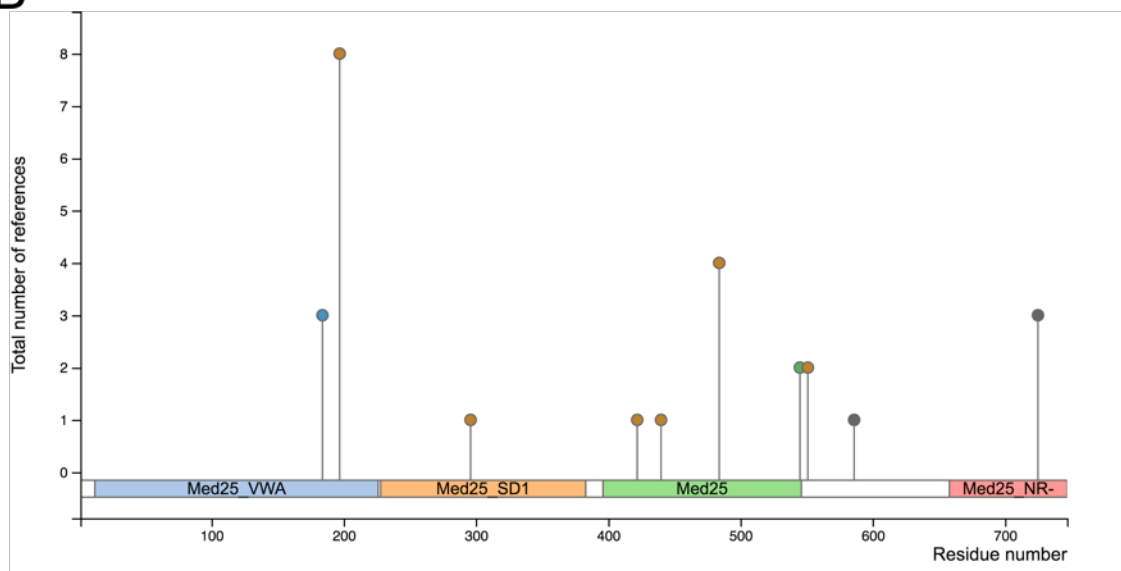


Figure 5.2. Potential PTMs within Med25. Online predicting tools identify potential sites of Med25 modification. Top: SUMOylation predictor shows two sites within AcID that have a high confidence score (>0.5), indicative of a hypothesized sumo modification. Bottom: PhosphoSite reveals the potential for a variety of PTMs at multiple sites. Blue dots indicate phosphorylation, orange dots ubiquitination, gray methylation, and green acetylation. Of note, K197 ubiquitination is the only PTM with >5 reference datasets supporting the modification. K484 ubiquitination has 4.

Ubiquitin, while smaller than SUMO proteins at 8.6 kDa, is often added as a chain (polyubiquitination) and thus could account for the change in observed mass.⁴⁰ Full

characterization of Med25 PTMs using proteomics would be extremely beneficial. Not only would this potentially solve the issue of the multiple species observed via western blot, this could point to novel points of modulation with small molecules. As detailed in Chapter 1, PTMs have been identified as critical to stabilize other coactivator protein-protein interactions, and effective small molecules have been developed to exploit this mechanism.

To date, Med25 has only been studied mechanistically in the context of the AcID domain. There are substantial differences between studying a single domain in isolation compared to the full-length protein. Future work to express and isolate full length Med25, using either insect cells or mammalian cells, will be key to understand if the conclusions made by studying AcID are applicable in a biological context. Preliminary data comparing the thermal stability between the AcID domain purified from *E. coli* and endogenous Med25 (see Appendix B for more details) have already shown discrepancies. There is a difference in the T_m observed for AcID using differential scanning fluorimetry (DSF) ($\sim 70^\circ\text{C}$) compared to full length Med25 using CETSA ($\sim 65^\circ\text{C}$). Most surprisingly, ligands binding to the AcID domain lead to a decrease in the T_m observed using DSF. This includes NA (Figure 5.3), which using CETSA was shown to stabilize Med25 by $>10^\circ\text{C}$.

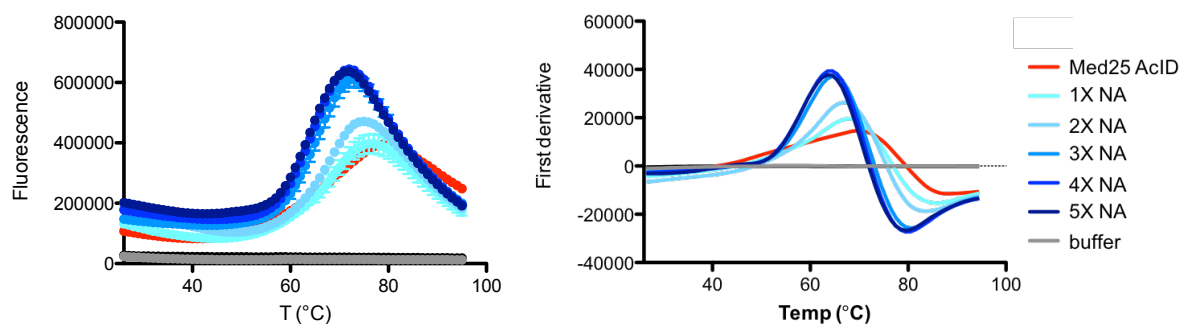


Figure 5.3. NA leads to dose dependent decrease in AcID thermal stability. Melting curves (left) and the first derivative (change in fluorescence over time) of the melting curves (right) obtained using DSF are shown. The maximum change in T_m is observed when concentrations of NA in greater than 3-fold excess of AcID protein are used. Largest ΔT_m observed with NA: -5°C .

Elucidation of chemical scaffolds selective for ABDs

Chapter 3 presented the first generation of a screening method that enabled rapid identification of selective modulators of specific ABD-TAD interactions. At the end of the chapter, this workflow was adapted for use with a disordered domain, IBiD, for which there have been no inhibitor discovery efforts to date. Thus, we plan to use the duplex method to screen a library of

compounds available through the center of chemical genomics to identify the first inhibitors of IBiD.

Further, secondary screening elements are being added to this workflow to enable fine-tuned profiling of chemical scaffold selectivity. DSF has already been adapted for use with multiple coactivator proteins, included AcID and KIX (See Appendix B). We hypothesize that DSF will allow us to distinguish molecular phenotypes that are found to engage ABDs broadly from those that are coactivator-specific. We will further expand the DSF approach presented in Appendix B into an array format utilizing select fluorescent dye reporters to further differentiate inhibitor mechanisms. In addition to SYPRO orange, other dyes used in DSF such as ANS, Nano Orange, 4-(dicyanovinyl)julolidine (DCVJ), Nile Red, bis-ANS, the rotor dye DASPMI (4-(4-(dimethylamino)styryl)-N-methylpyridinium iodide), and thiol-reactive dyes such as BODIPY-Fl cysteine, will be tested (86, 88)(89).⁴¹⁻⁴⁴ Dye sensor arrays have been used for chemical fingerprinting in a variety of contexts,⁴⁵⁻⁴⁷ with application to PPIs and small-molecule•intrinsically disordered protein (IDP) interactions representing an innovative direction. By analyzing the sensor array perturbations recorded across the high-throughput screening (HTS) hits and ABDs from different coactivators in combination with orthogonal binding and structural information, future screening work will identify chemotypes that favor particular binding modes for each ABD. Ultimately this would enable the development of pharmacophore models for selective, pan-selective and non-specific inhibitors of activator-ABD complexes.

Targeting TMPRSS2: profiling mechanism and identifying selective scaffolds

With expression, purification, and activity assay protocols now developed for TMPRSS2 peptidase domain, work in the immediate future will be dedicated to continued identification of novel inhibitor scaffolds with a particular emphasis on selectivity. From the work presented here, we have some natural starting points for novel inhibitor discovery. Since our first efforts in inhibitor discovery for TMPRSS2 were focused on molecules that were already in drug development pipelines or readily available to us, we will begin by screening the remainder of the computationally curated library (~4000 compounds). This will lead to the identification of novel TMPRSS2 inhibitor scaffolds. The preliminary hits from the enamine covalent library will also be pursued in secondary screening and evaluated in a time and concentration dependent manner. It will be necessary to optimize TMPRSS2 for mass spectrometry, as efforts to date have failed likely

due to low concentration of protein as well as high buffer and salt components that lead to ion suppression. Lead covalent fragments will provide great starting points for derivatization and structure activity relationship (SAR) campaigns. Additionally, debrisquinone was identified as a scaffold suitable for elaboration and SAR, with a LE of 0.42. The docking studies with inhibitors presented in Chapter 4 using the TMPRSS2 homology model suggest a potential mechanism to improve these fragments. Ligand binding at the active site led to structural rearrangement of a non-catalytic residue, Lys-342, within a proximal dynamic loop. Lys-342 appears to be unique to TMPRSS2 and not conserved in other similar serine proteases such as human plasma kallikrein, hepsin, and trypsin (Figure 5.4). In fact, the entire loop where this residue resides differs greatly in both length and conformation among the four proteases. Thus, building the fragments presented in this chapter to interact with Lys-342 or engage that unique dynamic loop may confer additional selectivity for TMPRSS2 across serine proteases.^{48,49}

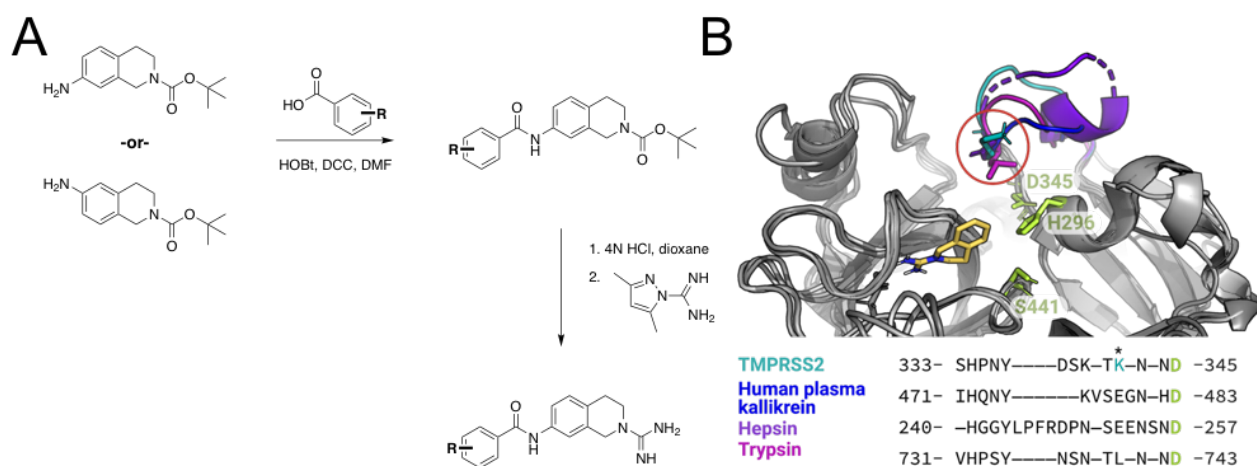


Figure 5.4. Enhancing potency and selectivity of TMPRSS2 targeting fragments. A) Proposed schematic to generate debrisquinone analogs. Starting with 6- and 7-amino-1,2,3,4-tetrahydro-*N*-Boc-isoquinoline will enable a route for rapid synthesis. Coupling a carboxylic acid to the isoquinoline, followed by guanidinylation will afford a novel scaffold for TMPRSS2 inhibitors. B) Comparison of Lys-342 (circled in red) containing loop in TMPRSS2 among similar serine proteases. Structure shown with debrisquinone docked. Loops in the superimposed structures are colored to match the protease name listed next to the amino acid sequence.

There are also a number of published analogs of the pentamidine and propamidine scaffolds that could be further explored for selectivity and potency.⁵⁰ Further profiling of the identified inhibitors, as well as derivative scaffolds, against a panel of representative serine proteases (for example: Factor Xa, Matriptase 2, Plasma Kallikrein, Kallikrein 12, Plasmin, Thrombin A, hepsin) will be conducted, as the work presented in this dissertation only used trypsin as a point of

comparison. This is pivotal, since the inhibitor discovery work flow is currently geared towards identification of active site inhibitors, which are classically more promiscuous, and novel selective TMPRSS2 inhibitors are desired.

Another important future direction will be identification of non-active site allosteric inhibitors. For this work, a larger construct of TMPRSS2, including the SRCR and peptidase domain, will need to be used. Expression protocols in *S. cerevisiae* have been reported, providing a starting point for optimization.⁵¹ Additionally, we currently are working with the Center for Structural Biology to develop and expression protocol for this construct in insect cells. There is evidence that the SRCR domain of TMPRSS2 can influence the activity of the peptidase motif. For example, the peptidase is inactive until cleavage occurs at Arg-255 between the two domains but the SRCR domain remains associated via a disulfide bond.^{52,53} Additionally, there are a number of TMPRSS2 variants with mutations that exhibit altered peptidase activity (Figure 5.5).⁵⁴ Seven mutations appear to be particularly deleterious for TMPRSS2 function (Val160Met, Gly181Arg, Arg240Cys, Gly259Ser, Pro335Leu, Gly432Ala, and Asp435Tyr), three of which are located in the SRCR domain (Figure 5.5).^{54,55} Interrogating the effects of these mutations on TMPRSS2 enzymatic function will be important to understand if there are potential sites that could be exploited for allosteric inhibition with a small molecule. Additionally, development of orthogonal screening methods such as thermal stability assays or fluorescence active site labelling and quenching⁵⁶ will facilitate identification of allosteric inhibitors.

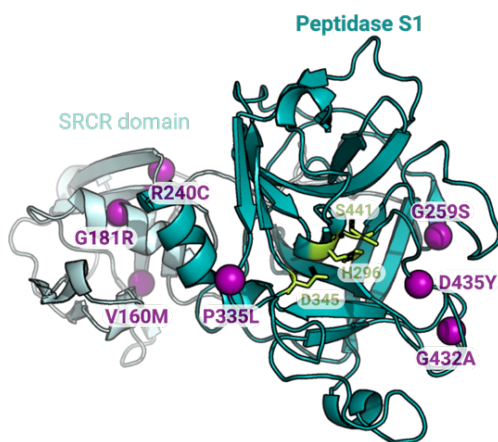


Figure 5.5. Common TMPRSS2 variants known to elicit a decrease in enzymatic function. Residues in magenta affect the peptidase domain via unknown mechanisms and would provide excellent starting points for mutational analysis. The catalytic triad is highlighted in lime green.

The inhibitors presented in this thesis as well as inhibitors emerging from future studies will need to be tested in a biological context. Well-established protocols exist for SARS-CoV-2 pseudotype particle entry assay in Calu-3 cells, thus this would provide a good starting point for testing the ability of the molecules to effect viral infectivity.¹⁶ Additionally, through a collaboration with Dr. Jonathan Sexton and the Center for Drug Repurposing, inhibitors will be assessed for anti-SARS-CoV-2 activity.^{57,58} Other biological experiments will also be conducted for evaluation of TMPRSS2 target engagement. Specifically, others have shown it is possible to identify different components of TMPRSS2 via Western blot,^{16,55,59} thus with thermal shift experiments using TMPRSS2 expressing cells such as Calu-3, we should be able to determine if our inhibitors are stabilizing the active peptidase domain (post-autocleavage), the full-length protein, or if they are preventing autocleavage and activation of protease activity.

5.3 References

- (1) Santamaría, A.; Fernández, P. L.; Farré, X.; Bénédict, P.; Reventós, J.; Morote, J.; Paciucci, R.; Thomson, T. M. PTOV-1, a Novel Protein Overexpressed in Prostate Cancer, Shuttles between the Cytoplasm and the Nucleus and Promotes Entry into the S Phase of the Cell Division Cycle. *Am. J. Pathol.* **2003**, *162* (3), 897–905. [https://doi.org/10.1016/S0002-9440\(10\)63885-0](https://doi.org/10.1016/S0002-9440(10)63885-0).
- (2) Vojnic, E.; Mourão, A.; Seizl, M.; Simon, B.; Wenzel, L.; Larivière, L.; Baumli, S.; Baumgart, K.; Meisterernst, M.; Sattler, M.; Cramer, P. Structure and VP16 Binding of the Mediator Med25 Activator Interaction Domain. *Nat. Struct. Mol. Biol.* **2011**, *18* (4), nsmb.1997. <https://doi.org/10.1038/nsmb.1997>.
- (3) Bontems, F.; Verger, A.; Dewitte, F.; Lens, Z.; Baert, J.-L.; Ferreira, E.; Launoit, Y. de; Sizun, C.; Guittet, E.; Villeret, V.; Monté, D. NMR Structure of the Human Mediator MED25 ACID Domain. *J. Struct. Biol.* **2011**, *174* (1), 245–251. <https://doi.org/10.1016/j.jsb.2010.10.011>.
- (4) Sela, D.; Konkright, J. J.; Chen, L.; Gilmore, J.; Washburn, M. P.; Florens, L.; Conaway, R. C.; Conaway, J. W. Role for Human Mediator Subunit MED25 in Recruitment of Mediator to Promoters by Endoplasmic Reticulum Stress-Responsive Transcription Factor ATF6a. *J. Biol. Chem.* **2013**, *288* (36), 26179–26187. <https://doi.org/10.1074/jbc.M113.496968>.
- (5) Hetz, C.; Zhang, K.; Kaufman, R. J. Mechanisms, Regulation and Functions of the Unfolded Protein Response. *Nat. Rev. Mol. Cell Biol.* **2020**, *21* (8), 421–438. <https://doi.org/10.1038/s41580-020-0250-z>.
- (6) Verger, A.; Baert, J.-L.; Verreman, K.; Dewitte, F.; Ferreira, E.; Lens, Z.; de Launoit, Y.; Villeret, V.; Monté, D. The Mediator Complex Subunit MED25 Is Targeted by the N-Terminal Transactivation Domain of the PEA3 Group Members. *Nucleic Acids Res.* **2013**, *41* (9), 4847–4859. <https://doi.org/10.1093/nar/gkt199>.
- (7) Oh, S.; Shin, S.; Janknecht, R. ETV1, 4 and 5: An Oncogenic Subfamily of ETS Transcription Factors. *Biochim. Biophys. Acta* **2012**, *1826* (1), 1–12. <https://doi.org/10.1016/j.bbcan.2012.02.002>.
- (8) Shepherd, T. G.; Kockeritz, L.; Szrajber, M. R.; Muller, W. J.; Hassell, J. A. The Pea3 Subfamily Ets Genes Are Required for HER2/Neu-Mediated Mammary Oncogenesis. *Curr. Biol. CB* **2001**, *11* (22), 1739–1748. [https://doi.org/10.1016/S0960-9822\(01\)00536-x](https://doi.org/10.1016/S0960-9822(01)00536-x).
- (9) Henley, M. J.; Linhares, B. M.; Morgan, B. S.; Cierpicki, T.; Fierke, C. A.; Mapp, A. K. Unexpected Specificity within Dynamic Transcriptional Protein–Protein Complexes. *Proc. Natl. Acad. Sci.* **2020**, *117* (44), 27346–27353. <https://doi.org/10.1073/pnas.2013244117>.
- (10) Henderson, A. R.; Henley, M. J.; Foster, N. J.; Peiffer, A. L.; Beyersdorf, M. S.; Stanford, K. D.; Sturlis, S. M.; Linhares, B. M.; Hill, Z. B.; Wells, J. A.; Cierpicki, T.; Brooks, C. L.; Fierke, C. A.; Mapp, A. K. Conservation of Coactivator Engagement Mechanism Enables Small-Molecule Allosteric Modulators. *Proc. Natl. Acad. Sci.* **2018**, *115* (36), 8960–8965. <https://doi.org/10.1073/pnas.1806202115>.
- (11) Majmudar, C. Y.; Højfeldt, J. W.; Arevang, C. J.; Pomerantz, W. C.; Gagnon, J. K.; Schultz, P. J.; Cesa, L. C.; Doss, C. H.; Rowe, S. P.; Vásquez, V.; Tamayo-Castillo, G.; Cierpicki, T.; Brooks, C. L.; Sherman, D. H.; Mapp, A. K. Sekikaic Acid and Lobaric

- Acid Target a Dynamic Interface of the Coactivator CBP/P300. *Angew. Chem. Int. Ed.* **2012**, *51* (45), 11258–11262. <https://doi.org/10.1002/anie.201206815>.
- (12) Wang, N.; Majmudar, C. Y.; Pomerantz, W. C.; Gagnon, J. K.; Sadowsky, J. D.; Meagher, J. L.; Johnson, T. K.; Stuckey, J. A.; Brooks, C. L.; Wells, J. A.; Mapp, A. K. Ordering a Dynamic Protein via a Small-Molecule Stabilizer. *J. Am. Chem. Soc.* **2013**, *135* (9), 3363–3366. <https://doi.org/10.1021/ja3122334>.
 - (13) Lambertz, R. L. O.; Gerhauser, I.; Nehlmeier, I.; Leist, S. R.; Kollmus, H.; Pöhlmann, S.; Schughart, K. 2019. Tmprss2 Knock-out Mice Are Resistant to H10 Influenza A Virus Pathogenesis. *J. Gen. Virol.* **100** (7), 1073–1078. <https://doi.org/10.1099/jgv.0.001274>.
 - (14) Hatesuer, B.; Bertram, S.; Mehnert, N.; Bahgat, M. M.; Nelson, P. S.; Pöhlman, S.; Schughart, K. Tmprss2 Is Essential for Influenza H1N1 Virus Pathogenesis in Mice. *PLOS Pathog.* **2013**, *9* (12), e1003774. <https://doi.org/10.1371/journal.ppat.1003774>.
 - (15) Kim, T. S.; Heinlein, C.; Hackman, R. C.; Nelson, P. S. Phenotypic Analysis of Mice Lacking the Tmprss2-Encoded Protease. *Mol. Cell. Biol.* **2006**, *26* (3), 965–975. <https://doi.org/10.1128/MCB.26.3.965-975.2006>.
 - (16) Hoffmann, M.; Kleine-Weber, H.; Schroeder, S.; Krüger, N.; Herrler, T.; Erichsen, S.; Schiergens, T. S.; Herrler, G.; Wu, N.-H.; Nitsche, A.; Müller, M. A.; Drosten, C.; Pöhlmann, S. SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor. *Cell* **2020**. <https://doi.org/10.1016/j.cell.2020.02.052>.
 - (17) Matsuyama, S.; Nagata, N.; Shirato, K.; Kawase, M.; Takeda, M.; Taguchi, F. Efficient Activation of the Severe Acute Respiratory Syndrome Coronavirus Spike Protein by the Transmembrane Protease TMPRSS2. *J. Virol.* **2010**, *84* (24), 12658–12664. <https://doi.org/10.1128/JVI.01542-10>.
 - (18) Shirato, K.; Kawase, M.; Matsuyama, S. Middle East Respiratory Syndrome Coronavirus Infection Mediated by the Transmembrane Serine Protease TMPRSS2. *J. Virol.* **2013**, *87* (23), 12552–12561. <https://doi.org/10.1128/JVI.01890-13>.
 - (19) Böttcher, E.; Matrosovich, T.; Beyerle, M.; Klenk, H.-D.; Garten, W.; Matrosovich, M. Proteolytic Activation of Influenza Viruses by Serine Proteases TMPRSS2 and HAT from Human Airway Epithelium. *J. Virol.* **2006**, *80* (19), 9896–9898. <https://doi.org/10.1128/JVI.01118-06>.
 - (20) Hempel, T.; Raich, L.; Olsson, S.; P. Azouz, N.; M. Klingler, A.; Hoffmann, M.; Pöhlmann, S.; E. Rothenberg, M.; Noé, F. Molecular Mechanism of Inhibiting the SARS-CoV-2 Cell Entry Facilitator TMPRSS2 with Camostat and Nafamostat. *Chem. Sci.* **2021**, *12* (3), 983–992. <https://doi.org/10.1039/D0SC05064D>.
 - (21) Shrimp, J. H.; Kales, S. C.; Sanderson, P. E.; Simeonov, A.; Shen, M.; Hall, M. D. An Enzymatic TMPRSS2 Assay for Assessment of Clinical Candidates and Discovery of Inhibitors as Potential Treatment of COVID-19. *ACS Pharmacol. Transl. Sci.* **2020**, *3* (5), 997–1007. <https://doi.org/10.1021/acspsci.0c00106>.
 - (22) Hoffmann, M.; Hofmann-Winkler, H.; Smith, J. C.; Krüger, N.; Arora, P.; Sørensen, L. K.; Søgaard, O. S.; Hasselstrøm, J. B.; Winkler, M.; Hempel, T.; Raich, L.; Olsson, S.; Danov, O.; Jonigk, D.; Yamazoe, T.; Yamatsuta, K.; Mizuno, H.; Ludwig, S.; Noé, F.; Kjolby, M.; Braun, A.; Sheltzer, J. M.; Pöhlmann, S. Camostat Mesylate Inhibits SARS-CoV-2 Activation by TMPRSS2-Related Proteases and Its Metabolite GBPA Exerts Antiviral Activity. *EBioMedicine* **2021**, 103255. <https://doi.org/10.1016/j.ebiom.2021.103255>.

- (23) Breining, P.; Frølund, A. L.; Højen, J. F.; Gunst, J. D.; Staerke, N. B.; Saedder, E.; Cases-Thomas, M.; Little, P.; Nielsen, L. P.; Søgaard, O. S.; Kjolby, M. Camostat Mesylate against SARS-CoV-2 and COVID-19—Rationale, Dosing and Safety. *Basic Clin. Pharmacol. Toxicol.* **2021**, *128* (2), 204–212. <https://doi.org/10.1111/bcpt.13533>.
- (24) Sun, W.; Zhang, X.; Cummings, M. D.; Albarazani, K.; Wu, J.; Wang, M.; Alexander, R.; Zhu, B.; Zhang, Y.; Leonard, J.; Lanter, J.; Lenhard, J. Targeting Enteropeptidase with Reversible Covalent Inhibitors to Achieve Metabolic Benefits. *J. Pharmacol. Exp. Ther.* **2020**. <https://doi.org/10.1124/jpet.120.000219>.
- (25) Ramjee, M. K.; Henderson, I. M. J.; McLoughlin, S. B.; Padova, A. The Kinetic and Structural Characterization of the Reaction of Nafamostat with Bovine Pancreatic Trypsin. *Thromb. Res.* **2000**, *98* (6), 559–569. [https://doi.org/10.1016/S0049-3848\(00\)00206-1](https://doi.org/10.1016/S0049-3848(00)00206-1).
- (26) Backus, K. M.; Correia, B. E.; Lum, K. M.; Forli, S.; Horning, B. D.; González-Páez, G. E.; Chatterjee, S.; Lanning, B. R.; Teijaro, J. R.; Olson, A. J.; Wolan, D. W.; Cravatt, B. F. Proteome-Wide Covalent Ligand Discovery in Native Biological Systems. *Nature* **2016**, *534* (7608), 570–574. <https://doi.org/10.1038/nature18002>.
- (27) Lanning, B. R.; Whitby, L. R.; Dix, M. M.; Douhan, J.; Gilbert, A. M.; Hett, E. C.; Johnson, T. O.; Joslyn, C.; Kath, J. C.; Niessen, S.; Roberts, L. R.; Schnute, M. E.; Wang, C.; Hulce, J. J.; Wei, B.; Whiteley, L. O.; Hayward, M. M.; Cravatt, B. F. A Road Map to Evaluate the Proteome-Wide Selectivity of Covalent Kinase Inhibitors. *Nat. Chem. Biol.* **2014**, *10* (9), 760–767. <https://doi.org/10.1038/nchembio.1582>.
- (28) Lum, K. M.; Sato, Y.; Beyer, B. A.; Plaisted, W. C.; Anglin, J. L.; Lairson, L. L.; Cravatt, B. F. Mapping Protein Targets of Bioactive Small Molecules Using Lipid-Based Chemical Proteomics. *ACS Chem. Biol.* **2017**, *12* (10), 2671–2681. <https://doi.org/10.1021/acscchembio.7b00581>.
- (29) Lomenick, B.; Jung, G.; Wohlschlegel, J. A.; Huang, J. Target Identification Using Drug Affinity Responsive Target Stability (DARTS). *Curr. Protoc. Chem. Biol.* **2011**, *3* (4), 163–180. <https://doi.org/10.1002/9780470559277.ch110180>.
- (30) Pai, M. Y.; Lomenick, B.; Hwang, H.; Schiestl, R.; McBride, W.; Loo, J. A.; Huang, J. Drug Affinity Responsive Target Stability (DARTS) for Small Molecule Target Identification. *Methods Mol. Biol. Clifton NJ* **2015**, *1263*, 287–298. https://doi.org/10.1007/978-1-4939-2269-7_22.
- (31) Huber, K. V. M.; Olek, K. M.; Müller, A. C.; Tan, C. S. H.; Bennett, K. L.; Colinge, J.; Superti-Furga, G. Proteome-Wide Drug and Metabolite Interaction Mapping by Thermal-Stability Profiling. *Nat. Methods* **2015**, *12* (11), 1055–1057. <https://doi.org/10.1038/nmeth.3590>.
- (32) Mateus, A.; Kurzawa, N.; Becher, I.; Sridharan, S.; Helm, D.; Stein, F.; Typas, A.; Savitski, M. M. Thermal Proteome Profiling for Interrogating Protein Interactions. *Mol. Syst. Biol.* **2020**, *16* (3), e9232. <https://doi.org/10.15252/msb.20199232>.
- (33) Molina, D. M.; Jafari, R.; Ignatushchenko, M.; Seki, T.; Larsson, E. A.; Dan, C.; Sreekumar, L.; Cao, Y.; Nordlund, P. Monitoring Drug Target Engagement in Cells and Tissues Using the Cellular Thermal Shift Assay. *Science* **2013**, *341* (6141), 84–87. <https://doi.org/10.1126/science.1233606>.
- (34) Jafari, R.; Almqvist, H.; Axelsson, H.; Ignatushchenko, M.; Lundbäck, T.; Nordlund, P.; Molina, D. M. The Cellular Thermal Shift Assay for Evaluating Drug Target Interactions in Cells. *Nat. Protoc.* **2014**, *9* (9), 2100–2122. <https://doi.org/10.1038/nprot.2014.138>.

- (35) Mittler, G.; Stühler, T.; Santolin, L.; Uhlmann, T.; Kremmer, E.; Lottspeich, F.; Berti, L.; Meisterernst, M. A Novel Docking Site on Mediator Is Critical for Activation by VP16 in Mammalian Cells. *EMBO J.* **2003**, *22* (24), 6494–6504. <https://doi.org/10.1093/emboj/cdg619>.
- (36) Yang, F.; DeBeaumont, R.; Zhou, S.; Näär, A. M. The Activator-Recruited Cofactor/Mediator Coactivator Subunit ARC92 Is a Functionally Important Target of the VP16 Transcriptional Activator. *Proc. Natl. Acad. Sci.* **2004**, *101* (8), 2339–2344. <https://doi.org/10.1073/pnas.0308676100>.
- (37) Han, Z.-J.; Feng, Y.-H.; Gu, B.-H.; Li, Y.-M.; Chen, H. The Post-Translational Modification, SUMOylation, and Cancer (Review). *Int. J. Oncol.* **2018**, *52* (4), 1081–1094. <https://doi.org/10.3892/ijo.2018.4280>.
- (38) Chang, C.-C.; Tung, C.-H.; Chen, C.-W.; Tu, C.-H.; Chu, Y.-W. SUMOgo: Prediction of Sumoylation Sites on Lysines by Motif Screening Models and the Effects of Various Post-Translational Modifications. *Sci. Rep.* **2018**, *8*. <https://doi.org/10.1038/s41598-018-33951-5>.
- (39) Hornbeck, P. V.; Zhang, B.; Murray, B.; Kornhauser, J. M.; Latham, V.; Skrzypek, E. PhosphoSitePlus, 2014: Mutations, PTMs and Recalibrations. *Nucleic Acids Res.* **2015**, *43* (Database issue), D512–520. <https://doi.org/10.1093/nar/gku1267>.
- (40) Komander, D.; Rape, M. The Ubiquitin Code. *Annu. Rev. Biochem.* **2012**, *81* (1), 203–229. <https://doi.org/10.1146/annurev-biochem-060310-170328>.
- (41) Gao, K.; Oerlemans, R.; Groves, M. R. Theory and Applications of Differential Scanning Fluorimetry in Early-Stage Drug Discovery. *Biophys. Rev.* **2020**, *12* (1), 85–104. <https://doi.org/10.1007/s12551-020-00619-2>.
- (42) DeSantis, K.; Reed, A.; Rahhal, R.; Reinking, J. Use of Differential Scanning Fluorimetry as a High-Throughput Assay to Identify Nuclear Receptor Ligands. *Nucl. Recept. Signal.* **2012**, *10* (1), nrs.10002. <https://doi.org/10.1621/nrs.10002>.
- (43) Kirley, T. L.; Norman, A. B.; Wetzel, H. N. A Novel Differential Scanning Fluorimetry Analysis of a Humanized Anti-Cocaine MAb and Its Ligand Binding Characteristics. *J. Immunol. Methods* **2020**, *476*, 112676. <https://doi.org/10.1016/j.jim.2019.112676>.
- (44) Bai, N.; Roder, H.; Dickson, A.; Karanicolas, J. Isothermal Analysis of ThermoFluor Data Can Readily Provide Quantitative Binding Affinities. *Sci. Rep.* **2019**, *9* (1), 2650. <https://doi.org/10.1038/s41598-018-37072-x>.
- (45) Wu, J.; Kwon, B.; Liu, W.; Anslyn, E. V.; Wang, P.; Kim, J. S. Chromogenic/Fluorogenic Ensemble Chemosensing Systems. *Chem. Rev.* **2015**, *115* (15), 7893–7943. <https://doi.org/10.1021/cr500553d>.
- (46) Kangas, M. J.; Burks, R. M.; Atwater, J.; Lukowicz, R. M.; Garver, B.; Holmes, A. E. Comparative Chemometric Analysis for Classification of Acids and Bases via a Colorimetric Sensor Array. *J. Chemom.* **2018**, *32* (2), e2961. <https://doi.org/10.1002/cem.2961>.
- (47) Kangas, M.; Wilson, C.; Burks, R.; Atwater, J.; Lukowicz, R.; Garver, B.; Mayer, M.; Havenridge, S.; Holmes, A. An Improved Comparison of Chemometric Analyses for the Identification of Acids and Bases With Colorimetric Sensor Arrays. *Int. J. Chem.* **2018**, *10* (2), p36. <https://doi.org/10.5539/ijc.v10n2p36>.
- (48) Faucher, F.; Bennett, J. M.; Bogoy, M.; Lovell, S. Strategies for Tuning the Selectivity of Chemical Probes That Target Serine Hydrolases. *Cell Chem. Biol.* **2020**, *27* (8), 937–952. <https://doi.org/10.1016/j.chembiol.2020.07.008>.

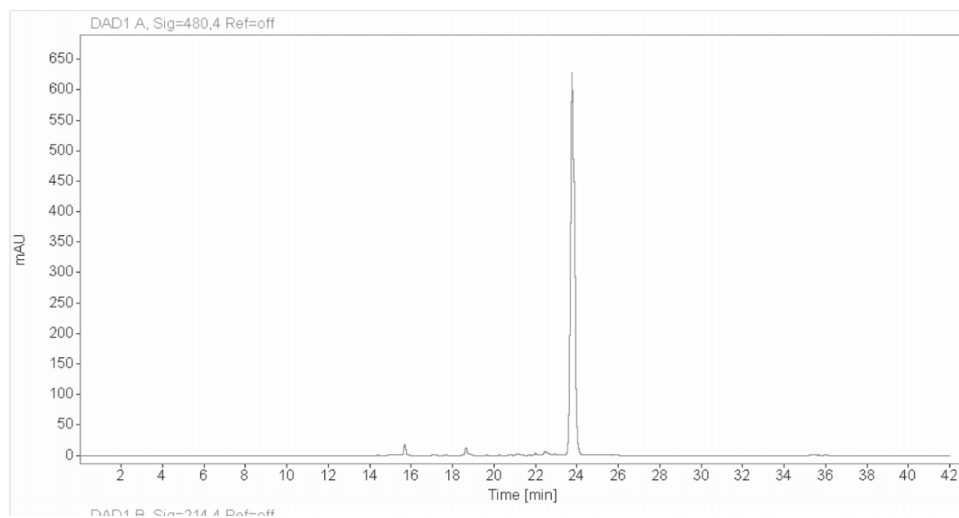
- (49) Chen, S.; Yim, J. J.; Boggyo, M. Synthetic and Biological Approaches to Map Substrate Specificities of Proteases. *Biol. Chem.* **2019**, *401* (1), 165–182. <https://doi.org/10.1515/hsz-2019-0332>.
- (50) Parkesh, R.; Childs-Disney, J. L.; Nakamori, M.; Kumar, A.; Wang, E.; Wang, T.; Hoskins, J.; Tran, T.; Housman, D.; Thornton, C. A.; Disney, M. D. Design of a Bioactive Small Molecule That Targets the Myotonic Dystrophy Type 1 RNA via an RNA Motif-Ligand Database and Chemical Similarity Searching. *J. Am. Chem. Soc.* **2012**, *134* (10), 4731–4742. <https://doi.org/10.1021/ja210088v>.
- (51) Lucas, J. M.; Heinlein, C.; Kim, T.; Hernandez, S. A.; Malik, M. S.; True, L. D.; Morrissey, C.; Corey, E.; Montgomery, B.; Mostaghel, E.; Clegg, N.; Coleman, I.; Brown, C. M.; Schneider, E. L.; Craik, C.; Simon, J. A.; Bedalov, A.; Nelson, P. S. The Androgen-Regulated Protease TMPRSS2 Activates a Proteolytic Cascade Involving Components of the Tumor Microenvironment and Promotes Prostate Cancer Metastasis. *Cancer Discov.* **2014**, *4* (11), 1310–1325. <https://doi.org/10.1158/2159-8290.CD-13-1010>.
- (52) Paoloni-Giacobino, A.; Chen, H.; Peitsch, M. C.; Rossier, C.; Antonarakis, S. E. Cloning of the TMPRSS2 Gene, Which Encodes a Novel Serine Protease with Transmembrane, LDLRA, and SRCR Domains and Maps to 21q22.3. *Genomics* **1997**, *44* (3), 309–320. <https://doi.org/10.1006/geno.1997.4845>.
- (53) Afar, D. E. H.; Vivanco, I.; Hubert, R. S.; Kuo, J.; Chen, E.; Saffran, D. C.; Raitano, A. B.; Jakobovits, A. Catalytic Cleavage of the Androgen-Regulated TMPRSS2 Protease Results in Its Secretion by Prostate and Prostate Cancer Epithelia. *Cancer Res.* **2001**, *61* (4), 1686–1692.
- (54) Hou, Y.; Zhao, J.; Martin, W.; Kallianpur, A.; Chung, M. K.; Jehi, L.; Sharifi, N.; Erzurum, S.; Eng, C.; Cheng, F. New Insights into Genetic Susceptibility of COVID-19: An ACE2 and TMPRSS2 Polymorphism Analysis. *BMC Med.* **2020**, *18* (1), 216. <https://doi.org/10.1186/s12916-020-01673-z>.
- (55) David, A.; Parkinson, N.; Peacock, T. P.; Pairo-Castineira, E.; Khanna, T.; Cobat, A.; Tenesa, A.; Sancho-Shimizu, V.; Investigators, G.; Investigators, I.; Casanova, J.-L.; Abel, L.; Barclay, W. S.; Baillie, J. K.; Sternberg, M. J. A Common TMPRSS2 Variant Protects against Severe COVID-19. *medRxiv* **2021**, 2021.03.04.21252931. <https://doi.org/10.1101/2021.03.04.21252931>.
- (56) Verespy Iii, S.; Mehta, A. Y.; Afosah, D.; Al-Horani, R. A.; Desai, U. R. Allosteric Partial Inhibition of Monomeric Proteases. Sulfated Coumarins Induce Regulation, Not Just Inhibition, of Thrombin. *Sci. Rep.* **2016**, *6* (1), 24043. <https://doi.org/10.1038/srep24043>.
- (57) Mirabelli, C.; Wotring, J. W.; Zhang, C. J.; McCarty, S. M.; Fursmidt, R.; Frum, T.; Kadambi, N. S.; Amin, A. T.; O'Meara, T. R.; Pretto, C. D.; Spence, J. R.; Huang, J.; Alysandratos, K. D.; Kotton, D. N.; Handelman, S. K.; Wobus, C. E.; Weatherwax, K. J.; Mashour, G. A.; O'Meara, M. J.; Sexton, J. Z. Morphological Cell Profiling of SARS-CoV-2 Infection Identifies Drug Repurposing Candidates for COVID-19. *bioRxiv* **2020**, 2020.05.27.117184. <https://doi.org/10.1101/2020.05.27.117184>.
- (58) Qiao, Y.; Wang, X.-M.; Mannan, R.; Pitchaiya, S.; Zhang, Y.; Wotring, J. W.; Xiao, L.; Robinson, D. R.; Wu, Y.-M.; Tien, J. C.-Y.; Cao, X.; Simko, S. A.; Apel, I. J.; Bawa, P.; Kregel, S.; Narayanan, S. P.; Raskind, G.; Ellison, S. J.; Parolia, A.; Zelenka-Wang, S.; McMurry, L.; Su, F.; Wang, R.; Cheng, Y.; Delekta, A. D.; Mei, Z.; Pretto, C. D.; Wang, S.; Mehra, R.; Sexton, J. Z.; Chinnaiyan, A. M. Targeting Transcriptional Regulation of

- SARS-CoV-2 Entry Factors ACE2 and TMPRSS2. *Proc. Natl. Acad. Sci.* **2021**, *118* (1). <https://doi.org/10.1073/pnas.2021450118>.
- (59) Chen, Y.-W.; Lee, M.-S.; Lucht, A.; Chou, F.-P.; Huang, W.; Havighurst, T. C.; Kim, K.; Wang, J.-K.; Antalis, T. M.; Johnson, M. D.; Lin, C.-Y. TMPRSS2, a Serine Protease Expressed in the Prostate on the Apical Surface of Luminal Epithelial Cells and Released into Semen in Prostatomes, Is Misregulated in Prostate Cancer Cells. *Am. J. Pathol.* **2010**, *176* (6), 2986–2996. <https://doi.org/10.2353/ajpath.2010.090665>.

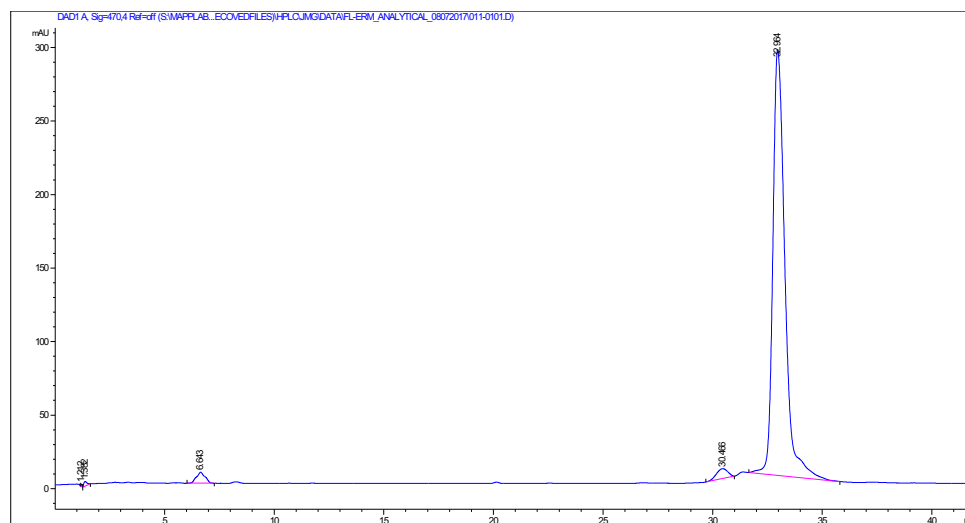
APPENDIX A

Characterization of Synthesized Peptides

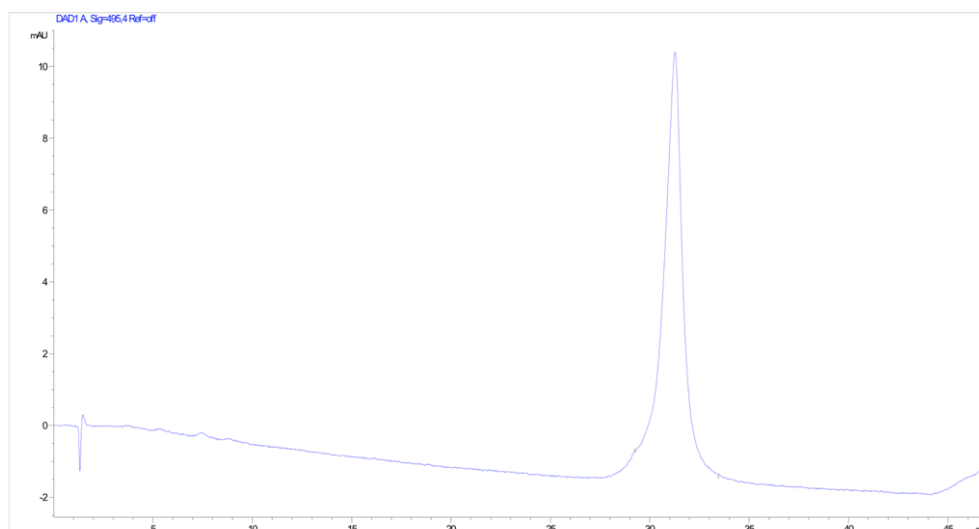
This appendix contains the analytical HPLC chromatograms of transcriptional activation domain peptides used in various experiments throughout this dissertation.



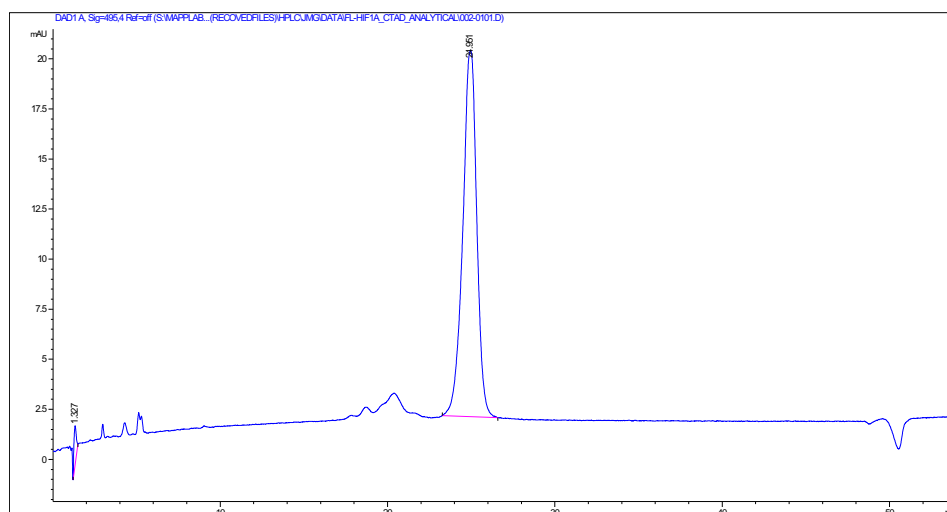
Analytical HPLC UV/Vis trace of FITC-VP16(465-490) monitored at 480 nm. Obtained by Dr. Steve Sturlis.



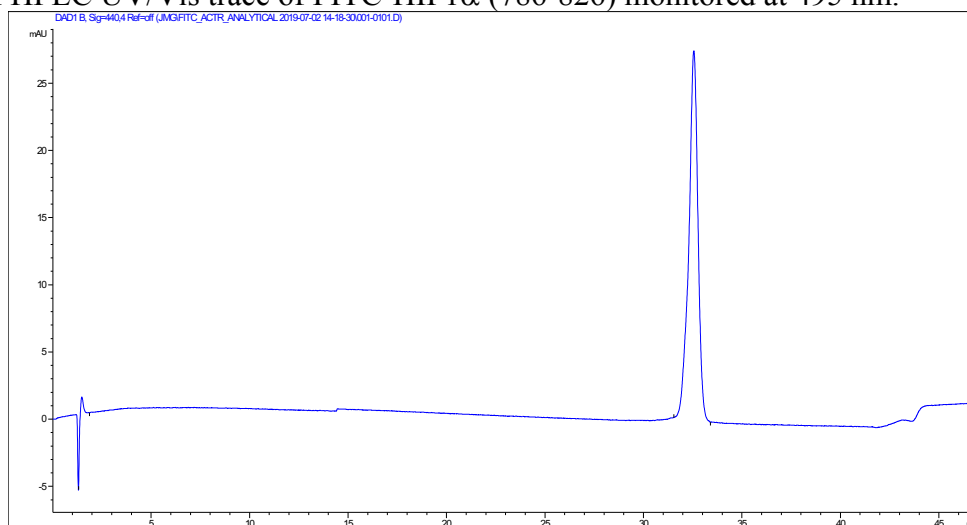
Analytical HPLC UV/Vis trace of FITC-ETV5 (38-68) monitored at 480 nm.



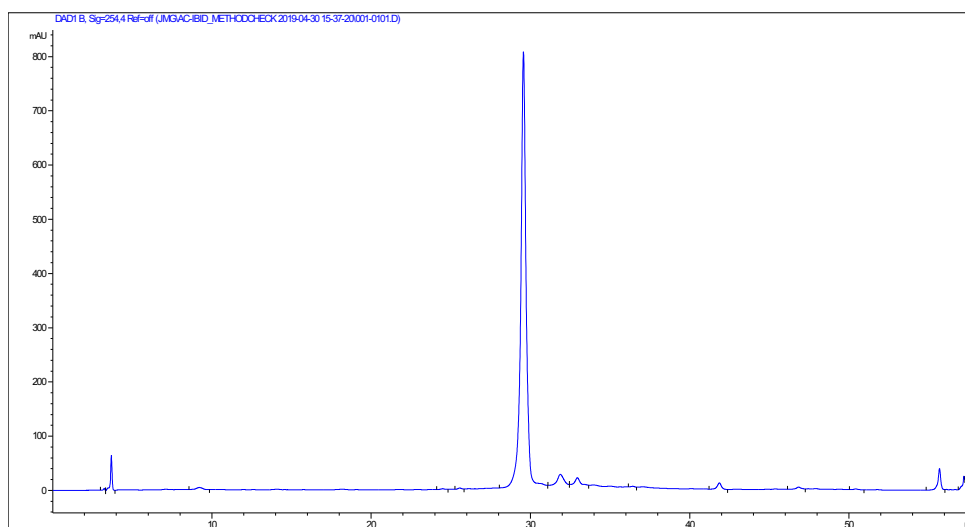
Analytical HPLC UV/Vis trace of FITC-ATF6 α (40-66) monitored at 495 nm.



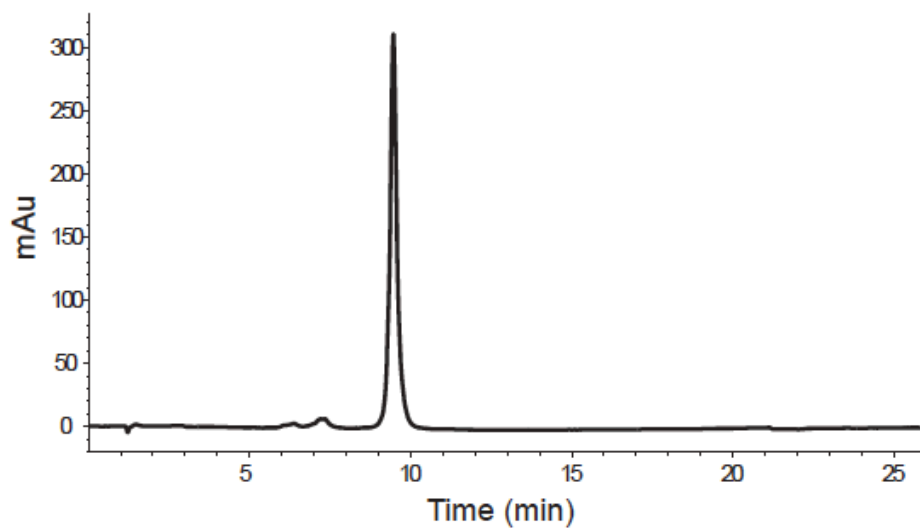
Analytical HPLC UV/Vis trace of FITC-HIF1 α (786-826) monitored at 495 nm.



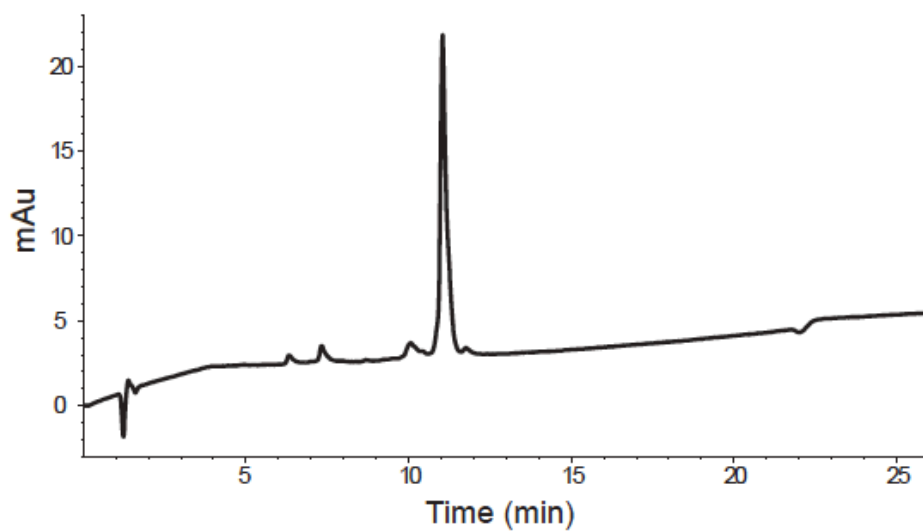
Analytical HPLC UV/Vis trace of FITC-ACTR (1041-1088) monitored at 440 nm.



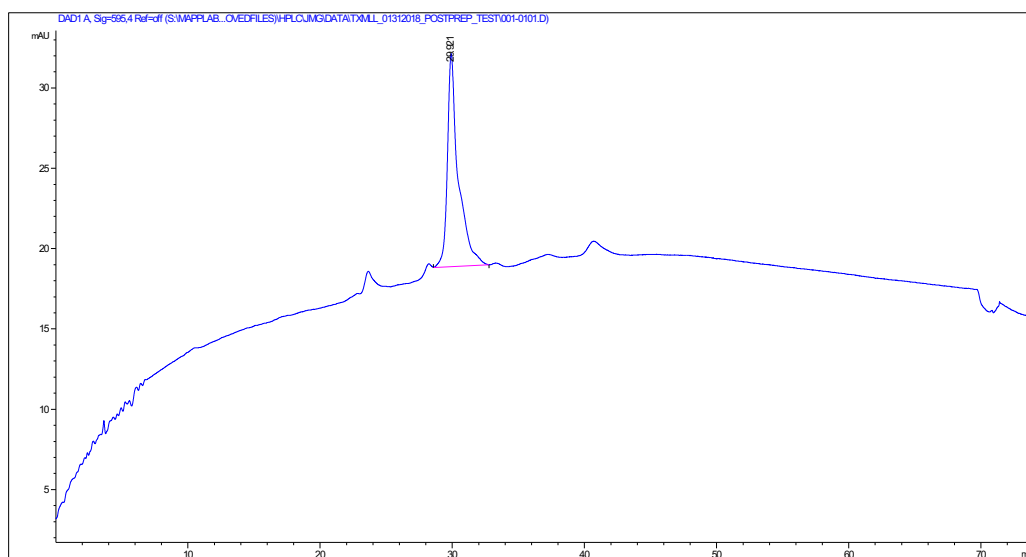
Analytical HPLC UV/Vis trace of Ac-IBiD (2063–2111) monitored at 254 nm.



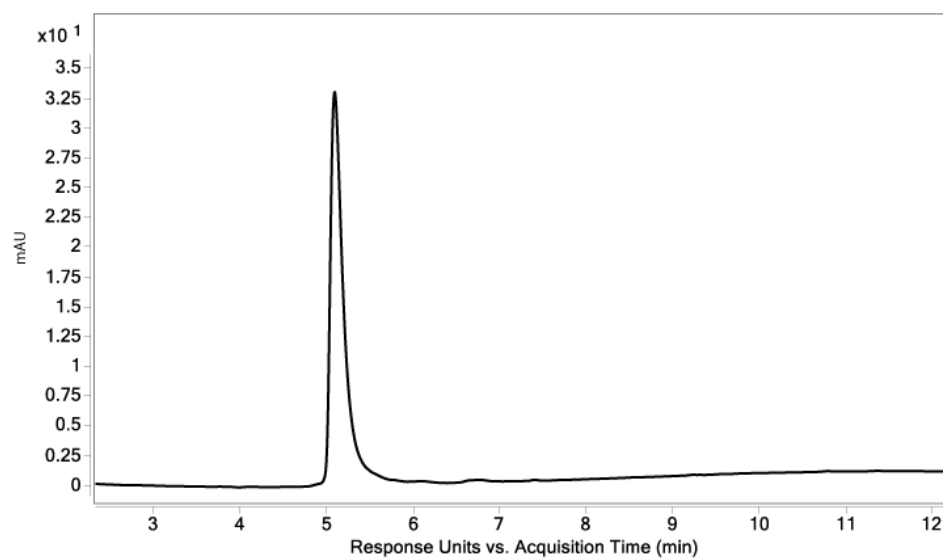
Analytical HPLC UV/Vis trace of FITC-Myb (291-316) monitored at 425 nm. Obtained by Dr. Stephen Joy



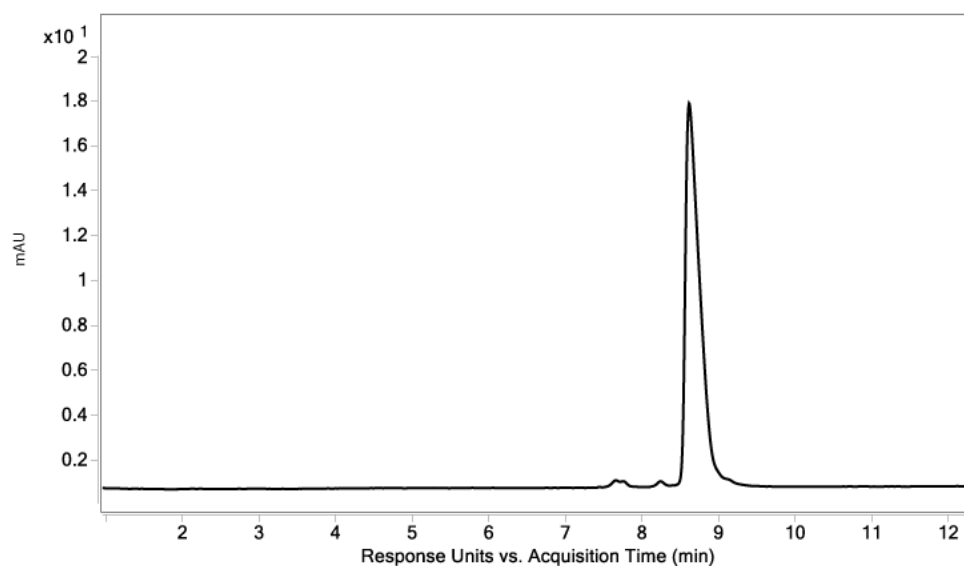
Analytical HPLC UV/Vis trace of FITC-MLL(840-858) monitored at 425 nm. Obtained by Dr. Stephen Joy.



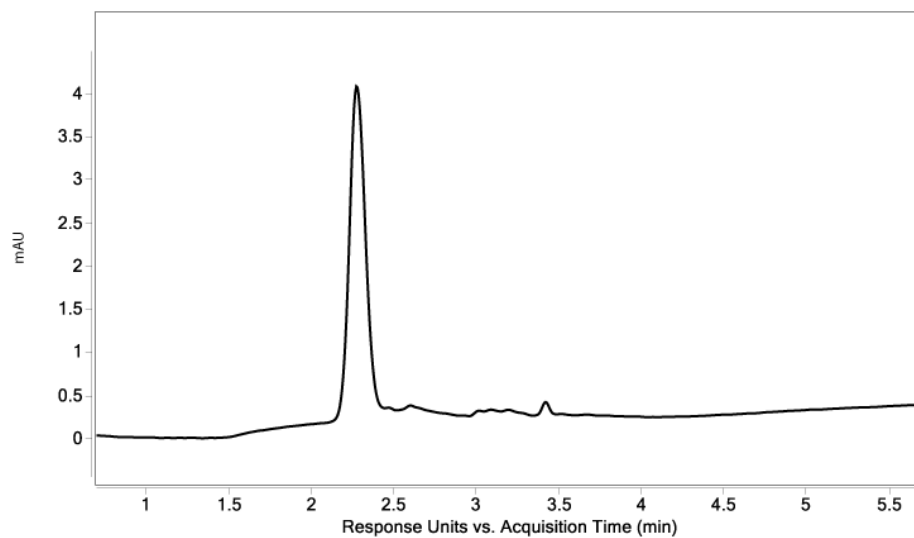
Analytical HPLC UV/Vis trace of TR-MLL(840-858) monitored at 595 nm.



Analytical HPLC UV/Vis trace of Ac-Myb(291-396) monitored at 280 nm. Obtained by Dr. Stephen Joy.



Analytical HPLC UV/Vis trace of Ac-MLL(840-858) monitored at 280 nm. Obtained by Dr. Stephen Joy.



Analytical HPLC UV/Vis trace of Ac-pKID(119-147) monitored at 280 nm. Obtained by Dr. Stephen Joy.

APPENDIX B

Differential Scanning Fluorimetry to Profile Dynamic Coactivators and Their Complexes

This appendix summarizes preliminary data using differential scanning fluorimetry to study Med25 AcID and CBP KIX, providing observations and hypotheses for future work.

B.1 Abstract

Differential scanning fluorimetry (DSF) is a classic technique that allows for the effect of small and large molecule ligands on thermal stability of protein targets to be measured. We hypothesize that this technique will be uniquely suited to identify allosteric modulators of dynamic protein complexes. Towards that end, DSF is used to characterize the effects of ligands, both small molecules and peptides, on the thermal stability of coactivator proteins.

B.2 Introduction

Differential scanning fluorimetry (DSF) is a robust, rapid, and high-content technique that measures ligand binding to a protein of interest and the resulting effect on thermal stability.¹⁻³ The unfolding of a protein is measured by subjecting the sample to a temperature gradient in the presence of a solvatochromic dye, which has affinity to specific parts of the protein that are revealed as the protein unfolds (Figure B.1). The most common dyes used in these experiments have affinity for hydrophobic parts of the protein, such as bis-ANS or SYPRO orange (Figure B.1).⁴⁻⁷

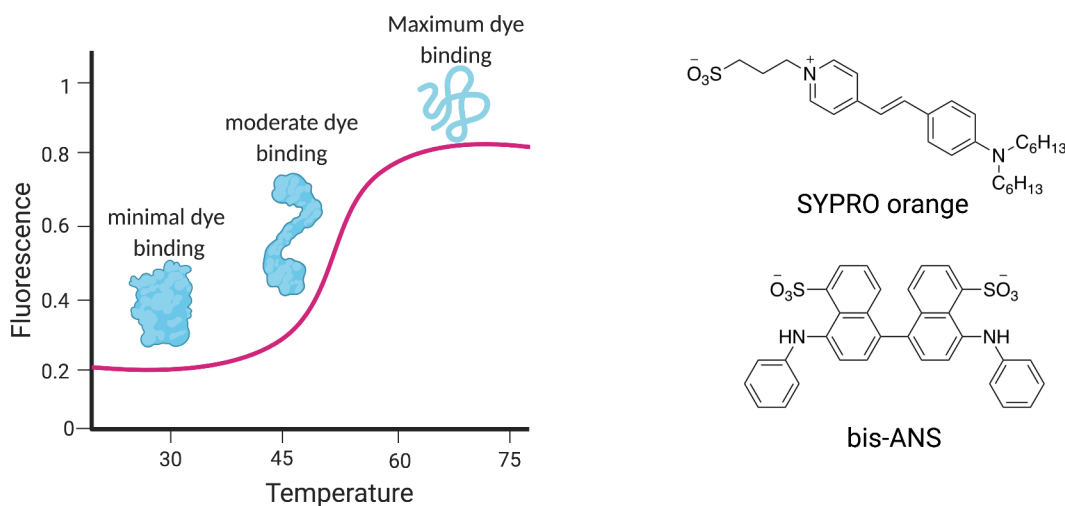


Figure B.1. Use of a reporter dye is a key component of DSF. Left: Overview of dye binding and effects on fluorescence as protein unfolding occurs over a temperature gradient. Right: The structure of two solvatochromic dyes commonly used in DSF.

These dyes have low fluorescence in an aqueous environment, such as the buffer in which a protein sample is stored, and high fluorescence in non-polar environments, such as the hydrophobic sites on unfolded proteins. SYPRO orange is most often the dye of choice because, compared to others, it has the greatest fluorescence quantum yield in a non-polar environment, meaning only a minimal

concentration is needed to elicit a large fluorescent signal.⁸ This results in a good signal to noise ratio for screening. Additionally, SYPRO orange excites near 500 nm, which is far from the wavelengths common for small molecules in screening libraries, decreasing the chances that a small molecule would interfere with the optical properties of the assay and lead to fluorescence quenching.

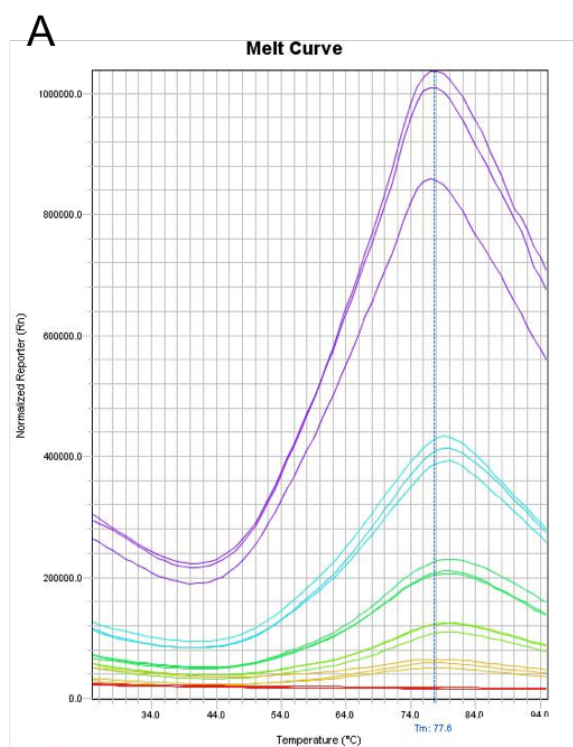
The data readout for DSF experiments is temperature-dependent fluorescence intensity. Typically, fluorescence intensity is normalized to the maximum value obtained for a sample, giving relative fluorescence units (RFU).² Plotting RFU as a function of temperature generates a sigmoidal curve that, for most proteins, can be described by a two-state transition.^{9,10} The inflection point of the curve corresponds to the protein melting temperature (T_m), or the temperature at which the concentrations of folded and unfolded protein are equal and the Gibbs free energy of unfolding is zero.⁹ T_m can be calculated from the melting curve by fitting to the Boltzmann equation.² Alternatively, the first derivative of RFU vs temperature can be calculated and plotted, and the T_m can be identified as the temperature at which the maximum of the first derivative is observed. There are also multiple online resources available for analysis of raw DSF data.

DSF is an attractive technique to identify and characterize inhibitors, particularly when complemented with experiments such as competition fluorescence polarization or FRET. Identifying a functional effect from inhibitor binding is important, however IC_{50} values do not provide much mechanistic insight. To determine mechanism, experiments such as NMR or stopped-flow kinetics can be utilized. However, due to the low throughput of these methods, typically only a few lead molecules are typically profiled. To date, DSF has been successfully applied to a wide range of protein targets^{11–14} including transcription factors.^{15,16} Others have proposed that DSF can be used to obtain more information than just melting temperature, such as inhibition constants, thermodynamic parameters, and even mechanistic information on ligand binding.^{17–19} Thus, DSF has may have the potential for detecting the small conformational differences in ABDs in complex with small-molecule or cognate activator binding partners. Profiling the changes in thermal stability that occurs upon binding of validated inhibitors and TAD ligands will provide insight into properties desirable for future ligand discovery. This may be especially relevant considering a potential mechanism of coactivator PPI inhibition is altering the distribution of conformational states away from those that can participate in TAD binding.²⁰

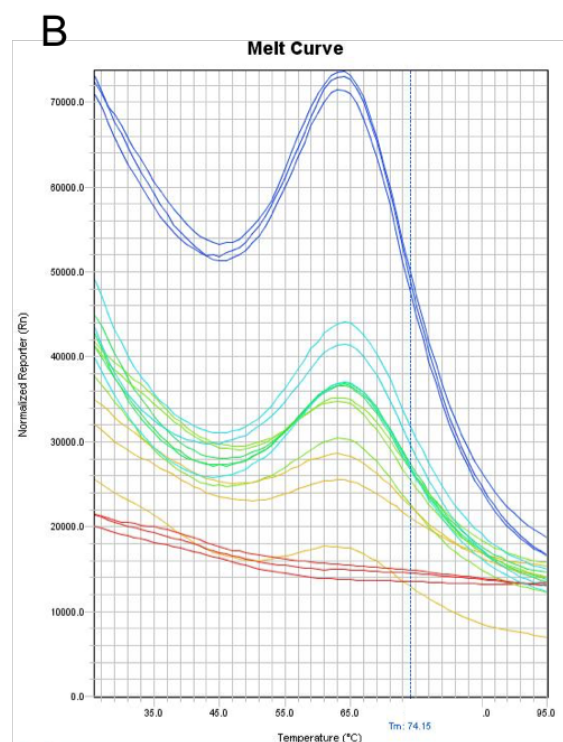
B.3 Results & Discussion

Development of DSF method for Med25 AcID and CBP KIX

To develop a method for screening AcID and KIX, it was first necessary to ensure that the T_m could be easily calculated for these proteins. This meant testing that the selected fluorescent reporter, SYPRO orange, was effective with these proteins and that the concentration of protein used in the assay was optimal for signal to noise. Typical DSF protocols recommend using a protein concentration of 75 $\mu\text{g/mL}$ for accurate T_m calculation, so a range of concentrations around 75 $\mu\text{g/mL}$ were tested for both KIX and Med25. This translates to 7 μM for KIX and 4 μM for AcID, so concentrations both higher and lower were tested (Figure B.2). DSF with AcID shows a classic melt curve, with low initial fluorescence that eventually increased to a maximum, followed by a short decrease at the higher temperatures, indicative of unfolded protein aggregation and/or precipitation which prevents dye binding. Notably, there appears to be two separate unfolding events occurring, as the melting transition is not smooth. Similar melting is observed via CD with AcID, and is hypothesized to be a result of initial helix fraying at lower temperatures, where



[Med25 AcID]	1 μM	2 μM	4 μM	8 μM	16 μM
Tm (1)	69.7	71.4	70.3	69.8	70.5
Tm (2)	69.2	71.4	70.1	70.2	70.2
Tm (3)	69.0	71.5	70.4	70.2	70.3
Avg.	69.3	71.5	70.3	70.1	70.3
St. Dev.	0.4	0.1	0.2	0.2	0.2



[CBP KIX]	5 μM	10 μM	15 μM	20 μM	25 μM
Tm (1)	56.6	56.7	56.7	56.5	56.3
Tm (2)	56.2	56.4	56.5	56.7	56.4
Tm (3)	56.7	56.5	56.3	56.3	56.0
Avg.	56.5	56.5	56.5	56.5	56.2
St. Dev.	0.3	0.2	0.2	0.2	0.2

Figure B.2. Initial melting temperature calculations for Med25 AcID (A) and CBP KIX (B) using DSF. Melting curves are shown as processed from the qPCR (see Methods). Colors of correspond to as follows: red = blank, from lowest to highest concentration = yellow, green, turquoise, light blue, dark blue. T_m values were calculated by manually determining the temperature at the maximum of the first derivative plot provided by the qPCR software.

the beta barrel is still stable, then eventual melting of the entire structure at higher temperatures (Figure B.3). The T_m determined for AcID using DSF agrees with the T_m determined using CD, 70 ± 1 °C compared to 69.3 ± 0.4 °C. Interestingly, the T_m determined for full length Med25 extracted from HeLa lysates using CETSA was lower, 64.8 ± 0.1 °C (see Chapter 2). This could be due to a few different reasons. First, the reporters used in the assays are not the same. CETSA utilizes antibody, which binds to a consensus sequence on the protein of interest whereas DSF utilizes a dye, which binds to hydrophobic areas. Second, CETSA utilizes endogenous, full length Med25, containing other domains in addition to AcID and unstructured interdomain regions and possibly PTMs. These additional elements likely change the thermodynamics of unfolding, leading to a different observed T_m . For assay development, this data indicates that 4-8 μ M AcID would be effective for small molecule screening. Because AcID is easy to express and purify in high yields, 8 μ M was selected as the optimal screening concentration. KIX is a protein that gives high initial fluorescence in DSF. This trend is observed at all concentrations and even at the highest concentration tested, 25 μ M, the initial fluorescence is slightly greater than the fluorescence at 100% unfolded. However, the signal to noise comparing the minimum that occurs at ~ 45 °C to the max at ~ 65 °C should be acceptable for small molecule screening at concentrations >15 μ M.

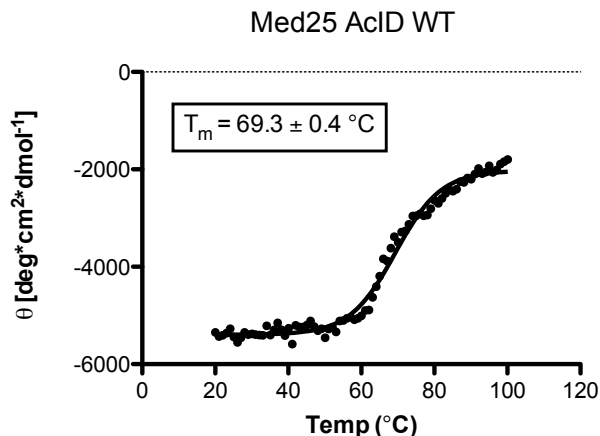
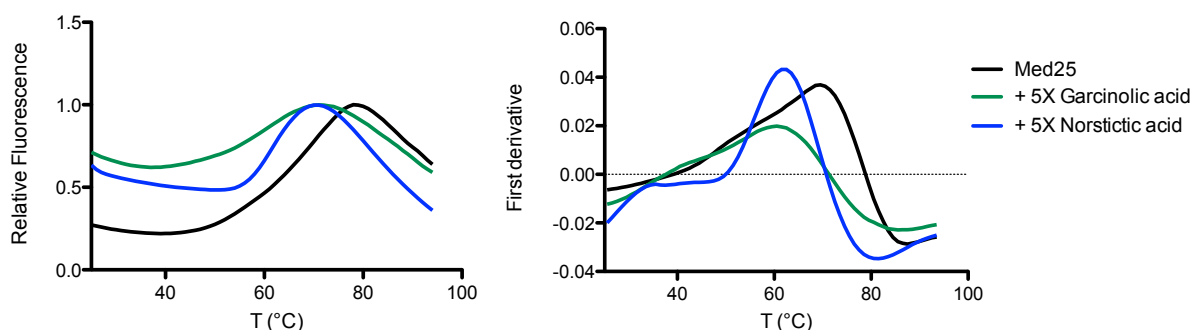


Figure B.3. CD thermal melt of Med25 AcID. Data obtained at 214 nm.

Because compounds are typically screened in 5 to 10-fold excess, it is beneficial to select a lower concentration of protein, thus 15 μM was selected for DSF with KIX.

Analysis of small molecule effects on Med25 AcID thermal stability

From Chapter 2, it is already known that norstictic acid (NA) engages with and thermally stabilized endogenous Med25. Additionally, in unpublished data from Dr. Meg Breen, garcinolic acid (GA) has been shown to thermally stabilize endogenous CBP, and from Chapter 3, GA was also identified as a Med25 AcID inhibitor. Thus, these two compounds should be great tools for initial validation of DSF to look at the effects of inhibitors on coactivator thermal stability. Upon incubating AcID with 5X NA or GA in technical triplicate, a decrease in AcID T_m was observed (Figure B.4). This was repeated multiple times with the same result, with both compounds leading to a change in T_m of -5°C .



Ligand	IC ₅₀ (μM)	ΔT_m ($^\circ\text{C}$)
garcinolic acid	8-10	- 5
norstictic acid	2-3	- 5

Figure B.4. DSF data for Med25 AcID and known ligands Garcinolic acid and Norstictic acid. Data was obtained in technical triplicate, and the average fluorescence is shown. T_m was calculated using GraphPad Prism. IC₅₀ values are shown from experiments reported elsewhere in this thesis.

Theoretically, “destabilization” upon ligand binding is not possible - the free energy contribution of ligand binding results in an increase in the Gibbs free energy of unfolding of a protein. However, observed decreases in T_m are sometimes observed using DSF, and thought to be due to ligand binding to and stabilization of a less populated state.⁹ In other words, the ligand preferably binds to a state other than the most populated, shifting the ensemble away from the “higher T_m ” and instead stabilizing a conformer that has a lower T_m .

While a similar ΔT_m was observed, each inhibitor leads to a unique change in the shape of the melting curve. For example, GA broadens the melt curve whereas NA sharpens the transition from folded to unfolded. This can also be seen with the first derivative plot: the value of the maximum of the first derivative is higher with NA and lower with GA. While more experimental evidence would be needed to confirm, this could reflect the way in which each small molecule inhibitor alters the number of accessible AcID conformations. Since the melting curve and T_m are representative of an ensemble of protein conformers, a broader melting curve could mean that there are more possible states being accessed and less specific interactions, while a sharp transition would indicate narrowing down of accessible populations and more specific interactions. Alternatively, this could be reflective of the binding mechanisms; NA is covalent whereas GA is noncovalent. We know that NA binds, via a specific mechanism mostly at one site on AcID. It is possible that GA binds to multiple sites on AcID, leading to an increase in potentially accessible substrates and a broader melting transition. However, further experiments would be necessary to test this hypothesis.

The stability of proteins is known to increase with addition of ligand in a dose dependent way, indicating that the thermal stability of the protein is proportional to the concentration of the ligand. Dosing AcID with concentrations of NA from 1X-5X led to a dose dependent decrease in T_m (Figure B.5). Not only did increasing [NA] lead to a more significant change in T_m , it also led to a sharper slope in the melting curve (as seen by higher values for the maximum of the first derivative). In addition, this experiment shows that maximum stabilizing effects of NA are obtained with 3X compound. The melt curves obtained at 3X, 4X, and 5X are identical, suggesting that at 3X NA compared to AcID maximum occupancy is obtained. DSF experiments looking at titration of ligands with ABDs could be useful in the future for testing compound concentrations at which to attempt protein crystallization.

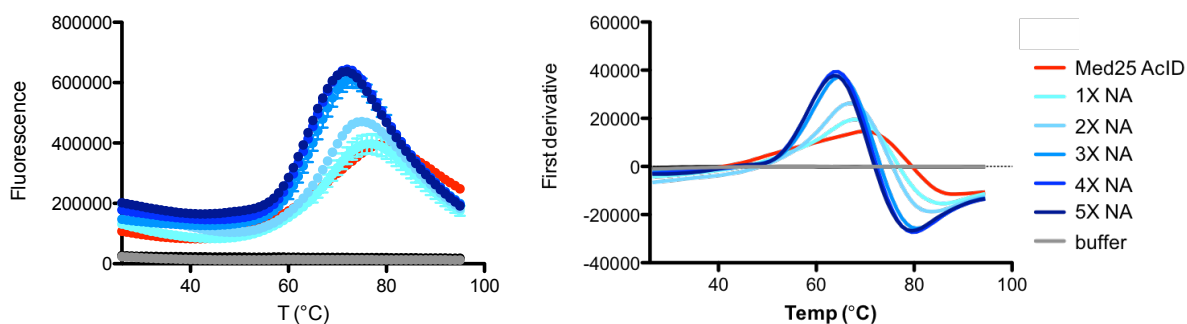


Figure B.5. Concentration dependence of NA on AcID T_m . Data was obtained in technical triplicate, and the average fluorescence is shown, with error bars representing the standard deviation of the mean.

While the results showing AcID destabilization via DSF are replicable, since observed T_m decreases upon ligand binding are rare, looking at the effects of NA and GA on AcID thermal stability using an alternate method was of interest. Inspired by the thermal shift assays conducted with NA in Chapter 2, a similar assay was developed utilizing purified AcID rather than nuclear extracts. After dosing AcID with NA and subjecting samples to various temperatures, the samples could be run on a protein gel and protein visualized via coumassie staining. For both GA and NA, a decrease in thermal stability was observed using this method, similar to the decrease observed with DSF (Figure B.6). To accurately determine the T_m using this method, an extended

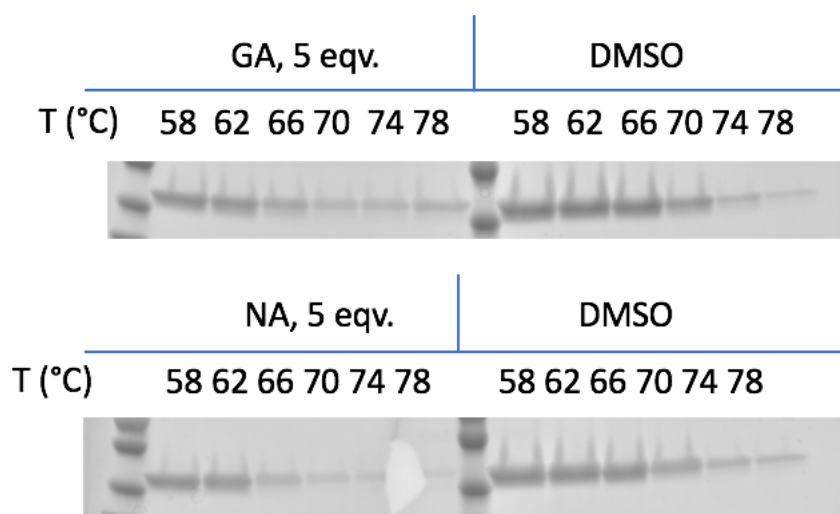


Figure B.6. Protein gel based thermal shift assay using purified Med25 AcID and small molecules.

temperature range was tested with NA (Figure B.7). The ΔT_m for AcID with NA was determined to be about -6 °C, similar to the ΔT_m calculated using DSF. Two different methods showing NA leads to a decrease in thermal stability of AcID validates this result.

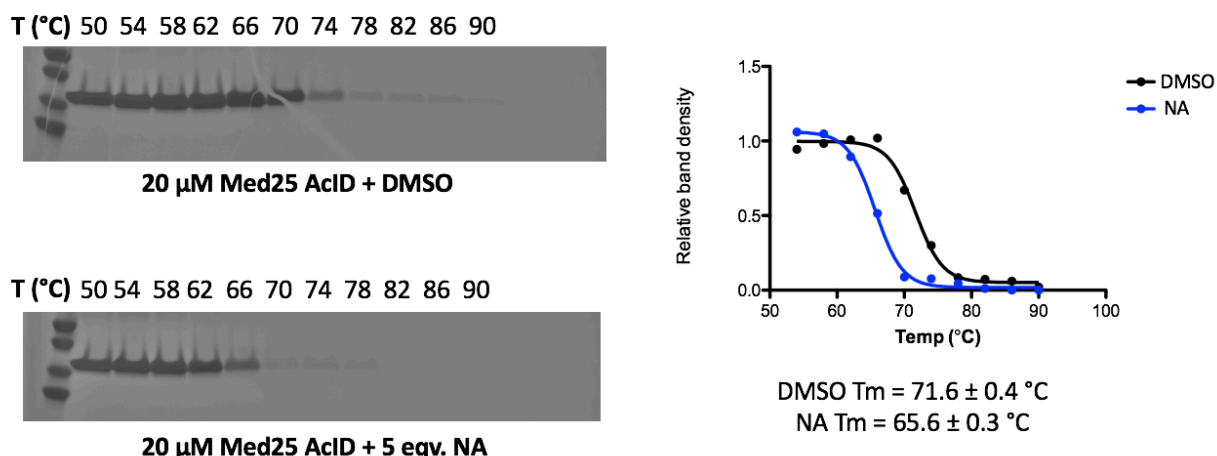
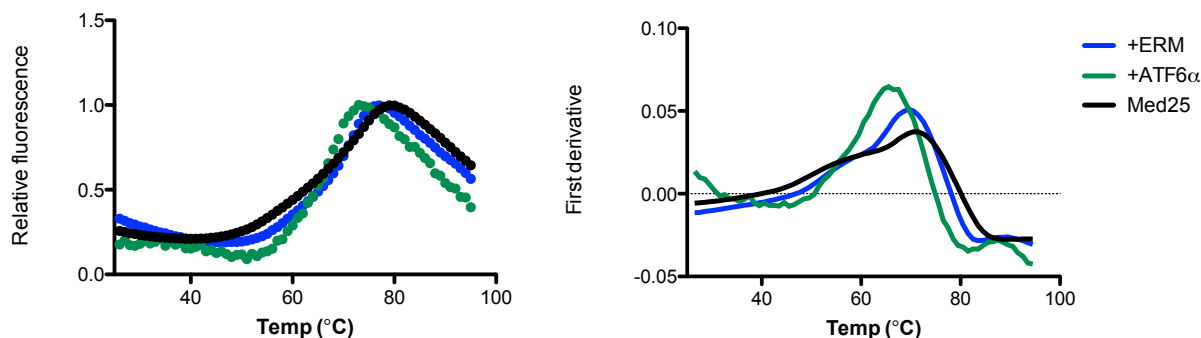


Figure B.7. Protein gel based thermal shift assay using purified Med25 AcID and NA. The gel images shown on the right were analyzed using ImageJ software. Relative band densities were plotted in GraphPad Prism and fit to a sigmoidal dose response equation to determine the T_m .

Analysis of TAD ligand effects on Med25 AcID thermal stability

Profiling of known TADs with DSF could provide additional insights into coactivator inhibitor discovery. With two known AcID inhibitors showing a decrease in T_m via DSF, it was initially hypothesized that the natural ligands would increase AcID T_m because they bind orthosterically. Since ATF6 α and ETV5 have defined binding sites on different faces of AcID, they were selected as the primary ligands to test. Unexpectedly, both ATF6 α and ETV5 binding led to an observed decrease in AcID T_m (Figure B.8). Additionally, both ligands led to an increase the slope of the melting curve. Notably, ATF6 α led to a larger decrease in T_m as well a more significant change in the melting curve slope and shape.



Ligand	K_d (μ M)	ΔT_m ($^{\circ}$ C)
ATF6 α	0.15-0.2	- 5
ETV5	0.5-0.7	- 2

Figure B.8. DSF data with Med25 AcID and TAD peptides. Data was obtained in technical triplicate, and the average fluorescence is shown. T_m was calculated using GraphPad Prism. K_D values were obtained from fluorescence polarization experiments.

To confirm the results observed using DSF, the effects of ATF6 α and ETV5 on AcID T_m were also tested using the previously described gel-based thermal shift assay (Figure B.9). Again, the results of the gel-based assay were similar to those observed via DSF: all ligands tested with AcID led to a decrease in thermal stability. Here a construct of VP16 that could interact with both binding faces simultaneously was also tested, with the hypothesis that engaging both binding faces would thermally stabilize AcID. However, this was not the case, and analysis of the gel images using ImageJ revealed a T_m decrease similar to that observed with ETV5.

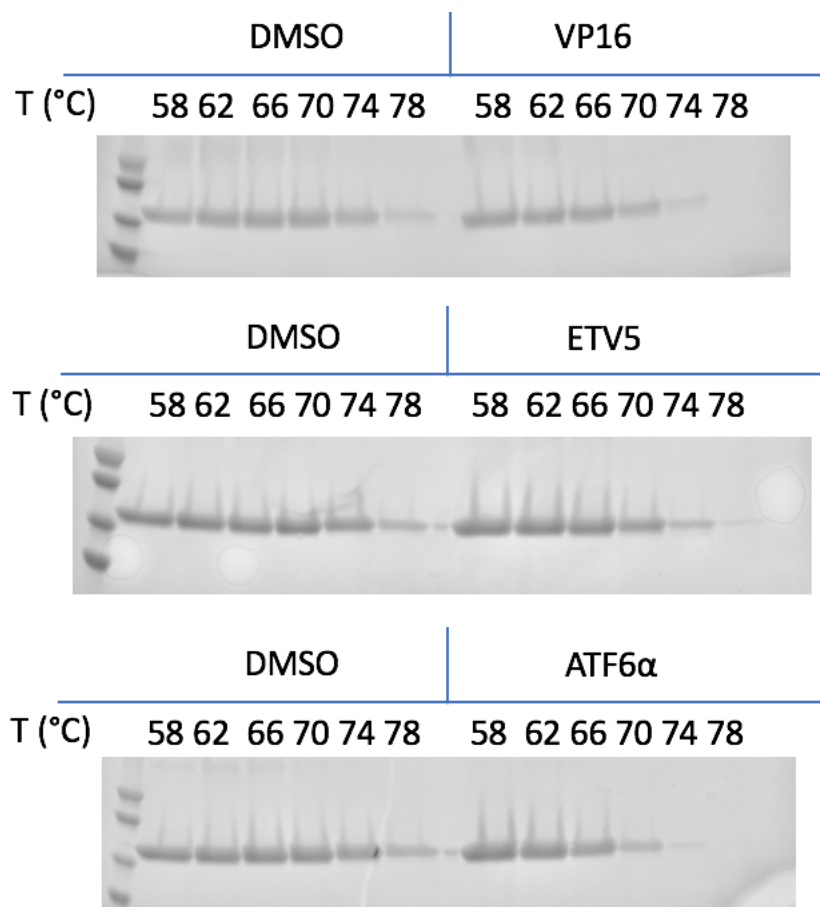
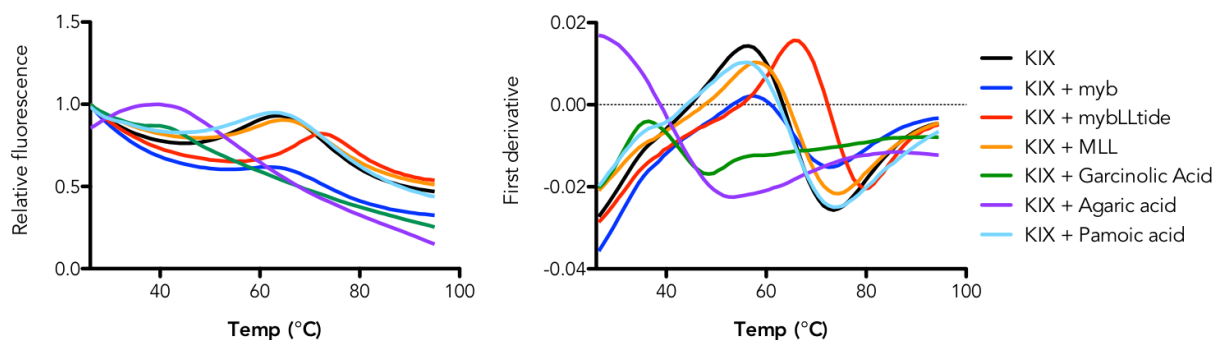


Figure B.9. Protein gel based thermal shift assay using purified Med25 AcID and TAD peptides.

Analysis of known ligand effects on CBP KIX thermal stability using DSF

With the unexpected observation that no tested ligands increase thermal stability of AcID, it was important to analyze other ABDs to determine if this was unique to AcID or widely observed. Analysis of KIX via DSF may provide some additional insights. While AcID has not been thoroughly characterized using thermal stability based methods, there is data in the literature on the influence of ligands on KIX thermal stability.²¹ Additionally, allostery and cooperativity in the KIX system has been thoroughly studied.^{22–27} First, a small panel of known ligands were tested with KIX using DSF (Figure B.10). A wide variety of ΔT_m values were observed, both positive and negative. Incubation with MLL led to an increase in thermal stability, which is not surprising given data from the literature suggesting the MLL binding stabilizes KIX structure.^{22,27} The most dramatic increase in thermal stability was observed with the inhibitor MybLLtide, a bivalent peptidic inhibitor comprised of the MLL and Myb TADs connected by a PEG linker (created by Dr. Stephen Joy). This large increase in thermal stability had previously been observed using CD. The remaining ligands tested all led to an observed decrease in thermal stability. Data with agaric acid appeared to show the most dramatic change in thermal stability; however there is likely some assay interference due to dye binding to the small molecule, as agaric acid is very hydrophobic. A control experiment testing a sample containing only agaric acid and SYPRO orange confirmed high background fluorescence. The data with KIX and garcinolic acid was also challenging to interpret. The melting curve was almost completely flattened, and the first derivative plot shows that at no point does the melt curve even have a positive slope. Like GA, Myb has a flattening effect on the KIX melt curve, decreasing the maximum slope by greater than 2-fold.



Ligand		ΔT_m (°C)
Myb	$K_d = 1-2 \mu\text{M}$	- 2.1
MLL	$K_d = 0.3-0.4 \mu\text{M}$	+ 1.2
mybLLtide	$IC_{50} = > 10 \text{ nM}$	+ 10.1
Garcinolic acid	$IC_{50} = 2-3 \mu\text{M}$	- 17.7
Agaric acid	$IC_{50} = 6.5 \mu\text{M}$	- 30*
Pamoic acid	$IC_{50} = 35 \mu\text{M}$	- 3.15

Figure B.10. DSF data with CBP KIX and ligands. Data was obtained in technical triplicate, and the average fluorescence is shown. T_m was calculated using GraphPad Prism. K_D and IC_{50} values were obtained from fluorescence polarization experiments.

DSF as a tool to analyze KIX ternary systems

CBP KIX is one of the most well studied ABDs. Further, there is an abundance of experimental and computational data published on ligand binding. KIX has two binding sites that interact with distinct TADS, and allostery between these two interfaces is well documented, particularly with regards to formation of the ternary complexes with transcriptional activators. KIX binds both the mixed lineage leukemia (MLL) transcription factor and the proto-oncogene transcription factor Myb in a cooperative manner, with ternary complex formation critical for hematopoiesis.^{23,25,26,28} Similar allosteric effects have been observed in KIX with MLL and a different activator, the phosphorylated kinase-inducible domain of CREB (pKID).^{22,29} In these ternary systems, initial binding of KIX and MLL leads to enhancement of binding of either Myb or pKID at the other KIX site. With DSF, we hypothesize that these effects could be observable as changes in the melting transition and T_m between the systems. Thus, the ternary systems of Myb•KIX•MLL and pKID•KIX•MLL were analyzed using DSF (Figure B.11). As expected, the greatest increase in thermal stability was observed for the ternary complex formation in both systems. For the pKID system, a similar T_m increase is observed with either MLL or pKID binding, 0.8 ± 0.2 °C and 0.7 ± 0.2 °C respectively. For the ternary complex, the T_m increases by 1.5 ± 0.3 °C, which happens to be the sum of the two individual melting temperature changes. For this experiment Myb binding to KIX led to ΔT_m of 0.3 ± 0.5 °C. Notably with the error incorporated this is practically no change. Also, this is in contrast to another experiment conducted with Myb (Figure B.10) where Myb led to a negative ΔT_m . Because the melt curve flattens with Myb, there is a lot of error associated with the determined T_m . MLL binding to KIX here lead to a ΔT_m of 1.0

± 0.2 °C and the ternary complex had a ΔT_m of 1.5 ± 0.3 °C. Again, there appears to be an almost directly additive effect on the T_m change when you compare the binary and ternary complexes.

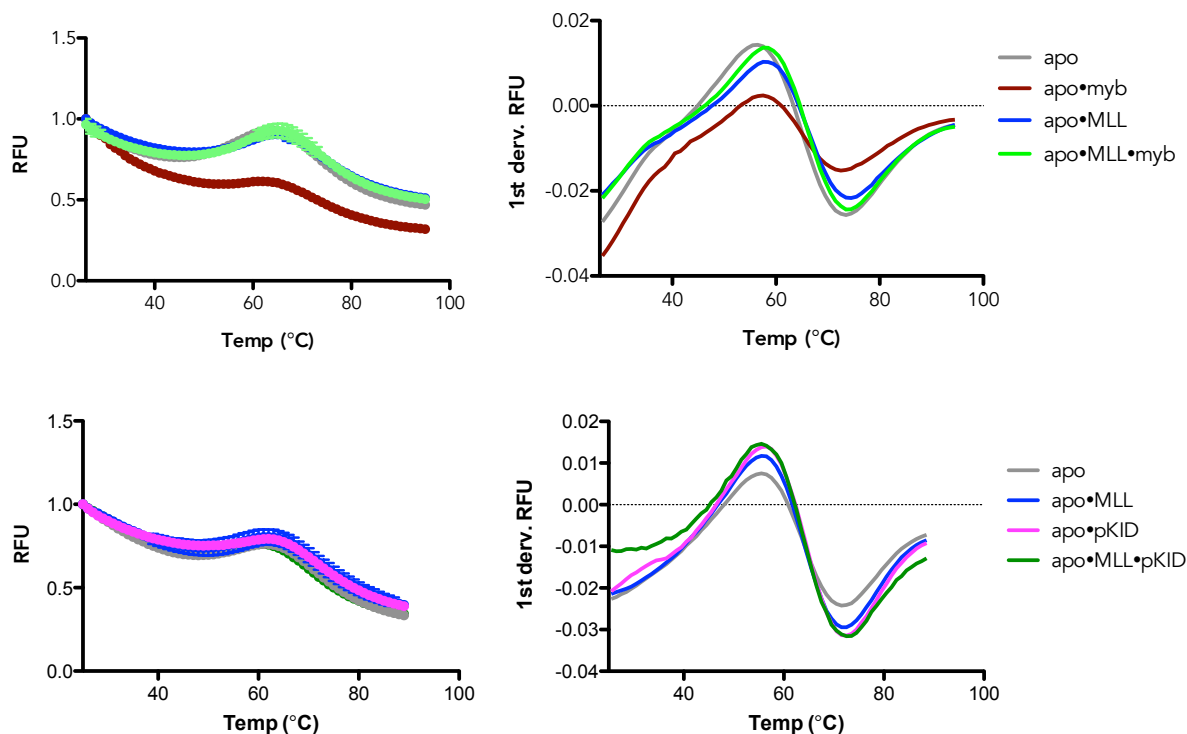


Figure B.11. DSF with the apo KIX ternary systems. Data was obtained in technical triplicate, and the average fluorescence is shown. T_m values were calculated using DSFworld (<https://gestwickilab.shinyapps.io/dsfworld/>).

Previous studies have found that the mutation KIX_{I660V} “turns on” the allosteric communication in the pKID ternary system; that is to say that with the I660V mutant, pKID binding is enhanced regardless of whether MLL is pre-bound or not.³⁰ Thus, it was hypothesized that for this mutant, a larger increase in thermal stability would be observed in the pKID binary complex. In agreement with this hypothesis, the ΔT_m of KIX_{I660V} with pKID increased to 1.6 ± 0.2 °C, which was the exact same ΔT_m for the ternary system. The ΔT_m of KIX_{I660V} with MLL actually decreased slightly from 1.0 ± 0.2 °C in the apo system to 0.6 ± 0.2 °C. So, we observe a similar enhancement of stabilization KIX with pKID regardless of if MLL is also bound or not, similar to the binding enhancement previously observed using this model.

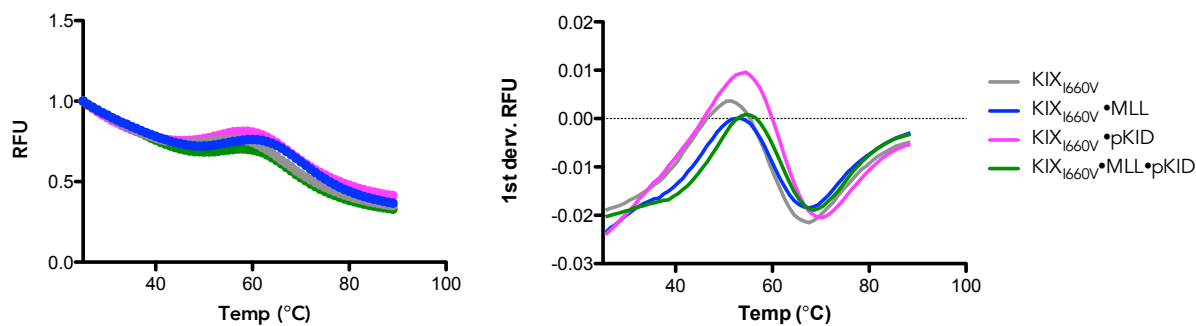


Figure B.12. DSF with the apo KIX and KIX_{L660V} ternary systems. Data was obtained in technical triplicate, and the average fluorescence is shown. T_m values were calculated using DSFworld (<https://gestwickilab.shinyapps.io/dsfworld/>).

Molecule 1-10 has been shown to bind at the MLL site and allosterically inhibit pKID binding at the distal site.^{21,30} The crystal structure of 1-10 Tethered to KIX_{L664C} has been solved and 1-10 has been shown to thermally stabilize KIX_{L664C} using CD.²¹ In line with this data, we observe a large increase in KIX thermal stability with 1-10 using DSF (Figure B.13). The T_m of the modified complex increases the melting temperature by 2.3 ± 0.4 °C in comparison to the apo protein and 3.9 ± 0.3 °C in comparison to KIX_{L664C}. The unfolding transition is sharper with the molecule bound, as seen in the plot of first derivative of changing fluorescence by temperature, that it is for unbound KIX_{L664C}. The increase in thermal melting temperature as determined by CD was ~ 15 °C, much larger than that observed with DSF, further illustrating the variability between different techniques.

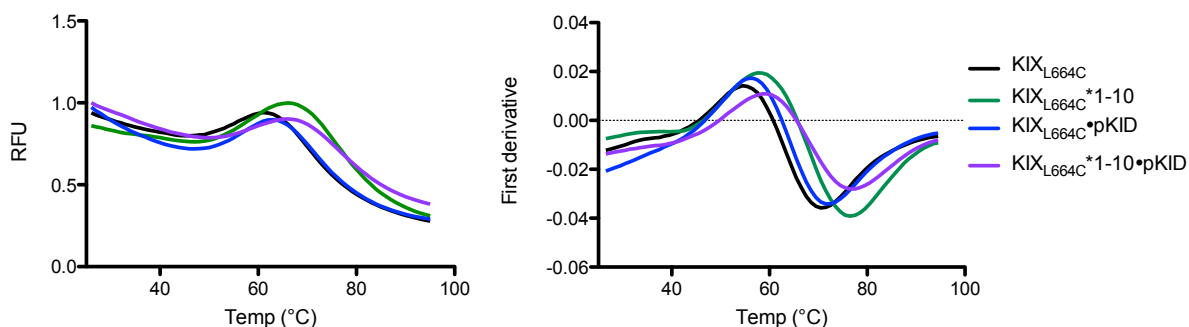


Figure B.13. DSF with the KIX_{L664C} and small molecule stabilizer 1-10. Data was obtained in technical triplicate, and the average fluorescence is shown. T_m 's were calculated using DSFworld (<https://gestwickilab.shinyapps.io/dsfworld/>).

One issue with KIX DSF experiments is the high background due to dye binding, the max fluorescence is actually at the lowest temperatures. KIX is a flexible three helix bundle and with abundant hydrophobic surface area, so this is not surprising. However, this leads to an extremely small observed melting transition, and much more error in T_m determination than compared to a protein with a traditional melt curve such as AcID. For future work with KIX using DSF it will be beneficial to explore alternative dyes, hopefully leading to a lower background and more defined melting transitions.

B.4 Conclusions

DSF is a tool that could be applied for HTS of coactivator proteins. Current observations suggest a hypothesis that small molecules that bind and decrease T_m are acting allosterically, binding to a conformational state other than the native state/altering the distribution of conformational states. Known orthosteric molecules increase T_m , although no molecules have led to stabilization of Med25 AcID. Individual molecules can have distinct effects not only on T_m but on melting curve shape. Current work analyzing dye dependency on thermal melt curves, as well as extracting additional thermodynamic parameters such as ΔS and ΔH , and looking for trends using principle component analysis (PCA) will make DSF an even more invaluable tool for dynamic protein inhibitor profiling in future studies.

B.5 Materials and Methods

Protein expression and purification

Med25 AcID and CBP KIX were expressed and purified as described in Chapter 2. Plasmids for KIX mutants were obtained from Dr. Jean Lodge and expressed and purified using the same method as WT (cite paper). All proteins were tested for purity using SDS-page gel (>95% pure) and identity was confirmed using mass spectrometry.

DSF protocol

Experiments were conducted utilizing 20 μ L sample volumes in 96 well PCR plates sealed with clear cap strips. To determine T_m , protein (8 μ M AcID; 15 μ M KIX – unless otherwise noted) in the presence of 5X SYPRO orange dye (1:1000 dilution of purchased 5000X stock; Invitrogen) was incubated with ligand (5X for small molecules, or 1X for TAD peptides unless otherwise

noted) at RT for 30 minutes. An Applied Biosystems StepOnePlus qPCR instrument was utilized to obtain melting curves by exciting at 488 nm and monitoring emission at 602 nm over a temperature gradient of 25-95°C with a 1°C/min ramp. Raw fluorescence data was converted to relative fluorescence units (RFU) by normalizing each individual melt curve to its maximum fluorescence. RFU was imported into the online data analysis program, DSFworld, and T_m was calculated by determining the maximum of the first derivative (dRFU). For data visualization, both RFU and dRFU are plotted as a function of temperature using GraphPad Prism software. The maximum of the first derivative is the reported T_m , with ΔT_m of each ligand calculated as the difference between the T_m of the protein and the T_m of the protein + ligand.

Gel-based thermal shift assay

Norstictic acid (dissolved in DMSO) or an equivalent volume of DMSO was added to purified Med25 AcID and samples were incubated at RT for 30 minutes. Final concentration of DMSO was 0.1% v/v. After incubation, samples were aliquoted into thin-walled PCR tubes (15 μ l per tube).

A Labnet Multigene OPTIMAX PCR was used to heat each sample for 3 minutes at various temperatures. Contents were transferred to epitubes and centrifuged at 17000 g for 1 minute at 4°C to remove precipitated proteins. Contents of each epitube were carefully transferred to a clean epitube, leaving precipitated protein behind. LDS loading dye was added and samples were boiled for 5 minutes at 95°C. 10 μ L of each sample was loaded onto a 4-20% mini-PROTEAN TGX gel (BioRad, 4561096); gel was run at 180V for 45 minutes. To visualize protein, the gel was stained with Coumassie quick stain and an image was taken using an Azure Biosystems c600 imager in the visible light spectrum. Analysis was conducted using ImageJ software.

Synthesis of transcriptional activation domain peptides

All peptides were synthesized automatically by Dr. Stephen Joy using a Liberty Blue peptide synthesizer on Protide resin from CEM. Peptides were deprotected and cleaved from the resin for 4 hours in 90% trifluoroacetic acid (TFA), 5% thioanisole, 3% ethanedithiol (EDT) and 2% anisole unless otherwise noted. Crude peptides were filtered to remove resin, dried under nitrogen stream, and precipitated from cold ether. Peptide suspensions were transferred to a 15 mL falcon tube, centrifuged at 4000 g for 5 minutes at 4 C, and ether decanted. Pellets were

resuspended in 20% acetonitrile, frozen and lyophilized. Dry, crude peptides were resuspended again in 20% acetonitrile, purified via HPLC on an Agilent 1260 analytical HPLC using a semi-prep C18 column (Phenomenex) over a 10-50% acetonitrile gradient in 0.1% TFA. Pure fractions were collected and lyophilized to afford pure peptides unless otherwise noted. Analytical traces and mass spectra were obtained using an Agilent 6230 LC/TOF and an Agilent 6545 LC/Q-TOF. Myb(291-396) was synthesized and purified as described above with no modifications and isolated in >98% purity (Figure S7a).

Myb sequence: Ac-KEKRIKELELLLMSTENELKGQQALW-NH₂.

Myb calculated mass [M+H]⁺: 3168.74. Mass observed [M+H]⁺: 3168.76.

MLL(840-858) was synthesized as described above but purification was modified slightly. Peptide was purified once on a semi-prep C18 column over a 40 min 10-50% acetonitrile gradient in 20 mM ammonium acetate to afford a mix of MLL and partially oxidized versions of MLL containing both disulfides and methionine oxide products. MLL and oxidized MLL could not be readily separated, and were instead combined, frozen, and lyophilized. Dried MLL peptides were then resuspended in 20% acetonitrile in 50 mM TRIS (pH = 8.0) and 10 mM dithiothreitol (DTT) and agitated at room temperature for 2 hours. The DTT/peptide solution was purified directly on 10-50% acetonitrile gradient in 0.1% TFA to afford MLL in 98% purity (Figure S7b).

MLL sequence: Ac-DCGNILPSDIMDFVLKNTPY-NH₂.

MLL calculated mass [M+H]⁺: 2296.09. Mass observed [M+H]⁺: 2296.10.

pKID(119-147) was synthesized and purified as described above except deprotection and resin cleavage was performed for only 2 hours in 95% TFA, 2.5% water and 2.5% triisopropylsilane. HPLC purification afforded pKID in >90% purity (Figure S7c).

pKID sequence: Ac-TDSQKRREILSRRPS(Phos)YRKILNDLSSDAPG-NH₂.

pKID calculated mass [M+H]⁺: 3479.78. Mass observed [M+H]⁺: 3479.81.

B.6 References

- (1) Gao, K.; Oerlemans, R.; Groves, M. R. Theory and Applications of Differential Scanning Fluorimetry in Early-Stage Drug Discovery. *Biophys. Rev.* **2020**, *12* (1), 85–104. <https://doi.org/10.1007/s12551-020-00619-2>.
- (2) Niesen, F. H.; Berglund, H.; Vedadi, M. The Use of Differential Scanning Fluorimetry to Detect Ligand Interactions That Promote Protein Stability. *Nat. Protoc.* **2007**, *2* (9), 2212–2221. <https://doi.org/10.1038/nprot.2007.321>.
- (3) Pantoliano, M. W.; Petrella, E. C.; Kwasnoski, J. D.; Lobanov, V. S.; Myslik, J.; Graf, E.; Carver, T.; Asel, E.; Springer, B. A.; Lane, P.; Salemme, F. R. High-Density Miniaturized Thermal Shift Assays as a General Strategy for Drug Discovery. *J. Biomol. Screen.* **2001**, *6* (6), 429–440. <https://doi.org/10.1177/108705710100600609>.
- (4) Smoot, A. L.; Panda, M.; Brazil, B. T.; Buckle, A. M.; Fersht, A. R.; Horowitz, P. M. The Binding of Bis-ANS to the Isolated GroEL Apical Domain Fragment Induces the Formation of a Folding Intermediate with Increased Hydrophobic Surface Not Observed in Tetradecameric GroEL. *Biochemistry* **2001**, *40* (14), 4484–4492. <https://doi.org/10.1021/bi001822b>.
- (5) Stopa, B.; Konieczny, L.; Piekarska, B.; Roterman, I.; Rybarska, J.; Skowronek, M. Effect of Self Association of Bis-ANS and Bis-Azo Dyes on Protein Binding. *Biochimie* **1997**, *79* (1), 23–26. [https://doi.org/10.1016/S0300-9084\(97\)87621-3](https://doi.org/10.1016/S0300-9084(97)87621-3).
- (6) Lo, M.-C.; Aulabaugh, A.; Jin, G.; Cowling, R.; Bard, J.; Malamas, M.; Ellestad, G. Evaluation of Fluorescence-Based Thermal Shift Assays for Hit Identification in Drug Discovery. *Anal. Biochem.* **2004**, *332* (1), 153–159. <https://doi.org/10.1016/j.ab.2004.04.031>.
- (7) Steinberg, T. H.; Jones, L. J.; Haugland, R. P.; Singer, V. L. SYPRO Orange and SYPRO Red Protein Gel Stains: One-Step Fluorescent Staining of Denaturing Gels for Detection of Nanogram Levels of Protein. *Anal. Biochem.* **1996**, *239* (2), 223–237. <https://doi.org/10.1006/abio.1996.0319>.
- (8) Epps, D. E.; Sarver, R. W.; Rogers, J. M.; Herberg, J. T.; Tomich, P. K. The Ligand Affinity of Proteins Measured by Isothermal Denaturation Kinetics. *Anal. Biochem.* **2001**, *292* (1), 40–50. <https://doi.org/10.1006/abio.2001.5047>.
- (9) Cimmperman, P.; Baranauskienė, L.; Jachimovičiūtė, S.; Jachno, J.; Torresan, J.; Michailovienė, V.; Matulienė, J.; Sereikaitė, J.; Bumelis, V.; Matulis, D. A Quantitative Model of Thermal Stabilization and Destabilization of Proteins by Ligands. *Biophys. J.* **2008**, *95* (7), 3222–3231. <https://doi.org/10.1529/biophysj.108.134973>.
- (10) Daumantas Matulis, ‡; James K. Kranz; F. Raymond Salemme, and; Todd*, M. J. Thermodynamic Stability of Carbonic Anhydrase: Measurements of Binding Affinity and Stoichiometry Using ThermoFluor <https://pubs.acs.org/doi/abs/10.1021/bi048135v> (accessed Apr 15, 2020). <https://doi.org/10.1021/bi048135v>.
- (11) Shao, H.; Olton, K.; Wu, T.; E. Gestwicki, J. Differential Scanning Fluorimetry (DSF) Screen to Identify Inhibitors of Hsp60 Protein–Protein Interactions. *Org. Biomol. Chem.* **2020**. <https://doi.org/10.1039/D0OB00928H>.
- (12) DeSantis, K.; Reed, A.; Rahhal, R.; Reinking, J. Use of Differential Scanning Fluorimetry as a High-Throughput Assay to Identify Nuclear Receptor Ligands. *Nucl. Recept. Signal.* **2012**, *10* (1), nrs.10002. <https://doi.org/10.1621/nrs.10002>.
- (13) Fedorov, O.; Niesen, F. H.; Knapp, S. Kinase Inhibitor Selectivity Profiling Using Differential Scanning Fluorimetry. In *Kinase Inhibitors: Methods and Protocols*; Kuster,

- B., Ed.; Methods in Molecular Biology; Humana Press: Totowa, NJ, 2012; pp 109–118. https://doi.org/10.1007/978-1-61779-337-0_7.
- (14) Makley, L. N.; McMenimen, K. A.; DeVree, B. T.; Goldman, J. W.; McGlasson, B. N.; Rajagopal, P.; Dunyak, B. M.; McQuade, T. J.; Thompson, A. D.; Sunahara, R.; Klevit, R. E.; Andley, U. P.; Gestwicki, J. E. Pharmacological Chaperone for α -Crystallin Partially Restores Transparency in Cataract Models. *Science* **2015**, *350* (6261), 674–677. <https://doi.org/10.1126/science.aac9145>.
 - (15) Attarha, S.; Reithmeier, A.; Busker, S.; Desroses, M.; Page, B. D. G. Validating STAT Protein-Inhibitor Interactions Using Biochemical and Cellular Thermal Shift Assays. *ACS Chem. Biol.* **2020**. <https://doi.org/10.1021/acscchembio.0c00046>.
 - (16) STAT3 Differential Scanning Fluorimetry and Differential Scanning Light Scattering Assays: Addressing a Missing Link in the Characterization of STAT3 Inhibitor Interactions. *J. Pharm. Biomed. Anal.* **2018**, *160*, 80–88. <https://doi.org/10.1016/j.jpba.2018.07.018>.
 - (17) Sun, C.; Li, Y.; Yates, E. A.; Fernig, D. G. SimpleDSFviewer: A Tool to Analyze and View Differential Scanning Fluorimetry Data for Characterizing Protein Thermal Stability and Interactions. *Protein Sci.* **2020**, *29* (1), 19–27. <https://doi.org/10.1002/pro.3703>.
 - (18) Wu, T.; Yu, J.; Gale-Day, Z.; Woo, A.; Suresh, A.; Hornsby, M.; Gestwicki, J. E. Three Essential Resources to Improve Differential Scanning Fluorimetry (DSF) Experiments. *bioRxiv* **2020**, 2020.03.22.002543. <https://doi.org/10.1101/2020.03.22.002543>.
 - (19) Wright, T. A.; Stewart, J. M.; Page, R. C.; Konkolewicz, D. Extraction of Thermodynamic Parameters of Protein Unfolding Using Parallelized Differential Scanning Fluorimetry. *J. Phys. Chem. Lett.* **2017**, *8* (3), 553–558. <https://doi.org/10.1021/acs.jpclett.6b02894>.
 - (20) Henley, M. J.; Linhares, B. M.; Morgan, B. S.; Cierpicki, T.; Fierke, C. A.; Mapp, A. K. Unexpected Specificity within Dynamic Transcriptional Protein–Protein Complexes. *Proc. Natl. Acad. Sci.* **2020**, *117* (44), 27346–27353. <https://doi.org/10.1073/pnas.2013244117>.
 - (21) Wang, N.; Majmudar, C. Y.; Pomerantz, W. C.; Gagnon, J. K.; Sadowsky, J. D.; Meagher, J. L.; Johnson, T. K.; Stuckey, J. A.; Brooks, C. L.; Wells, J. A.; Mapp, A. K. Ordering a Dynamic Protein via a Small-Molecule Stabilizer. *J. Am. Chem. Soc.* **2013**, *135* (9), 3363–3366. <https://doi.org/10.1021/ja3122334>.
 - (22) 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.* **2002**, *277* (45), 43168–43174. <https://doi.org/10.1074/jbc.M207660200>.
 - (23) Jin, S.; Zhao, H.; Yi, Y.; Nakata, Y.; Kalota, A.; Gewirtz, A. M. C-Myb Binds MLL through Menin in Human Leukemia Cells and Is an Important Driver of MLL-Associated Leukemogenesis. *J. Clin. Invest.* **2010**, *120* (2), 593–606. <https://doi.org/10.1172/JCI38030>.
 - (24) 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.* **2001**, *21* (7), 2249–2258. <https://doi.org/10.1128/MCB.21.7.2249-2258.2001>.
 - (25) 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. Publ. Protein Soc.* **2014**, 23 (7), 962–969.
<https://doi.org/10.1002/pro.2480>.
- (26) Kasper, L. H.; Fukuyama, T.; Lerach, S.; Chang, Y.; Xu, W.; Wu, S.; Boyd, K. L.; Brindle, P. K. Genetic Interaction between Mutations in C-Myb and the KIX Domains of CBP and P300 Affects Multiple Blood Cell Lineages and Influences Both Gene Activation and Repression. *PloS One* **2013**, 8 (12), e82684.
<https://doi.org/10.1371/journal.pone.0082684>.
 - (27) 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.* **2014**, 111 (33), 12067–12072.
<https://doi.org/10.1073/pnas.1405831111>.
 - (28) Graf, T. Myb: A Transcriptional Activator Linking Proliferation and Differentiation in Hematopoietic Cells. *Curr. Opin. Genet. Dev.* **1992**, 2 (2), 249–255.
[https://doi.org/10.1016/s0959-437x\(05\)80281-3](https://doi.org/10.1016/s0959-437x(05)80281-3).
 - (29) Campbell, K. M.; Lumb, K. J. Structurally Distinct Modes of Recognition of the KIX Domain of CBP by Jun and CREB. *Biochemistry* **2002**, 41 (47), 13956–13964.
<https://doi.org/10.1021/bi026222m>.
 - (30) 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. U. S. A.* **2014**, 111 (33), 12061–12066.
<https://doi.org/10.1073/pnas.1406033111>.

APPENDIX C

Techniques Used for *in silico* Study of TMPRSS2

This appendix provides additional detail on the computational methodology utilized in Chapter 4 to study and screen TMPRSS2. This work was conducted by Yujin Wu and Amanda Peiffer, and the details in this appendix were reproduced from Peiffer A.L.*, Garlick J.M.*, Wu Y., Soellner M.B., Brooks C.L. III, Mapp A.K. TMPRSS2 inhibitor discovery facilitated through an *in silico* and biochemical screening platform (DOI: 10.1101/2021.03.22.436465).

Construction of homology model

The homology model of TMPRSS2 was generated using SWISS-Model based on the serine protease Hepsin (PDB 5CE1), which has 34% similarity and 70% coverage of the TMPRSS2 sequence. Included in the Hepsin structure is a 100 nM inhibitor, 2-[6-(1-hydroxycyclohexyl)pyridin-2-yl]-1H-indole-5-carboximidamide, which is bound in the active site. The inhibitor is utilized as one of the pharmacophore targets in our *fastdock* protocol. The SWISS-MODEL structure of TMPRSS2 was further “conditioned” through the application of molecular dynamics in an implicit solvent (GBMV) model to facilitate better packing and configurational relaxation.^{1,2,3}

Virtual screening to identify preliminary hits for in vitro assays

Extensive virtual screening was performed to obtain putative hits for follow-up testing via *in vitro* inhibition assays (Figure 4). Over 100,000 molecules were collected from multiple databases, including the ZINC database,⁴ SWEETLEAD,⁵ and the Center for Chemical Genomics (CCG), which were subjected to a hierarchical refinement of docking poses. In the first stage, rigid receptor docking was performed exploring two means of initially positioning the small molecules. One utilized pharmacophores based on ligands in other bound serine proteases (see Methods), and the other initiated from a random generation of molecular conformations and random positioning inside the pocket. The second relied upon a novel 3D pharmacophore *fastdock* framework, which has the potential to perform a preliminary screen of millions of compounds by superposing pharmacophores onto compounds bound in experimentally solved structures.

Level 1 screening candidates were subjected to GPU accelerated Flexible-CDOCKER methods that were recently developed as part of the CHARMM molecular modeling package cite.⁶⁻⁹ This approach utilizes flexible side chains for residues in or near the binding pocket while using a grid representation for the remaining receptor. Multiple copies of each set of side chains and initial ligand poses are created, which allows for parallel, multiple copy processing of multiple flexible ligands-flexible receptor trials simultaneously on GPUs. The flexible docking searching algorithm combines molecular dynamics (MD) based simulated annealing and a continuous genetic algorithm search protocol to enhance the sampling of differing receptor conformations.

A novel scoring methodology was utilized, performing conformational clustering of the flexible side chains and the ligand, which provided key contributions to the ligand scoring from

the entropic variation of the side chains to accommodate various ligand poses. The ligands were rescored in the protein binding site a final time using an implicit solvent model that captures aspects of the desolvation costs not generally accessible in typical docking methods.¹⁰ The rescoring was accomplished by minimizing the docked poses from the flexible side chains and flexible ligand in the context of the rigid protein, while also considering the total energy of the solvated docked and undocked systems.

General flexible docking setup

The *fastdock* protocol is a python-based workflow that integrates the *align-it* software to search across our curated library of compounds for 3D pharmacophore matches to an inhibitor from a solved structure.^{11,12} The *fastdock* ligand templates are taken from the Hepsin structure used in the initial generation of the model (PDB 5CE1) as well as from a plasma kallikrein structure with the 1 nM inhibitor N-[(6-amino-2,4-dimethylpyridin-3-yl)methyl]-1-({4-[(1H-pyrazol-1-yl)methyl]phenyl}methyl)-1H-pyrazole-4-carboxamide bound (PDB 6O1G; 43% sequence similarity and 51% sequence coverage). Scoring of the pharmacophore matches is based on a volumetric Tanimoto value of the target ligand pharmacophore map and the reference ligand map. Based on this initial selection of potential ligands for exploration, we harvested 1-10% of the top hits.

The MMTSB tool set was used to cluster binding poses and prepare pdb files.¹³ Open Babel was used to generate ligand random conformations.¹⁴ MOE was used to predict the correct protonation state for the ligands at pH 7.4.¹⁵ ParamChem was used to prepare the ligand topology and parameter files with the CGenFF force field.¹⁶⁻¹⁸ Clustering used the tool cluster.pl with a 1 Å cutoff radius for the K-means clustering. The CHARMM C36 force fields were used and docking was performed in CHARMM with the CHARMM/OpenMM parallel simulated annealing feature.^{9,19}

Flexible docking setup

Flexible CDOCKER with a hybrid searching algorithm combining molecular dynamics (MD) based simulated annealing and continuous genetic algorithm was used to dock and rank the top hits.⁶ Flexible CDOCKER uses a physics-based scoring function and allows both ligand and protein side chains to explore their conformational space simultaneously. The following amino

acid side chains are considered flexible : His 296, Tyr 337, Lys 342, Asp 435, Ser 436, Gln 438, Ser 441, Thr 459, Trp 461 and Cys 465.

Each docking measurement represents 500 genes (docking trials). The coordinates of the ligand-protein flexible side chains are used to assemble a gene (potential docking pose). Each ligand in the dataset is first aligned to the pharmacophore model with *align-it*. In the initial generation, half of the genes have the ligand starting with the aligned position. The rest of the genes are constructed by generating a random conformation of the ligand with Open Babel and centering at the binding pocket. A random translation (within a volume with a 2 Å edge length) and rotation (maximum 360°) are performed on ligands in each gene. An energy cutoff is applied to avoid potential collision between ligand atoms and protein atoms due to the random translation and rotation. The protein flexible side chains are initialized with the coordinates from the input homology model. Then these genes are optimized by an MD based simulated annealing algorithm. Detailed values for softness parameter E_{max} used in flexible receptor docking are summarized in Table C.1.

Table C.1: Soft-core potentials used in flexible receptor docking

name	$E_{max}^{*}(\text{vdw})$	$E_{max}^{*}(\text{att})$	$E_{max}^{*}(\text{rep})$
Soft-core potential I	15.0	-120.0	-2.0
Soft-core potential II	3.0	-20.0	40.0
Soft-core potential III	10000	-10000	10000

* $E_{max}(\text{vdw})$, $E_{max}(\text{att})$ and $E_{max}(\text{rep})$ in the unit of kcal/mol are parameters for the Van der Waals, electrostatic attractive, and electrostatic repulsive interactions, respectively.

The docking poses (optimized genes) are then K-means clustered based on ligand heavy atom RMSD with a radius cutoff of 1 Å. We then select the best individuals (minimum energy pose) from the top 10 largest clusters to construct the second generation. In our previous study, we show that using two generations is adequate and the average computer time for each docking measurement is around 30~45 mins. After the second generation, the docking poses are clustered and the best individuals from the top 15 largest clusters are saved. These docking poses are then rescored using the FACTS implicit solvent model.¹⁰

References

- (1) Gong, X.; Chiricotto, M.; Liu, X.; Nordquist, E.; Feig, M.; Brooks III, C. L.; Chen, J. Accelerating the Generalized Born with Molecular Volume and Solvent Accessible Surface Area Implicit Solvent Model Using Graphics Processing Units. *J. Comput. Chem.* **2020**, *41* (8), 830–838. <https://doi.org/10.1002/jcc.26133>.
- (2) Lee, M. S.; Salsbury, F. R.; Brooks III, C. L. Novel Generalized Born Methods. *J. Chem. Phys.* **2002**, *116* (24), 10606–10614. <https://doi.org/10.1063/1.1480013>.
- (3) Lee, M. S.; Feig, M.; Salsbury, F. R.; Brooks III, C. L. New Analytic Approximation to the Standard Molecular Volume Definition and Its Application to Generalized Born Calculations. *J. Comput. Chem.* **2003**, *24* (11), 1348–1356. <https://doi.org/10.1002/jcc.10272>.
- (4) Sterling, T.; Irwin, J. J. ZINC 15 – Ligand Discovery for Everyone. *J. Chem. Inf. Model.* **2015**, *55* (11), 2324–2337. <https://doi.org/10.1021/acs.jcim.5b00559>.
- (5) Novick, P. A.; Ortiz, O. F.; Poelman, J.; Abdulhay, A. Y.; Pande, V. S. SWEETLEAD: An in Silico Database of Approved Drugs, Regulated Chemicals, and Herbal Isolates for Computer-Aided Drug Discovery. *PloS One* **2013**, *8* (11), e79568. <https://doi.org/10.1371/journal.pone.0079568>.
- (6) Gagnon, J. K.; Law, S. M.; Brooks III, C. L. Flexible CDOCKER: Development and Application of a Pseudo-Explicit Structure-Based Docking Method within CHARMM. *J. Comput. Chem.* **2016**, *37* (8), 753–762. <https://doi.org/10.1002/jcc.24259>.
- (7) Ding, X.; Wu, Y.; Wang, Y.; Vilseck, J. Z.; Brooks III, C. L. Accelerated CDOCKER with GPUs, Parallel Simulated Annealing, and Fast Fourier Transforms. *J. Chem. Theory Comput.* **2020**, *16* (6), 3910–3919. <https://doi.org/10.1021/acs.jctc.0c00145>.
- (8) Wu, G.; Robertson, D. H.; Brooks III, C. L.; Vieth, M. Detailed Analysis of Grid-Based Molecular Docking: A Case Study of CDOCKER-A CHARMM-Based MD Docking Algorithm. *J. Comput. Chem.* **2003**, *24* (13), 1549–1562. <https://doi.org/10.1002/jcc.10306>.
- (9) Brooks, B. R.; Brooks III, C. L.; Mackerell, A. D.; Nilsson, L.; Petrella, R. J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S.; Caflisch, A.; Caves, L.; Cui, Q.; Dinner, A. R.; Feig, M.; Fischer, S.; Gao, J.; Hodoseck, M.; Im, W.; Kuczera, K.; Lazaridis, T.; Ma, J.; Ovchinnikov, V.; Paci, E.; Pastor, R. W.; Post, C. B.; Pu, J. Z.; Schaefer, M.; Tidor, B.; Venable, R. M.; Woodcock, H. L.; Wu, X.; Yang, W.; York, D. M.; Karplus, M. CHARMM: The Biomolecular Simulation Program. *J. Comput. Chem.* **2009**, *30* (10), 1545–1614. <https://doi.org/10.1002/jcc.21287>.
- (10) Haberthür, U.; Caflisch, A. FACTS: Fast Analytical Continuum Treatment of Solvation. *J. Comput. Chem.* **2008**, *29* (5), 701–715. <https://doi.org/10.1002/jcc.20832>.
- (11) Taylor, R.; Cole, J. C.; Cosgrove, D. A.; Gardiner, E. J.; Gillet, V. J.; Korb, O. Development and Validation of an Improved Algorithm for Overlaying Flexible Molecules. *J. Comput. Aided Mol. Des.* **2012**, *26* (4), 451–472. <https://doi.org/10.1007/s10822-012-9573-y>.
- (12) Sanders, M. P. A.; Barbosa, A. J. M.; Zarzycka, B.; Nicolaes, G. A. F.; Klomp, J. P. G.; de Vlieg, J.; Del Rio, A. Comparative Analysis of Pharmacophore Screening Tools. *J. Chem. Inf. Model.* **2012**, *52* (6), 1607–1620. <https://doi.org/10.1021/ci2005274>.
- (13) Feig, M.; Karanicolas, J.; Brooks III, C. L. MMTSB Tool Set: Enhanced Sampling and Multiscale Modeling Methods for Applications in Structural Biology. *J. Mol. Graph. Model.* **2004**, *22* (5), 377–395. <https://doi.org/10.1016/j.jmgm.2003.12.005>.

- (14) O'Boyle, N. M.; Banck, M.; James, C. A.; Morley, C.; Vandermeersch, T.; Hutchison, G. R. Open Babel: An Open Chemical Toolbox. *J. Cheminformatics* **2011**, 3 (1), 33.
<https://doi.org/10.1186/1758-2946-3-33>.
- (15) Vilar, S.; Cozza, G.; Moro, S. Medicinal Chemistry and the Molecular Operating Environment (MOE): Application of QSAR and Molecular Docking to Drug Discovery. *Curr. Top. Med. Chem.* **2008**, 8 (18), 1555–1572.
<https://doi.org/10.2174/156802608786786624>.
- (16) Vanommeslaeghe, K.; MacKerell, A. D. Automation of the CHARMM General Force Field (CGenFF) I: Bond Perception and Atom Typing. *J. Chem. Inf. Model.* **2012**, 52 (12), 3144–3154. <https://doi.org/10.1021/ci300363c>.
- (17) Vanommeslaeghe, K.; Raman, E. P.; MacKerell, A. D. Automation of the CHARMM General Force Field (CGenFF) II: Assignment of Bonded Parameters and Partial Atomic Charges. *J. Chem. Inf. Model.* **2012**, 52 (12), 3155–3168.
<https://doi.org/10.1021/ci3003649>.
- (18) Vanommeslaeghe, K.; Hatcher, E.; Acharya, C.; Kundu, S.; Zhong, S.; Shim, J.; Darian, E.; Guvench, O.; Lopes, P.; Vorobyov, I.; Mackerell, A. D. CHARMM General Force Field: A Force Field for Drug-like Molecules Compatible with the CHARMM All-Atom Additive Biological Force Fields. *J. Comput. Chem.* **2010**, 31 (4), 671–690.
<https://doi.org/10.1002/jcc.21367>.
- (19) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. CHARMM: A Program for Macromolecular Energy, Minimization, and Dynamics Calculations. *J. Comput. Chem.* **1983**, 4 (2), 187–217.
<https://doi.org/10.1002/jcc.540040211>.