Interrogation of Dynamic Proteins to Expand the Druggable Proteome

by

Amanda Lee Peiffer

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

Doctoral Committee:

Professor Charles L. Brooks III, Co-Chair
Professor Anna K. Mapp, Co-Chair
Associate Professor Tomasz Cierpicki
Assistant Professor Aaron T. Frank
Associate Professor Bruce A. Palfey

Amanda Lee Peiffer
apeiffer@umich.edu
ORCID iD: 0000-0001-8113-9458

Acknowledgements

I never pictured what it would be like to sit down and write my acknowledgements for my PhD thesis, but as I sit here typing this up (in the single outfit I've worn for what feels like weeks while writing), I'm finding it one of the more difficult pages to produce, mostly because I'm not convinced that I can truly convey in words the level of gratitude I feel.

First and foremost, I have to thank my advisors, Professor Charlie Brooks and Professor Anna Mapp — first for agreeing to take me on as a student, and second for not firing me along the way. I feel so lucky to have found two unbelievable scientists who beyond everything else, are human. You both have allowed me so much creative freedom in my work, and you've been supportive and guiding forces along the way.

Charlie, thank you for constantly pushing me to do better — to *be* better. You were always encouraging while making sure to challenge me; and anyone who knows me knows I thrive on being challenged. You changed the ways in which I think about science, which is an invaluable gift. I will forever be grateful for ridiculously funny conversations we had, like when you called me "nose ring" from across the room just to see if I would respond. It's moments like these that stick with me and always make me smile. In addition, I personally will always appreciate the mentality you have for mentoring graduate students: if your students are well mentored and succeed, they will be your future collaborators. I hope that this is true for me (and if not, I take back all of the nice things I said before this).

Anna, your unwavering support for the past five years has been truly incredible and way beyond what I could have ever expected. I have never met someone so open and willing to hear ideas being proposed. I am particularly grateful for the time you answered my Friday evening email about the COVID conspiracy theories I saw on Fox News, and that you subsequently went on CNN with me to discuss it. And for answering

the other email about wanting to start a COVID research project. I have always felt like you believed in me, and these two emails turned out to be the highlight of my time here. So thank you, truly.

To my dissertation committee, thank you for the guidance and feedback along the way. Your input was instrumental in my success. And of course, thank you to everyone in the Mapp and Brooks labs. I have always felt so lucky to be surrounded by such a great group of people, who have always provided encouragement and feedback along the way. Thank you Yujin for being an awesome collaborator and friend in the Brooks lab. And thank you, Rachel, for teaching me how to use Illustrator, and for getting me into graphic design. You opened up a whole new creative world for me, and it made my thesis so much more fun. Thank you.

Julie, thank you for being my scientific counterpart and one of my closest friends. I have never met anyone who has been more excited to talk science or Survivor. Our near-daily walks at the beginning of the pandemic were where we really developed the TMPRSS2 project, which will always be one of my most valued memories from grad school. You're a phenomenal scientist. And your friendship is one of the best things I am leaving Ann Arbor with. So thank you.

My family has been so important in my PhD journey. I want to thank my parents for always encouraging me to do what makes me happy, and to carve my own path in life. To my sisters, who have become my closest friends — thank you for honestly just being you. Lindsey, thank you for being that most amazing big sister. Our FaceTime sessions during my time in Michigan were always the best way to boost my mood on any day. You're an unbelievable mom, and your kids are my heart. And Erin, you may be my little sister, but I swear I look up to you. Your ability to hold your own in any situation, and your unwavering moral compass are truly incredible. Thank for your continuous encouragement and compassion. And importantly, to my practically-sister, Kara. You may be half my height, by you count twice as much. You've always been the most incredible friend, and your Ann Arbor visits were always the best.

To literally all of my educators along the way — you have molded me into the person I am today. I have never related more to a mission statement than my alma mater, Denison University: "Our purpose is to inspire and educate our students to become

autonomous thinkers, discerning moral agents and active citizens of a democratic society." Thank you, Denison for pushing me to live by this. My time there made me a better human being. Thank you to my unbelievable undergraduate mentors, Dr. Specht and Dr. Ludwig. You both were absolutely instrumental in my success. You taught me how to critically think and, more important, how to question. Thank you for everything.

Lastly, Landon. Life has been so unbelievably strange, yet your support has never wavered. I never would have expected to fall in love while in college, and I will always appreciate our chance encounter in the rain. You have changed my life, and we've made a home. Thank you for cooking me meals when I was living off of peanut butter; thank you for constantly asking me what I need during this crazy week as I wrap up my dissertation; and thank you for loving our cats as much as two humans possibly can. I can't wait to marry you. I love you.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

ABD	Activator binding domain
ACE2	Angiotensin-converting enzyme 2
AcID	Activator interaction domain
AMC	7-amino-4-methylcoumarin
ATF6a	Activator transcription factor 6a
BD	Bromodomain
BET	Bromodomain and extra terminal
BME	β-mercaptoethanol
c-Myb	Cellular MYB
CD	Circular dichroism
CBP	CREB binding protein
CDOCKER	CHARMM docking
CGENFF	CHARMM General Force Field
CH2	Cysteine-histidine-rich region 2
CHARMM	Chemistry at HARvard Macomolecular Mechanics
CMK	Chloromethylketone
COVID-19	Coronavirus disease 2019; 2019-nCoV
Cryo-EM	Cryo-electrom microscopy
CSP	Chemical shift perturbation
DBD	DNA binding domain
DIPEA	N,N-Diisopropylethylamine
DMF	Dimethylformamide
DMN	4-*N*,*N*-dimethylamino-1,8-naphthalimide
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid

DSF	Differential scanning fluorimetry
DTT	Dithiothreitol
EDT	Ethanedithiol
EDTA	Ethylenediaminetetraacetic acid
ETV	Ets translocation variant
FOY251	4-(4-guanidinobenzoyl-oxy)phenylacetic acid
FPLC	Fast protein liquid chromatography
GACKIX	Gal11, Arc105, CBP/p300, kinase-inducible domain interacting (KIX)
GBMV	Generalized Born using Molecular Volume
GBSW	Generalized Born with simple switching
GPU	Graphic processing unit
GSH	Reduced glutathione
GSSG	Oxidized glutathione
HAT	Histone acetyltransferase
HPLC	High performance liquid chromotagraphy
HRP	Horseradish peroxidase
HSQC	Heteronuclear single quantum coherence
HTS	High throughput screening
iBiD	IRF-3 binding domain
IC_{50}	Half-maximal inhibitory concentration
IDP	Intrinsically disordered protein
IMAC	Immobilized metal iron affinity chromatography
ITC	Isothermal calorimetry
IPTG	Isopropyl-β-D-1-thiogalactopyranoside
KIX	Kinase inducible domain interacting domain
KNF	Koshland, Nemethy and Filmer
LC/TOF	Liquid chromatography/ time-of-flight
LE	Ligand efficiency
L-S	Lipari-Szabo
MD	Molecular dynamics
Med15	Mediator subunit 15

MED25	Mediator subunit 25
MERS	Middle Eastern respiratory syndrome
MLL	Multi-lineage leukemia
MMTSB	Multiscale Modeling Tools for Structural Biology
MOE	Molecular Operating Environment
mRNA	Messenger RNA
MW	Molecular weight
MWC	Monod-Wyman-Changeux
Ni-NTA	Nickel-Nitrilotriacetic Acid
NMR	Nuclear magnetic resonance
NR	Nuclear receptor
OD	Optical density
OP	Order parameter
PC4	Positive Cofactor 4
PDB	Protein DataBank
PEA3	Polyoma enhancer activator 3
PHD	Plant homeodomain
pKID	Phosphorylated kinase inducible domain (CREB)
PPACK	FPR-Chloromethylketone
PPI	Protein-protein interaction
PPM	Part per million
R state	Relaxed state
RFU	Relative fluorescence units
RMSD	Root mean square deviation
RMSF	Root mean square fluctuations
RNA	Ribonucleic acid
RNA Pol	RNA Polymerase II
RT	Room temperature
S protein	Spike protein
SARS	Severe Acute Respiratory Syndrome
SARS-CoV-2	Severe acute respiratory coronavirus 2

SD	Standard deviation
SRCR	Scavenger receptor cysteine-rich
T state	Tense state
RPM	Revolutions per minute
TAD	Transcriptional activation domain
TAZ	Transcriptional adapter zinc finger
TBST	Tris-Buffered Saline with Tween
TFA	Trifluoracetic acid
TMPRSS2	Transmembrane protease, serine 2
UV/Vis	Ultraviolet/visible spectroscopy
VP16	Herpes simplex virus protein 16
VWA	Von Willebrand Factor Type A
WT	Wild-type

List of Appendices

Abstract

The human proteome is vastly complex, and our understanding of it is constantly evolving. There are roughly 20,000 protein-coding genes in the human genome, yet only about 10% of the resultant proteins are deemed “druggable” targets. And, only half of those have disease relevance. Thus, the druggable proteome is surprisingly narrow, consisting largely of structured proteins with defined binding pockets. With so many disease signatures residing in the “undruggable” portion of the proteome, there is much work to be done to expand the druggable landscape. An area rich with disease relevance is dynamic protein-protein interactions (PPIs), which underpin many regulatory cellular functions both in healthy and diseased states. However, devoid of typical binding pockets that enable traditional drug discovery approaches (i.e. substrate mimicry), dynamic PPIs occur over large, flat surface areas, which is why they have remained “undrugged.”

A disproportionate number of dynamic proteins can be found in transcriptional regulation. As such, it provides an interesting avenue for chemical probe development and therapeutic intervention. For instance, a hallmark of cancerous cells are rampant growth and proliferation, with many proteins being overexpressed. While many research efforts have focused on targeting the overexpressed proteins themselves, halting the overexpression at the transcriptional level could stop the disease progression at its initiation.

This dissertation works towards expanding the druggable proteome by establishing principles of molecular recognition that guide native PPIs. By primarily using molecular dynamics simulations, with complementary biophysical experimentation, I dissect coactivators and establish rules of activator recognition and engagement. In doing so, I demonstrate the utility of disorder in transcriptional regulation. In particular, I identify ways in which allostery manifests in dynamic coactivator proteins. Further, I explore how

inhibition / enhancement of particular PPIs can be achieved using small molecules that attenuate fluctuations and disrupt binding allosterically.

The connections between allostery and disorder are explored in Chapter 1. There, I go through the historical context for how allostery was discovered as well as the models proposed around the underlying physics. I highlight how experimental advances over many decades has demonstrated that dynamic proteins have a high propensity for cooperative interactions, where binding of first ligand at one site stabilizes the dynamic protein for enhanced/inhibited binding of a second ligand at a distal site. This introductory chapter thus sets the foundation for exploring ways in which allostery manifests in the highly dynamic PPIs between transcriptional coactivators and activators.

In Chapter 2, I utilize simulations to interrogate the wide distribution of conformational micro-states available in the apo form of the coactivator domain CBP KIX. I dissect the ways ternary complex formation and allosteric communication can be traced through redistributing relative populations of KIX configurational states. In doing so, I demonstrate that high conformational entropy allows for KIX to recognize multiple partners, and that subsequent binding winnows down the number of accessible micro-states. This in turn results in allostery, where a redistribution of micro-states increases/decreases the likelihood of a second binding event. Importantly, I investigate how a small molecule can allosterically inhibit binding, finding that it does so by winnowing down the distribution of conformational states to ones that disfavor a second binding event. Thus, I finish by demonstrating ways in which druggability could be achieved with this difficult target using an allosteric small molecule instead of an orthosteric inhibitor.

In Chapter 3, I utilize enhanced sampling techniques to explore binary complex formation with the coactivator Med25 AcID. While this protein has been structurally elucidated through NMR experiments, there are no current structures of it in complex with any activators. Through my simulations, I demonstrate how a structurally unique coactivator like Med25 AcID, containing a stable β-barrel core, follows similar principles of activator engagement as the helical bundle KIX; ultimately, I illustrate how Med25 utilizes flanking loops and helices to accommodate binding partners. I use molecular dynamics to dissect ways in which the covalent small molecule norstictic acid is able to bind to the H2 face and allosterically inhibit PPIs on the H1 face. In doing so, I establish

that altering protein dynamics through a small molecule can alter fluctuations across the protein, lending insights into future drug development strategies.

With the emergence of the COVID-19 pandemic, I switch focus in Chapter 4 to a transmembrane serine protease that has been implicated in coronaviral infections named TMPRSS2. TMPRSS2 has been shown to prime the viral spike protein, which promotes membrane fusion of the viral particle and the host. With little known about the protease, I utilize a combined computational and *in vitro* experimental workflow for identifying new inhibitors for the protease. In doing so, I find several clinically approved compounds that effectively inhibit proteolytic cleavage of substrates, which can be explored in drug repurposing for COVID-19.

In Chapter 5, I summarize the findings discussed throughout this thesis and establish future directions for the projects. I describe strategies for “drugging” dynamic protein-protein interactions namely between coactivators activators, highlighting how dynamic substructures on the proteins can be targeted for allosteric inhibition/enhancement. Through this work, I expand the feasibility of drugging the “undruggable” classes of proteins.

Chapter 1

Molecular Recognition in Dynamic Transcriptional Protein-Protein Interactions

Abstract

One of the major limitations in drug discovery is the small number of proteins that are deemed “druggable.” Intrinsically disordered proteins (IDPs) rank high as some of the most unfeasible targets, and yet their abundance in regulatory cellular processes, both healthy and diseased, make them highly desirable drug targets. In particular, a disproportionate number of IDPs appear in transcriptional regulation. Activators and coactivators exhibit high conformational plasticity, which has made characterization of the binding mechanisms challenging. By better understanding the molecular recognition in dynamic PPIs, we can better strategize for targeted inhibition. As allostery often appears in conjunction with dynamic proteins, understanding how allostery manifests in activator-coactivator interactions would provide key insights into chemical probe development, which is vital first step in expanding the druggable proteome.

1.1 Protein structure-function relationship

In their simplest forms, proteins are merely strings of amino acids. What distinguishes them from one another involves how the varying patterns of primary sequences manifests a structural form through secondary, tertiary and quaternary interactions to ultimately perform a cellular function. Once properly folded, a protein can catalyze reactions,[1] join other macromolecules and form higher structured complexes,[2] and so on (signaling,[3] structural support,[4] etc.). Thus, a protein’s function is determined through its three-dimensional fold.[5]

Theories surrounding the relationship between protein structure and function can be traced back to the earliest days of protein crystallography.[6] Myoglobin, a monomeric protein responsible for binding molecular oxygen, was the first protein structure to be

solved (*Figure 1.1A*).[7] Shortly thereafter, the functionally similar protein hemoglobin, a heterotetrameric protein responsible for transporting molecular oxygen, had its structure solved using crystallography (*Figure 1.1B*).[8] Immediately evident is the structural similarities between the two proteins; superimposing their structures shows that the two deviate by only 2.3 Å backbone root-mean-square-deviation (RMSD). Surprisingly, however, the two proteins share only 17% of sequence similarities. The obvious similarities connecting protein fold to biological roles led researchers to deduce that a protein's structure can be explained by its structure.

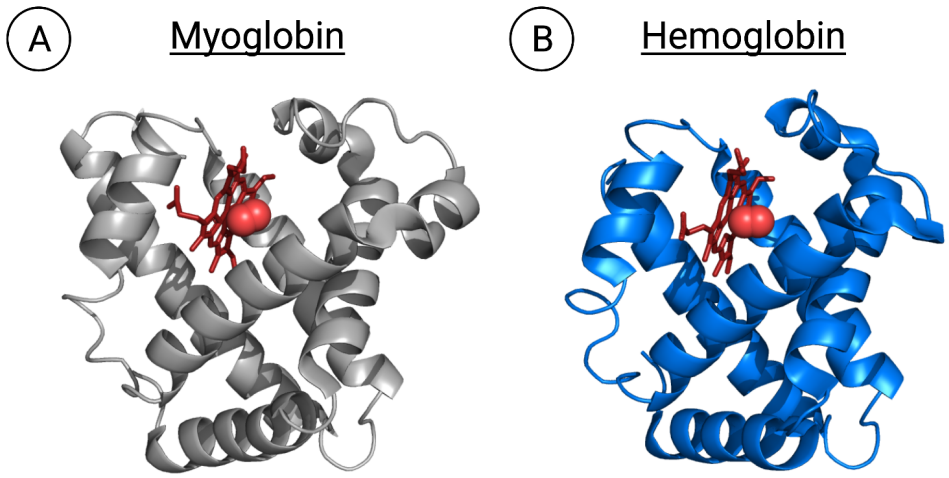


Figure 1.1. *Myoglobin and hemoglobin x-ray crystallography structures.* (A) The structure of myoglobin (PDB 1MBN)[9] and (B) a monomer of hemoglobin (PDB 1HHO)[10]. Despite the significant lack of sequence similarities between the two proteins (17% shared sequence identity), they share a both a similar protein fold and functional purpose, i.e., binding molecular oxygen. When superimposed, the two structures have a 2.3 Å backbone RMSD difference.

1.2 Allostery in well-folded proteins

Even though hemoglobin and myoglobin exhibit similar functions (i.e. binding molecular oxygen), they do have significant differences, which can explain why the two proteins differ in sequence by 83%. Myoglobin is primarily located inside muscle cells, and it shows a high affinity towards molecular oxygen, as its role is oxygen *storage*.[11] Alternatively, the major role of hemoglobin is for oxygen *transport*, meaning that it needs to not only be able to bind oxygen, but also (and perhaps more importantly), it needs to be able to release it.[6,12] Thus, hemoglobin shows a lower affinity for oxygen, allowing for it to be released as needed.

In trying to better understand the relationships between protein structure and function, hemoglobin was found to occupy two structural states: the Tense state (T state), which is the deoxy-form of the protein, and the Relaxed state (R state), which is the fully oxygenated form.[13] Thus, substrate binding was connected to conformational changes, however it was not yet clear how. As a tetrameric protein, hemoglobin has four subunits, each being able to bind molecular oxygen, with a fully oxygenated hemoglobin binding four molecules of oxygen. Further, what researchers found was that if one hemoglobin subunit bound oxygen, it increased the affinity for subsequent oxygen binding.[13–15] Hence, there was some cooperativity that occurs, called allostery, where binding of a molecule increases the affinity for a subsequent binding event at a different location.[13,16]

Two general models emerged to explain allostery in hemoglobin, and both relied on the two physical states determined from crystallography.[16,17] The Monod-Wyman-Changeux (MWC) model suggests that the T and R states exist in some equilibrium. Subsequent binding of oxygen would change the equilibrium towards the R state (i.e. bind more O_2), whereas unbinding of an oxygen would shift it towards the T state (*Figure 1.2A*).[16] Alternatively, the Sequential of KNF model developed by Koshland, Nemethy, and Filmer, was used to describe “induced fit” binding, where cooperative effects with oxygen binding could be attributed to induced structural changes that occur in the protein to make binding more oxygens more favorable (*Figure 1.2B*).[17] As hemoglobin is a well-folded protein, the KNF model attributed cooperativity to structural changes and mechanical coupling from one protein site to another.[17] Binding sites were thought to be linked by communication networks of amino acid contacts, where an initial binding event would cause structural rearrangements that increase the affinity for subsequent binding.

The two different models for allostery were both able to account for measured cooperative effects in hemoglobin, making it difficult to truly dissect what could be going on. In the MWC model, differences in binding interactions came from changing the equilibrium between two states. In the KNF model, mechanical coupling caused physical structuring at a distal site. While slight variations on the two models were put forth, many shared one central component: the protein’s highly structured nature was a necessary component for allostery.

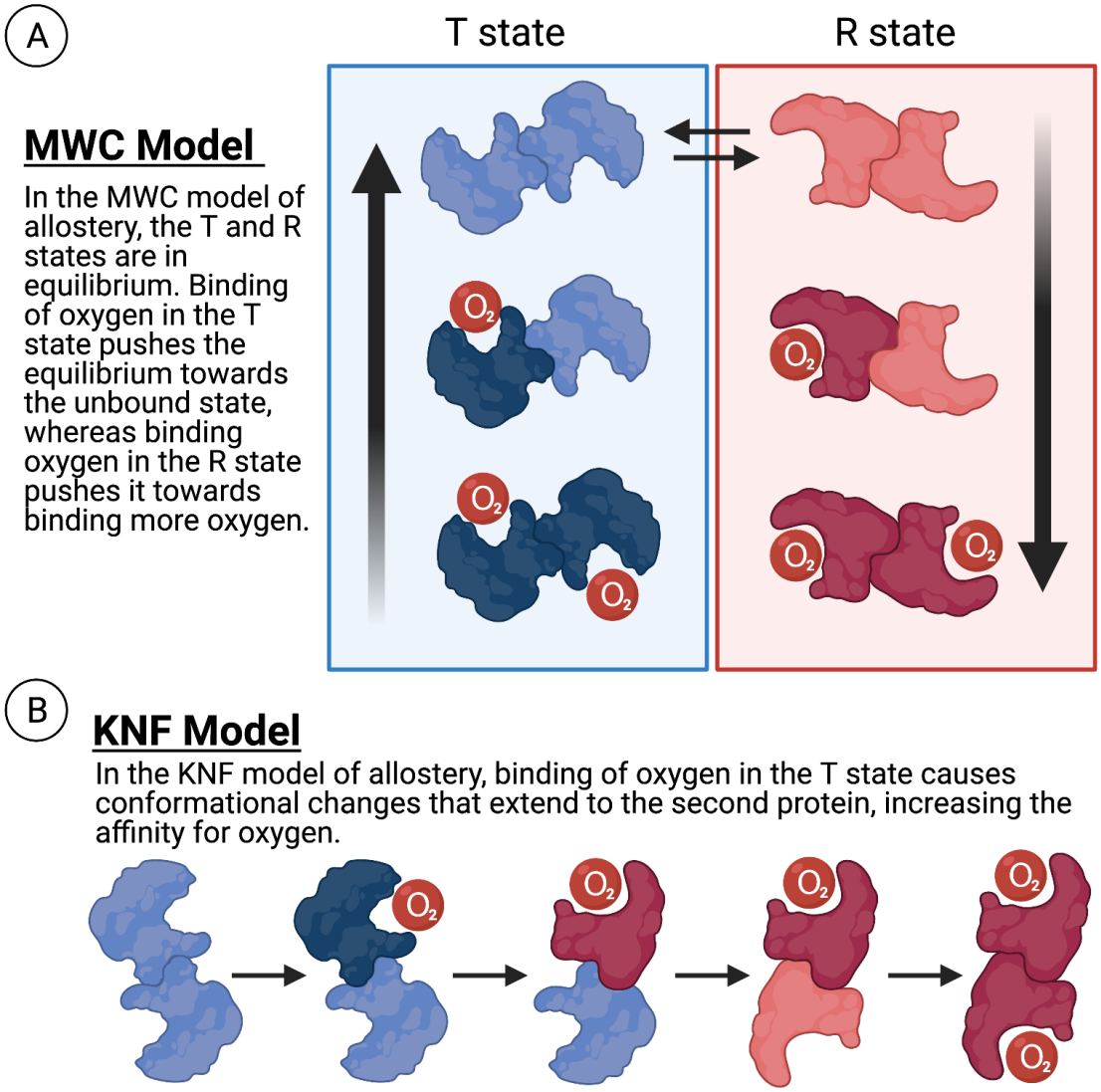


Figure 1.2. *The Monod-Wyman-Changeux (MWC) and Sequential (KNF) models for allostery in hemoglobin.* Shown here is an example of cooperativity using two domains, where oxygen-bound states are darker colored and unbound states are lighter. (A) In the MWC model, the T state (blue) and the R state (red) exist in equilibrium. Binding of oxygen molecules shifts the equilibrium towards the R state, increasing the affinity for additional oxygen molecules. (B) In the KNF or Sequential model, allostery is a result of conformational changes that occur both in the domain that binds oxygen as well as the allosterically connected domain. These conformational changes towards the R state increase the affinity for oxygen.

1.3 Intrinsically disordered proteins: breaking the structure-function rules

Today, over 170,000 protein structures have been elucidated by experimental methods, which has excelled our understanding of cellular biology and the proteome.[18] However, such experiments also led scientists to believe protein folding is a prerequisite for functional output (i.e. disordered proteins lack function).[19] Through decades of new

experimental and theoretical developments, protein science has grown beyond highly structured domains, recognizing the many roles such proteins play in essential biology.[20,21] In doing so, competing theories of molecular recognition have been put forward.

At the center of the allosteric models of hemoglobin are questions about conformational changes involved in ligand binding, an idea that extends to other protein types. Protein structural changes can be thought to exist on a spectrum, from no changes with ligand binding to significant conformational rearrangements. For instance, enzyme-substrate interactions were initially thought to subscribe to “Lock and Key” mechanisms of engagement, first proposed by Emil Fisher in 1894. In this model, conformational changes are essentially nonexistent, as enzymes were believed to have well-formed binding pockets that complement substrate shape (*Figure 1.3A*). Alternatively, the “Induced Fit” model for binding aligns with the KNF model for allostery, where ligand binding induces some structural changes in a protein (*Figure 1.3A*).[22] However, in both the “Lock and Key” as well as the “Induced Fit” models, a protein’s structure is relatively fixed before binding.

While the development of crystallographic techniques in the mid-1950s allowed for atomic-level visualization of complicated biomolecules, it led many to falsely believe that proteins are rigid. Fortunately, the first molecular dynamics (MD) simulations brought forth a new picture of proteins that included dynamics.[23] Simulations showed that even rigid proteins demonstrate atomic-level fluctuations in a thermal distribution, suggesting that the models of molecular recognition were overly simplistic and misleading.

The implementation of statistical mechanics into models of protein interactions views proteins not as static structures, but instead as a distribution of conformational micro-states that are in some equilibrium before binding. The MWC model for hemoglobin lies within this thinking, deemed “Conformational Selection” (*Figure 1.3A*).[24,25] In it, ligands select for states that enable the most favorable interactions, which redistributions relatively populations of states. Lastly, “fuzzy” interactions defy easy classification, with disordered proteins showing the propensity to sometimes fold when binding (*Figure 1.3A*).[21,22,26]

While each of these classifications appear to be distinct (Lock and Key; Induced Fit; Conformational Section; “Fuzzy” interactions), the ensemble framework for protein conformational states from statistical mechanics can be applied to all theories (*Figure 1.3B*).[27,28] There are constant atomic movements in proteins, even enzymes who follow a general “Lock and Key” mechanism of binding. Thus, even the most rigid proteins still occupy a thermal distribution of conformational states; however, it is a narrow distribution of few highly populated structures (*Figure 1.3B*, solid pink line). Further, as the protein is already rigid, the conformational distribution does not change much with substrate binding (*Figure 1.3B*, dashed lines). The “Induced Fit” model is similar, except now instead of minorly narrowing the distribution of states, the distribution shifts to a new (or lowly populated) state (*Figure 1.3B*, orange distribution). In “Conformational Selection,” states exist in equilibrium, and as ligands have preferably protein conformations they bind, the distribution of protein states shifts (*Figure 1.3*, yellow distribution). Lastly, “fuzzy” interactions show the propensity for a protein to occupy a wide distribution of lowly populated conformational states, with binding resulting in some structural ordering. Thus, protein ensembles demonstrate how multiple mechanisms for ligand engagement can be reflected in the redistribution of conformational states, a direct reflection of protein conformational entropy.

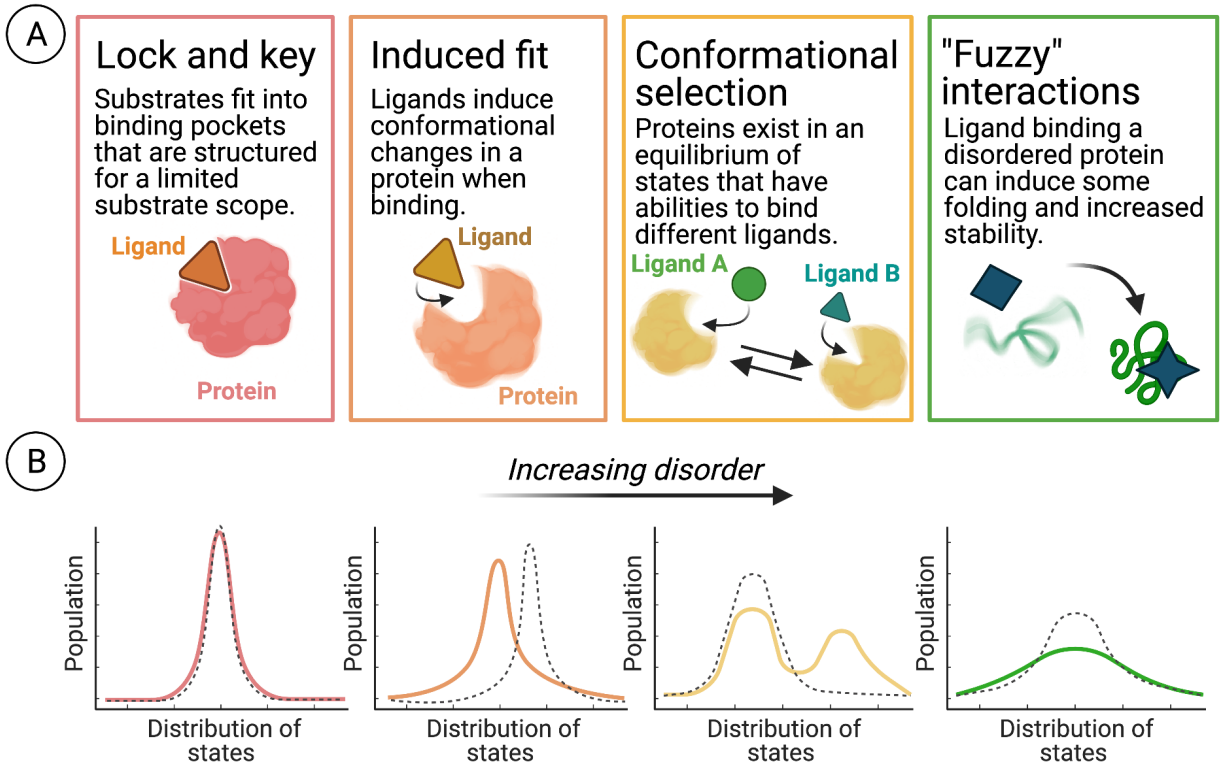


Figure 1.3 *Binding modes in the ensemble framework of molecular recognition and allostery*. (A) Binding modes of proteins and ligands can be classified by conformational changes associated with binding, starting with the smallest conformational changes (Lock and Key) and going all the way to interactions with intrinsically disordered proteins ("fuzzy" interactions). (B) The ensemble framework of protein conformational states can be applied to all classifications. The apo protein distributions are shown with solid-colored lines corresponding to each classification. Dashed lines show how the distribution of protein structural states are limited through binding.

1.4 Intrinsic disorder and allostery

Over the years, experimental advancements have led to an abundance of disordered proteins being discovered.[20,21,29,30] The emergence of unstructured/dynamic proteins has challenged previous intuitions about protein structure-function relationships, shedding light on the diverse and critical roles dynamic proteins take on. This shift, in part, can be attributed to the high number of disordered proteins that are exclusive to higher-order organisms. In fact, over 25% of proteins in high order organisms are deemed disordered, indicating an evolutionary benefit.[30–33] However, this appears to be inconsistent with initial structure-function models of proteins in which only well-folded domains are functional. The question that follows is how disorder contributes to cellular/organismal complexity.

As with many other classifications, dynamic proteins come in different “flavors.” Intrinsically disordered proteins (IDPs) lacky bulky hydrophobic groups, and thus do not follow basic protein folding rules, where proteins form three dimensional shapes by burying hydrophobic residues internally. Instead, IDPs are very dynamic and do not possess any measurable structure.[29] Intrinsically disordered domains can occur within a protein, where long unstructured linker regions separate folded domains. Disorder like this can allow for binding to multiple partners. Further, while some interactions between IDPs maintain disorder, other interactions induce sequential binding followed by folding.

In exploring questions around how disordered proteins function, scientists discovered that disordered domains could exhibit high levels of allostery, challenging the two models for cooperative binding interactions established from hemoglobin studies.[28,30,34,35] For instance, the MWC model states that a protein exists in an equilibrium that allosterically couples two sites. More conflicting is the KNF model, which attributes ligand binding and mechanical coupling as the driver of allostery, thus requiring a network of amino acid contacts in a well-structured protein.

The ensemble framework for understanding protein structures has challenged previous notions about molecular recognition, and it has propelled our understanding of allostery. As seen in *Figure 1.4*, proteins exist in equilibrium in a distribution of states, with binding interactions narrowing and/or shifting the distribution. Thus, allostery can be viewed as altering the populations of distributions to states that allow for more favorable interactions with a second ligand.[27] This framework has been able to account for a greater range of allosteric processes, especially those that emerge from more disordered proteins, which have been shown to exhibit high levels of allostery.[34,36]

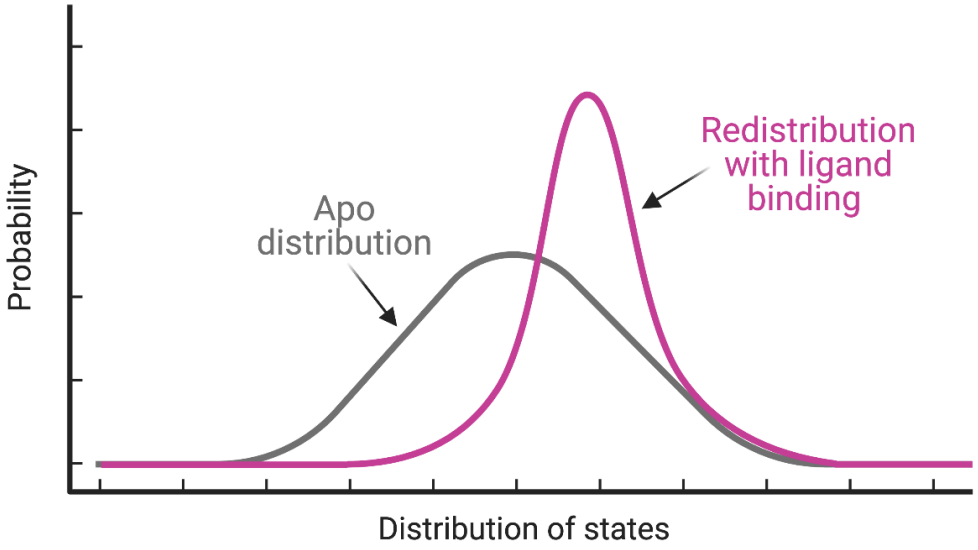


Figure 1.4. *The ensemble framework of allostery.* While the MWC and KNF models of allostery are reliant on having two structural states associated with high and low affinity interactions, the energetic basis of allostery shows how disordered proteins can exhibit cooperative binding by reweighting the distribution of accessible conformational micro-states to ones that offer more favorable interactions with a ligand.

1.5 Functional disorder: the utility of dynamics in transcriptional regulation

One of the ways disordered proteins operate is by acting as signaling "hubs" in protein-protein interactions.[20,20,29] Transcription is an essential and heavily regulated cellular process that involves a disproportionate number of disordered proteins.[32,37,38] This is particularly true of coactivators and activators.[29,39,40] While these proteins do contain folded domains, they are notorious for containing large portions of completely unstructured regions separating folded domains.[40] Further, the folded domains themselves have exhibited high conformational plasticity, with many interactions between activators-coactivators showing structural fuzziness, having multiple bound conformational states separated by low energy barriers.

Through a complicated network of interactions, transcriptional proteins coalesce to the promoter region of a gene and initiate expression (*Figure 1.5*).[37] Coactivator proteins serve as communication hubs in the transcriptional process, playing an important role in RNA polymerase II (RNA Pol II)-mediated transcription.[41] At the promoter or enhancer region of DNA, the DNA-Binding Domain (DBD) of a transcription factor recognizes a specific promoter sequence and binds (*Figure 1.5*). If the transcription factor

is an activator, it contains a transcriptional activation domain (TAD), which is used to assemble the remaining transcriptional machinery to turn on gene expression. The recruitment of RNA Pol II to the promoter region of a gene is of particular importance for coactivators, which transcribes the gene of interest into RNA.

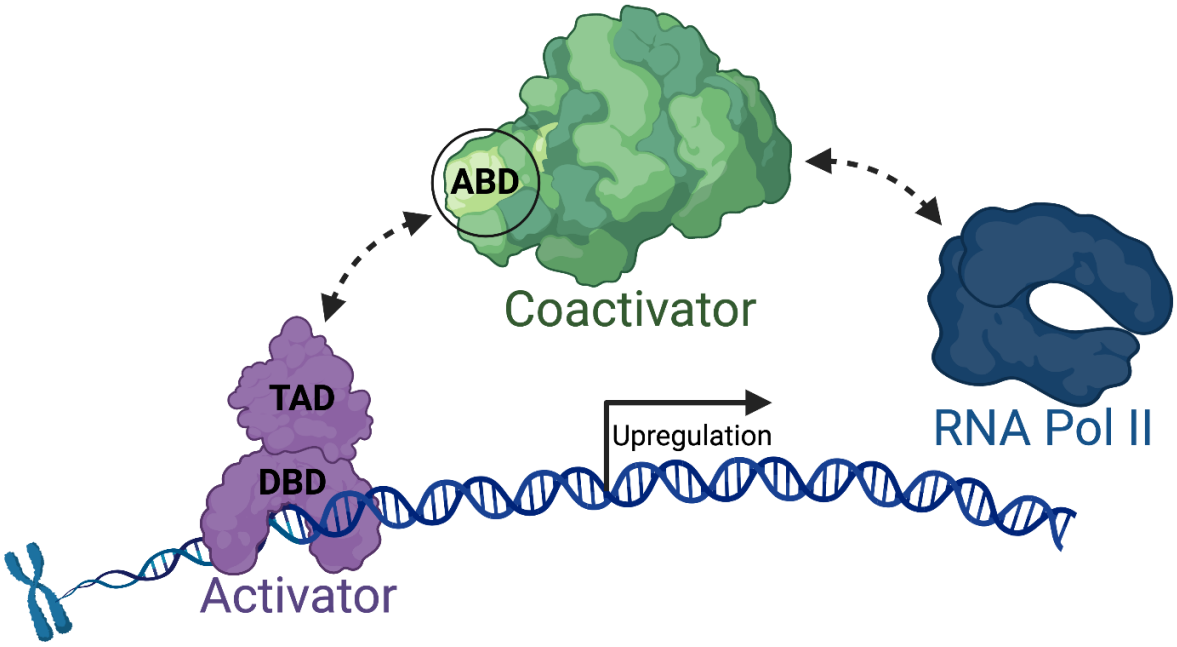


Figure 1.5. *Transcription is initiated by assembling all of the necessary cellular machinery*. Transcriptional activators (purple) bind at the promoter region using a DNA binding domain (DBD). The transcriptional activation domain (TAD) interacts with activator binding domains (ABDs) of coactivators (green), which aid in the recruitment of RNA polymerase II (blue) and ultimately starts the gene expression.

1.6 Coactivators serve as communication hubs

The role of dynamic/disordered domains in transcription is to interact with multiple binding partners depending on the cellular context.[37,40,41] For example, a person contains multiple cell types that all contain the same genetic composition (i.e. DNA), yet they are differentiated by levels of gene expression, which results from particular transcriptional PPIs prompting expression for cell-specific needs.[42] Coactivators are an important component of this, interacting with a wide array of protein classes that includes RNA Polymerase II, epigenetic modulators, transcription factors, and even other coactivators (*Figure 1.6*). Hence, coactivator proteins coordinate the necessary transcriptional machinery through these interactions with many types of proteins.

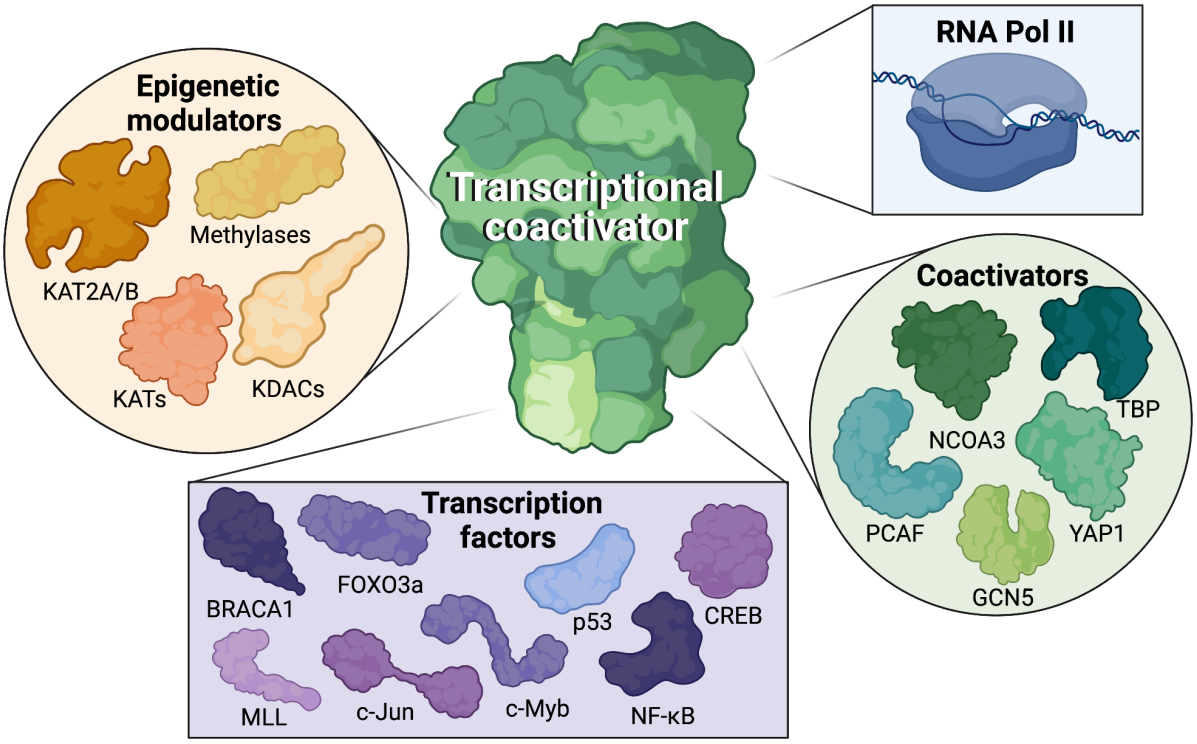


Figure 1.6. *Coactivators are interaction hubs in transcriptional regulation.* Coactivators possess the ability to interact with a wide range of different protein types, including epigenetic modulators (yellow), RNA Pol II (red), transcription factors (purple), and other coactivators (green). The specific proteins listed in each category correspond to binding partners for the coactivator CBP.

The CREB Binding Protein (CBP) along with the homolog p300 are coactivators that are important in regulating genes.[43] Both are multi-domain proteins, containing activator binding domains (ABDs) like KIX and iBid; zinc-finger domains like TAZ1, TAZ2, PHD, ZZ, and RING; a histone acetyl-transferase (HAT); a bromodomain (BD); and a nuclear receptor.[44,45] Through all of these domains, these proteins recognize over 400 binding partners and elicit multiple roles in the transcriptional process (*Figure 1.6*).[46,47] While the individual domains in these coactivators do have secondary and tertiary structures, there are very long disordered regions separating the folded domains (*Figure 1.7*).[40] The domains of CBP/p300 have been studied in isolation,[42,48,49] demonstrating high flexibility in many portions. However, the long linker regions that are completely unstructured have only recently been explored to elucidate their function.[40] And while initially thought to be merely the connecting pieces to hold each folded domain together,

the disordered regions have now been shown to have binding interactions with other proteins and are important for overall function of CBP/p300.

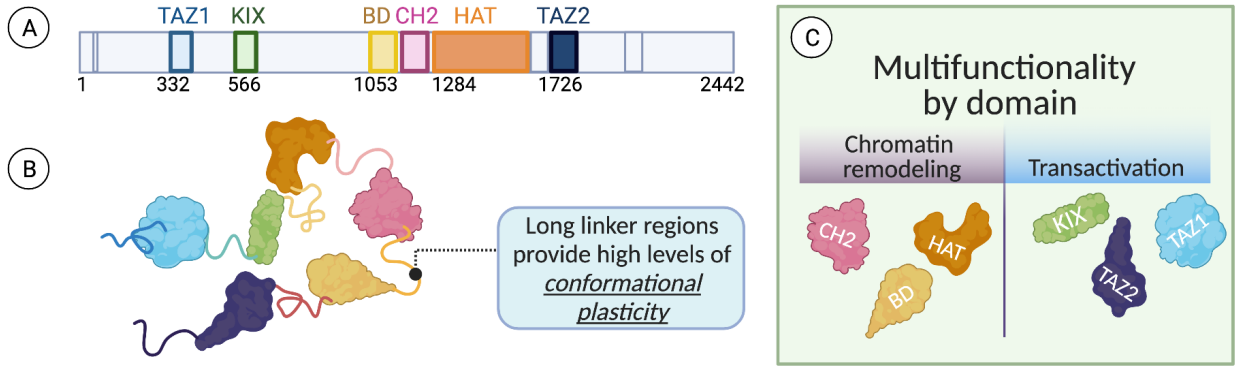


Figure 1.7. *Domains of CBP*. (A) The gene corresponding to CBP with domains annotated shows how almost half of the protein is linked by long, intrinsically-disordered regions. (B) Cartoon representation showing how long linker regions separate the folded domains of CBP. (C) Each domain of coactivators like CBP elicit independent functions and have their own set of cognate binding partners.

Similar to CBP/p300, the coactivator known as the Mediator complex that has been deemed as a "master coordinator" in gene regulation.[50–52] However, unlike CBP/p300, the Mediator is a massive protein complex (>1 mDa) composed of around 30 proteins.[53] In addition, the Mediator only exists in eukaryotes.[54] With 3 general subunits along with a cyclin kinase module, this massive protein complex is critical in assembling necessary biomolecules to initiate transcription, relying on each domain for critical interactions (*Figure 1.8*). Of the four subunits, the tail subunit has proven to be the most difficult to obtain structural information for, as it has been shown to undergo major conformational rearrangement. The tail subunit is of particular interest, as it contains domains that directly interact with transcriptional activators and repressors, and portions of it are only found in higher eukaryotes, including Med25.[50,51,55]

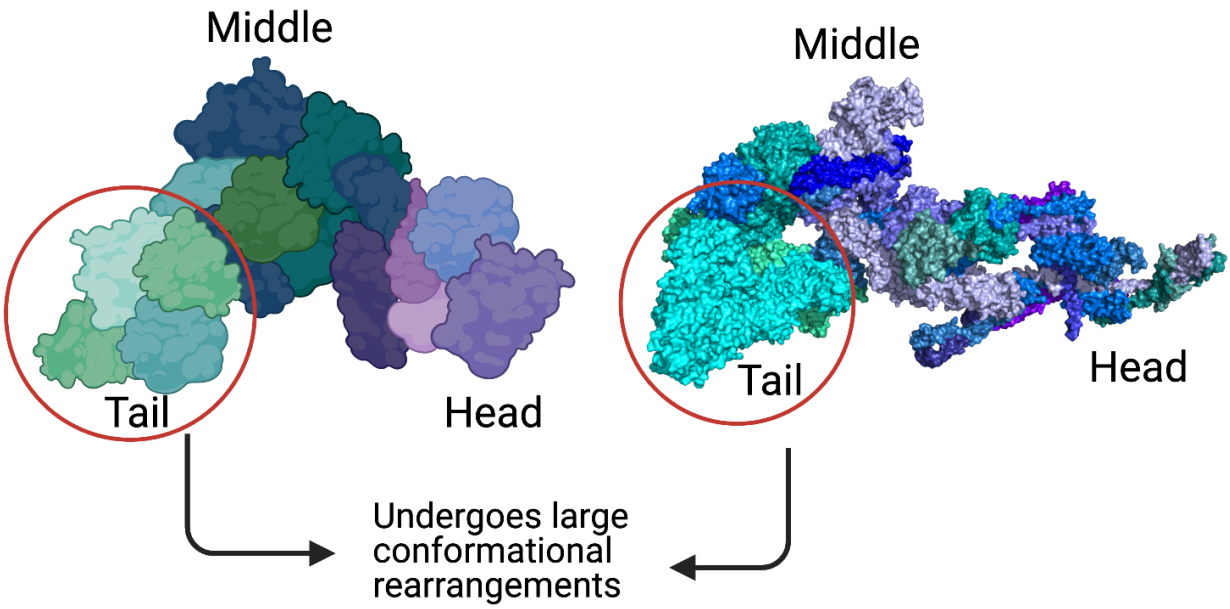


Figure 1.8. *Composition of the Mediator complex*. A cartoon representation (left) of the Mediator complex along with a cryo-EM structure of the complex (right; PDB 7LBM). The experimentally solved structure contains more tail subunits than previously-solved Mediator structures, however it is still not completely resolved because of the large conformational changes that have been shown to occur in that portion of the complex. For example, the activator binding domain of Med25 is missing.

1.7 Activator binding domains

Activator binding domains (ABDs) of coactivators are responsible for directly interacting with the DNA-bound transcriptional activators.[37] ABDs have been shown to typically be structurally dynamic, comprised mostly of helices and loops. This is immediately clear upon looking at solution NMR structures that have been solved for various ABDs (*Figure 1.9*). Several ABDs have been shown to take on similar folding patterns, with the prototypical example being helical bundles like KIX (*Figure 1.9*). Ultimately, the intrinsic mobility inherent to this class of proteins is important functionally, as ABDs need to be able to bind multiple partners that share little sequential overlap.

ABDs in the Mediator complex are localized to the tail subunit, which is the most conformationally dynamic portion of the complex.[55] While the structure of the Mediator complex has been successfully solved via cryo-EM, the structure lacks full details for the tail subunit (*Figure 1.8*).[53] Recent advancements have yielded structural models for portions of the tail subunit, showing that like most ABDs, the proteins in the tail portion adopt predominantly helical structures.[55] Interestingly, a missing protein from the cryo-EM structure is Med25, which was unable to be resolved due to long linker regions

separating its independent domains. However, the structure of the activator binding domain (ABD) of Med25 termed AcID (Activator Interacting Domain) has been solved via NMR,[56,57] showing that AcID is structurally unique with a stable β-barrel core that is flanked with three α-helices (*Figure 1.9*).

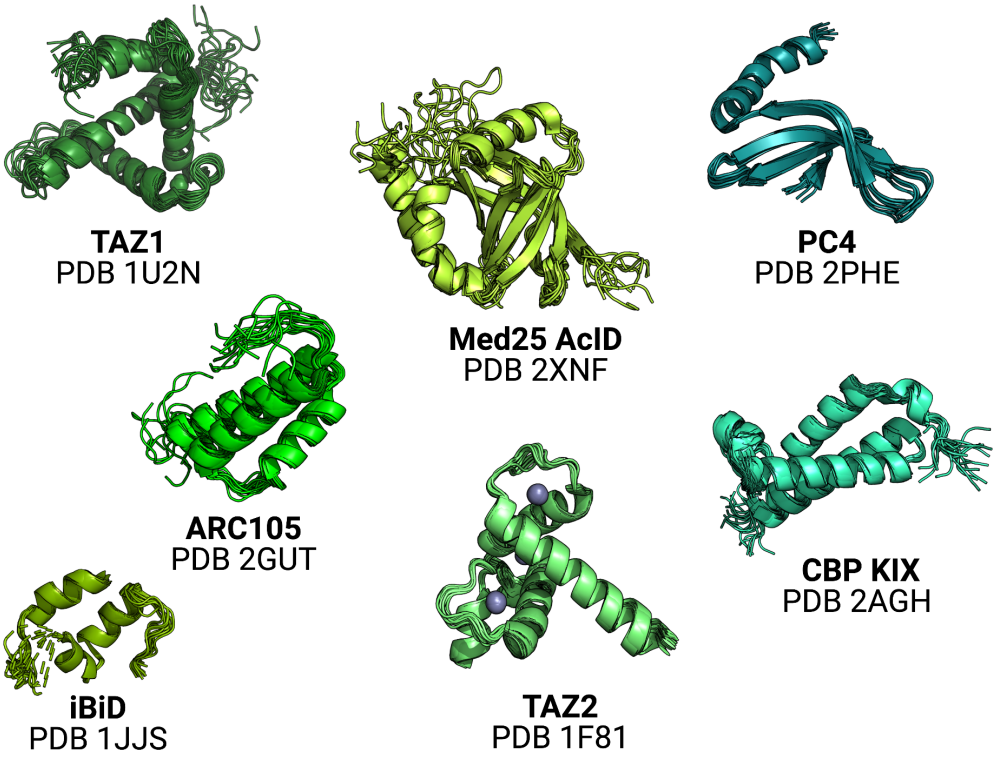


Figure 1.9. *Structures of activator binding domains*. Solution NMR structures showing all identified conformational states for the following ABDs: TAZ1 (PDB 1U2N), ARC105 (PDB 2GUT), iBiD (PDB 1JJS), Med25 AcID (PDB 2XNF), TAZ2 (PDB 1F81), PC4 (PDB 2PHE), and CBP KIX (PDB 2AGH).

1.8 ABDs and disease

As coactivators serve critical roles in gene regulation, it comes as no surprise that *misregulation* of particular interactions has been linked to disease.[58–61] The ABD KIX of the coactivator CBP has been shown to interact with over 15 partners, with direct implications in diseases such as cancer, metabolic disorders, and viral infections (*Figure 1.10*). The Mediator also has relevance in many diseases. For example, the ABD of Med25, termed AcID, is very structurally distinct from KIX, yet it too is implicated in a wide range of diseases from cancer to viral infection. This has prompted an urgent need for

the development of new chemical probes to interrogate these interactions for potential therapeutic purposes.

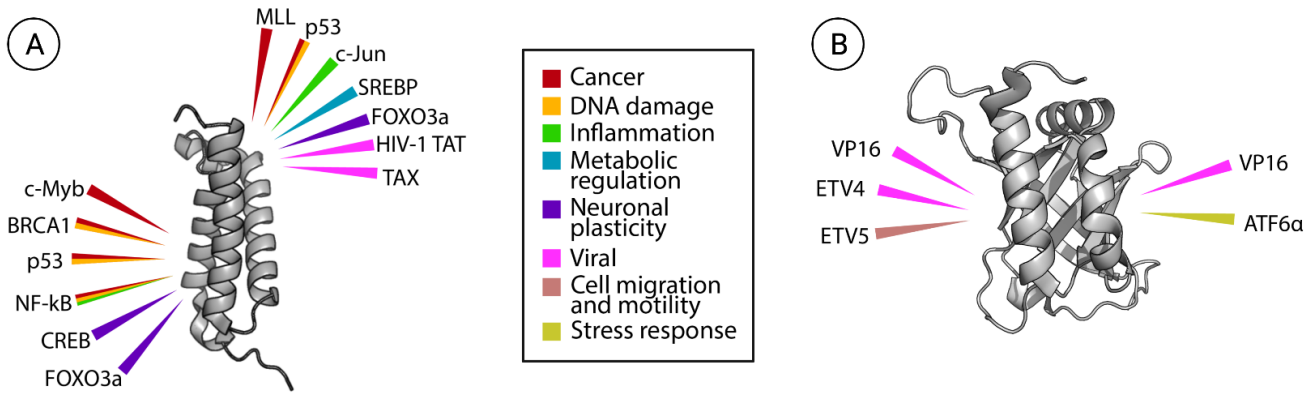


Figure 1.10. *Disease relevance in activator binding domains*. (A) The KIX domain of CBP (PDB 2AGH) is able to interact with a multitude of binding partners, and misregulation of particular interactions has been linked to a variety of diseases. (B) Med25 AcID (PDB 2XNF) is structurally unique among typical activator binding domains, yet interactions been the protein and binding partners has been connected to disease states.

1.9 Experimental challenges surrounding "fuzzy" PPIs

Even as ABDs like CBP KIX and Med25 AcID show high disease relevance, chemical probe development for these targets has been stalled by challenges inherent in dynamic PPIs. Mechanistic studies have shown native activator-coactivator interactions to be "fuzzy," defying many standard biophysical techniques. This is because binding typically occurs over large, flat surface areas that are transient and dynamic in nature. Immediately clear is the difficulties this presents in trying to obtain structural information about such PPIs. For instance, x-ray crystallography necessitates a single low energy bound state when crystalized. Thus, a major tool to delineate structural information is through solution NMR, which can allow for multiple states to be elucidated.

NMR approaches have been utilized to obtain structural information about the dynamic class of proteins, ABDs. For instance, the structure of CBP KIX has been solved for a variety of binary complexes and ternary complexes. Alternatively, Med25 AcID has only been elucidated in the apo state. The reason for that is because binding of activators occurs in a "fuzzy" manner, where bound complexes can occupy many conformational states that are separated by low energy barriers and exchange on the intermediate

timescale. Thus, NMR experiments performed with Med25 and activators yield extensive line broadening, preventing structural elucidation of the bound complex.

Molecular dynamics (MD) simulations are an alternative approach to dissecting complicated binding interactions, which can prove more fruitful. However there are inherent challenges present in MD as well. The association of two proteins could take upwards of microseconds, which would require a large amount of computing power, as many simulations are used on the nanosecond timescales. Further, as these PPIs are deemed “fuzzy,” quantifying independent binding conformations can be extremely challenging. Thus, in order to build a comprehensive understanding of activator-coactivator interactions, a combination of techniques is typically required.

1.10 Allostery as an avenue to expand the druggable proteome

In developing new chemical probes for interrogating coactivator-activator interactions relevant in disease states, we need to better understand the mechanisms of molecular recognition. The transient interactions occur over a large surface area, which has made finding small molecules that can block certain interactions very challenging. Thus, the conformational plasticity of the proteins makes orthosteric inhibition intractable. Alternatively, many ABDs have been shown to form ternary complexes, and moreover, complex formation has been demonstrated in multiple domains to occur through allosteric communication, where binding at one surface increases the affinity for a second PPI.

In addition to allostery being inherent to many dynamic proteins, allosteric inhibition provides an interesting avenue for targeted probe development. Allosteric small molecules have been used with other proteins, such as kinases, and they have been vital for achieving higher levels of selectivity. As previously mentioned, ABDs interact with a suite of cognate binding partners, impacting the expression of thousands of genes. Because of this, it would be advantageous to block specific activator-coactivator interactions over others. With success in allosteric small molecules with other proteins, selectivity could be achieved in ABDs through perturbing the distribution of states towards ones that disfavor only a particular PPI. Hence, it is integral that we not only understand native allostery, but that we can find ways to exploit it for chemical probe development and therapeutic purposes.

The approaches used throughout this thesis highlight the value in combining MD/molecular modeling with biophysical experimentation. Using various modeling approaches, we are able to study the ways ABDs utilize conformational plasticity in recognition at an atomic level.

1.11 Dissertation summary

There is a need for better understanding of native activator-coactivator interactions, as it could aid in chemical probe development and ultimately therapeutic development. My thesis work focuses on utilizing a set of biophysical tools, both *in vitro* and *in silico*, to determine the principles of molecular recognition used by ABDs.

In the first data chapters of my thesis (Chapters 2 and 3), we identify key molecular recognition principles that guide activator•coactivator interactions. Starting with a prototypical activator binding domain CBP KIX (Chapter 2), we explore how the small and malleable domain can recognize a multitude of different binding partners using only two distinct binding surfaces. Further, we investigate how allostery manifests through KIX, demonstrating the utility of fluctuations and dynamics in cooperative binding. In building towards a model of inhibiting said interactions, we determine how small molecules can mimic native properties to allosterically inhibit specific binding events through modulating/attenuating dynamic movements.

In Chapter 3, we transition to another activator binding domain, Med25 AcID, that is structurally disparate from other coactivators, comprised of a seven stranded β barrel with three flanking helices. Using molecular dynamics simulations alongside disulfide Tethering experiments, we construct a model of how the activator VP16 binds. While structurally distinct, we determine that Med25 AcID abides by similar molecular recognition principles as KIX by using its most malleable regions to engage with activators and participate in allosteric communication. Thus, this work builds a foundation for future drug development, and presents a platform of biophysical experiments that can be used for challenging problems like these.

Chapter 4 addresses a protein unrelated to transcription that possesses immense human health consequences. With the emergence of COVID-19, I identified a human transmembrane serine protease TMPRSS2 as a valuable antiviral drug target. TMPRSS2

has been shown to directly aid in allowing SARS-CoV-2 to enter human cells, and yet it is deemed functionally redundant in knockout mouse models. With almost no drugs on the market targeting this protease, there is an urgent need to identify new compounds or repurposed drugs that can inhibit TMPRSS2. We develop a fluorescence-based kinetic assay to assess inhibition of compounds that show up as hits in a virtual screen, giving promise to an alternative route of preventing SARS-CoV-2 infection as well as a multitude of other viral infections.

1.12 References

(1) Benkovic, S. J.; Hammes-Schiffer, S. A Perspective on Enzyme Catalysis. *Science* **2003**, *301* (5637), 1196–1202. https://doi.org/10.1126/science.1085515.

(2) Spirin, V.; Mirny, L. A. Protein Complexes and Functional Modules in Molecular Networks. *Proc. Natl. Acad. Sci.* **2003**, *100* (21), 12123–12128. https://doi.org/10.1073/pnas.2032324100.

(3) Lee, M. J.; Yaffe, M. B. Protein Regulation in Signal Transduction. *Cold Spring Harb. Perspect. Biol.* **2016**, *8* (6). https://doi.org/10.1101/cshperspect.a005918.

(4) Mithieux, S. M.; Weiss, A. S. Elastin. *Adv. Protein Chem.* **2005**, *70*, 437–461. https://doi.org/10.1016/S0065-3233(05)70013-9.

(5) Ouzounis, C. A.; Coulson, R. M. R.; Enright, A. J.; Kunin, V.; Pereira-Leal, J. B. Classification Schemes for Protein Structure and Function. *Nat. Rev. Genet.* **2003**, *4* (7), 508–519. https://doi.org/10.1038/nrg1113.

(6) Ahmed, M. H.; Ghatge, M. S.; Safo, M. K. Hemoglobin: Structure, Function and Allostery. *Subcell. Biochem.* **2020**, *94*, 345–382. https://doi.org/10.1007/978-3-030-41769-7_14.

(7) Kendrew, J. C.; Bodo, G.; Dintzis, H. M.; Parrish, R. G.; Wyckoff, H.; Phillips, D. C. A Three-Dimensional Model of the Myoglobin Molecule Obtained by x-Ray Analysis. *Nature* **1958**, *181* (4610), 662–666. https://doi.org/10.1038/181662a0.

(8) Perutz, M. F.; Rossmann, M. G.; Cullis, A. F.; Muirhead, H.; Will, G.; North, A. C. T. Structure of Hæmoglobin: A Three-Dimensional Fourier Synthesis at 5.5-Å. Resolution, Obtained by X-Ray Analysis. *Nature* **1960**, *185* (4711), 416–422. https://doi.org/10.1038/185416a0.

(9) Watson, H. C.; Kendrew, J. C. The Stereochemistry of the Protein Myoglobin. *Prog. Stereochem.* **1969**, *4*, 299–333.

(10) Shaanan, B. Structure of Human Oxyhaemoglobin at 2.1 A Resolution. *J. Mol. Biol.* **1983**, *171* (1), 31–59. https://doi.org/10.1016/s0022-2836(83)80313-1.

(11) Wilson, M. T.; Reeder, B. J. MYOGLOBIN. In *Encyclopedia of Respiratory Medicine*; Laurent, G. J., Shapiro, S. D., Eds.; Academic Press: Oxford, 2006; pp 73–76. https://doi.org/10.1016/B0-12-370879-6/00250-7.

(12) Marengo-Rowe, A. J. Structure-Function Relations of Human Hemoglobins. *Proc. Bayl. Univ. Med. Cent.* **2006**, *19* (3), 239–245.

(13) Perutz, M. F. Stereochemistry of Cooperative Effects in Haemoglobin. *Nature* **1970**, *228* (5273), 726–739. https://doi.org/10.1038/228726a0.

(14) Changeux, J.-P.; Edelstein, S. J. Allosteric Mechanisms of Signal Transduction. *Science* **2005**, *308* (5727), 1424–1428. https://doi.org/10.1126/science.1108595.
(15) Dickerson, R. E. X-Ray Studies of Protein Mechanisms. *Annu. Rev. Biochem.* **1972**, *41* (1), 815–842. https://doi.org/10.1146/annurev.bi.41.070172.004123.
(16) Monod, J.; Wyman, J.; Changeux, J.-P. On the Nature of Allosteric Transitions: A Plausible Model. *J. Mol. Biol.* **1965**, *12* (1), 88–118. https://doi.org/10.1016/S0022-2836(65)80285-6.
(17) Koshland, D. E.; Némethy, G.; Filmer, D. Comparison of Experimental Binding Data and Theoretical Models in Proteins Containing Subunits. *Biochemistry* **1966**, *5* (1), 365–385. https://doi.org/10.1021/bi00865a047.
(18) ServiceNov. 30, R. F.; 2020; Am, 10:30. 'The game has changed.' AI triumphs at solving protein structures https://www.sciencemag.org/news/2020/11/game-has-changed-ai-triumphs-solving-protein-structures (accessed May 5, 2021).
(19) Folding to Function. *Nat. Struct. Mol. Biol.* **2009**, *16* (6), 573–573. https://doi.org/10.1038/nsmb0609-573.
(20) Dunker, A. K.; Cortese, M. S.; Romero, P.; Iakoucheva, L. M.; Uversky, V. N. Flexible Nets. The Roles of Intrinsic Disorder in Protein Interaction Networks. *FEBS J.* **2005**, *272* (20), 5129–5148. https://doi.org/10.1111/j.1742-4658.2005.04948.x.
(21) Arai, M.; Sugase, K.; Dyson, H. J.; Wright, P. E. Conformational Propensities of Intrinsically Disordered Proteins Influence the Mechanism of Binding and Folding. *Proc. Natl. Acad. Sci.* **2015**, *112* (31), 9614–9619. https://doi.org/10.1073/pnas.1512799112.
(22) Csermely, P.; Palotai, R.; Nussinov, R. Induced Fit, Conformational Selection and Independent Dynamic Segments: An Extended View of Binding Events. *Trends Biochem. Sci.* **2010**, *35* (10), 539–546. https://doi.org/10.1016/j.tibs.2010.04.009.
(23) McCammon, J. A.; Gelin, B. R.; Karplus, M. Dynamics of Folded Proteins. *Nature* **1977**, *267* (5612), 585–590. https://doi.org/10.1038/267585a0.
(24) Vogt, A. D.; Pozzi, N.; Chen, Z.; Di Cera, E. Essential Role of Conformational Selection in Ligand Binding. *Biophys. Chem.* **2014**, *186*, 13–21. https://doi.org/10.1016/j.bpc.2013.09.003.
(25) Michel, D. Conformational Selection or Induced Fit? New Insights from Old Principles. *Biochimie* **2016**, *128–129*, 48–54. https://doi.org/10.1016/j.biochi.2016.06.012.
(26) 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.
(27) Motlagh, H. N.; Wrabl, J. O.; Li, J.; Hilser, V. J. The Ensemble Nature of Allostery. *Nature* **2014**, *508* (7496), nature13001. https://doi.org/10.1038/nature13001.
(28) Motlagh, H. N.; Hilser, V. J. Agonism/Antagonism Switching in Allosteric Ensembles. *Proc. Natl. Acad. Sci.* **2012**, *109* (11), 4134–4139. https://doi.org/10.1073/pnas.1120519109.
(29) Dyson, H. J.; Wright, P. E. Intrinsically Unstructured Proteins and Their Functions. *Nat. Rev. Mol. Cell Biol.* **2005**, *6* (3), 197–208. https://doi.org/10.1038/nrm1589.

(30) Berlow, R. B.; Dyson, H. J.; Wright, P. E. Expanding the Paradigm: Intrinsically Disordered Proteins and Allosteric Regulation. *J. Mol. Biol.* **2018**, *430* (16), 2309–2320. https://doi.org/10.1016/j.jmb.2018.04.003.
(31) Oldfield, C. J.; Cheng, Y.; Cortese, M. S.; Brown, C. J.; Uversky, V. N.; Dunker, A. K. Comparing and Combining Predictors of Mostly Disordered Proteins. *Biochemistry* **2005**, *44* (6), 1989–2000. https://doi.org/10.1021/bi047993o.
(32) Wright, P. E.; Dyson, H. J. Intrinsically Disordered Proteins in Cellular Signalling and Regulation. *Nat. Rev. Mol. Cell Biol.* **2015**, *16* (1), 18–29. https://doi.org/10.1038/nrm3920.
(33) Uversky, V. N. Intrinsically Disordered Proteins and Their “Mysterious” (Meta)Physics. *Front. Phys.* **2019**, *7*. https://doi.org/10.3389/fphy.2019.00010.
(34) Hilser, V. J.; Thompson, E. B. Intrinsic Disorder as a Mechanism to Optimize Allosteric Coupling in Proteins. *Proc. Natl. Acad. Sci.* **2007**, *104* (20), 8311–8315. https://doi.org/10.1073/pnas.0700329104.
(35) Mannige, R. V. Dynamic New World: Refining Our View of Protein Structure, Function and Evolution. *Proteomes* **2014**, *2* (1), 128–153. https://doi.org/10.3390/proteomes2010128.
(36) Gunasekaran, K.; Ma, B.; Nussinov, R. Is Allostery an Intrinsic Property of All Dynamic Proteins? *Proteins Struct. Funct. Bioinforma.* **2004**, *57* (3), 433–443. https://doi.org/10.1002/prot.20232.
(37) Shammas, S. L. Mechanistic Roles of Protein Disorder within Transcription. *Curr. Opin. Struct. Biol.* **2017**, *42*, 155–161. https://doi.org/10.1016/j.sbi.2017.02.003.
(38) Berlow, R. B.; Dyson, H. J.; Wright, P. E. Hypersensitive Termination of the Hypoxic Response by a Disordered Protein Switch. *Nature* **2017**, *543* (7645), 447–451. https://doi.org/10.1038/nature21705.
(39) Liu, J.; Perumal, N. B.; Oldfield, C. J.; Su, E. W.; Uversky, V. N.; Dunker, A. K. Intrinsic Disorder in Transcription Factors. *Biochemistry* **2006**, *45* (22), 6873–6888. https://doi.org/10.1021/bi0602718.
(40) Contreras-Martos, S.; Piai, A.; Kosol, S.; Varadi, M.; Bekesi, A.; Lebrun, P.; Volkov, A. N.; Gevaert, K.; Pierattelli, R.; Felli, I. C.; Tompa, P. Linking Functions: An Additional Role for an Intrinsically Disordered Linker Domain in the Transcriptional Coactivator CBP. *Sci. Rep.* **2017**, *7* (1), 4676. https://doi.org/10.1038/s41598-017-04611-x.
(41) 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.
(42) 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.
(43) Roy, A.; George, S.; Palli, S. R. Multiple Functions of CREB-Binding Protein during Postembryonic Development: Identification of Target Genes. *BMC Genomics* **2017**, *18* (1), 996. https://doi.org/10.1186/s12864-017-4373-3.

(44) Radhakrishnan, I.; Pérez-Alvarado, G. C.; Parker, D.; Dyson, H. J.; Montminy, M. R.; Wright, P. E. Structural Analyses of CREB-CBP Transcriptional Activator-Coactivator Complexes by NMR Spectroscopy: Implications for Mapping the Boundaries of Structural Domains11Edited by F. E. Cohen. *J. Mol. Biol.* **1999**, *287* (5), 859–865. https://doi.org/10.1006/jmbi.1999.2658.
(45) Torchia, J.; Rose, D. W.; Inostroza, J.; Kamei, Y.; Westin, S.; Glass, C. K.; Rosenfeld, M. G. The Transcriptional Co-Activator p/CIP Binds CBP and Mediates Nuclear-Receptor Function. *Nature* **1997**, *387* (6634), 677–684. https://doi.org/10.1038/42652.
(46) Holmqvist, P.-H.; Boija, A.; Philip, P.; Crona, F.; Stenberg, P.; Mannervik, M. Preferential Genome Targeting of the CBP Co-Activator by Rel and Smad Proteins in Early Drosophila Melanogaster Embryos. *PLoS Genet.* **2012**, *8* (6), e1002769. https://doi.org/10.1371/journal.pgen.1002769.
(47) Philip, P.; Boija, A.; Vaid, R.; Churcher, A. M.; Meyers, D. J.; Cole, P. A.; Mannervik, M.; Stenberg, P. CBP Binding Outside of Promoters and Enhancers in Drosophila Melanogaster. *Epigenetics Chromatin* **2015**, *8*. https://doi.org/10.1186/s13072-015-0042-4.
(48) Law, S. M.; Gagnon, J. K.; Mapp, A. K.; Brooks III, 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.
(49) Brüschweiler, S.; Konrat, R.; Tollinger, M. Allosteric Communication in the KIX Domain Proceeds through Dynamic Repacking of the Hydrophobic Core. *ACS Chem. Biol.* **2013**, *8* (7), 1600–1610. https://doi.org/10.1021/cb4002188.
(50) Borggrefe, T.; Yue, X. Interactions between Subunits of the Mediator Complex with Gene-Specific Transcription Factors. *Semin. Cell Dev. Biol.* **2011**, *22* (7), 759–768. https://doi.org/10.1016/j.semcdb.2011.07.022.
(51) Yin, J.; Wang, G. The Mediator Complex: A Master Coordinator of Transcription and Cell Lineage Development. *Dev. Camb. Engl.* **2014**, *141* (5), 977–987. https://doi.org/10.1242/dev.098392.
(52) Cantin, G. T.; Stevens, J. L.; Berk, A. J. Activation Domain–Mediator Interactions Promote Transcription Preinitiation Complex Assembly on Promoter DNA. *Proc. Natl. Acad. Sci.* **2003**, *100* (21), 12003–12008. https://doi.org/10.1073/pnas.2035253100.
(53) Chadick, J. Z.; Asturias, F. J. Structure of Eukaryotic Mediator Complexes. *Trends Biochem. Sci.* **2005**, *30* (5), 264–271. https://doi.org/10.1016/j.tibs.2005.03.001.
(54) 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), 728–737. https://doi.org/10.1038/nchembio.127.
(55) Ansari, S. A.; Morse, R. H. Selective Role of Mediator Tail Module in the Transcription of Highly Regulated Genes in Yeast. *Transcription* **2012**, *3* (3), 110–114. https://doi.org/10.4161/trns.19840.
(56) 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.

(57) Eletsky, A.; Ruyechan, W. T.; Xiao, R.; Acton, T. B.; Montelione, G. T.; Szyperski, T. Solution NMR Structure of MED25(391–543) Comprising the Activator-Interacting Domain (ACID) of Human Mediator Subunit 25. *J. Struct. Funct. Genomics* **2011**, *12* (3), 159. https://doi.org/10.1007/s10969-011-9115-1.
(58) 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.
(59) Cong, S.-Y.; Pepers, B. A.; Evert, B. O.; Rubinsztein, D. C.; Roos, R. A. C.; van Ommen, G.-J. B.; Dorsman, J. C. Mutant Huntingtin Represses CBP, but Not P300, by Binding and Protein Degradation. *Mol. Cell. Neurosci.* **2005**, *30* (1), 12–23. https://doi.org/10.1016/j.mcn.2005.05.003.
(60) Karamouzis, M. V.; Konstantinopoulos, P. A.; Papavassiliou, A. G. Roles of CREB-Binding Protein (CBP)/P300 in Respiratory Epithelium Tumorigenesis. *Cell Res.* **2007**, *17* (4), 324–332. https://doi.org/10.1038/cr.2007.10.
(61) Uversky, V. N. Chapter Four - Intrinsic Disorder, Protein–Protein Interactions, and Disease. In *Advances in Protein Chemistry and Structural Biology*; Donev, R., Ed.; Protein-Protein Interactions in Human Disease, Part A; Academic Press, 2018; Vol. 110, pp 85–121. https://doi.org/10.1016/bs.apcsb.2017.06.005.

Chapter 2

The KIX Domain Uses High Conformational Plasticity for Molecular Recognition

Abstract

The coactivator KIX of CBP uses two binding surfaces to recognize multiple activators and exhibits allostery in ternary complex formation. Activator•coactivator interactions are central to transcriptional regulation, yet the microscopic origins of allostery in dynamic proteins like KIX are largely unknown. Here, we investigate the molecular recognition and allosteric manifestations involved in two KIX ternary systems c-Myb•KIX•MLL and pKID•KIX•MLL. Exploring the hypothesis that binary complex formation prepays an entropic cost for positive cooperativity, we utilize molecular dynamics simulations, side chain methyl order parameters, and differential scanning fluorimetry (DSF) to calculate conformational entropy changes in KIX. The protein's configurational micro-states from structural clustering highlight the utility of protein plasticity in molecular recognition and allostery. We find that apo KIX occupies a wide distribution of lowly-populated configurational states. Each binding partner has its own suite of KIX states that it selects, building a model of molecular recognition fingerprints. Allostery is maximized with MLL pre-binding, which corresponds to the observation of a significant reduction in KIX micro-states observed when MLL binds. Experimentally capturing KIX stabilization is challenging, particularly because of the disordered nature of particular activators. However, DSF melting curves allow for inference of relative entropic changes that occur across complexes. Mutations and a small molecule allosterically KIX interactions, which do so by attenuating loop movements, demonstrating how the most dynamic region tunes overall conformational entropy. Thus, we demonstrate how the small flexible domain recognizes multiple partners, and further, how this native property can be exploited for targeted allosteric inhibition.

2.1 Introduction

Allostery and disorder are both hallmarks of transcriptional regulation, suggesting a link between the two.[1–3] Allostery describes events where a ligand binding to a central protein impacts the binding or catalytic success of a second ligand at a non-overlapping site.[4–10] Early models attributed allosteric communication to a network of amino acid contacts in a well-structured protein.[11–16] However, the prevalence of allostery in disordered proteins has led to a shift in understanding from mechanical coupling and towards an ensemble framework, viewing allostery instead as a reweighting of micro-states in a thermal distribution.[3,17–19] While ensembles provide a useful conceptual framework for understanding allostery, the question remains as to how this manifests in real biological systems.

A prototypical example of structural disorder and allostery occurs within the KIX domain of CREB Binding Protein (CBP), a multidomain coactivator that acts as a bridge between activators and other transcriptional machinery components.[20] KIX is comprised of three helices (*Figure 2.1*), and it has been shown to bind more than fifteen different partners with only two binding surfaces.[21–27] KIX binds both the mixed lineage leukemia (MLL) transcription factor and the proto-oncogene transcription factor c-Myb in a cooperative manner, with ternary complex formation deemed critical for hematopoiesis.[28–31] Misregulation of the c-Myb•KIX•MLL ternary complex has been implicated in leukemogenesis, motivating a detailed atomic understanding of the molecular mechanisms of complex formation in order to discover chemical probes and therapeutic agents.[28–31] Similar allosteric effects have been observed in KIX with MLL and a different c-Myb-site activator, the phosphorylated kinase-inducible domain of CREB (pKID) (*Figure 2.1*).[23] Solution NMR structures of the c-Myb•KIX•MLL and the pKID•KIX•MLL ternary complexes (called the c-Myb and pKID system hereon) show that while pKID and c-Myb bind on the same general KIX surface, they form structurally distinct KIX ternary complexes (Figure 1).[32,33] The critical question that emerges from this data is whether the allosteric coupling is the same in both the c-Myb and pKID systems, or if dynamic domains such as KIX can employ multiple mechanisms.

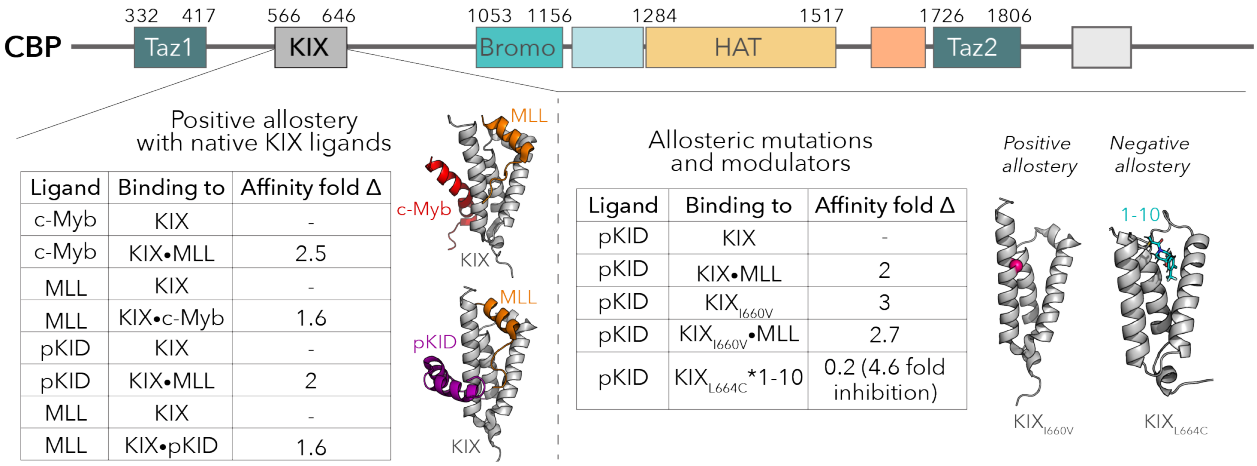


Figure 2.1. *The KIX domain of CBP exhibits allostery.* (Top) The various domains of CBP. (Left) Positive cooperativity seen in the two native KIX systems (c-Myb and pKID, respectively), with affinity fold changes calculated using K_ds obtained from ITC.[23] The two solution NMR structures of the ternary complexes (c-Myb•KIX•MLL from PDB 2agh[32], and pKID•KIX•MLL from PDB 2lxt)[33] were used to construct all of the native KIX complexes for simulations (apo KIX, KIX•c-Myb, KIX•MLL, c-Myb•KIX•MLL, KIX•pKID, and pKID•KIX•MLL). (Right) KIX mutations and ligands that perturb allostery in the pKID system affinity fold changes calculated using K_ds obtained from fluorescence polarization and stopped flow experiments.[34] All KIX_{I660V} complexes were constructed using the original coordinates for the unmutated ternary structure (pKID•KIX•MLL, PDB 2lxt), with the location of I660 highlighted as a pink sphere. The crystal structure of KIX_{L664C}*1-10 (PDB 4i9o)[35] was used to construct simulations for the Tethered complex.

Here we test the hypothesis that reduction of conformational entropy is the microscopic origin of the allostery observed in these key transcriptional regulatory systems. Molecular dynamics simulations are utilized to construct a detailed model of the KIX conformational landscape in ternary complex formation. Structural clustering of each complex yields independent distributions of KIX micro-states, from which we can estimate entropic changes. Comparisons across distributions of micro-states results in clear pathways of conformational selection. Results show that the distribution for the apo KIX protein contains many lowly-populated micro-states. Each activator peptide selects a unique suite of cognate micro-states, yielding a molecular recognition "fingerprint" for each binding partner. Examination of synthetic modulators and mutant KIX motifs (inhibition with molecule 1-10;[35] enhancement with KIX_{I660V} mutation,[34] *Figure 2.1*) indicate allosteric enhancement /inhibition can be achieved by perturbing the distribution of KIX micro-states, which is consistent with observed changes in unfolding rates and melting temperatures. The data supports a model in which maximum positive cooperativity is achieved with decreases in conformational entropy, whereas inhibition can be achieved through a redistribution of states. Moreover, the largest entropy changes

in KIX are attributed to altered dynamics in the most mobile region of the protein, the L_{12}-G_2 loop (residues 614-621). Thus, the molecular recognition principles guiding KIX interactions, both in native and mutant systems, rely heavily on the protein's innate malleability to occupy various structural states in a wide distribution, which allows for accommodation of multiple binding partners despite the lack of significant sequence similarities.

2.2 Results and discussion

KIX adopts a diminishing number of micro-states as substrates bind

Using the atomic coordinates from experimentally solved KIX complexes, all-atom molecular dynamics simulations were performed on each system independently (apo, binary/ternary complexes, and mutants) (*Figure 2.2, Panels 1-2*). In order to obtain information about the distribution of KIX conformational states for a given system, KIX structures from the trajectories are superposed and subjected to K-means clustering (2.5 Å RMSD cutoff using C_α atoms; *Figure 2.2, Panel 3*). Hence, a given system (apo, binary/ternary, mutants) results in *W* structural clusters, with each individual cluster C_1, C_2, … , C_W having occupancy P_n ($\sum_{n=1}^{W} P_n = 1$) based on the number of KIX frames in a given cluster (*Figure 2.2, Panel 4*; *W* clusters for each complex represented as colored circles with sizes corresponding to relative occupancy/population). Averaged centroid structures are generated for each cluster, with each representing a given conformational micro-state "basin".

To test whether KIX conformational states in bound complexes arise from micro-states sampled in the apo distribution, centroid structures from the binary distributions are iteratively compared by RMSD to the distribution of states in the apo protein; The binary complex is said to have originated from the apo distribution if there exists an apo KIX centroid that is within the cutoff used for clustering (2.5 Å, C_α atoms; the cluster corresponding to the minimum RMSD is shown with a connecting line in *Figure 2.2, Panel 4*). This process is repeated for the ternary complexes in comparison to the binary distribution of micro-states. There are two ways in which ternary complex formation can happen with KIX: either MLL binds first, or c-Myb/pKID binds first. As allosteric measurements have been tested in both directions, we test the impact on the distribution

of KIX states by binding order (MLL binding first shown in *Figure 2.3*; c-Myb/pKID binding first in *Figure 2.4*).

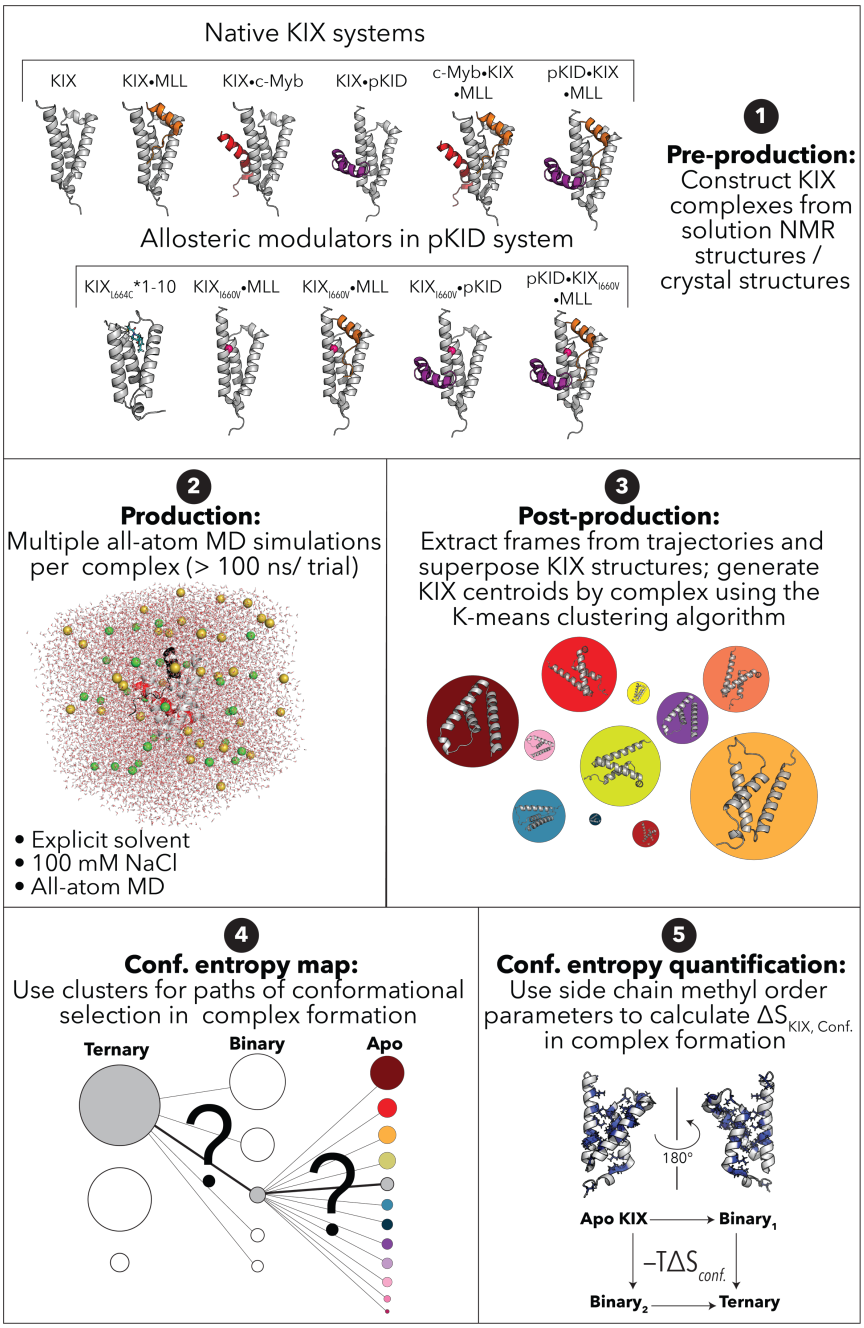


Figure 2.2. *Molecular dynamics methods for dissecting KIX conformational entropy.* [*Panel 1*] Six native systems are constructed using the solution NMR for c-Myb•KIX•MLL (PDB 2AGH) for ternary, binary, and apo starting structures, and pKID•KIX•MLL (PDB 2LXT) for ternary and binary structures. The KIX_{I660V} mutation systems were created using the coordinates for the WT protein and mutating the residue to valine using CHARMM. The KIX_{L664C}*1-10 structure used coordinates from the crystal structure of KIX_{L664C} Tethered to molecule 1-10 through a disulfide bond. [*Panel 2*] All atom molecular dynamics simulations were performed in replicates for each of the system described in panel 1. Simulations contained 100 mM

NaCl and were performed using CHARMM. [*Panel 3*] Structural frames of KIX from simulations were superposed subjected to K-means clustering using 2.5 Å RMSD cutting with C_α atoms. [*Panel 3*] The distribution of KIX clusters was projected in two dimensions where each circle represents a given KIX cluster (i.e. conformational microstate), with size corresponding to occupancy (i.e. number of frames in given cluster). [*Panel 4*] In answering how the distribution of KIX states shifts with binding partners, centroids structures generated for each cluster in the ternary complex were iteratively compared binary complex centroids. This was repeated from binary to apo micro-states, which allowed for conformational selection maps to be constructed. [*Panel 5*] In quantifying KIX conformational entropy changes that occur in activator binding, side chain methyl order parameters were used.

Immediately evident is the wide distribution of lowly-populated micro-states in the apo protein, shown in the central portion of *Figure 2.3* and *Figure 2.4*. When MLL binds to apo KIX (*Figure 2.3*), there is a substantial reduction in conformational states sampled; Subsequent binding of c-Myb/pKID in ternary complex formation further winnows down the number of accessible KIX conformational states. Full pathways of conformational selection show that 87% of the distribution of KIX microstates in c-Myb•KIX•MLL and 100% of KIX microstates in pKID•KIX•MLL originate from the same lowly-populated apo micro-state (*Figure 2.3*, coral colored micro-state).

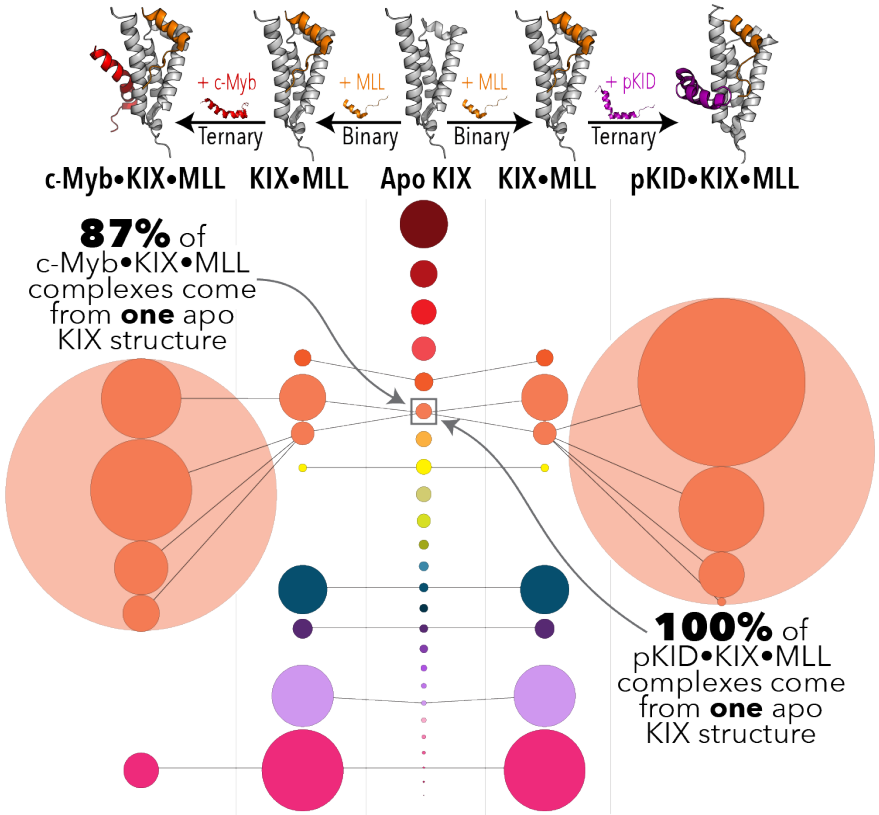


Figure 2.3. *Diminishing KIX micro-states in ternary complex formation when MLL binds first*. Circles represent KIX micro-state basins that come from K-means clustering analysis (see Methods). The diameter

of each circle is proportional to the relative population. Circle colors correspond to the cluster origin from the apo KIX distribution (shown in the center). The diminishing number of states are shown through these pathways of conformational selection for the (left) c-Myb ternary complex and (right) the pKID ternary complex when MLL binds first.

In the reverse binding order, c-Myb and pKID bind to the same KIX binding face, yet they have almost no sequential overlap, indicating differing mechanisms of molecular recognition. Indeed, we see that the two peptides select for different subsets of KIX micro-states, which in turn both differ from selected states with MLL binding. However, the vast majority of c-Myb•KIX•MLL micro-states arise from the same coral colored apo KIX state seen in either binding direction (*Figure 2.4*). When pKID binds to KIX first (KIX•pKID), all of the resultant ternary KIX micro-states originate from an alternative apo KIX conformational basin (yellow). These data illustrate an analogous mechanism of decreasing conformational micro-states, while also demonstrating that order of binding may lead to differing final state selection (*Figure 2.4*).

There are multiple KIX micro-states found in the KIX•c-Myb binary complex that cannot be “traced back” to the distribution of apo KIX micro-states — That is to say that the RMSD of those KIX•c-Myb centroids to all KIX micro-states is greater than the 2.5 Å used in clustering (37% of the total KIX micro-states in the KIX•c-Myb complex, shown in gray in *Figure 2.4*). This suggests that c-Myb alters the conformational landscape of KIX by inducing a conformational shift. Additionally, as is evident in these maps, the number of accessible states in apo KIX is dramatically reduced when the first partner binds, and thus the search for those states that are most “binding competent” when the last partner interacts is much smaller, representing a decrease in configurational entropy for binding this last activator. While these findings are not quantitative with regards to entropy measurements, the observations are consistent with the stronger allostery in both systems occurring with MLL binding first. These results highlight a key feature of the molecular recognition used by KIX: Each activator selects for a unique suite of cognate micro-states, demonstrating the utility of disorder for a multi-partner protein hub. KIX has over 15 native binding partners, which necessitates specialized molecular recognition. The wide distribution of conformational states coupled with the protein’s ability to undergo conformational changes is integral in accommodating each partner and ultimately manifests in allostery.

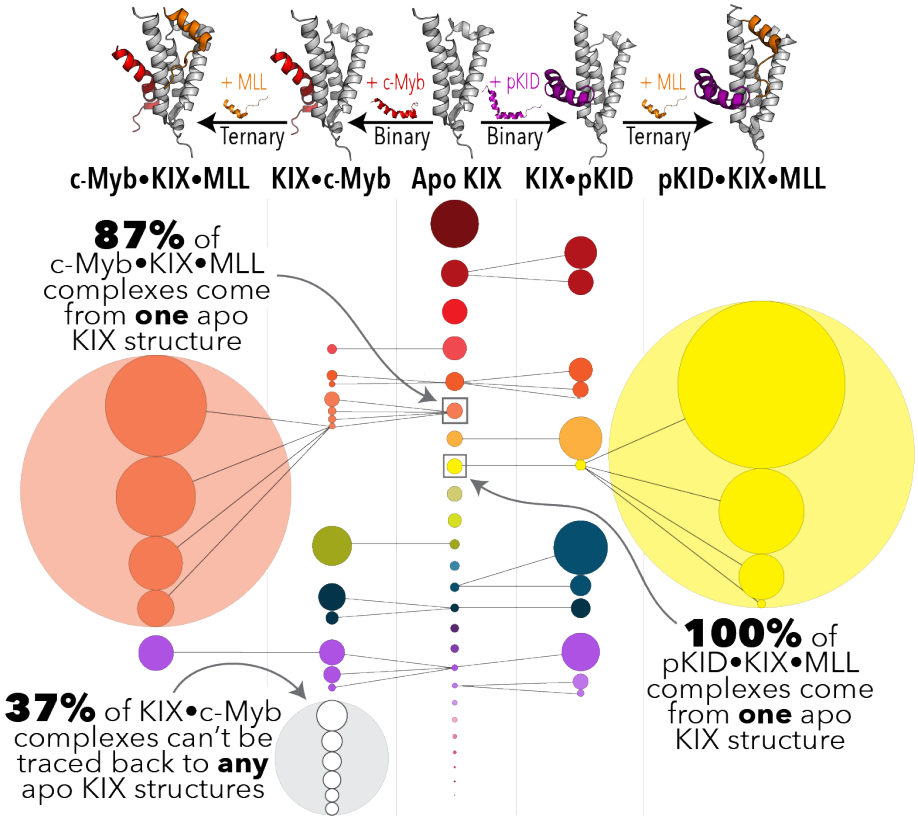


Figure 2.4. *Diminishing KIX micro-states in ternary complex formation when c-Myb/pKID bind first*. Circles represent KIX micro-state basins that come from clustering (see *Figure 2.3*). The diminishing number of states are shown through these pathways of conformational selection for the (left) c-Myb ternary complex when c-Myb binds first and (right) the pKID ternary complex when pKID binds first.

Previous studies have found that the mutation KIX_{I660V} "turns on" the allosteric communication in the pKID system; that is to say that with the I660V mutant, pKID binding is enhanced regardless of whether MLL is pre-bound or not (*Figure 2.1*).[34] Our analysis of molecular dynamics trajectories using structural clustering demonstrate that the I660V mutation winnows down the number of attainable apo KIX micro-states (*Figure 2.5*), much like MLL binding does in the wild type system (*Figure 2.3*). With the limited number of KIX_{I660V} micro-states, pKID binding to form the binary KIX_{I660V}•pKID complex no longer selects for the yellow KIX state (*Figure 2.5*), but instead follows a more "MLL-like" pathway by binding to the coral-colored state (*Figure 2.3*). Thus, the I660V mutation winnows down and pre-organizes the distribution of KIX micro-state basins to enhance pKID binding without MLL binding. In fact, we see that 93% of KIX_{I660V}•pKID complexes stem from the same two apo KIX basins as 97% of the ternary pKID•KIX_{I660V}•MLL complexes when MLL binds before pKID (*Figure 2.5*).

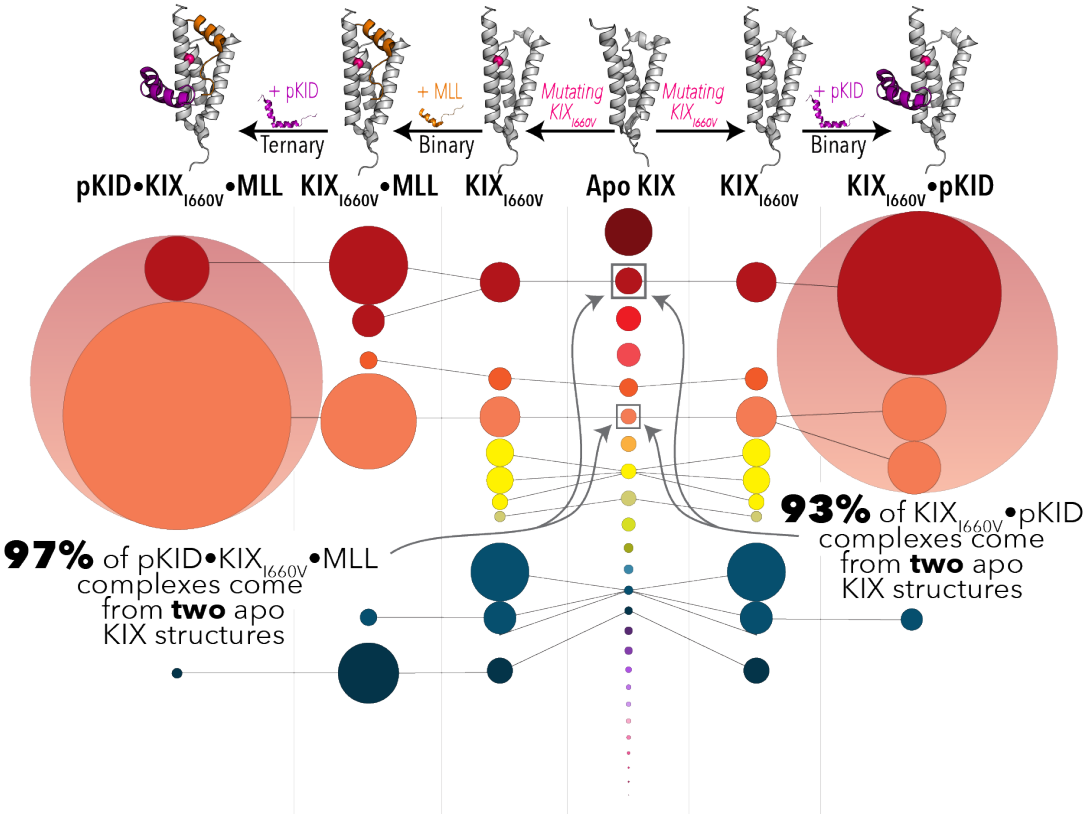
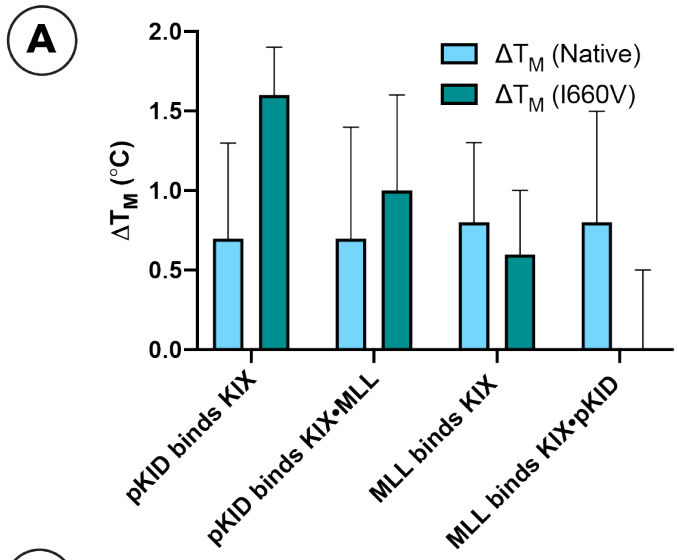


Figure 2.5. *Pathways of KIX conformational selection in the mutant (I660V) pKID system*. Circles represent KIX micro-state basins that come from K-means clustering analysis, and the diameter of each circle is proportional to the relative population. Circle colors correspond to the cluster origin from the apo KIX distribution. (Left) The diminishing states pathways in conformational selection when MLL binds first in ternary complex formation and (right) when the KIX_{I660V}•pKID forms the binary complex.

Directly monitoring conformational entropy changes in experiments presents a challenge. Calometric approaches cannot provide access to conformational entropy since they are global measurements, convoluted with entropy changes from solvation. Thus, we opted to measure changes in thermal stability across complexes using differential scanning fluorimetry (DSF).[36,37] Pertinent thermodynamics can be qualitatively extrapolated by monitoring changes in the fluorescence when KIX is in various bound complexes by using a dye, SYPRO Orange, that binds to hydrophobic regions of proteins that become more exposed during unfolding.[36–38] In performing these experiments, we find that while pKID and MLL in the wild-type system induce similar increases in T_M, pKID greatly stabilizes the mutant (*Figure 2.6*), which aligns with the conformational selection map shown in *Figure 2.6*. More significantly, when pKID is pre-bound to KIX_{I660V} (KIX_{I660V}•pKID), MLL binding has no detectable change in KIX melting temperature — a

finding consistent with the “ternary-like” distribution of structural basins seen in KIX_{I660V}•pKID.



B

Complex	T_M (°C), WT	T_M (°C), I660V
Apo KIX	57.3 ± 0.4	52.7 ± 0.2
KIX•pKID	58.0 ± 0.4	54.3 ± 0.2
KIX•MLL	58.1 ± 0.3	53.3 ± 0.4
pKID•KIX•MLL	58.8 ± 0.6	54.3 ± 0.5

Figure 2.6. *DSF experiments for thermal stability changes in WT KIX and K_{I660V} systems.* (A) Changes in melting temperatures determined via DSF in the native KIX system (light blue) and the mutant KIX_{I660V} system (teal), with individual melting temperatures of both the wild type (WT) KIX and the mutant KIX_{I660V} systems shown in (B).

Molecule 1-10, when Tethered to the MLL site, has been shown to allosterically inhibit pKID binding at the distal site.[34] The crystal structure of 1-10 Tethered to KIX_{L664C} has been solved (*Figure 2.1*),[35] which allows for MD and clustering to be performed to probe how the Tethered small molecule perturbs the distribution of micro-states. In doing so, we find that the bound adduct alters the distribution of micro-state basins such that 63% of KIX_{L664C}*1-10 structures cannot be traced back to any structures in the apo KIX distribution, suggesting that the molecule causes notable conformational changes and forces KIX into states unseen in the native distribution (*Figure 2.7*). Moreover, pKID does not appear to be able to bind any of these “new” states, and in fact 15% of the states seen in the KIX•pKID complex can no longer be traced back through structures in the KIX_{L664C}*1-10 distribution (*Figure 2.7*). Thus, the allosteric inhibition observed with 1-10

can be, in part, attributed to modulating the KIX distribution of micro-states to ones that no longer favor pKID interactions, aligning with the partial inhibition seen *in vitro*.[34]

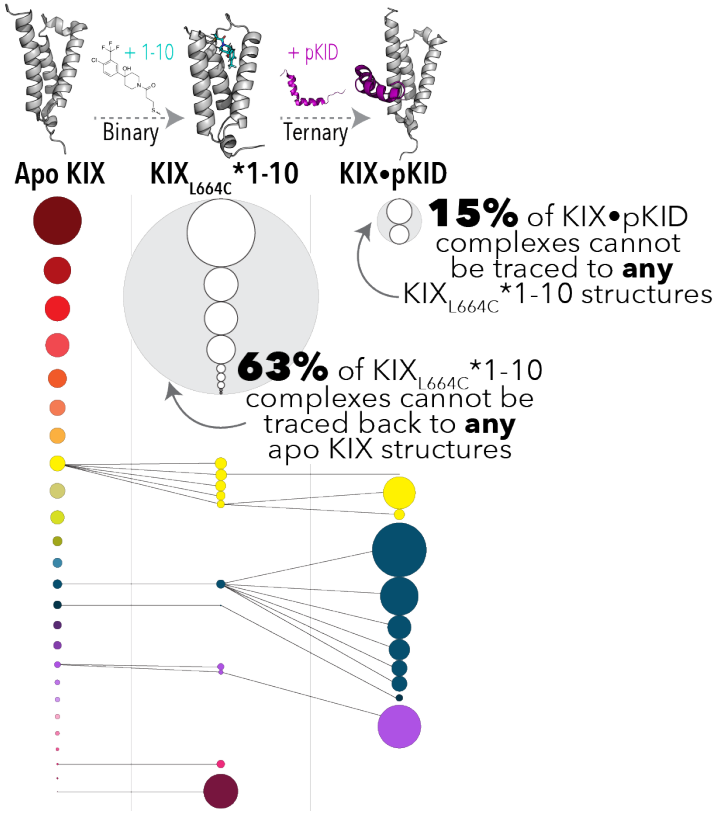
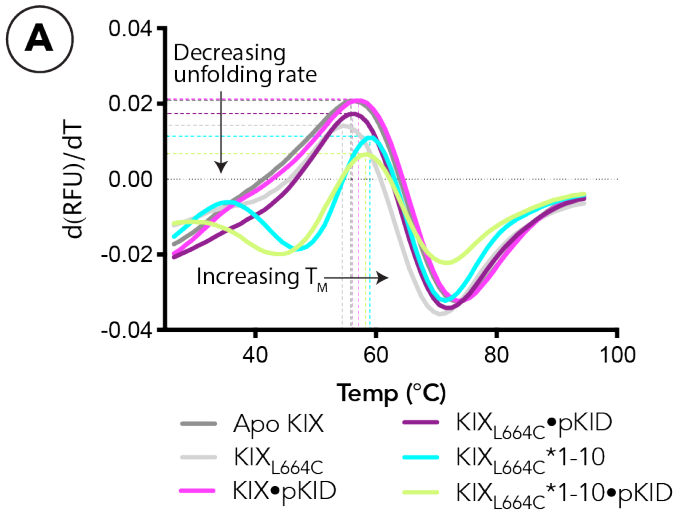


Figure 2.7. *Pathways of KIX conformational selection in the pKID system with allosteric modulator molecule 1-10*. Circles represent KIX micro-state basins that come from K-means clustering analysis, and the diameter of each circle is proportional to the relative population. Circle colors correspond to the cluster origin from the apo KIX distribution. Shown are the diminishing states pathways when molecule 1-10 is Tethered to KIX$_{L664C}$. The gray dashed lines indicate that there are more steps involved in forming the complex than are depicted here (for instance, we did not perform MD on apo KIX$_{L664C}$, nor did we perform MD on pKID•KIX$_{L664C}$*1-10).

In DSF experiments, we observe a large increase in stability with 1-10, which is determined by finding the temperature corresponding to the maximum change in relative fluorescence units (RFU) (*Figure 2.8*). The T_M of the Tethered complex increases the melting temperature by 2.3 °C in comparison to the apo protein and 3.9 °C in comparison to KIX$_{L664C}$ (*Figure 2.8*). However, the unfolding transition is sharper in unbound KIX/KIX$_{L664C}$ than with the molecule bound, as seen as higher maximum first derivatives of changing fluorescence by temperature (*Figure 2.8*). This finding could suggest a less-

cooperative unfolding transition, which would correspond to the new KIX states induced from 1-10 binding as seen in MD.



B

Complex	T_M (°C)	Max. d(RFU)/dT
Apo KIX	57.3 ± 0.4	0.021
KIX_{L664C}	55.7 ± 0.1	0.014
KIX•pKID	58.0 ± 0.4	0.021
KIX_{L664C}•pKID	57.0 ± 0.2	0.017
KIX_{L664C}*1-10	59.6 ± 0.7	0.011
KIX_{L664C}*1-10•pKID	58.5 ± 0.7	0.007

Figure 2.8. *First derivative plots of relative fluorescence units (RFU) by temperature (°C) from DSF*. (A) The change in RFU by temperature, (d(FRU)/dT), is shown for various complexes in KIX•pKID systems with allosteric modulators. (B) Changes in melting temperature measured by DSF for each complex as well as the maximum height of the first derivative.

Conformational entropy changes can be masked in macroscopic measurements

The conformational selection maps presented in previous figures show a diminishing number of KIX micro-states upon binding, which is consistent with earlier coarse-grained modeling studies.[39,40] However, while compelling, these data alone do not provide a fully quantitative explanation of the observed allostery. Thus, for quantification of configurational entropy changes associated with these binding events in MD simulations, side chain methyl order parameters are utilized. Order parameters describe individual side chain dynamics, ranging from 0 (completely disordered) to 1 (completely rigid). Average methyl order parameters across an entire protein can be used to calculate overall conformational entropy (Equation 1 in Methods). As KIX has high coverage and

good dispersion of methyl amino acids (*Figure 2.9*), we use order parameters to calculate changes in KIX conformational entropy in complex formation (*Table 2.1*).

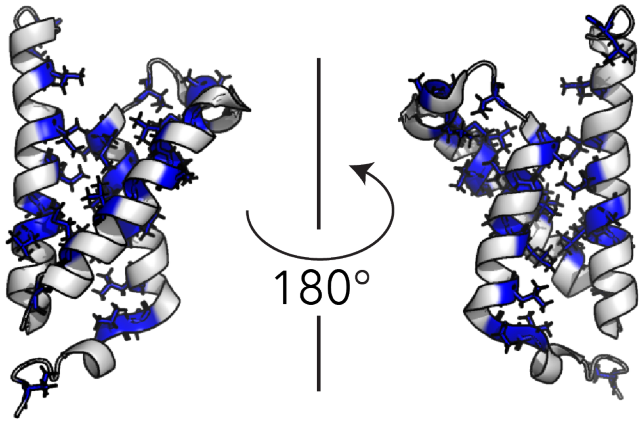


Figure 2.9. *Coverage of methyl-bearing amino acids on KIX*. KIX show high coverage of methyl amino acids (Ala, Leu, Ile, Thr, Val, Met; shown in blue) that are evenly dispersed across the protein, indicating that methyl order parameters would be appropriate to use in this system.

Table 2.1. KIX conformational entropy changes in wild type systems that occur with activator binding at 298 K calculated using methyl order parameters from MD.

Ligand	Binding to	$-T\Delta S_{\text{conf}}$ (kcal/mol)
c-Myb	KIX	-2.1
c-Myb	KIX•MLL	-0.5
MLL	KIX	3.7
MLL	KIX•c-Myb	5.4
pKID	KIX	-0.7
pKID	KIX•MLL	0.6
MLL	KIX	3.7
MLL	KIX•pKID	5.1

Conformational entropy is only one of the many components in the full thermodynamic expression (i.e. total enthalpy, ligand entropy, rotational/translational entropy, solvent entropy, etc.). Thus, thermodynamic measurements in experiment oftentimes report on a combination of the listed components, which makes it difficult to determine the true driving forces behind binding interactions. Isothermal calorimetry (ITC), a common tool to measure free energy changes upon binding events, has been utilized in the native KIX binding interactions. Comparing total entropy changes from ITC to KIX conformational entropy changes calculated here using order parameters, we

observe that the measurements follow the same trend for the c-Myb system (*Figure 2.10*). Further, changes in melting temperatures measured via DSF also follow a similar trend; c-Myb binding causes an increase in KIX conformational entropy, which corresponds to the minimal changes seen in melting temperatures (*Figure 2.10*). MLL binding greatly decreases KIX conformational entropy, aligning with the increase in thermal stability. Thus, in the c-Myb system, it appears that conformational entropy is at the root of the thermodynamic driving forces in ternary complex formation.

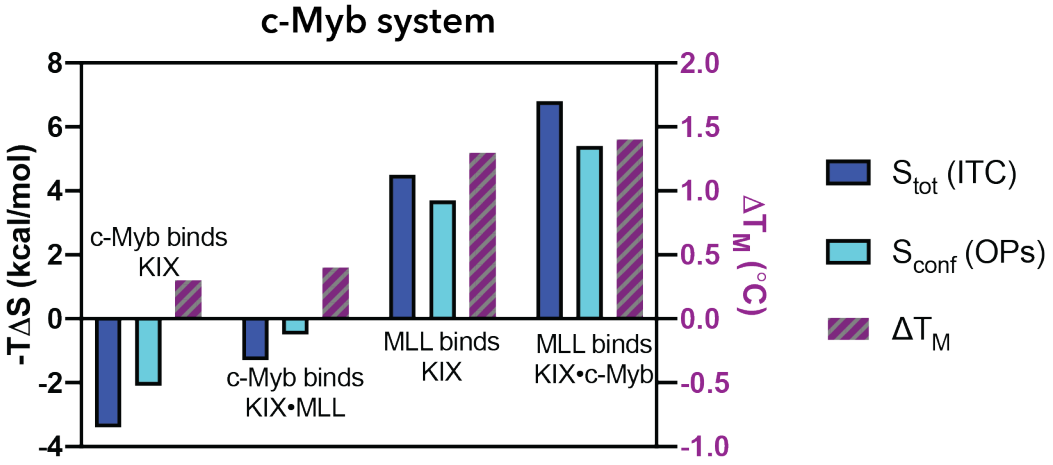


Figure 2.10. *Entropy changes and melting temperature changes in the c-Myb system*. Changes in total entropy upon binding from previously-published ITC experiments (dark blue)[23] compared to KIX conformational entropy changes calculated here using methyl order parameters (OPs) from molecular dynamics simulations (light blue) for the c-Myb system. Changes in melting temperatures determined from DSF are shown on the right y-axis (purple).

In contrast, while calculations from order parameters show pKID slightly increases KIX conformational entropy like c-Myb binding does, the total entropy greatly decreases (*Figure 2.11*). Further, changes in melting temperatures align with neither the calculated changes in KIX conformational entropy nor the total entropy changes, suggesting a more complex mechanism. Several studies have compared the binding mechanisms of c-Myb and pKID, finding that they differ with regards to the amount of secondary structural content each peptide contains prior to binding.[41–43] The pKID activator is the largest peptide used in this study (pKID has 34 residues, c-Myb has 25, and MLL has 19), and unlike c-Myb, shows little to no propensity for pre-forming any helical content prior to binding KIX.[41–43] The entropic cost affiliated with the folding transition of pKID upon binding would greatly contribute to the total entropy measured via ITC/DSF, and this could

outweigh the shifts in the ensemble of folded KIX micro-states we observe in our work, thus dominating the overall thermodynamics of binding but masking the origins of the observed allostery.

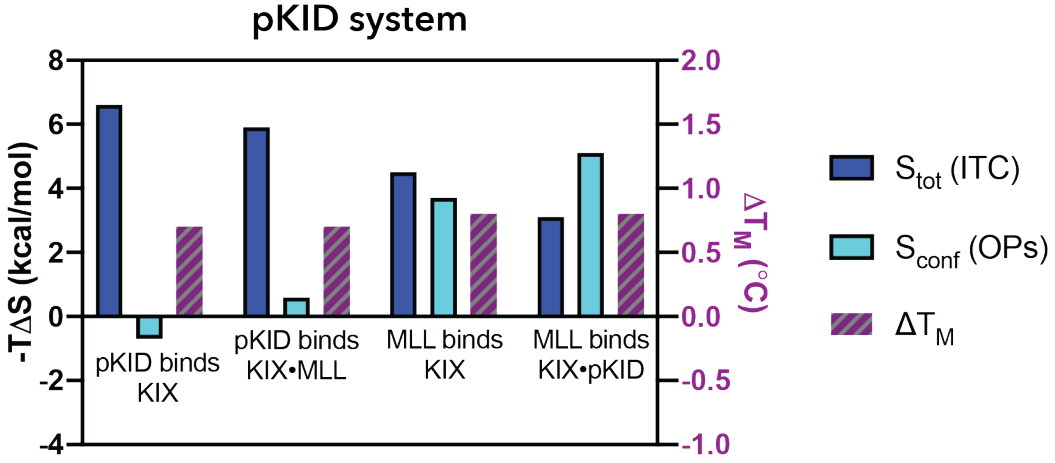


Figure 2.11. *Entropy changes and melting temperature changes in the pKID system*. Changes in total entropy upon binding from previously-published ITC experiments (dark blue)[23] compared to KIX conformational entropy changes calculated here using methyl order parameters (OPs) from molecular dynamics simulations (light blue) for the pKID system. Changes in melting temperatures determined from DSF are shown on the right y-axis (purple).

In vitro binding assays have demonstrated for both the c-Myb and pKID systems that the highest cooperativity occurs when MLL binds first.[23] Our results are consistent with this finding; using methyl order parameters to calculate entropy, MLL binding prepays the largest entropic cost ($-T\Delta S_{conf} = 3.7$ kcal/mol). It also causes the largest increase in T_M (0.8 °C) in both systems, significantly limiting the number of attainable KIX micro-states to increase affinity for the second activator. Quantifying the entropic effects of mutating apo KIX to KIX_{I660V}, we find that the mutation alone causes a reduction in entropy similar to when MLL binds apo KIX ($-T\Delta S_{conf} = 2.3$ kcal/mol). Thus, a single mutation can alter the conformational landscape of the apo protein, which in turn affects binding and allostery, once more highlighting the significance of disorder/conformational entropy in dynamic allosteric interactions. Interestingly, the L664C and 1-10 combination decreases KIX entropy, but to a lesser degree ($-T\Delta S_{conf} = 0.5$ kcal/mol), suggesting a more complicated mechanism for inhibition that includes a redistribution of accessible micro-states, which we observe from our clustering (*Figure 2.5*). Thus, the inhibition mechanism is achieved not through a quantitative reduction in conformational entropy (i.e. inhibition

through conformational “trapping” into a single unfavorable state), but instead through a large redistribution of micro-states to ones that are less “binding competent.”

L_{12}-G_2 loop dynamics dictate KIX conformational entropy changes

To attribute changes in conformational entropy to specific structural elements on KIX, root-mean-square fluctuations (RMSF) are computed from MD simulations and compared across structural elements of KIX. For all complexes, residue-based fluctuations highlight that the L_{12}-G_2 loop provides the only major dynamical region outside of the protein termini (*Figure 2.12A*) — a finding supported by multiple studies.[35,39,40,44] Averaging the RMSF by secondary structural elements within KIX, we examined the RMSF as a function of KIX conformational entropy to explore which region most strongly correlates with the calculated entropy changes (see Methods, *Figure 2.17*). By comparing the slope of the correlation line as well as the regression coefficient (R^2), the L_{12}-G_2 loop exhibits both the largest slope as well as the highest R^2 value (*Figure 2.12B*), suggesting that loop dynamics can tune the protein’s conformational entropy.

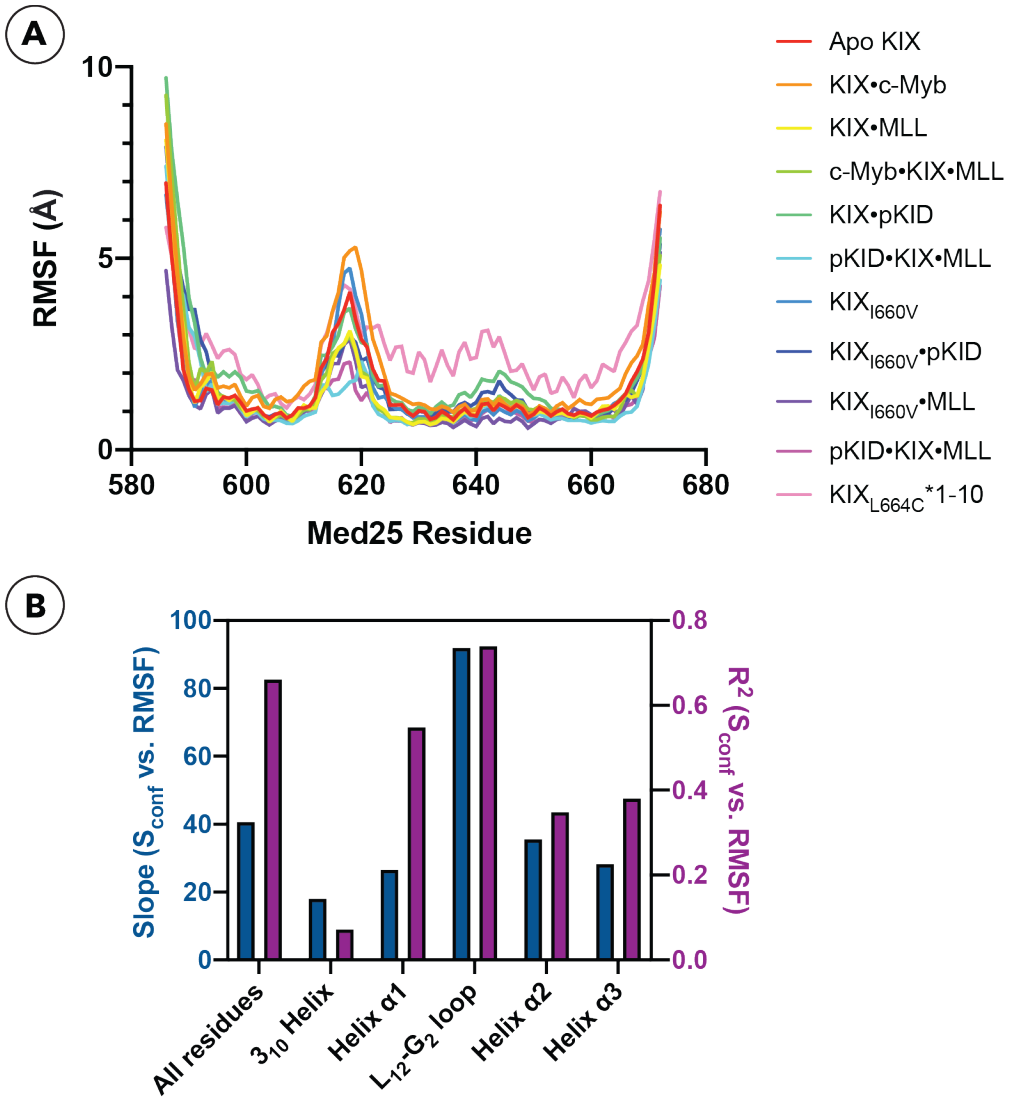


Figure 2.12. *Root mean square fluctuations (RMSF) and conformational entropy*. (A) RMSF by KIX residues for all systems tested here. (B) The correlation between RMSF and configurational entropy derived from side chain methyl order parameters were determined. Both the magnitude (slope, dark blue) as well as the goodness of fit (R^2, purple) of the correlation across different secondary structural elements of KIX show that fluctuations in the L_{12}-G_2 loop best correlate to KIX entropy.

Looking at residue-based methyl order parameters in KIX, the L_{12}-G_2 loop shows the largest differences between the unbound and bound complexes (*Figure 2.13*). In both of the wild type KIX systems, the loop becomes most stabilized in the ternary complex (*Figure 2.13*). However, in the c-Myb system, MLL binding alone rigidifies the loop to the same extent as the ternary complex (*Figure 2.13A*). Of particular significance, we see that the I660V mutation slightly stabilizes the loop in the apo distribution of states (*Figure 2.13B*, dashed lines). Further, pKID binding to KIX_{I660V} causes the loop to rigidify to the

same extent as when MLL binds, demonstrating how the mutation impacts loop dynamics and thus conformational entropy. On the inhibitory side, the opposite is observed — molecule 1-10 binding *increases* some loop motions, providing mechanistic insights into how allostery can be tuned through modulation of loop dynamics.

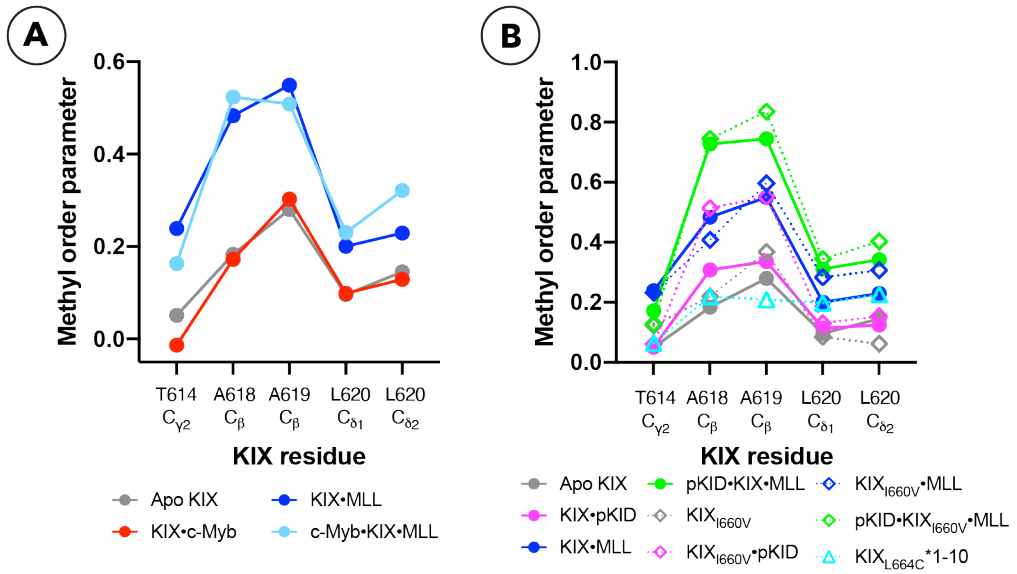


Figure 2.13. *Methyl order parameters of the L_{12}-G_2 loop by residue.* (A) Order parameters for (A) the c-Myb system and (B) the pKID system. Solid lines represent the wild type system, whereas dashed lines correspond to the allosteric modulators (I660V mutation and molecule 1-10).

2.3 Conclusions

Activator binding domains (ABDs) bind an impressive range of diverse partners that have unique primary sequences. While seemingly chaotic, these interactions are vital for gene regulation in eukaryotes, proving that plasticity is evolutionarily advantageous. Indeed, like other ABDs, CBP KIX can bind more than 15 diverse activators, and its role in diseases like cancer have prompted thorough investigations into the mechanisms of molecular recognition. Here we show how inherent malleability in KIX enables interactions with a diverse set of partners by having a wide spread of lowly-populated configurational states. Activators have a suite of cognate conformational states they bind, which in turn reweights the distribution of states. Allostery in turn is a manifestation of this reweighting process, where a second binding partner simply has a higher or lower chance of forming favorable interactions with KIX based on how the distribution of states was reweighted.

The conformational selection maps provide a qualitative illustration for decreasing KIX conformational entropy states, yet they are qualitative and do not aid in calculating entropy changes. NMR experiments have been innovatively applied to complicated problems like measuring protein conformational entropy. While the physical experiments were not performed here, MD simulations have proven to align well with experimental measurements. In calculating KIX conformational entropy changes using methyl order parameters, the quantitative measurements of allostery show differences between the c-Myb and pKID systems studied here. Conformational entropy changes in the c-Myb system align with total entropy changes measured via ITC, demonstrating how KIX configurational landscape dictates positive cooperativity. Alternatively, conformational entropy in the pKID system aligns with neither the total entropy changes nor the thermal stability changes measured via ITC. From this, we demonstrate how subtle changes in conformational entropy can be masked in global measurements, making techniques like methyl order parameters very valuable for detecting such changes.

Where methyl order parameters fall short is in being unable to detect shifts in conformational states that don't quantitatively change entropy measurements. Thus, the combination of mapping these redistributions with quantitative methods like order parameters gives the broadest mechanistic understanding of the critical role conformational entropy plays in dynamic allosteric systems like KIX.

By establishing the importance of conformational entropy in native KIX allostery, we hypothesized that perturbating the distribution of states through mutation and/or small molecule binding would result in attenuated allostery. Indeed, the two allosteric modulators tested here on the pKID system (K_{I660V} "turns on" pKID allostery; KIX_{L664C}*1-10 allosterically inhibits pKID) align with this model. KIX_{I660V} winnows down the distribution of states to look more "MLL-like", and in doing so the mutation pays an entropic cost ($-T\Delta S_{\text{conf}} = 2.3$ kcal/mol), which eliminates the need to have MLL pre-bound for enhanced binding. Alternatively, molecule 1-10 induces conformational changes unseen in the apo distribution of micro-states, providing mechanistic insights into its ability to allosterically inhibit pKID binding. Thus, this work highlights at the microscopic level strategies for future chemical probe development to enhance or inhibit particular binding interactions by perturbing the distribution of KIX micro-states.

2.4 Materials and methods

Pre-production and production of MD simulations

All simulations were done on GPUs using the CHARMM/OpenMM interface in the CHARMM molecular simulation package.[45–47] The atomic coordinates of the solution NMR structure the c-Myb•KIX•MLL ternary complex (PDB 2AGH[32]) were used to construct the starting structures for the four simulations (apo KIX, KIX•c-Myb, KIX•MLL, and c-Myb•KIX•MLL). Two other complexes (KIX•pKID and pKID•KIX•MLL) were constructed using the solution NMR structure of the pKID•KIX•MLL ternary complex (PDB 2LXT[33]). The KIX structures used for this work included 87 residues (586-672); c-Myb contained 25 residues (291-315); MLL contained 19 residues (839-857); and pKID contained 34 residues (116-149) with residue S133 being phosphorylated. KIX_{I660V} mutants were constructed using the initial coordinates from the solution NMR structure (PDB 2LXT[33]), with the mutations being made using CHARMM.[46] Simulations of KIX_{L664C}*1-10 were constructed using the crystal structure (PDB 4I9O[35]), with missing residues being built in through CHARMM. Molecule 1-10 was parameterized using CGENFF,[48] and the disulfide bond to KIX was reformed using the PATCH command in CHARMM. All systems were solvated with TIP3P water and neutralized with 100 mM NaCl using the MMTSB Toolset[49] so that each complex was in a cubic box with a minimum distance cutoff of 8-10 Å from the box edges (initial simulations were performed with an 8 Å cutoff, and the later simulations were run with the larger box sizes). Simulations were run using the CHARMM36 force field (and CGENFF when performing simulations using 1-10) in the NVT ensemble at 298 K using a Langevin dynamics algorithm with a friction coefficient of 5 ps^{-1}, and the SHAKE algorithm was used to fix bond lengths during simulations. PME and vswitch were used for nonbonded interactions using a 12 Å cutoff.

After the systems were constructed and solvated, the solvent was minimized by fixing the protein(s) and performing a 200-step minimization using the steepest descent algorithm. An additional 200 minimization steps using the steepest descent algorithm with a force constant of 40 kcal mol^{-1} $Å^{-2}$ on the protein(s) was performed prior to running MD simulations. In order to ensure that the systems were equilibrated, an initial 10 ns of restrained MD was run using a force constant of 10 kcal mol^{-1} $Å^{-2}$ on all protein heavy

atoms and using 2 fs timesteps. MD simulations were then initiated with no atomic restraints using periodic boundary conditions with 2 fs timesteps at 298 K. Each protein complex was simulated for a minimum of 100 ns (not including equilibration), and five independent trials were run per complex.

Post-production: structural clustering

After production of the simulations, water molecules and binding partners (c-Myb, MLL, and pKID, as well as molecule 1-10) were stripped away from each of the simulations. Combining all simulations trials by complex, 37,500 KIX coordinate frames were extracted. The remaining KIX structures were aligned and superposed by C_α backbone atoms to remove translational and rotational artifacts. The MMTSB Toolset function 'cluster.pl'[49] was used to first align C_α backbone atoms of all structure files and then cluster using the K-means clustering algorithm with varying cutoffs. For the clustering performed to estimate total KIX conformational entropy changes, all KIX residues (586-672) were used.

Conformational selection maps

Based on these extensive sampling data, we constructed conformational basin linkage maps, tracing the conformational states sampled in each complex and identifying the ones that represented the dominant basins sampled during subsequent binding processes. For each complex, KIX centroid structures were generated using a 2.5 Å cutoff on C_α atoms for residues 597-672. Populations of each KIX conformation were obtained using the number of frames that resided in a given cluster. The centroid of each cluster served as a reference to identify linked conformational basins between simulations of different complexes; we then took each centroid structure for a given bound complex and used it to determine where it originated from in the previous distribution (centroid structures were generated using the MMTSB Toolset function cluster.pl).[49] For instance, the KIX centroid structures in a given ternary complex were iteratively compared across all KIX binary centroids to find its structural origin, which was determined by calculating the minimum RMSD after superimposing the ternary centroid across all binary centroids. This procedure was then repeated for the binary KIX centroids to find their origins across

the distribution of apo KIX centroids. If a given centroid comparison did not yield a minimum RMSD ≤ the cutoff used for clustering (2.5 Å), we said that the centroid in the bound form induced a conformational change in KIX and thus was not traceable back to its structural origin.

Correlation functions and order parameters

Methyl order parameters were calculated for each methyl-bearing amino acid on KIX to calculate conformational entropy changes that occur with activator binding. Correlation functions of the second Legendre polynomials corresponding to the C-C axial bond were calculated using the CORREL module in CHARMM (**Eq 1**).

$$C(t) = \langle P_2(\hat{\mu}(0) \cdot \hat{\mu}(t)) \rangle \qquad \textbf{(Eq 1)}$$

In this, P_2 corresponds to the second order Legendre polynomial, and $\hat{\mu}$ corresponds to the unit vector along the C-C axial bond, which would be derived from relaxation experiments via NMR . The Lipari-Szabo (L-S) squared generalized order parameters can be calculated in the model-free formalism by:[50]

$$O^2 = \lim_{t\to\infty} C(t) = \lim_{t\to\infty} \langle P_2(\hat{\mu}(0) \cdot \hat{\mu}(t)) \rangle \qquad \textbf{(Eq 2)}$$

This can be rewritten using the Cartesian axes (x, y, z) of the unit vector of the C-C bond as such:

$$O^2 = \frac{3}{2}[\langle x^2 \rangle^2 + \langle y^2 \rangle^2 + \langle z^2 \rangle^2 + 2\langle xy \rangle^2 + 2\langle xz \rangle^2 + 2\langle yz \rangle^2] - \frac{1}{2} \quad \textbf{(Eq 3)}$$

Using this formula, side chain methyl order parameters were calculated on each side chain methyl group on KIX for all of the systems (amino acids with methyl side chains are Ala, Leu, Ile, Thr, Val, Met; *Figure 2.9*). Average side chain methyl order parameters are then used to calculate KIX conformational entropy for each bound state using the empirically derived equation (**Eq 4**):

$$S = 0.83 \sum N_\chi \, (0.91 - 0.74\langle O^2_{axis} \rangle) \qquad \textbf{(Eq. 4)}$$

where $\sum N_\chi$ is the sum of all side chain χ angles in the protein and $\langle O^2_{axis} \rangle$ is the L-S[51] squared generalized order parameter.[52] Side chain methyl order parameters are plotted by KIX residue (*Figure 2.14 and Figure 2.15*).

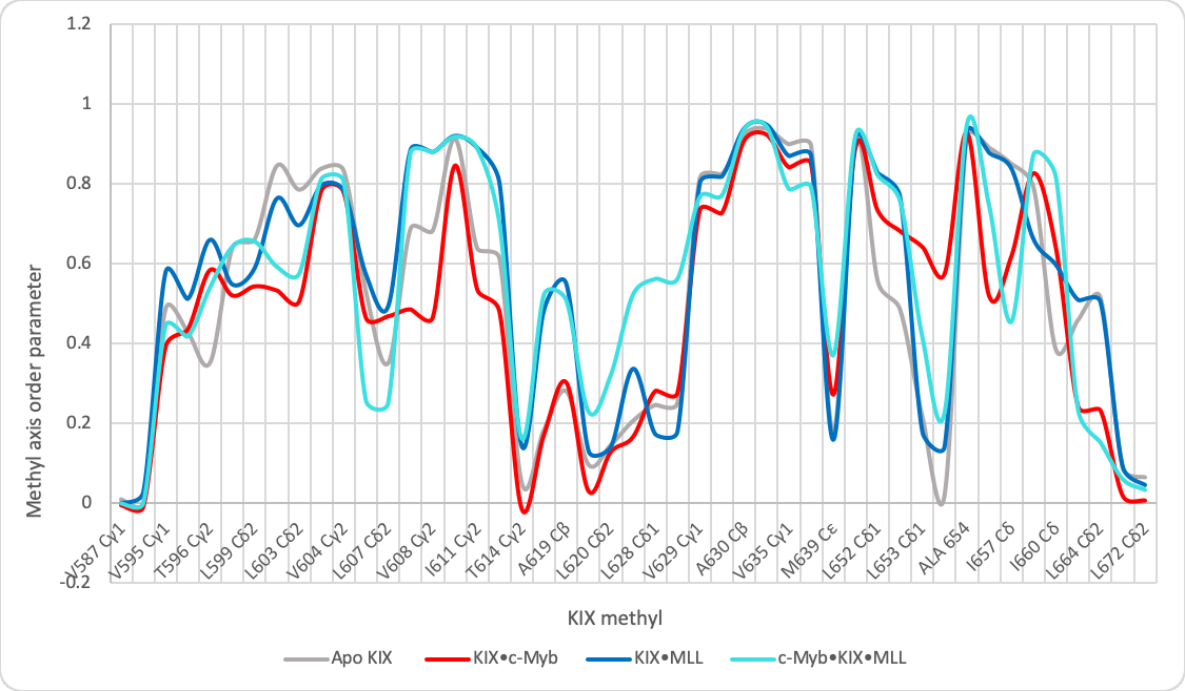


Figure 2.14. *Methyl order parameter by KIX residue for all complexes in the c-Myb system.* The complexes are colored as follows: apo KIX (gray), KIX•c-Myb (red), KIX•MLL (dark blue), and c-Myb•KIX•MLL (light blue). Order parameters for the L_{12}-G_2 region are shown in *Figure 2.13A*.

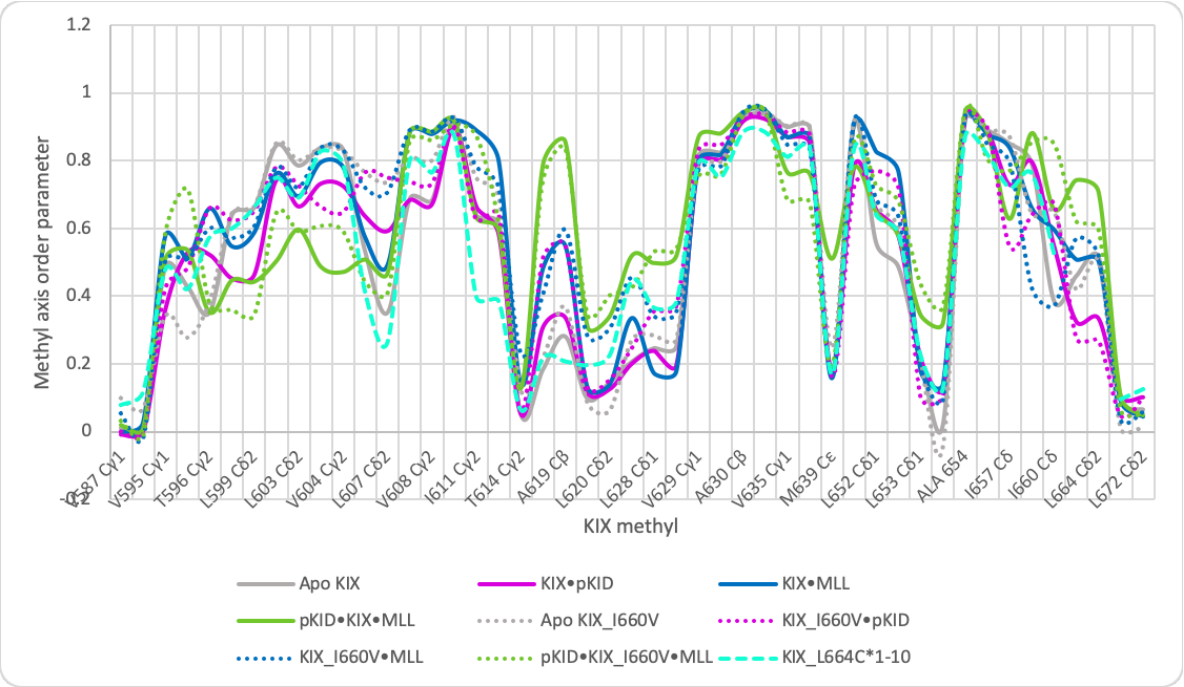


Figure 2.15. *Methyl order parameter by KIX residue for all complexes in the pKID system and mutant systems*. The complexes are colored as such: apo KIX (gray solid line), KIX•pKID (magenta solid line), KIX•MLL (dark blue solid line), pKID•KIX•MLL (green solid line), apo KIX_{I660V} (gray dotted line), KIX_{I660V}•pKID (magenta dotted line), KIX_{I660V}•MLL (dark blue dotted line), pKID•KIX_{I660V}•MLL (green dotted

line), and KIX_{L664C}*1-10 (cyan dashed line). Order parameters in the L_{12}-G_2 region are shown in *Figure 2.13B*.

Conformational entropy is associated with the number of attainable conformational micro-states of a molecule and its respective probabilities as such:

$$S = -k_B \sum_{n=1}^{W} P_n \ln P_n \qquad \textbf{(Eq. 5)}$$

where k_B is the Boltzmann constant, W is the number of conformational basins, and P_n is the probability of being in conformational basin n. Thus, if averaged methyl order parameters are reporting on conformational entropy, they should correlate with entropy values obtained using the K-means clustering algorithm of the KIX structures from a given trajectory. Indeed, we find that regardless of the radius cutoff on C_α atoms used for clustering, a high correlation to the averaged methyl order parameters is observed (*Figure 2.16*), indicating that the methyl order parameters consistently reflect conformational entropy changes in the KIX systems and would be useful tools in experimentally testing the ideas related here.

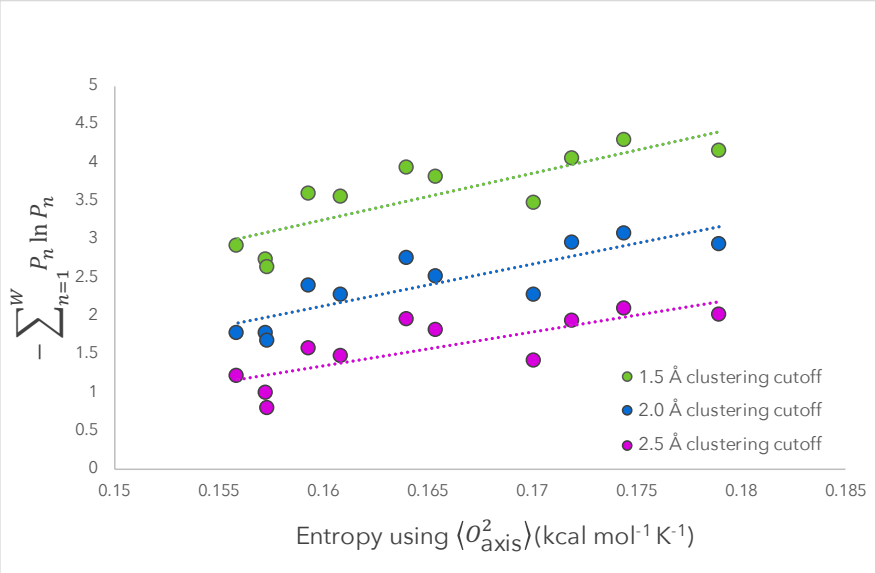


Figure 2.16. *Entropy calculations from clustering versus order parameters*. Entropy calculations of KIX from methyl order parameters versus calculations using the K-means clustering algorithm with varying cutoff. There is good agreement between the two entropy calculations regardless of the radius cutoff used for clustering (1.5 Å cutoff: $y = 60.7x - 6.5$, $R^2 = 0.68$; 2.0 Å cutoff: $y = 54.4x - 6.6$, $R^2 = 0.72$; 2.5 Å cutoff: $y = 43.9x - 5.7$, $R^2 = 0.61$).

Root mean squared fluctuations

Root mean squared fluctuations (RMSF) were calculated by residue on KIX using C_α backbone atoms. The average KIX structure for each system was taken from the average structure of the most highly populated cluster after K-means clustering with a 1.5 Å cutoff. Trajectories of KIX structures that were superposed by C_α atoms were used to calculate RMSF per residue of C_α atoms using Equation 6:

$$RMSF = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\delta_i^2} \qquad \textbf{(Eq 6)}$$

where *N* is the number of identical C_α atoms in the trajectory and δ_i^2 is the distance between atom *i* and the averaged structure. Average RMSF across the various secondary structural elements of KIX were calculated by averaging the mean squared fluctuations across the relevant number of residues and then taking the square root (*Figure 2.17*).

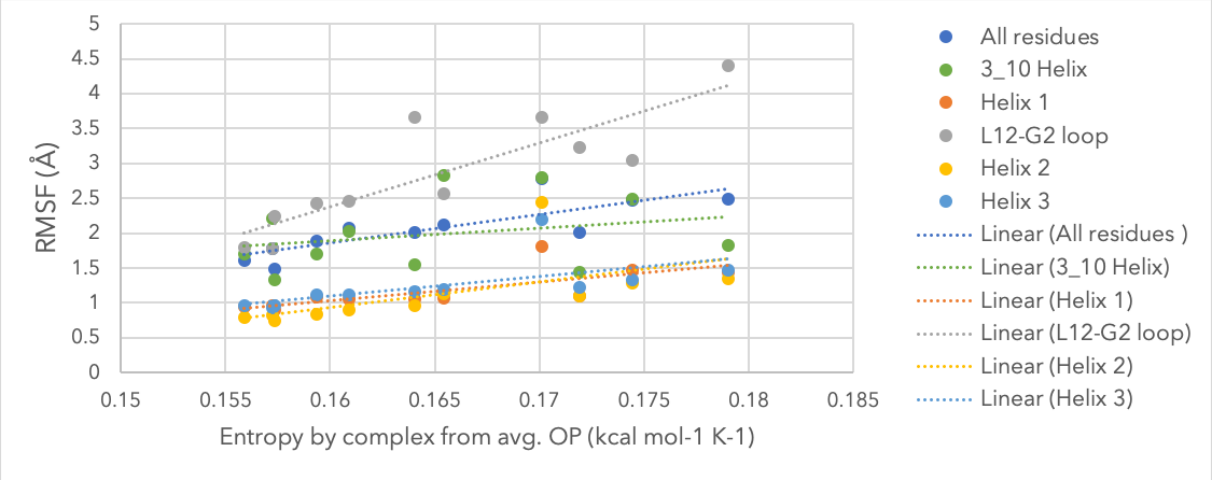


Figure 2.17. *Relationship between KIX substructures and overall protein conformational entropy*. In order to attribute the largest conformational entropy changes to specific KIX structural elements, RMSF was averaged by each secondary structure and plotted against total entropy for each complex. Entropy was calculated using average side chain methyl order parameters. The sections of KIX included were as follows: all residues (586-672 in dark blue), 3_{10} helix (591-594 in green), helix α1 (597-611 in red), L_{12}-G_2 loop (614-621 in gray), helix α2 (623-640 in yellow), and helix α3 (646-669 in light blue). Each secondary structural element listed was fit to a line to determine the correlation (slope) as well as goodness-of-fit (R^2), which is plotted in Figure 4a. The linear fits were as follows: All residues $y = 40.7x - 4.6, R^2 = 0.66$; 3_{10} helix $y = 18.1x - 1.0, R^2 = 0.07$; helix α1 $y = 26.6x - 3.2, R^2 = 0.55$; L_{12}-G_2 loop $y = 91.8x - 12.3, R^2 = 0.74$; helix α2 $y = 36.5x - 4.9, R^2 = 0.35$; and helix α3 $y = 28.2x - 3.4, R^2 = 0.38$.

Protein expression and purification

The DNA sequence encoding the KIX domain from mouse CBP, residues 586–672, was cloned into the bacterial expression pRSETB vector as previously described.[53]

The mutants at I660V and L664C were generated through site-directed mutagenesis, also as previously described.[54]

Apo KIX(586-672) and mutants (I660V and L664C) were expressed in BL21 DE3 *E. coli*. Cells were grown to an optical density (OD600 nm) of 0.8 (37°C, 250rpm), induced with 0.25 mM isopropyl b-d-1-thiogalactopyranoside (IPTG) for 16 hours (overnight) at 20°C, harvested by centrifugation (20 min, 6500 x g) 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. 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 10mM to 600mM 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 1 M. Purified protein was buffer-exchanged by dialysis (overnight, 4C) into 10 mM phosphate, 100 mM NaCl, 10% glycerol, pH 6.8. Purified protein samples were aliquoted and stored at -80°C.

Peptide synthesis and purification

All peptides were synthesized automatically with 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.

c-Myb(291-396) was synthesized and purified as described above with no modifications and isolated in >98% purity (See *Appendix A*).

c-Myb sequence: Ac-KEKRIKELELLLMSTENELKGQQALW-NH_2.

c-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 (See *Appendix A*).

MLL sequence: Ac-DCGNILPSDIMDFVLKNTPY-NH_2.

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% purify (See *Appendix A*).

pKID sequence: Ac-TDSQKRREILSRRPS(Phos)YRKILNDLSSDAPG-NH_2.

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

Differential scanning fluorimetry

Experiments were conducted utilizing 20 μL sample volumes in 96 well PCR plates sealed with clear cap strips. To determine T_m, 20 μM protein in the presence of 5X SYPRO orange dye (1:1000 dilution of purchased 5000X stock; Invitrogen) was incubated with ligand (4X 1-10, 1X TAD peptide) 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 (See *Appendix*

B). RFU was imported into the online data analysis program, DSFworld, and Tm 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.

2.5 References

(1) Garcia-Pino, A.; Balasubramanian, S.; Wyns, L.; Gazit, E.; De Greve, H.; Magnuson, R. D.; Charlier, D.; van Nuland, N. A. J.; Loris, R. Allostery and Intrinsic Disorder Mediate Transcription Regulation by Conditional Cooperativity. *Cell* **2010**, *142* (1), 101–111. https://doi.org/10.1016/j.cell.2010.05.039.

(2) Dyson, H. J.; Wright, P. E. Intrinsically Unstructured Proteins and Their Functions. *Nat. Rev. Mol. Cell Biol.* **2005**, *6* (3), 197–208. https://doi.org/10.1038/nrm1589.

(3) Hilser, V. J.; Thompson, E. B. Intrinsic Disorder as a Mechanism to Optimize Allosteric Coupling in Proteins. *Proc. Natl. Acad. Sci.* **2007**, *104* (20), 8311–8315. https://doi.org/10.1073/pnas.0700329104.

(4) Monod, J.; Jacob, F. Teleonomic Mechanisms in Cellular Metabolism, Growth, and Differentiation. *Cold Spring Harb. Symp. Quant. Biol.* **1961**, *26*, 389–401.

(5) Changeux, J. P. The Feedback Control Mechanisms of Biosynthetic L-Threonine Deaminase by L-Isoleucine. *Cold Spring Harb. Symp. Quant. Biol.* **1961**, *26*, 313–318.

(6) Monod, J.; Wyman, J.; Changeux, J.-P. On the Nature of Allosteric Transitions: A Plausible Model. *J. Mol. Biol.* **1965**, *12* (1), 88–118. https://doi.org/10.1016/S0022-2836(65)80285-6.

(7) Dickerson, R. E. X-Ray Studies of Protein Mechanisms. *Annu. Rev. Biochem.* **1972**, *41* (1), 815–842. https://doi.org/10.1146/annurev.bi.41.070172.004123.

(8) Perutz, M. F.; Wilkinson, A. J.; Paoli, M.; Dodson, G. G. The Stereochemical Mechanism of the Cooperative Effects in Hemoglobin Revisited. *Annu. Rev. Biophys. Biomol. Struct.* **1998**, *27*, 1–34. https://doi.org/10.1146/annurev.biophys.27.1.1.

(9) Changeux, J.-P.; Edelstein, S. J. Allosteric Mechanisms of Signal Transduction. *Science* **2005**, *308* (5727), 1424–1428. https://doi.org/10.1126/science.1108595.

(10) Laskowski, R. A.; Gerick, F.; Thornton, J. M. The Structural Basis of Allosteric Regulation in Proteins. *FEBS Lett.* **2009**, *583* (11), 1692–1698. https://doi.org/10.1016/j.febslet.2009.03.019.

(11) Tsai, C.-J.; Del Sol, A.; 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.

(12) del Sol, A.; Tsai, C.-J.; Ma, B.; Nussinov, R. The Origin of Allosteric Functional Modulation: Multiple Pre-Existing Pathways. *Struct. Lond. Engl. 1993* **2009**, *17* (8), 1042–1050. https://doi.org/10.1016/j.str.2009.06.008.

(13) Tsai, C.-J.; del Sol, A.; Nussinov, R. Allostery: Absence of a Change in Shape Does Not Imply That Allostery Is Not at Play. *J. Mol. Biol.* **2008**, *378* (1), 1–11. https://doi.org/10.1016/j.jmb.2008.02.034.
(14) Tsai, C.-J.; Nussinov, R. A Unified View of "How Allostery Works." *PLOS Comput. Biol.* **2014**, *10* (2), e1003394. https://doi.org/10.1371/journal.pcbi.1003394.
(15) Brünger, A. T.; Brooks III, C. L.; Karplus, M. Active Site Dynamics of Ribonuclease. *Proc. Natl. Acad. Sci.* **1985**, *82* (24), 8458–8462. https://doi.org/10.1073/pnas.82.24.8458.
(16) Radkiewicz, J. L. Protein Dynamics in Enzymatic Catalysis: Exploration of Dihydrofolate Reductase. *J. Am. Chem. Soc.* **2000**, *122* (2), 225–231. https://doi.org/10.1021/ja9913838.
(17) Hilser, V. J.; Wrabl, J. O.; Motlagh, H. N. Structural and Energetic Basis of Allostery. *Annu. Rev. Biophys.* **2012**, *41* (1), 585–609. https://doi.org/10.1146/annurev-biophys-050511-102319.
(18) Motlagh, H. N.; Hilser, V. J. Agonism/Antagonism Switching in Allosteric Ensembles. *Proc. Natl. Acad. Sci.* **2012**, *109* (11), 4134–4139. https://doi.org/10.1073/pnas.1120519109.
(19) Zimmermann, J.; Oakman, E. L.; Thorpe, I. F.; Shi, X.; Abbyad, P.; Brooks III, C. L.; Boxer, S. G.; Romesberg, F. E. Antibody Evolution Constrains Conformational Heterogeneity by Tailoring Protein Dynamics. *Proc. Natl. Acad. Sci.* **2006**, *103* (37), 13722–13727. https://doi.org/10.1073/pnas.0603282103.
(20) 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.
(21) 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.
(22) Radhakrishnan, I.; Pérez-Alvarado, G. C.; Parker, D.; Dyson, H. J.; Montminy, M. R.; Wright, P. E. Structural Analyses of CREB-CBP Transcriptional Activator-Coactivator Complexes by NMR Spectroscopy: Implications for Mapping the Boundaries of Structural Domains11Edited by F. E. Cohen. *J. Mol. Biol.* **1999**, *287* (5), 859–865. https://doi.org/10.1006/jmbi.1999.2658.
(23) 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.
(24) 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.
(25) Vendel, A. C.; McBryant, S. J.; Lumb, K. J. KIX-Mediated Assembly of the CBP−CREB−HTLV-1 Tax Coactivator−Activator Complex. *Biochemistry* **2003**, *42* (43), 12481–12487. https://doi.org/10.1021/bi0353023.
(26) 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.

(27) Vendel, A. C.; Lumb, K. J. NMR Mapping of the HIV-1 Tat Interaction Surface of the KIX Domain of the Human Coactivator CBP. *Biochemistry* **2004**, *43* (4), 904–908. https://doi.org/10.1021/bi035612l.
(28) 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.
(29) Hess, J. L.; Bittner, C. B.; Zeisig, D. T.; Bach, C.; Fuchs, U.; Borkhardt, A.; Frampton, J.; Slany, R. K. C-Myb Is an Essential Downstream Target for Homeobox-Mediated Transformation of Hematopoietic Cells. *Blood* **2006**, *108* (1), 297–304. https://doi.org/10.1182/blood-2005-12-5014.
(30) Graf, T. Myb: A Transcriptional Activator Linking Proliferation and Differentiation in Hematopoietic Cells. *Curr. Opin. Genet. Dev.* **1992**, *2* (2), 249–255.
(31) 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.
(32) De Guzman, R. N.; Goto, N. K.; Dyson, H. J.; Wright, P. E. Structural Basis for Cooperative Transcription Factor Binding to the CBP Coactivator. *J. Mol. Biol.* **2006**, *355* (5), 1005–1013. https://doi.org/10.1016/j.jmb.2005.09.059.
(33) Brüschweiler, S.; Konrat, R.; Tollinger, M. Allosteric Communication in the KIX Domain Proceeds through Dynamic Repacking of the Hydrophobic Core. *ACS Chem. Biol.* **2013**, *8* (7), 1600–1610. https://doi.org/10.1021/cb4002188.
(34) 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.* **2014**, *111* (33), 12061–12066. https://doi.org/10.1073/pnas.1406033111.
(35) Wang, N.; Majmudar, C. Y.; Pomerantz, W. C.; Gagnon, J. K.; Sadowsky, J. D.; Meagher, J. L.; Johnson, T. K.; Stuckey, J. A.; Brooks III, 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.
(36) 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.
(37) Huynh, K.; Partch, C. L. Analysis of Protein Stability and Ligand Interactions by Thermal Shift Assay. *Curr. Protoc. Protein Sci.* **2015**, *79* (1), 28.9.1-28.9.14. https://doi.org/10.1002/0471140864.ps2809s79.
(38) Wright, T. A.; Stewart, J. M.; Page, R. C.; Konkolewicz, D. Extraction of Thermodynamic Parameters of Protein Unfolding Using Parallelized Differential Scanning Fluorimetry https://pubs.acs.org/doi/pdf/10.1021/acs.jpclett.6b02894 (accessed Dec 18, 2020). https://doi.org/10.1021/acs.jpclett.6b02894.
(39) Law, S. M.; Gagnon, J. K.; Mapp, A. K.; Brooks III, 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.

(40) Wang, Y.; Brooks III, C. L. Enhanced Sampling Applied to Modeling Allosteric Regulation in Transcription. *J. Phys. Chem. Lett.* **2019**, *10* (19), 5963–5968. https://doi.org/10.1021/acs.jpclett.9b02226.
(41) Parker, D.; Rivera, M.; Zor, T.; Henrion-Caude, A.; Radhakrishnan, I.; Kumar, A.; Shapiro, L. H.; Wright, P. E.; Montminy, M.; Brindle, P. K. Role of Secondary Structure in Discrimination between Constitutive and Inducible Activators. *Mol. Cell. Biol.* **1999**, *19* (8), 5601–5607. https://doi.org/10.1128/MCB.19.8.5601.
(42) Radhakrishnan, I.; Pérez-Alvarado, G. C.; Dyson, H. J.; Wright, P. E. Conformational Preferences in the Ser133-Phosphorylated and Non-Phosphorylated Forms of the Kinase Inducible Transactivation Domain of CREB. *FEBS Lett.* **1998**, *430* (3), 317–322. https://doi.org/10.1016/S0014-5793(98)00680-2.
(43) Zor, T.; Mayr, B. M.; Dyson, H. J.; Montminy, M. R.; Wright, P. E. Roles of Phosphorylation and Helix Propensity in the Binding of the KIX Domain of CREB-Binding Protein by Constitutive (c-Myb) and Inducible (CREB) Activators. *J. Biol. Chem.* **2002**, *277* (44), 42241–42248. https://doi.org/10.1074/jbc.M207361200.
(44) Korkmaz, E. N.; Nussinov, R.; Haliloğlu, T. Conformational Control of the Binding of the Transactivation Domain of the MLL Protein and C-Myb to the KIX Domain of CREB. *PLOS Comput. Biol.* **2012**, *8* (3), e1002420. https://doi.org/10.1371/journal.pcbi.1002420.
(45) 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.
(46) 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.; Hodoscek, 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.
(47) Eastman, P.; Swails, J.; Chodera, J. D.; McGibbon, R. T.; Zhao, Y.; Beauchamp, K. A.; Wang, L.-P.; Simmonett, A. C.; Harrigan, M. P.; Stern, C. D.; Wiewiora, R. P.; Brooks, B. R.; Pande, V. S. OpenMM 7: Rapid Development of High Performance Algorithms for Molecular Dynamics. *PLOS Comput. Biol.* **2017**, *13* (7), e1005659. https://doi.org/10.1371/journal.pcbi.1005659.
(48) 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 (CGenFF): 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.
(49) 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.
(50) Lipari, G.; Szabo, A. Model-Free Approach to the Interpretation of Nuclear Magnetic Resonance Relaxation in Macromolecules. 2. Analysis of Experimental

Results. *J. Am. Chem. Soc.* **1982**, *104* (17), 4559–4570. https://doi.org/10.1021/ja00381a010.
(51) Lipari, G.; Szabo, A. Model-Free Approach to the Interpretation of Nuclear Magnetic Resonance Relaxation in Macromolecules. 1. Theory and Range of Validity. *J. Am. Chem. Soc.* **1982**, *104* (17), 4546–4559. https://doi.org/10.1021/ja00381a009.
(52) Kasinath, V.; Sharp, K. A.; Wand, A. J. Microscopic Insights into the NMR Relaxation-Based Protein Conformational Entropy Meter. *J. Am. Chem. Soc.* **2013**, *135* (40), 15092–15100. https://doi.org/10.1021/ja405200u.
(53) 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.
(54) 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.

Chapter 3

The Coactivator Med25 AcID Exhibits High Fluctuations in Activator Engagement[1,2]

Abstract

Transcriptional coactivators are a molecular recognition marvel because a single domain within these proteins, the activator binding domain or ABD, interacts with multiple compositionally diverse transcriptional activators. Also remarkable is the structural diversity among ABDs, which range from conformationally dynamic helical motifs to those with a stable core such as a β-barrel. A significant objective is to define conserved properties of ABDs that allow them to interact with disparate activator sequences. The ABD of the coactivator Med25 (activator interaction domain or AcID) is unique in that it contains secondary structural elements that are on both ends of the spectrum: helices and loops that display significant conformational mobility and a seven-stranded β-barrel core that is structurally rigid. Using biophysical approaches that span from molecular dynamics simulations to NMR and kinetic experiments, we build a mechanistic model of how AcID forms complexes with the activators VP16 and ERM, as well as a small molecule inhibitor norstictic acid. In doing so, we find that despite its static core, Med25 forms short-lived, conformationally mobile, and structurally distinct complexes with both cognate binding partners. We establish that Med25 utilizes conformational malleability through flanking loops and helices to uniquely accommodate both activators, forming

[1] Portions of this chapter were adopted from: Henderson, A. R. Henley, M. J.; Foster, N. J.; **Peiffer, A. L.**; Beyersdorf, M. D.; Standford, K. D., Sturlis, S. M.; Linhares, B. M.; Hill, Z. B.; Wells, J. A.; Cierpicki, T.; Brooks III, C. L.; Fierke, C. A.; Mapp, A. K. "Conservation of coactivator engagement mechanism enables small-molecule allosteric modulators," *PNAS*, **115**, 8960-8965 (2018).

[2] Portions of this chapter were adopted from: Garlick, J. M.; Sturlis, S. M.; Bruno, P. A.; Yates, J. A.; **Peiffer, A. L.**; Liu, Y.; Goo, L.; Bao, L.; De Salle, S. N.; Tamayo-Castillo, G.; Brooks III, C. L.; Merajver, S. D.; Mapp, A. K. "Norstictic acid is a selective allosteric transcriptional regulator," *J. Am. Chem. Soc.*, **143**, 9297-9302 (2021).

distinct complexes seen in modeling as well as NMR. Further, we find that norstictic inhibits ortho- and allosterically through modulating dynamics across the protein.

3.1 Introduction

The Mediator complex is a massive coactivator complex comprising 30 individual proteins.[1–3] Its role in transcription has deemed it as a "master coordinator" as it serves as a bridge between promoter-bound activators and remaining transcriptional machinery.[1,2,4] Independent activator binding domains (ABDs) occupy the tail portion of the Mediator and directly impact gene expression by forming protein-protein interactions with promoter-bound activators.[5] As such, understanding these interactions is great interest. Med25 anchors to the tail subunit of the Mediator using a VWA domain (*Figure 3.1*)[6]. A long linker connects the VWA to the activator interacting domain (AcID) of Med25, which is a binding partner of a diverse array of transcriptional activators, including VP16,[6,7] ATF6α,[8] and the ETV/PEA3 activators.[9,10] Through these interactions, Med25 plays significant roles in the unfolded protein response and in oncogenesis, generating significant interest into gaining an atomic level understanding these interactions.

While the structure of Med25 AcID has been solved via NMR,[11,12] structural elucidation has not been achieved with any of the AcID•activator interactions. This is in part because the activators cause extensive line broadening of many peaks, which is a result of the transient and conformationally dynamic interactions that are formed. This "fuzziness" inherent to the native interactions has made targeted drug discovery an unsolved problem. The VP16 transcriptional activation domain contacts a surface of ~1,800 Å^2 of AcID, wrapping around the topologically challenging β-barrel while also contacting two flanking helices.[11] The transcriptional activation domain of the ETV/PEA3 member ERM interacts with one face of the β-barrel, a binding surface referred to as H1 that is ~900 Å^2 in area[9,10] (*Figure 3.1*). The β-barrel core of AcID is unusual among ABDs, with helices more commonly observed, and raises the question of the role that the barrel might play in the molecular recognition of activators relative to the other substructures within AcID.

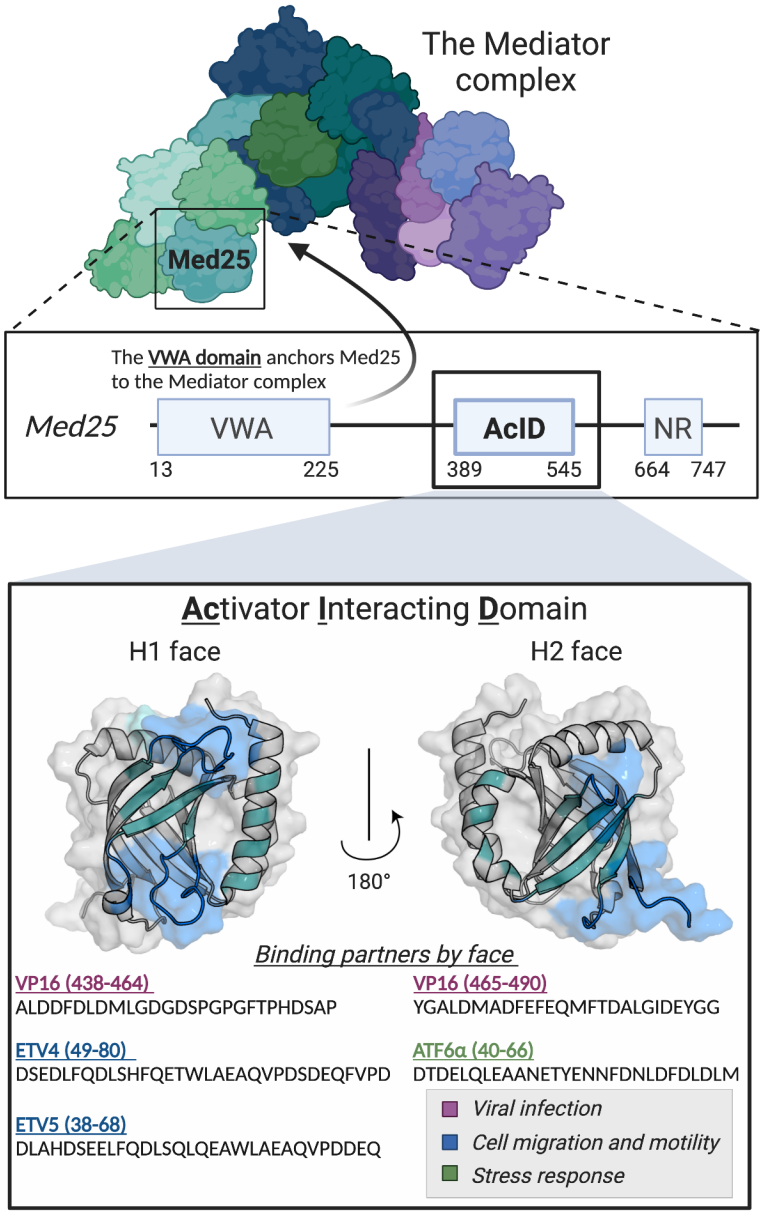


Figure 3.1. *Med25 is a part of the tail subunit of the human Mediator complex*. The **Ac**tivator-**I**nteraction **D**omain (AcID) is the binding partner of a growing number of transcriptional activators and contains at least two binding surfaces, termed H1 and H2. The sequences of the transcriptional activation domains of the three Med25-dependent activators used in this study are shown below the protein structure (PDB ID 2XNF).

The observation that a portion of VP16 and ERM utilize the same H1 binding surface in AcID despite their distinct sequences suggests that conformational plasticity within the ABD could play a role in its molecular recognition capabilities and, ultimately, function. If Med25 AcID indeed behaves like typical ABDs, we hypothesize that each

activator–AcID complex should be conformationally labile, with two or more conformation states energetically accessible. We assess this through a combination of biophysical techniques to dissect binding events with both VP16 and ERM, relying on both molecular dynamics simulations and *in vitro* experiments. Transient kinetic experiments with the activators revealed that in each case AcID exploits conformational lability to recognize the distinct sequences, aligning with multiple bound complexes seen in the structural models from MD. Simulations highlight the critical role that the flexible loops and helices play in the remodeling of one PPI surface, while also suggesting how these motions relate to the larger family of ABDs. Taken together, these data suggest a conserved mechanism for transcriptional activators despite considerable structural divergence.

3.2 Results and discussion

Disulfide Tethering with VP16 circumvents limitations of structural techniques

A signature of activator•coactivator interactions is fuzziness, wherein complexes adopt multiple conformational states that are separated by low energy barriers. As crystallographic techniques require a protein to form a single stable conformation, NMR is the major tool used to delineate structural information in these systems. However, multiple binding modes can be exchanging on the intermediate timescale, causing extensive line broadening and defying structural characterization.[13] Such is the case with Med25 AcID, where no structures have been solved of the protein in complex with binding partners. While NMR has thus far been unsuccessful in full structural characterization of the bound forms, chemical shift perturbations from NMR provide useful insights into binding interactions. Separate NMR studies of AcID in complex with the transcriptional activation domains of VP16 and ERM suggest that the two activators both contact the H1 binding surface, with the significantly larger VP16 also interacting with the H2 surface[7,11,14] (*Figure 3.1*). In order to build more complete structural models for these interactions, we utilized a suite of biophysical techniques that includes both theory/molecular dynamics simulations and *in vitro* experiments.

Disulfide Tethering was utilized to construct a preliminary model of how H1-binding partners, namely VP16 and ERM, interact on that binding surface. Disulfide Tethering works by measuring complex formation via mass spec in the presence of a competitor

thiol so that complexes will only form if there are favorable interactions between the protein and peptide.[15] Ultimately, by varying the position of the H1-binding transcriptional activation domain (TAD) of VP16(438-454), a model for the interaction can be deconstructed. Med25 AcID contains three native cysteine residues (*Figure 3.2B*). However, C249 is on a beta strand and points into the beta barrel, making it inaccessible for disulfide Tethering. Disulfide Tethering experiments are performed by systematically mutating VP16(438-454) residues with cysteines, which are capped using cystamine (*Figure 3.2A*). The VP16 TADs are incubated with Med25 AcID and a competitor thiol, β-mercaptoethanol, thus disulfide bonds only formed between Med25 and VP16 when there are favorable binding interactions (*Figure 3.2B*).

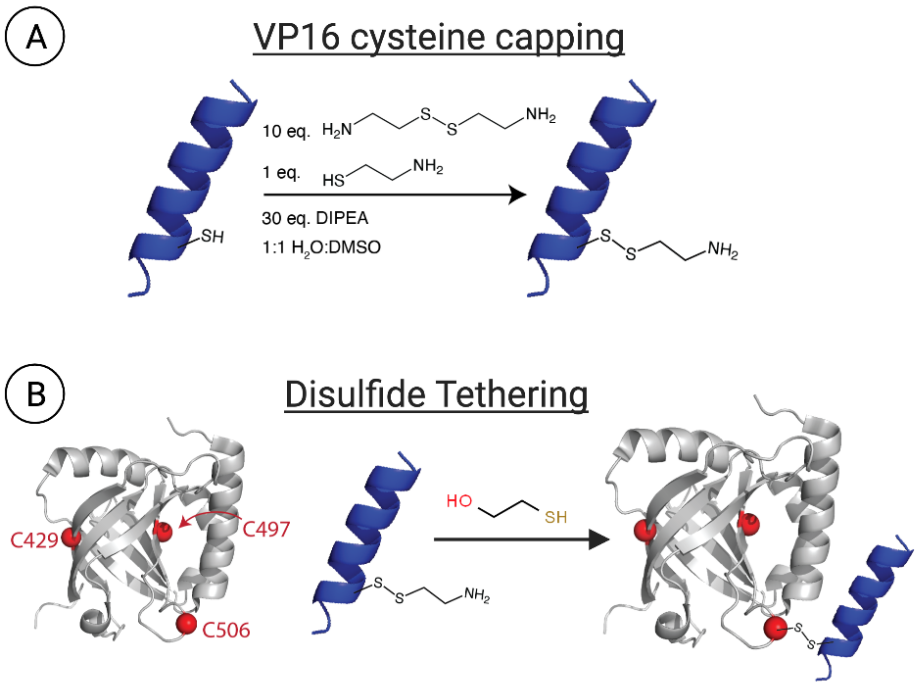


Figure 3.2. *Disulfide Tethering with H1-binding TADs*. (A) Cysteine residues are inserted in VP16 (438-454) (blue helix) and are capped using 10 equivalents of cystamine. (B) The capped TAD is used in disulfide Tethering with Med25 AcID (PDB 2XNF, shown in gray; the three native cysteines are shown as red spheres) in the presence of a competitor thiol, β-mercaptoethanol. Med25 AcID has three native cysteines: C429, which points into the beta barrel; C497, which is behind helix α2; and C506, which is on a loop on the H1 face.

Tethering percentages for all VP16(438-454)$_{cys.\ mutant}$ peptides were determined using mass spectrometry, where tethered adducts were compared to the apo protein

parent mass. In doing so, VP16(438-454)$_{cys.\ mutant}$ was found to only form stable covalent complexes when the cysteine was towards the C-terminal portion of the TAD (residues 450-454; *Figure 3.3A*). Tethering experiments performed using Med25 mutants at the two solvent accessible cysteines (C497A and C506A) indicate that VP16 only reacts with C506, with no detectable Tethering occurring at C497 (*Figure 3.3B*). This provided a general starting orientation for how VP16(438-454) interacts on the H1 binding surface.

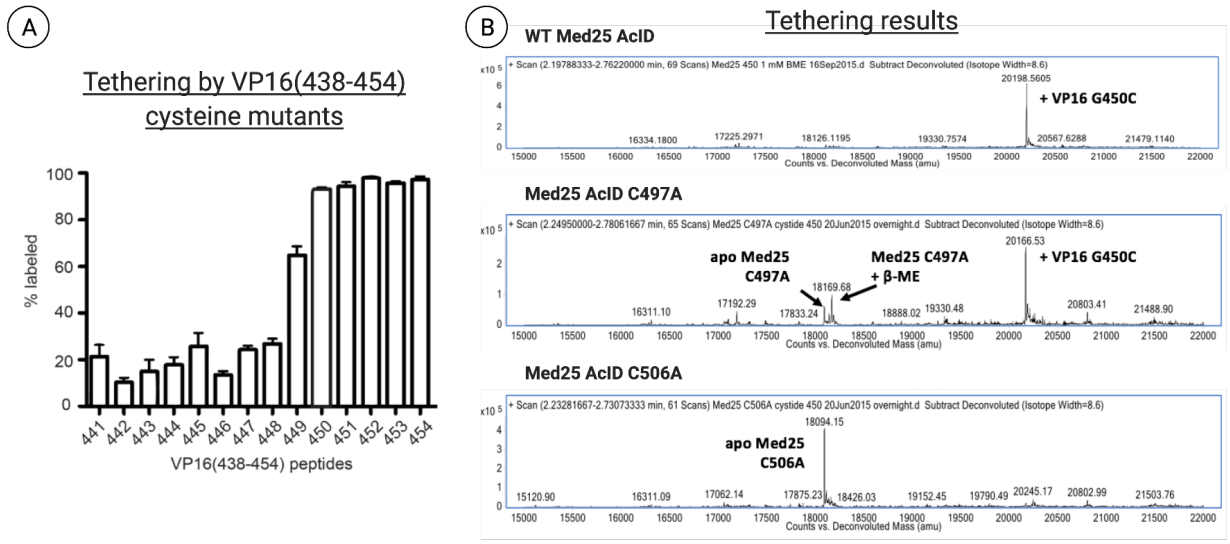


Figure 3.3. *Disulfide Tethering with VP16(438-454).* (A) Tethering percentages are determined via mass spectrometry at varying cysteine positions across VP16(438-454). (B) Tethering positions on Med25 AcID are determined using alanine mutagenesis on the two solvent-accessible cysteines, C497 and C506, which yields that Tethering only occurs at C506. Experiments performed by Dr. Andrew Henderson.

Temperature replica exchange overcomes energy barriers in simulations

Molecular dynamics (MD) simulations are a powerful tool for obtaining atomic-level data around binding interactions, yet much like structural experiments with NMR and crystallography, there are inherent challenges and limitations. For example, capturing a protein-protein interaction in an unbiased simulation and fully solvated water box would entail an inordinate amount of computing power for a single trial, let alone experimental replicates. The unbiased nature of these simulations allows for more exploration, and real mechanisms of association can potentially be captured. However, as binding can occur on the micro- or millisecond timescale, this method wouldn't be optimal or realistic for these protein systems. Further, we know from NMR that while Med25's binding partners do have preferential binding sites, they are capable and do in fact bind to other locations

on the protein. Thus, the transient nature of these associations suggests alternative and more advanced modeling methods would be more fruitful.

A major challenge in MD is overcoming large energy barriers that can be involved in binding (*Figure 3.4*). Thus, enhanced sampling methods have been developed to allow for more efficient exploration of conformational space.[16,17] Temperature replica exchange is an enhanced sampling technique where multiple simulations are run in parallel at increasing temperatures.[17,18] As see in *Figure 3.4*, while the simulation of interest occurs at the lowest temperature, higher temperature simulations allow for overcoming large energy barriers. Periodic exchanges between the replicates are attempted, so that these barriers are overcome, and we are able to collect data for structural states in the thermal distribution of interest.

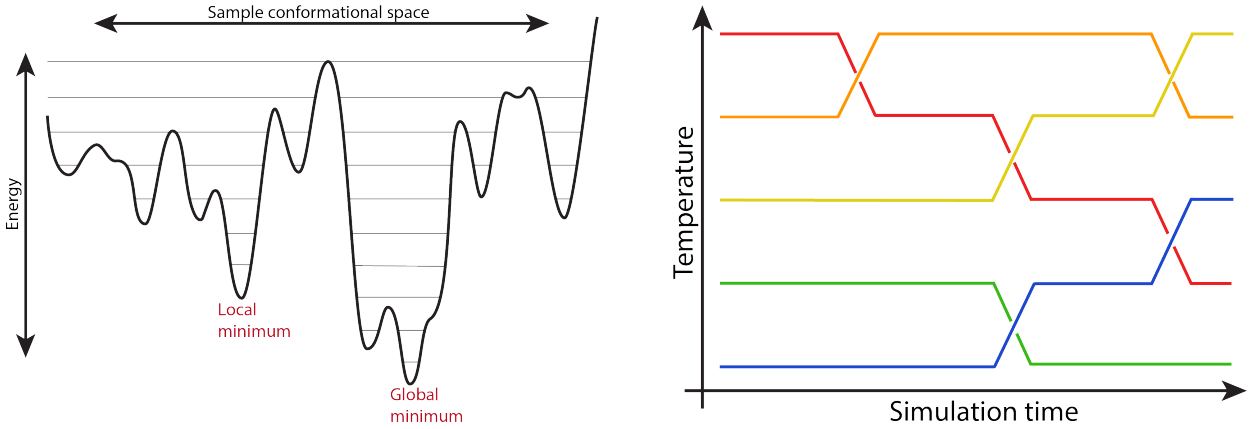


Figure 3.4. *Temperature replica exchange method as an ehanced sampling simulation technique*. (Left) A major challenge in simulation is overcoming large energy barriers associated with binding. However, simulations at higher temperatures can be used to overcome these barriers. (Right) Temperature replica exchange involves running multiple simulations in parallel and periodically exchanging coordinates between distributions, which allows for large energy barriers to be more easily overcome.

All-atom molecular dynamics simulations via temperature replica exchange were used for both apo Med25 as well as tethered to VP16(438-454)$_{cys.\ mutant}$. As many activators have shown a propensity for comprising of mostly helices, VP16 was constructed as a helix and tethered to C506 using a disulfide patch in CHARMM; the “unbound” peptide was oriented out into space, as to not bias simulations toward a particular bound pose (*Figure 3.5*).

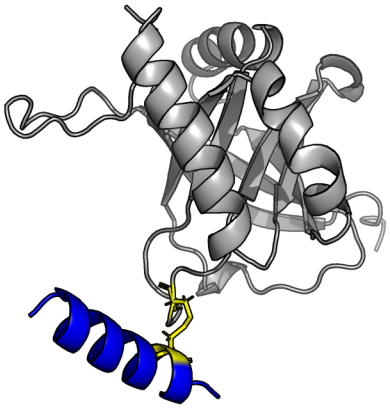


Figure 3.5. *Simulation setup for temperature replica exchange with a Tethered peptide.* The NMR structure of Med25 (PDB 2XNF; shown in gray) was utilized in constructing the initial structure of the VP16(438-454)$_{G450C}$ construct, where VP16 is shown as the blue peptide, and the disulfide bond to Med25$_{C506}$ is shown in yellow.

Simulations ran for 60ns with 12 replicas ranging in temperature from 280 to 500 K. Data from the 280 K replicas were ultimately used for analysis, where frames from those simulations were extracted (4,000 frames / simulation) and K-means clustered using a 1.5 Å RMSD cutoff on C_α atoms. In doing so, we found that VP16(438-454)$_{G450C}$ appears to most significantly stabilize the protein; while the apo protein simulations yields 20 clusters with many that are all relatively lowly populated, the VP16(438-454)$_{G450C}$ complex results in only 5 clusters with a clear majority for one of the clusters (72% of all frames; *Figure 3.6B*). This finding is even more apparent when looking at root mean square fluctuations (RMSF) of the apo protein in comparison to Med25+VP16(438-454)$_{G450C}$ (*Figure 3.7*). In particular, much like with KIX, we find that the most dynamic region on the protein, the upper loop on the H1 face, shows the greatest stabilization upon binding. While Tethering experiments suggest that any VP16 binds with cysteine mutants at any position from 450 to 454, we find that the most specific interactions occur at the G450C position. Hence, this particular VP16(438-454) construct was further pursued.

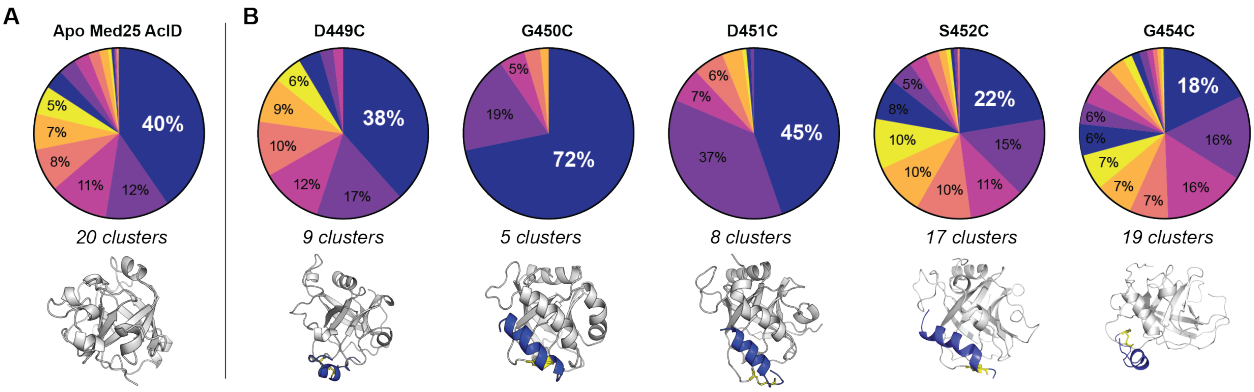


Figure 3.6. *Clustering results using a K-means clustering algorithm.* All clusters with ≥ 5% occupancy are labeled. (A) Pie chart showing the breakdown of clusters for apo Med25 AcID, with the centroid structure from the most populated cluster shown below. (B) Clustering results of Med25 AcID tethered to VP16(438-454) with varying cysteine mutations. Centroid structures of the most populated cluster for each bound complex are shown below.

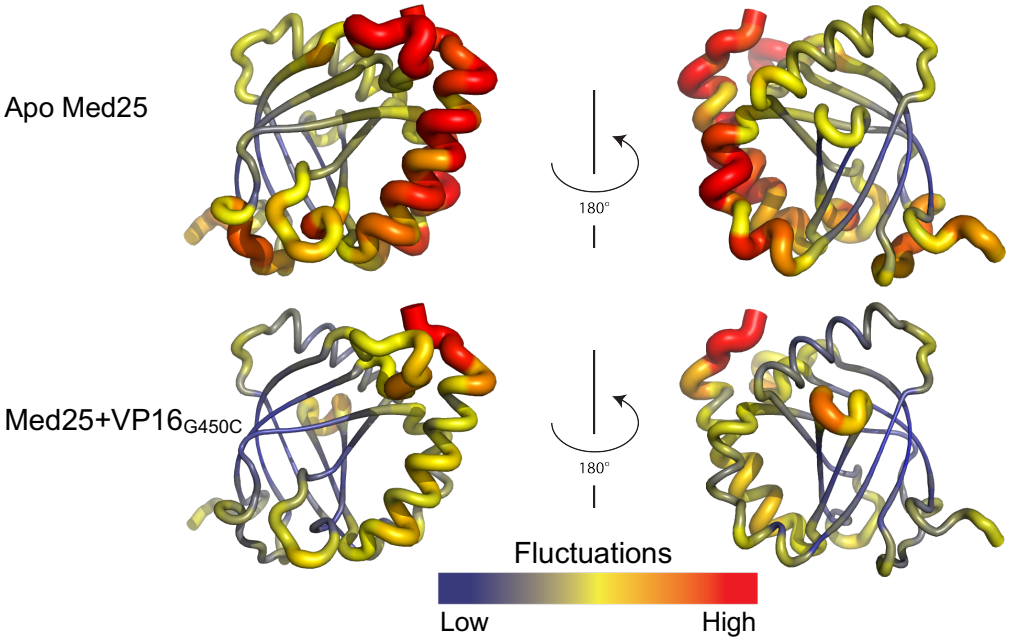


Figure 3.7. *Fluctuations in the emerging structural model for AcID-activtor complex formation.* Root mean square fluctuations (RMSF) were calculated for each Med25 AcID residue by overlaying Cα atoms for all coordinate files produced from simulations. The RMSF are projected onto the NMR structure for Med25 (PDB), where thickness and coloring correspond to fluctuations, for (top) the apo protein and (bottom) Med25+VP16(438-454)$_{G450C}$.

The stabilization seen in MD was tested through protein melting experiments via circular dichroism (CD). In doing so, we determined that there was an increase in protein melting temperature of $\Delta T_M = 0.9\,°C$ when VP16(438-454)$_{G450C}$ (*Figure 3.8*). The melting

experiments also highlighted the stabilization induced through activator binding in a different way; the transition of going from the folded to the unfolded state became much sharper when in complex with VP16 (*Figure 3.8B*). This could in part be due to the stabilization we see of the helices that occurs when VP16 binds on the H1 face.

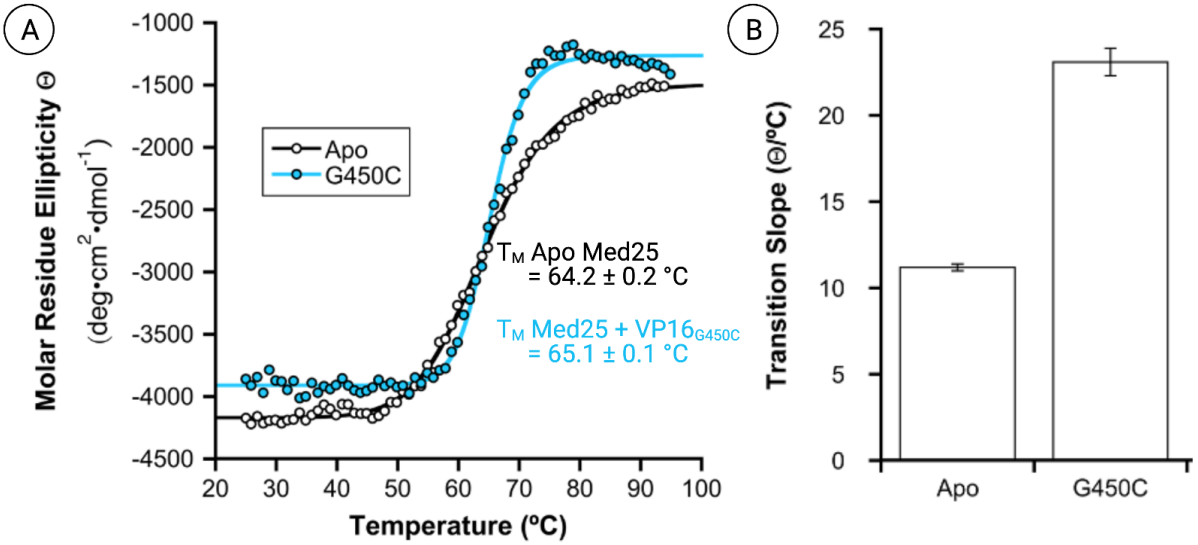


Figure 3.8. *Protein melting experiments for apo Med25 and Med25+VP16(438-454)$_{G450C}$.* (A) Thermal melting experiments for Apo Med25 (black; T_M = 64.2 ± 0.2 °C) and Med25+VP16$_{G450C}$ (light blue, T_M = 65.1 ± 0.1 °C) (B) The transition slope going from folded to unfolded is much sharper for the VP16(438-454)$_{G450C}$ tethered construct than for the apo protein. Experiments performed with Dr. Andrew Henderson and Dr. Matthew Henley.

In building a more complete model for AcID activator recognition, we transitioned to the larger VP16 construct (residues 438-490), which has been shown to interact with both faces on Med25.[7] Using the initial structure developed for the shorter construct, VP16(438-454)$_{G450C}$, the peptide was appended with the remaining sequence on the C-terminal (residues 438-490) and pointed the "unbound" helix outwards into space (*Figure 3.9*). Temperature replica exchange was again employed, where we were able to obtain a model of binding. Using K-means clustering with a 1.5 Å backbone RMSD cutoff, we find that there are 16 clusters that The two binding poses are differentiated by which side of the triple lysine loop they bind onto on the H2 face (*Figure 3.10*). The two binding modes show varying degrees of helical content in VP16, with Pose 1 showing VP16 comprised of almost exclusively helices and Pose 2 corresponding with more random

coils. Thus, these models suggest that multiple bound conformational states can exist, which aligns with the difficulties in obtaining solving for a bound structure experimentally.

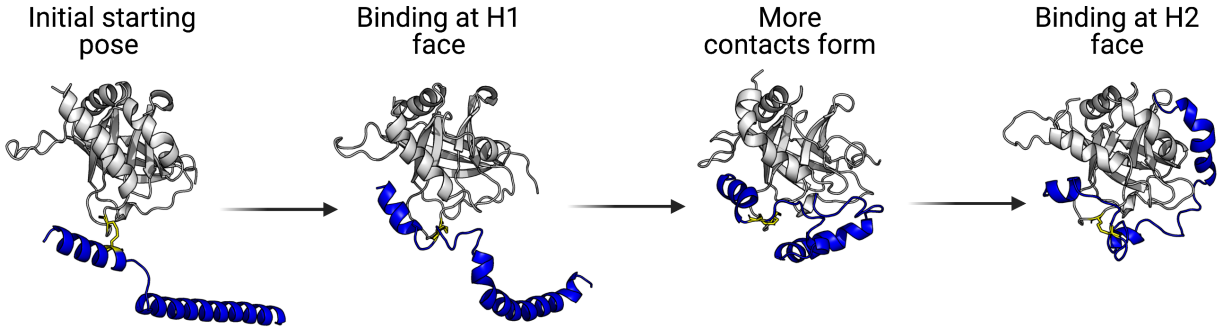


Figure 3.9. *Binding of the long VP16(438-490) Tethered construct using temperature replica exchange.* Snapshots from simulation demonstrating how VP16 (438-490)$_{G450C}$ (blue) binds to Med25 (gray) when tethered starting from (left) the initial construct made in CHARMM with the NMR coordinates of Med25 (PDB 2XNF) and a disulfide bond to the peptide.

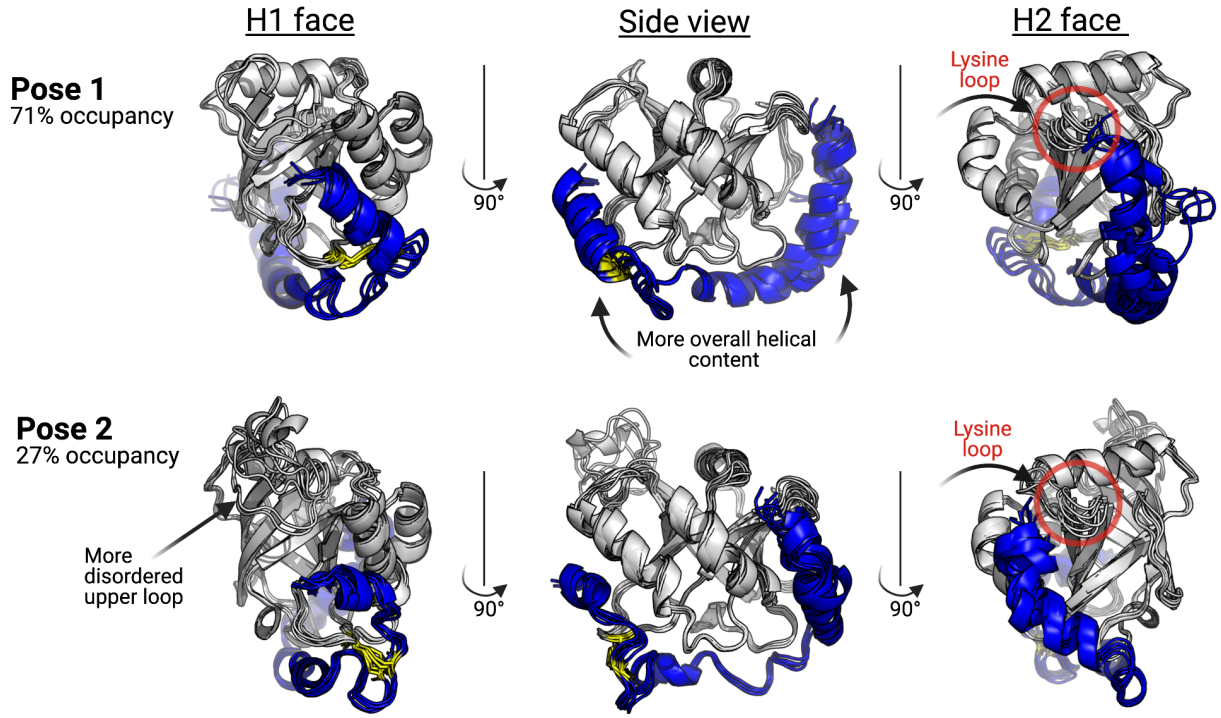


Figure 3.10. *Two binding modes for VP16(438-490)$_{G450C}$ Tethered to Med25$_{C506}$.* (Top) Pose 1 is the higher occupancy state, comprising of 71% of the total frames from the simulation, where the VP16 peptide maintains a higher degree of helical content. (Bottom) Pose 2 is the less occupied pose, comprising of 27% of the total states, where VP16 binds on the left side of the lysine loop on the H2 face.

We performed $^{1}H,^{15}N$ HSQC experiments on Med25+VP16(438-490), and the chemical shift perturbations (CSPs) align with structural models generated from our

modeling. Namely, we see most CSPs occur around the two side helices (*Figure 3.11*). Looking at the structure of generated through modeling of the bound complex, one large change that occurs is the bending the helix closest to the H1 face, α3 (Figure 3.12). This occurs not only in the longer VP16 construct but also in the shorter H1-binding TAD model.

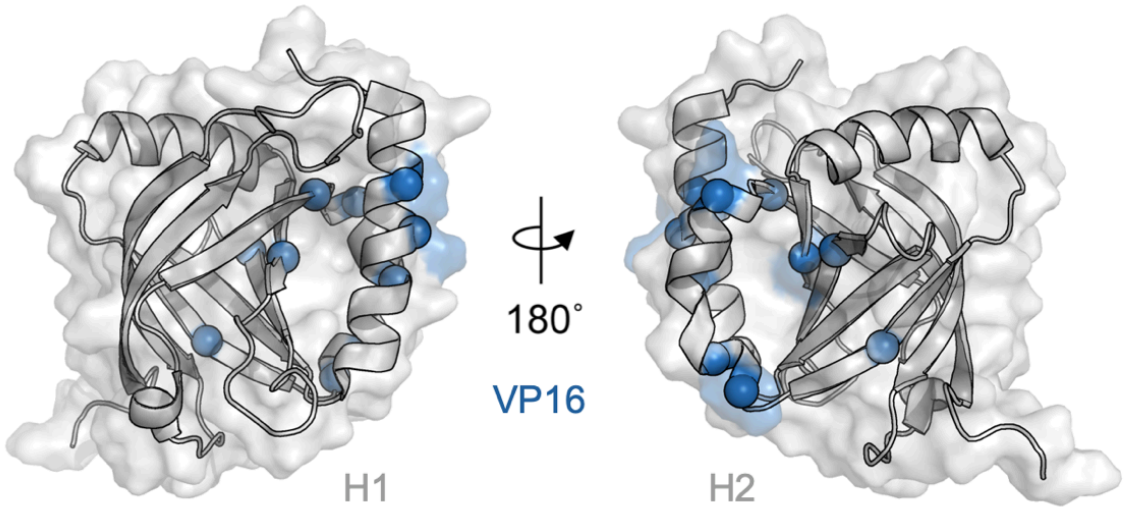


Figure 3.11. *Med25 chemical shift perturbations upon VP16 binding.* Results of VP16(438-490) chemical shift perturbation experiments superimposed upon the Med25 AcID NMR structure (PDB 2XNF). Residues displaying chemical shift perturbation greater than 2 SD upon VP16 binding are depicted in blue spheres. Experiments were performed by Dr. Andrew Henderson and Dr. Brian Linhares.

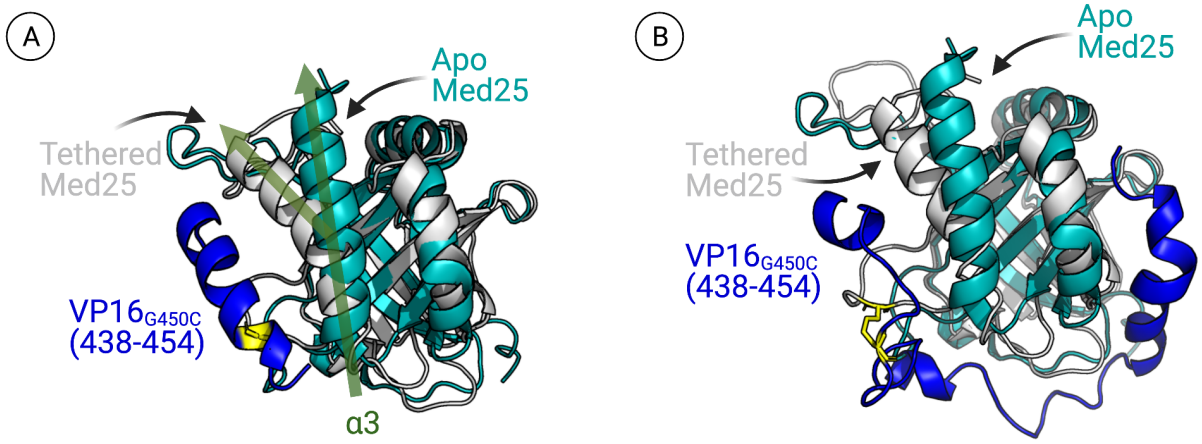


Figure 3.12. *Helix bending in Med25 caused by VP16 binding.* Centroid structures of Med25 tethered complexes for (A) VP16(438-454)$_{G450C}$ (blue) and (B) VP16(438-490)$_{G450C}$. Med25 structures from complexes formation are shown in light gray, with the disulfide bond used for Tethering shown in yellow.

The NMR structure of apo Med25 AcID (PDB 2XNF) is shown in teal. As seen with the green arrows, helix α3 bends towards the H1 face when VP16 binds.

The fact that the majority of CSPs upon VP16 binding occur *between* the H1 and H2 face (i.e. on the helices separating the two faces) rather than directly *on* the two binding surfaces suggests that binding induces conformational changes, as seen in the models generated here. Indeed, looking closely at the highest CSPs, we see that the sum effect of CSPs report on such conformational changes. The movement in the helix that neighbors the H1 face (α3) as seen in *Figure 3.12* corresponds to large CSPs on that helix. Starting with G456 (light pink spheres), we can see how subsequent chemical shifts include Q539 (dark green spheres) and T542 (light orange spheres), which are pulled towards the H1 face with binding (*Figure 3.13A*). Similar effects are seen with the neighboring helix on the H2 face (α1). VP16 binding on the H2 face causes pulling from the bottom of Med25's α1. R466 (cyan spheres) show large movement towards the H2 face, corresponding to a large CSPs in NMR (*Figure 3.13*). L464 (green spheres) on helix α1 also shows movement and high CSP, as well as Q456 (yellow spheres) on the top of that helix.

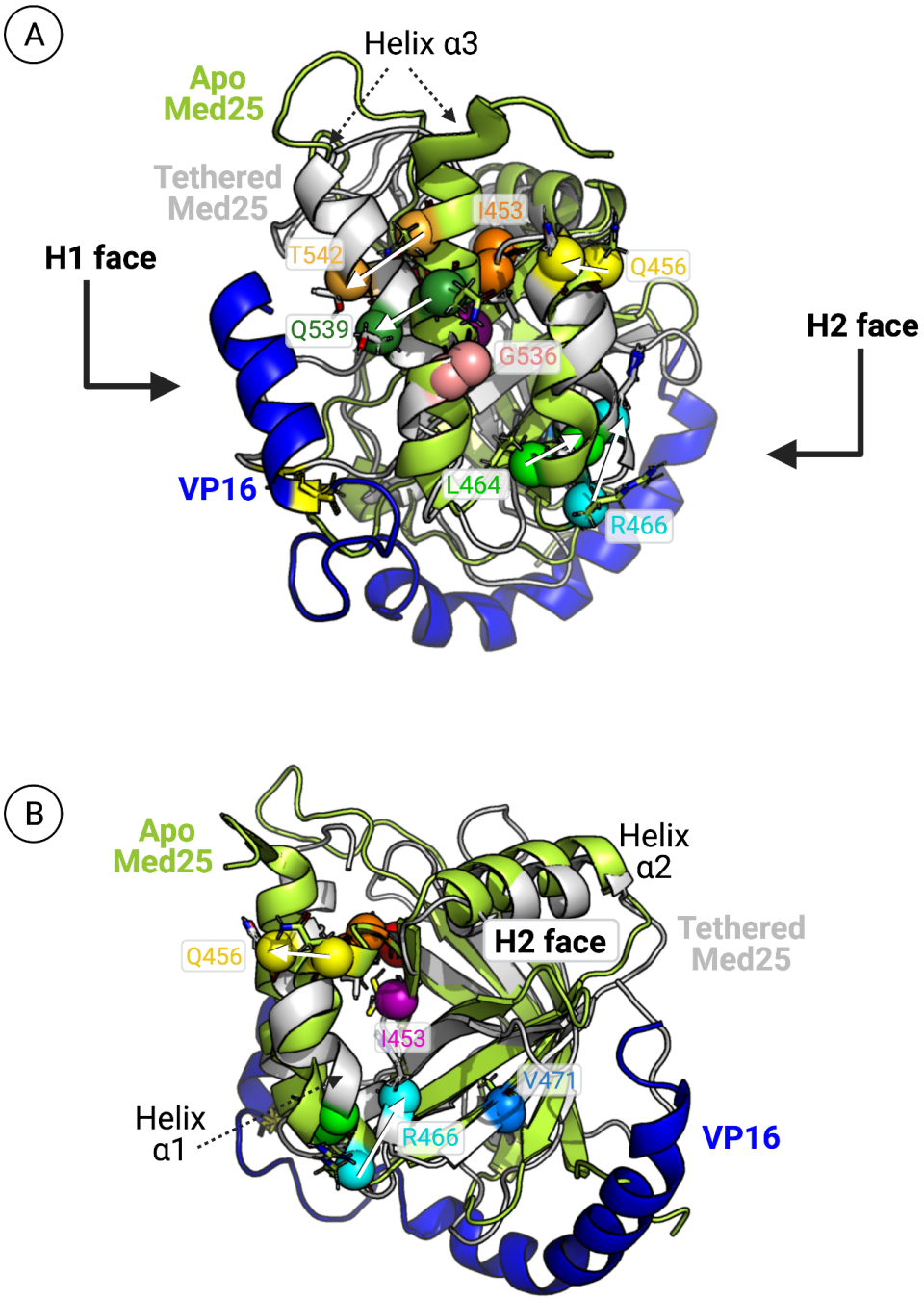


Figure 3.13. *Chemical shifts report on conformational changes with VP16 binding.* Results of VP16(438-490) chemical shift perturbation experiments superimposed upon the Med25 AcID NMR structure (lime green, PDB 2XNF) for residues displaying CSPs > 2 SD upon VP16 binding (shown as spheres). Superposed is the model for VP16(438-490)$_{G450C}$ (blue) Tethered to Med25$_{C506}$ (light gray). (A) The side view for large CSPs, and (B) CSPs from the H2 face, where residues with high CSPs are shown as colored spheres. NMR experiments were performed by Dr. Andrew Henderson and Dr. Brian Linhares.

Med25 AcID shows unique structural accommodation for activators

With success in generating structural models of VP16 Tethered to AcID through enhanced sampling methodologies, temperature replica exchange was utilized once more to dissect key interactions between ERM and Med25. However, this system lacked *in vitro* Tethering data. Nonetheless, a Tether was created using CHARMM by appending the C-terminal portion of the TAD with repeating glycine-serine pairs and a C-terminal cysteine, named ERM(38-68)$_{+GSGSGSGC}$ hereon (*Figure 3.14A*). Using K-means clustering (1.5 Å backbone RMSD cutoff), we find an ensemble of bound poses that all generally show the ERM peptide wrapping across the beta barrel (*Figure 3.14B*). In addition, all clusters show the top loop on the H1 face in a single down conformation.

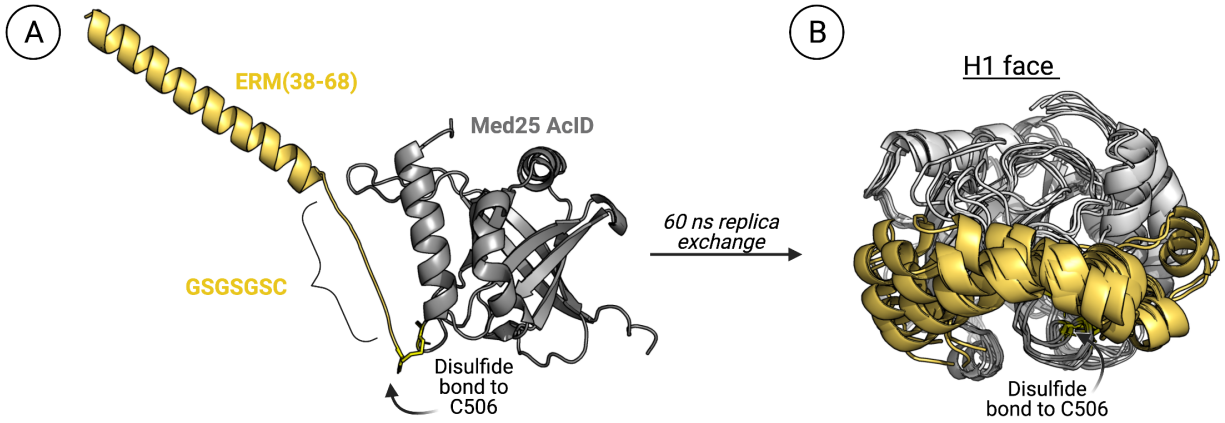


Figure 3.14. *Molecular dynamics simulations with a Tethered ERM construct.* (A) ERM(38-68) (gold) was appended with a linker (GSGSGSGC) and Tethered to Med25$_{C506}$ (PDB 2XNF, shown in gray) for simulations with temperature replica exchange. (B) Ensemble of ERM structures bound to the H1 face of Med25 after 60 ns of simulation. Structures shown are the centroids generated using the K-means clustering algorithm with a 1.5 Å backbone RMSD cutoff.

Much like with VP16, the CSPs from $^{1}H,^{15}N$ HSQC experiments with ERM(38-68) support the structural model of binding we see from modeling, where perturbations occur across the beta barrel (*Figure 3.15*). This, along with data corresponding to VP16, suggests that the models generated from replica exchange can be used to generate structures of Tethered binding interactions. Further, the distinct but overlapping chemical shift patterns observed upon binding of each of the activators to Med25 suggest several unique binding modes accommodated within AcID.

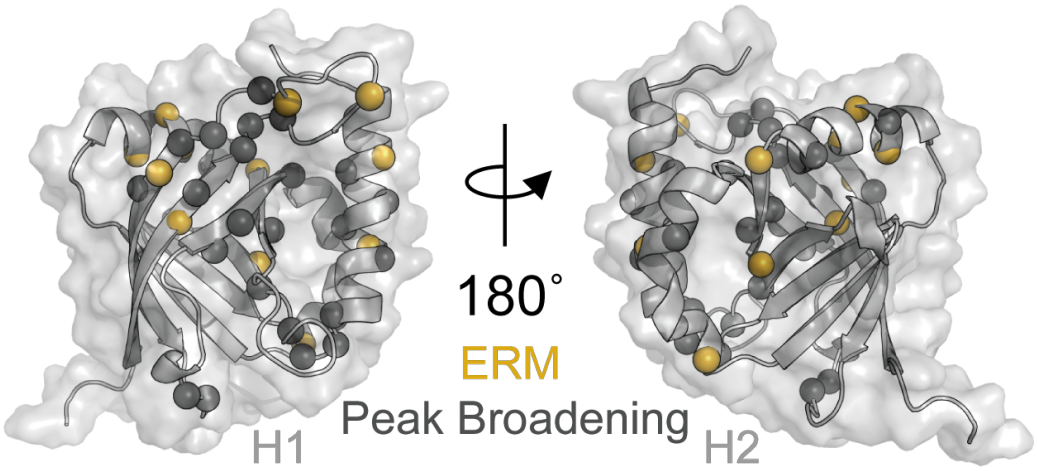


Figure 3.15. *Med25 chemical shift perturbations upon ERM binding.* Results of chemical shift perturbation experiments superimposed upon the Med25 AcID structure (PDB 2XNF). Residues displaying chemical shift perturbation greater than 2 SD upon ERM binding are depicted in maize spheres. Grey spheres indicate residues with chemical shifts that broaden upon ERM binding. Experiments were performed by Dr. Andrew Henderson and Dr. Brian Linhares.

The modeling aligns with the NMR experiments with ERM by showing the general direction of the activator engagement across the beta barrel (*Figures 3.14 and 3.15*). In addition, there are many residues in the upper loop on the H1 face that show either large CSPs or are in intermediate exchange. Looking at the generated model for the Tethered construct, we see that the upper loop participates in charged interactions with ERM (*Figure 3.16*). Thus, ERM uses a different mechanism of activator engagement than VP16, which did not show direct interactions with that upper loop. Further, as clearly shown, electrostatics play an important role in activator engagement.

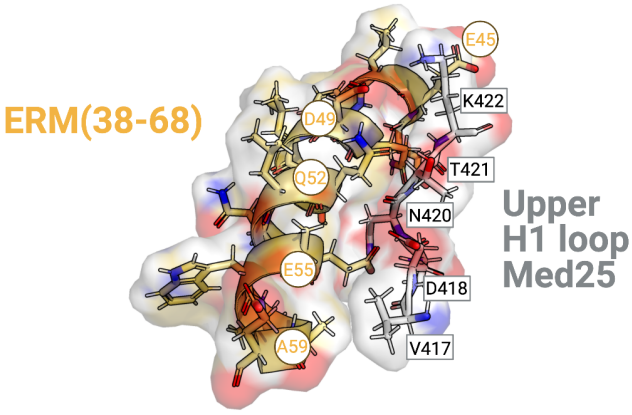


Figure 3.16. *Interactions between ERM(38-68) and the upper H1 loop*. Residues in ERM(38-68) (gold, with residues labeled in circles) show many charged interactions with the upper H1 loop of Med25 (gray, labels in rectangles).

The two bound complexes from modeling with VP16 and ERM demonstrate the inherent malleability in the seemingly rigid Med25 AcID domain; while both activator bind on the H1 face, they bind in and form two very unique binary complexes (*Figure 3.17*). This finding is supported by the amide proton perturbation patterns measured for the activator–AcID complexes via NMR, suggesting unique binding modes for each activator. This is analogous to helical activator binding domains such as GACKIX of CBP/p300, a three-helix bundle that contains at least two activator binding sites.[19]

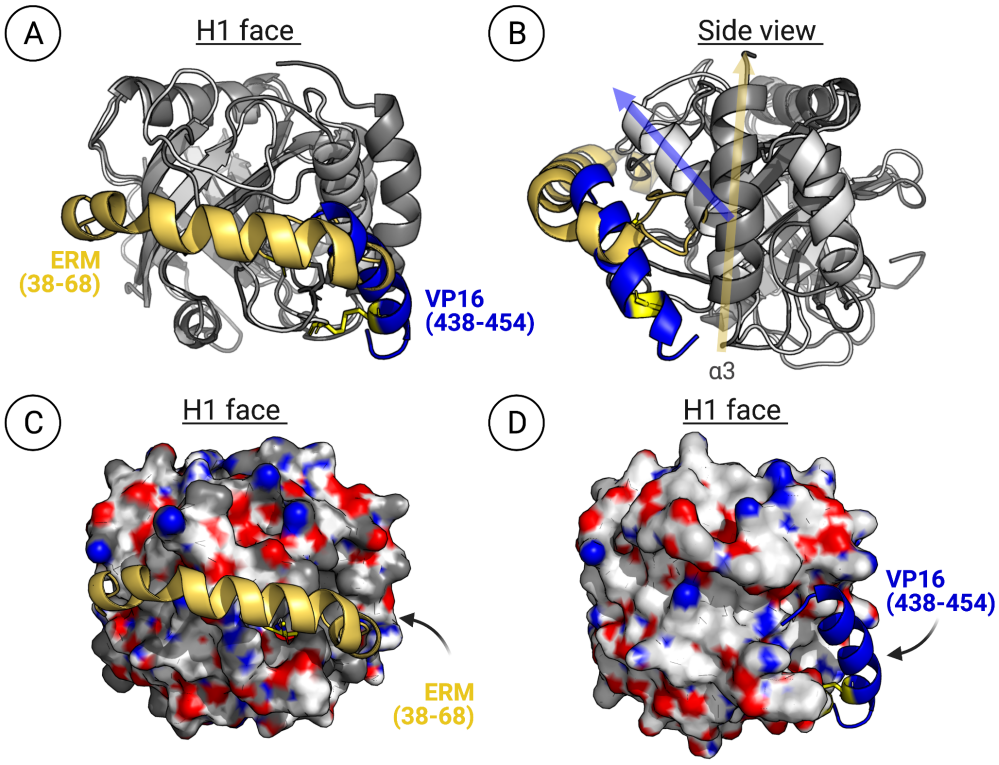


Figure 3.17. *H1-binding TADs form unique complexes with Med25 AcID.* (A) Centroid structures from simulation of Med25+VP16(438-454)$_{G450C}$ (light gray Med25, blue peptide) and Med25+ERM(38-68)$_{+GSGSGSGC}$ (dark gray Med25, gold peptide). The disulfide bond for both Tethered complexes are shown in yellow, and orientation of Med25 shows the H1 face. (B) The side view of the bound complexes for the two activators. The differences in helix α3 structure are highlighted by color corresponding to the activator that is bound. (C&D) The surface representation of Med25 on the H1 face with (C) ERM bound or (D) VP16 bound.

While the structures induced by activator binding is very different, both cause significant decreases Med25 dynamics (*Figure 3.18*). Much like with KIX, activator binding stabilizes the protein, with the most dynamic regions, i.e. upper loop on H1 face, particularly residues 415-421. Additionally, the long flexible region under the β-barrel (residues 436-444) shows a large reduction in movement. Thus, again similar to KIX, most dynamic region of the protein is able to tune the overall rigidity of the protein.

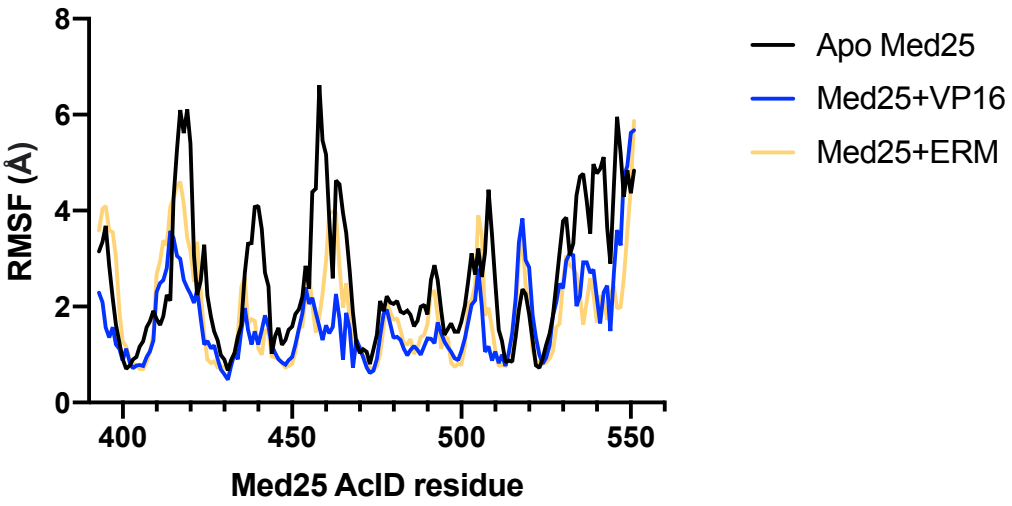
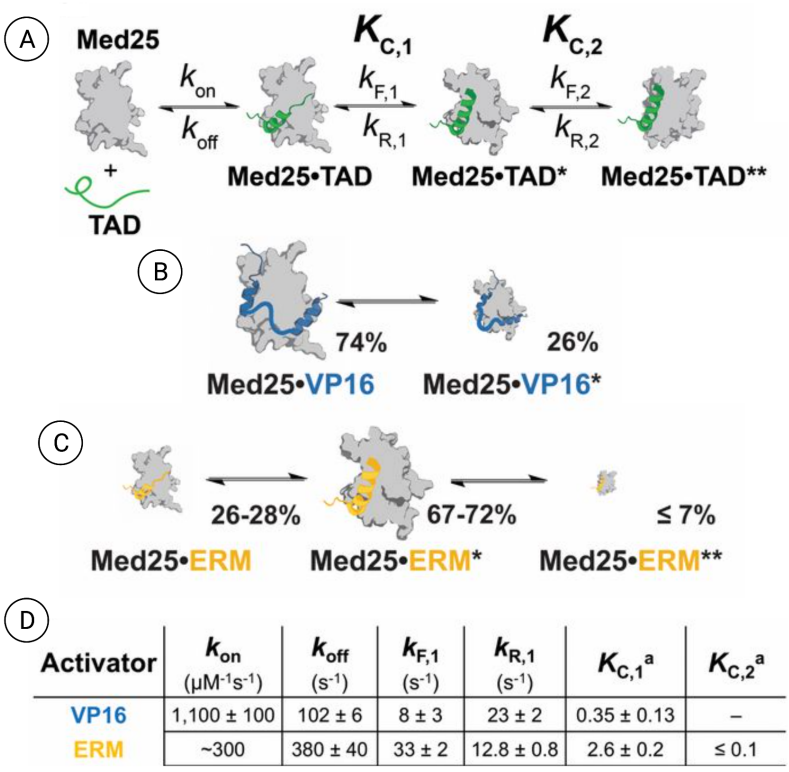


Figure 3.18. *Root mean square fluctuations (RMSF) by Med25 AcID residue*. RMSF by residue is shown for the apo protein (black), Med25+VP16(438-454)$_{G450C}$ (blue), and Med25+ERM(38-68$_{+GSGSGSGC}$) (gold). Residues that show the greatest stabilization include 415-421, which is the majority of the upper loop on the H1 face of Med25.

Kinetics data supports that activator–Med25 complexes are conformationally dynamic

The underlying mechanistic features of activator–AcID complex formation were examined by determining association mechanisms of AcID with the TADs of VP16 and ERM using stopped-flow fluorescence spectroscopy. These kinetic experiments allow calculation of microscopic rate constants for association and dissociation, as well as forward and reverse rate constants for any conformational changes that happen during the binding process[20,21] (*Figure 3.19*). Previous kinetic studies of helical coactivators revealed that complex formation with activators proceeds by fast association (k_{on}) and dissociation (k_{off}) rate constants.[21–26] Consistent with this observation, we found that activator–AcID complexes form with elevated k_{on} and k_{off} values, with k_{on} ranging between 300 and 1,100 $\mu M^{-1} \cdot s^{-1}$and k_{off} ranging between 100 and 400 s^{-1} (*Figure 3.19*). This behavior allows for activators to form tight interactions (K_D values 50–500 nM) with Med25 AcID that are short-lived, with activator residence times less than 10 ms. Additionally, at least one conformational change during the binding process was observed for both activators, with similar observed rate constants (k_{obs} = 10–40 s^{-1}) for each activator (*Figure 3.19*). These findings align with the multiple bound complexes determined from MD.



D

Activator	k_{on} ($\mu M^{-1}s^{-1}$)	k_{off} (s^{-1})	$k_{F,1}$ (s^{-1})	$k_{R,1}$ (s^{-1})	$K_{C,1}$[a]	$K_{C,2}$[a]
VP16	1,100 ± 100	102 ± 6	8 ± 3	23 ± 2	0.35 ± 0.13	–
ERM	~300	380 ± 40	33 ± 2	12.8 ± 0.8	2.6 ± 0.2	≤ 0.1

Figure 3.19. *Transient kinetic experiments define minimal mechanism of activator-AcID complexation.* (A) General kinetic mechanism for TAD-AcID complex formation as determined by these experiments for all activators. Microscopic equilibrium constants ($K_{C,n}$) are defined as the ratio of respective forward and reverse rate constants. (B) Representation of equilibrium population distributions of bound states, calculated from equilibrium constants, for VP16(438-490) and (C) ERM(38-68). Sizing of each complex is scaled according to the indicated percentage population. When one equilibrium constant is too small to measure, the values are given as ranges. (D) Measured kinetic and equilibrium constants for both activators. Kinetic constants $k_{F,2}$ and $k_{R,2}$ are unable to be reliably calculated. a, The conformational change equilibrium constant is too small to be measured with precision. Experiments performed by Dr. Matthew Henley.

A covalent small molecule inhibits ortho- and allosterically by stabilizing Med25 AcID

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). As previously reported, this VP16 sequence contains the minimal binding sequence for interaction with AcID (K_D = 0.60 ± 0.06 μM) and interacts with the H1 and H2 binding surfaces. 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 (Z' = 0.87; 1.6% hit rate). Compounds with activity > 3 S.D. 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 lichen-derived natural products norstictic acid (NA) and psoromic acid (PA) emerged as the best inhibitors, with apparent IC_{50} values of 2.3 ± 0.1 μM and 3.9 ± 0.3 μM (*Figure 3.20*).

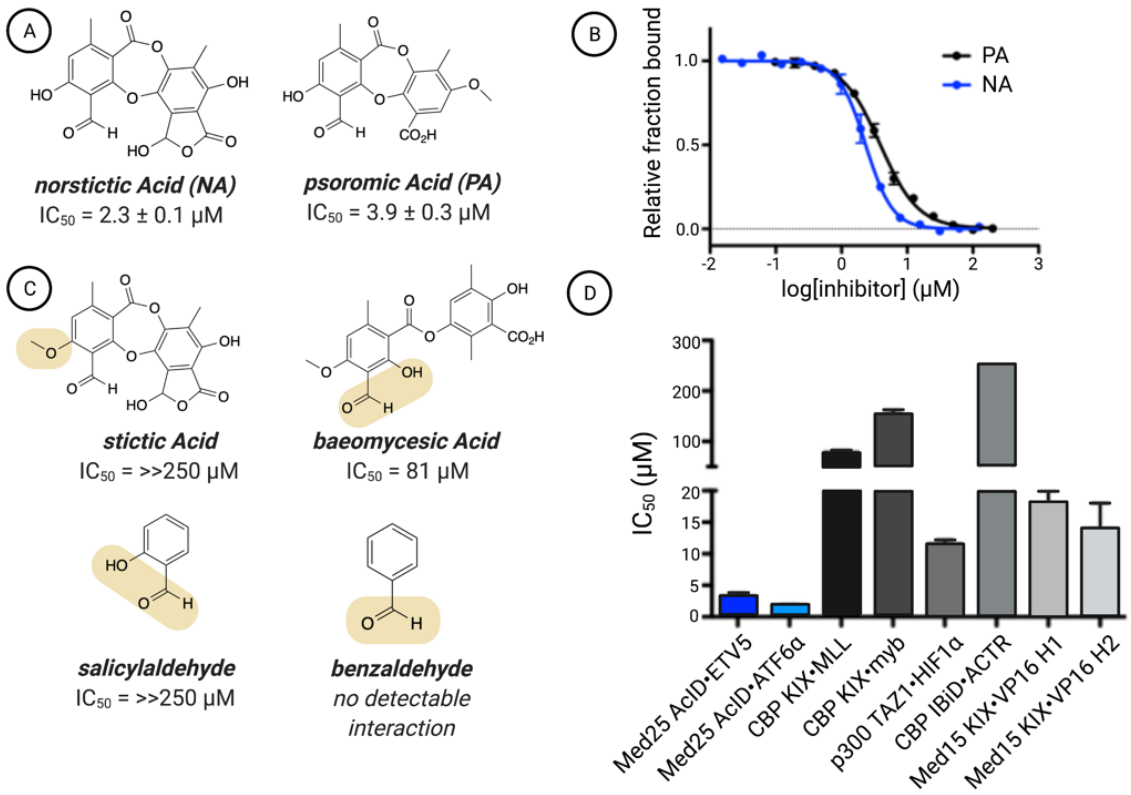


Figure 3.20. *Norstictic acid and psoromic acid inhibit Med25 AcID.* (A) Chemical structures of the top two hits emerging from the screen of the Med25 AcID•fl-VP16(465-490) along with (B) their apparent IC_{50} values. The apparent IC_{50} values were determined through titrations of either NA or PA against Med25 AcID•fl-VP16(465-490) performed in triplicate with the indicated error (SDOM). (C) Assessment of related structures shows that the orthophenoxyaldehyde moiety is important but not sufficient for inhibitory activity. IC_{50} values were determined via competition fluorescence polarization against Med25•VP16(465-490). (D) Inhibition of related PPI networks by NA. Apparent IC_{50} values were measured via fluorescence polarization against a suite of coactivator domains (CBP KIX, p300 TAZ1, CBP IBiD, Med15 KIX) bound to fluorescein-tagged activators. The values are the average of three independent experiments with the indicated error (SDOM). No error bars are shown for the IC50 against IBiD•ACTR because the IC_{50} was greater than the highest concentration of NA tested, 250 μM, and thus we can only accurately report the IC50 as > 250 μM. Experiments performed by Dr. Steven Sturlis and Dr. Julie Garlick.

Both NA and PA are natural products in the depsidone family containing an orthophenolic aldehyde moiety (*Figure 3.20A*). The presence of a reactive aldehyde functionality suggested a potential covalent mechanism of action, for example 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). Treatment with the reducing agent $NaBH_4$ led to incorporation of H_2 into the adduct, indicating initial formation of a Schiff base followed by reduction (*Figure 3.21*). Data from a time-course experiment revealed that at 5 minutes, significant inhibition is observed, with full activity after 30 minutes. An examination of related structures indicates that the orthophenolic aldehyde is necessary, but not sufficient for interaction with Med25 AcID or for inhibitory activity. Stictic acid, in which the phenol is masked as a methyl ether, inhibits Med25 interactions poorly (IC_{50} > 250 μM). Additionally, salicylaldehyde efficiently labels Med25 AcID, but does not impact binding of activators. These data suggest that noncovalent interactions play essential roles in the inhibitor function of NA. Consistent with this, NA exhibits remarkable selectivity for Med25 PPIs relative to other coactivators with similar binding surfaces (*Figure 3.20D*). Notably, NA inhibits Med25 PPIs at both binding surfaces, including those formed with transcriptional activators ETV5 (H1 binding surface) and ATF6α (H2 binding surface).

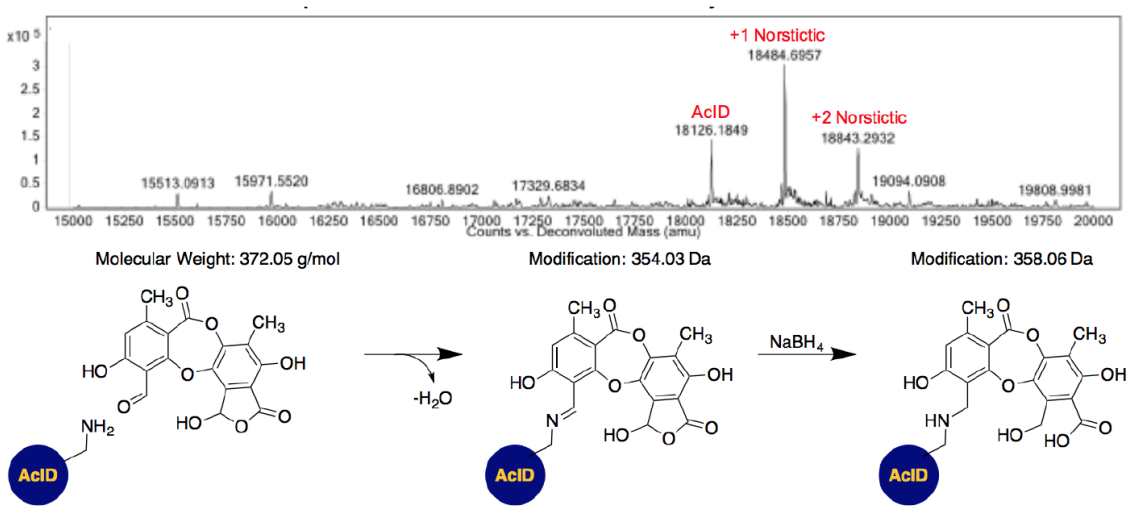
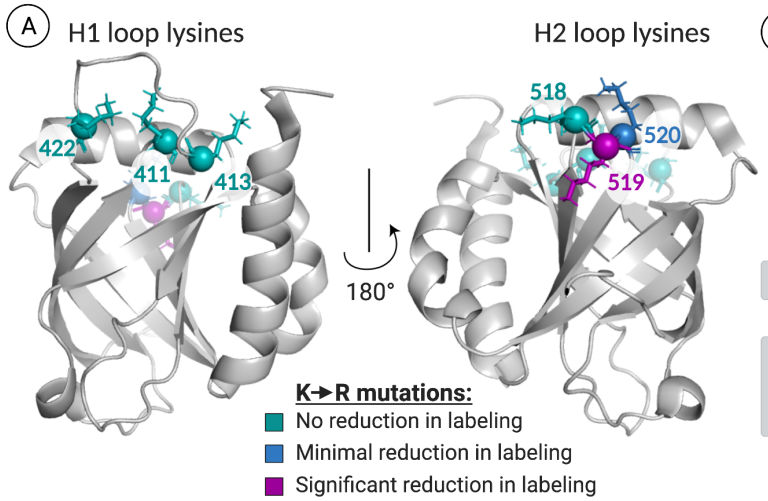


Figure 3.21. *Mass spectrometry of Med25 AcID with norstictic acid.* Addition of 4 equivalents $NaBH_4$ after incubation with NA for 30 minutes leads to formation of +358 adduct (top), indicating covalent modification through initial imine formation (bottom). Experiments performed by Dr. Steven Sturlis and Dr. Julie Garlick.

Several lines of evidence suggested that the engagement site of NA is a lysine-rich dynamic loop that borders the H2 binding surface (*Figure 3.22A*). There are 11 lysine residues within Med25 AcID, 6 of which are found on dynamic loop regions flanking the two known activator binding surfaces. Replacement of these lysines with arginine either alone or in combination had minimal effects on binding of the cognate transcriptional

activator binding partners. 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 (*Figure 3.22*). In contrast, mutation of K519 had a profound effect on NA binding and inhibition. This residue is part of a lysine-rich dynamic loop that flanks the H2 face and the mutational data indicates that NA can also interact with K520 and K518 within this loop.



B

	IC_{50} (μM)	Fold change
WT	2.7 ± 0.4	-
K411R, K413R	6 ± 3	2.2
K420R	2 ± 1	0.74
K518R	2.3 ± 0.4	0.9
K519R	23 ± 1	8.5
K520R	3.6 ± 0.6	1.3
K518R, K519R	39 ± 2	14.4
K518R, K519R, K520R	44 ± 2	16.3

Figure 3.22. *Med25 AcID lysine reactivity with norstictic acid.* (A) LC-MS analysis of norstictic acid covalent adduct formation with Med25 lysine-to-arginine mutants indicates that K519R leads to the most significant reduction of labelling. No reduction of labelling corresponds to a decrease in abundance of the NA covalent adduct of less than 10%. Minimal reduction in labelling, observed for K520R, corresponds to a 22% reduction in the mass abundance of the NA covalent adduct. Significant reduction in labelling, observed for K519R, corresponds to a 53% reduction in the mass abundance of the +1 covalent adduct. PDB 2XNF used to generate figure. (B) Inhibition of Med25 AcID•ETV5 interaction by norstictic acid measured using fluorescence polarization. Mutants containing K519R, highlighted in grey, demonstrate the most significant increase in apparent IC_{50}. Values represent the average of three independent experiments with the indicated error (SDOM). Experiments performed by Dr. Steven Sturlis and Dr. Julie Garlick.

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. Looking at the RMSF by Med25 residue, minimal restructuring in the lysine loop adjacent to the H2 binding interface is observed (*Figure 3.23A*). However, helix α1 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 (*Figure 3.23C*). Taken together, the data indicate that NA serves as both an orthosteric

inhibitor of H2-binding transcription factors (e.g. ATF6α) and an allosteric inhibitor of H1 binding transcriptional activators (e.g. ETV5).

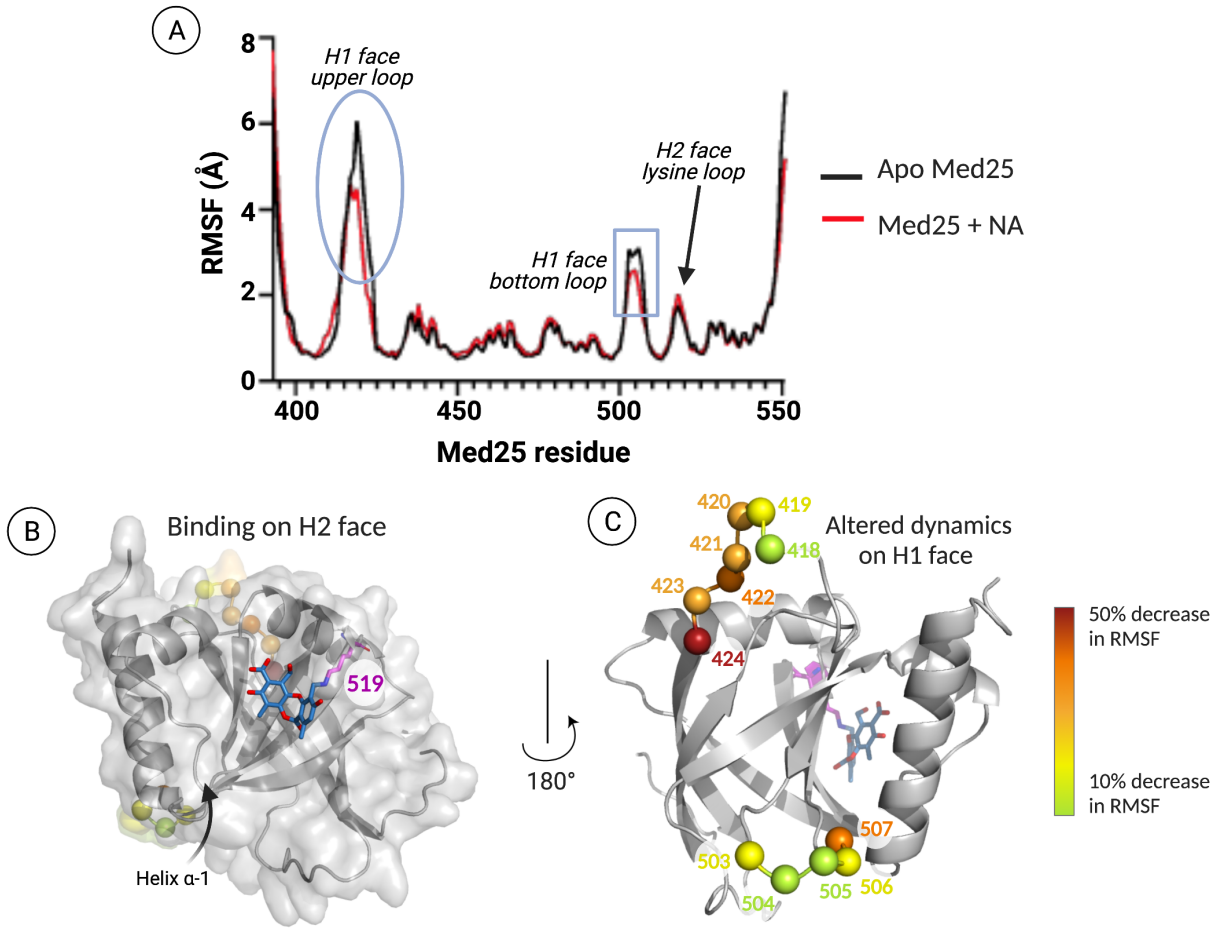


Figure 3.23. *Structural model and fluctuations of Med25 with norstictic acid.* (A) Root mean square fluctuations (RMSF) across Med25 residues. (B) Centroid structure of the most populated cluster from molecular dynamics simulations, where norstictic acid binds to the H2 face of Med25 and covalently links to K519. (C) The residues that showed the greatest reduction in fluctuations (RMSF) upon activator binding all occur on dynamic substructures on the H1 face.

3.3 Conclusion

ABDs are molecular recognition marvels, with a single ABD typically being the cognate binding partner for tens of different activators. Hence, ABDs require a significant degree of structural mobility to accommodate this diversity. This mobility likely corresponds to local folding-like transitions; it not only allows the binding interfaces to morph into unique conformations as part of binding different activators, but underlies the allosteric interactions between different binding sites in an individual domain.[27,28] Therefore, the "important" molecular recognition elements should be the most mobile

regions, which is in line with our results shown here with Med25 AcID. Despite the large surface area of the core β-barrel that is used for interacting with activators, it is changes in the flanking loops and helices that enable accommodation of the distinct cognate ligands. Further, since the first structural reports of AcID, the identity of activator and coactivator binding partners of Med25 has expanded,[29] and the molecular recognition model outlined here indicates that cooperative binding of Med25 to activators and/or coactivators such as CBP may be a key regulatory mechanism.

Taken together, the combined modeling and *in vitro* results detailed here point to a mode of interaction where Med25 AcID forms conformationally unique complexes with its activator interaction partners. While the activators bind to Med25 with similar affinities, they have different kinetic signatures characterized by a conformational change after binding that extends throughout the AcID structure. This is found not only in modeling, but also in CSPs measured from NMR. Finally, all of these complexes can be described as “fuzzy”; at equilibrium, each complex contains multiple populated conformational substates that are separated by low energy barriers.[30] Hence, we establish that while Med25 is structurally unique in its β-barrel core, it shares mechanisms of activator engagement with other more flexible domains like KIX — Both proteins utilize their most dynamic regions to individually accommodate multiple partners, thus again establishing the utility of disorder processes like transcription.

Beyond interrogating native interactions, we were able to identify an inhibitor for Med25, namely norstictic acid. We demonstrated how this small molecule covalently binds to a lysine loop on the H2 face, and yet it can inhibit not only orthosterically, but also allosterically on the H1 face. Using molecular dynamics simulations, we show that allosteric inhibition is achieved through changes in H1 face fluctuations. Thus, much like with KIX, inhibition of dynamic coactivators can be achieved through changes in the conformational landscape towards micro-states that disfavor binding. This consistency between Med25 and KIX — two coactivators that are structurally distinct from one another — builds on our understanding of this class of proteins (i.e. coactivators) and highlights how druggability can be achieved through targeting dynamic substructures on a protein to alter overall dynamics.

3.4 Materials and methods

Modeling experimental

The objectives of our modeling efforts were to predict an ensemble of putative structures for binding of the two activators, VP16 and ERM, tethered to the Med25 AcID domain C506 via a disulfide Tether. Our modeling was initiated from the published NMR coordinates for Med25 AcID (PDB 2XNF).[7] The protein structure of Med25 AcID was prepared for simulation in CHARMM using the Multiscale Modeling Tools for Structural Biology (MMTSB).[31–33] Activators have been shown to typically form a helical conformational states when in complex with coactivators, so both peptides were constructed in CHARMM as a contiguous helix, which were then bonded using the DISU patch in CHARMM to $Med25_{C506}$ through the formation of a disulfide bond. Prior to running the implicit solvent simulations, Med25 was fixed using harmonic restraints, and the complex was minimized with 1000 steps of a steepest descent algorithm. Using GBSW implicit solvent,[34,35] temperature replica exchange[17] was implemented using the CHARMM22 force field. These simulations were run for a total of 60 ns (2 fs time steps) using 12 replicas, sampling between 280-500 K and attempting coordinate exchanges every 5000 steps. The 12 replica trajectories were sorted by their respective temperatures, and the last 40 ns of the 280 K trajectories were then parsed into 4000 coordinate files. The MMTSB tool cluster.pl was used to cluster these structures, using K-means clustering[36] with 1.5 Å RMSD cutoff for the superposed C_α backbone atoms for all of the structures. The root-mean-square fluctuations (RMSF) for the last 40 ns of the 280 K trajectory were calculated for each Med25 AcID residue by superposing $C\alpha$ atoms of the coordinate files produced from the simulations.

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 into 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 2 ns 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 were performed using GPUs through the CHARMM compatible OpenMM interface. Five independent trials of simulations were performed for each molecule.

Peptide synthesis and preparation

The peptides listed in Table 3.1 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.

Table 3.1. Sequences of peptides used in experiments.

Entry	Peptide	Sequence
1	VP16 (438-490)	AcALDDFDLDMLGDGDSPGPGFTPHDSAPYGALDMADFEFEQMFTDALGIDEYGG
2	ERM (38-68)	AcDLAHDSEELFQDLSQLQEAWLAEAQVPDDEQ
3	4-DMN-VP16 (438-490)	4-DMN-βAla-DALDDFDLDMLGDGDSPGPGFTPHDSAPYGALDMADFEFEQMFTDALGIDEYGG
4	4-DMN-ERM (38-68)	4-DMN-βAla-DLAHDSEELFQDLSQLQEAWLAEAQVPDDEQ

5	VP16 (438-454) D441C*	AcY-βAla-ALDCFDLDMLGDGDSPG
6	VP16 (438-454) F442C*	AcY-βAla-ALDDCDLDMLGDGDSPG
7	VP16 (438-454) D443C*	AcY-βAla-ALDDFCLDMLGDGDSPG
8	VP16 (438-454) L444C*	AcY-βAla-ALDDFDCDMLGDGDSPG
9	VP16 (438-454) D445C*	AcY-βAla-ALDDFDLCMLGDGDSPG
10	VP16 (438-454) M446C*	AcY-βAla-ALDDFDLDCLGDGDSPG
11	VP16 (438-454) L447C*	AcY-βAla-ALDDFDLDMCGDGDSPG
12	VP16 (438-454) G448C*	AcY-βAla-ALDDFDLDMLCDGDSPG
13	VP16 (438-454) D449C*	AcY-βAla-ALDDFDLDMLGCGDSPG
14	VP16 (438-454) G450C*	AcY-βAla-ALDDFDLDMLGDCDSPG
15	VP16 (438-454) D451C*	AcY-βAla-ALDDFDLDMLGDGCSPG
16	VP16 (438-454) S452C*	AcY-βAla-ALDDFDLDMLGDGDCPG
17	VP16 (438-	AcY-βAla-ALDDFDLDMLGDGDSCG

	454) P453C*	
18	VP16 (438-454) G454C*	AcY-βAla-ALDDFDLDMLGDGDSPC
19	VP16 (465-490)	FITC--βAla-YGALDMADFEFEQMFTDALGIDEYGG
20	ETV5 (38-68)	FITC-βA-ALDMADFEFEQMFTDALG
21	ATF6a (40-66)	FITC-βA-DTDELQLEAANETYENNFDNLDFDLDLM
22	MLL (840-858)	FITC-βA- DCGNILPSDIMDFVLKNTP
23	Myb (291-316)	FITC-βA-KEKRIKELELLLMSTENELKGQQVLP
24	IBiD (2063-2111)	Ac-SPSALQDLLRTLKSPSSPQQQQQVLNILKSNPQLMAAFIKQRTAKYVAN
25	ACTR (1041-1088)	FITC-βA-PSNLEGQSDERALLDQLHTLLSNTDATGLEEIDRALGIPELVNQGQAL
26	HIF1a (786-826)	FITC-βA-SMDESGLPQLTSYDCEVNAPIQGSRNLLQGEELLRALDQVN

*The cysteine thiol has been converted to a mixed disulfide with 2-aminoethanethiol.

At the conclusion of the syntheses, peptides listed in entry numbers 1, 2, and 5-18 underwent a final FMOC deprotection and were treated with acetic anhydride (Fisher Scientific) in the presence of triethylamine (Fisher Scientific) to acetylate the amino terminus. At the conclusion of synthesis for entries 3 and 4, an additional coupling was carried out with 4-*N*,*N*-dimethylamino- 1,8-napththalimide (4-DMN) linked to b-alanine prior to cleavage. Peptides were cleaved from resin with 95% trifluoroacetic acid (Sigma Aldrich), 2.5% H_2O, 2.5% triisopropylsilane (Sigma Aldrich) and filtered. The resulting solution was concentrated, and the peptide then precipitated by addition of cold diethyl ether. After filtration the solid was dissolved in 3:1 100 mM ammonium acetate (pH 7.0, Fisher Scientific) and acetonitrile with minimal ammonium hydroxide to fully solubilize the peptide. Purification was accomplished by reversed- phase HPLC on an Agilent 1260 Series instrument with a 250 x 10 mm Luna Omega 5 µm PS C_{18} column (Phenomenex)

using a gradient elution of acetonitrile in 100 mM ammonium acetate (pH 7.0). Gradient conditions are included in the legends of the analytical traces for each peptide found at the end of this document. Final purity was assessed by analytical HPLC and peptide identity was confirmed via mass spectrometry under negative mode ionization conditions (Agilent Q- TOF). Purified peptides were reconstituted in a minimal volume of DMSO and quantified by UV/Vis spectroscopy using either tyrosine absorbance at 280 nm (ε_{280}=1280 M^{-1} cm^{-1}) or 4-DMN absorbance at 450 nm (ε_{450}=10,800 M^{-1} cm^{-1}). The 4-DMN extinction coefficient was measured by serial dilution of a known concentration of 4-DMN-β-Ala-OtBu in identical conditions as the measurement of peptide concentrations.

Preparation of disulfide-capped peptides for Tethering experiments

The VP16-derived cysteine mutants (Table 3.1, Entries 5-18) were cleaved from resin and following ether precipitation were dissolved in minimal 1:1 DMSO:H_2O solution followed by the addition of 10 eq of cystamine•HCl (Fisher Scientific), 30 eq of diisopropylethylamine, and 1 eq of cysteamine (Fisher Scientific). The resulting solution was incubated at room temperature for 18 h on a rotating carousel. Solvent was removed by lyophilization and peptides were resuspended in 3:1 100 mM ammonium acetate (pH 7.0):acetonitrile for HPLC purification. Purified peptides were dissolved in a minimal volume of DMSO, quantified by UV/Vis spectroscopy using tyrosine absorbance at 280 nm (ε_{280}=1280 M^{-1} cm^{-1}), and stored at -20 °C.

Plasmids for protein expression

The Med25 expression plasmid pET21b-Med25(394-543)-His_6 was generously provided by Prof. Patrick Cramer.[7] Variants of pET21b-Med25(394-543)-His_6 were prepared using site directed mutagenesis, as previously described, with the primers indicated in Table 3.2.[37] Plasmid sequence identity was confirmed via standard Sanger sequencing methods on an Applied Biosystems 3730xl DNA Analyzer at the University of Michigan DNA Sequencing Core and analyzed using SeqMan Pro from the Lasergene DNASTAR software suite.

Table 3.2. Primers used for site-directed mutagenesis of Med25 AcID.

Plasmid	Primer sequence
pET21b-Med25(394-543)C506A-His6	F: TCCCCCCACGGCGCCCGCCGAGGTGCGCGTGCTCATG R: CATGAGCACGCGCACCTCGGCGGGCGCCGTGGGGGGA
pET21b-Med25(394-543)C497A-His6	F: GGCCAACGGCTTCGCGGGCGCCGTGCACTTCCCCACACG R: CGTGTGGGGAAGTGCACGGCGCCCGCGAAGCCGTTGCC
pET21b-Med25(394-543)K518R-His6	F: TCATGCTCCTGTACTCGTCCAGGAAGAAGATCTTCATGGGCCTCATCCC R: GGGATGAGGCCCATGAAGATCTTCTTCCTGGACGAGTACAGGAGCATGA
pET21b-Med25(394-543)K519R-His6	F: TCATGCTCCTGTACTCGTCCAAGAGGAAGATCTTCATGGGCCTCATCCC R: GGGATGAGGCCCATGAAGATCTTCCTCTTGGACGAGTACAGGAGCATGA
pET21b-Med25(394-543)K520R-His6	F: TCATGCTCCTGTACTCGTCCAAGAAGAGGATCTTCATGGGCCTCATCCC R: GGGATGAGGCCCATGAAGATCCTCTTCTTGGACGAGTACAGGAGCATGA
pET21b-Med25(394-543)KK518RR-His6	F: TCATGCTCCTGTACTCGTCCAGGAGGAAGATCTTCATGGGCCTCATCCC R: GGGATGAGGCCCATGAAGATCTTCCTCCTGGACGAGTACAGGAGCATGA
pET21b-Med25(394-543)KKK518RRR-His6	F: TCATGCTCCTGTACTCGTCCAGGAGGAGGATCTTCATGGGCCTCATCCC R: GGGATGAGGCCCATGAAGATCCTCCTCCTGGACGAGTACAGGAGCATGA

pET21b-Med25(394-543)K411R/K413R-His6	F: GGGGTCCTGGAGTGGCAAGAGAGACCCAGACCTGCCTCAGTGGATGCCAAC R: GTTGGCATCCACTGAGGCAGGTCTGGGTCTCTCTTGCCACTCCAGGACCCC

Protein expression

Wild-type Med25(394-543) and variants were expressed in *E. coli* Rosetta pLysS cells (EMD Millipore) using standard protocols.2 Briefly, overnight starter cultures (50 mL) were grown in LB (Research Products International) in the presence of 1 mg/mL ampicillin (Gold Bio Technology) and 0.034 mg/mL chloramphenicol (Sigma Aldrich) at 37 °C at 150 RPM. Following overnight growth, 1 L of terrific broth (12 g tryptone, 24 g yeast extract, 4 mL glycerol, 100 mL 0.17M KH2PO4/0.72M K2HPO4, 900 mL water) supplemented with 1 mg/mL ampicillin and 0.034 mg/mL chloramphenicol was inoculated with 5 mL of the overnight culture and grown at 37 °C to an OD of 0.8, at which point the incubator temperature was lowered to 21 °C and isopropyl β-D-1-thiogalactopyranoside (IPTG, Research Products International) was added to a final concentration of 0.5 mM to induce protein expression. Agitation at 150 RPM was continued for 18 hours. Cells were harvested via centrifugation at 6500 RPM for 20 minutes, snap frozen in liquid nitrogen, and stored at -80 °C.

To isolate and purify Med25 AcID and mutants, frozen cell pellets were resuspended in lysis buffer (50 mM phosphate, 300 mM NaCl, 10 mM imidazole, pH 6.5, 0.7 μL/mL β-ME) and lysed by sonication. Cellular debris was then pelleted by centrifugation for 20 min at 10,000 rpm. The supernatant was filtered using a 0.45 μm syringe filter (CellTreat) and loaded directly on an AKTA Pure FPLC equipped with a Ni column (HisTrap HP, GE Healthcare) using a gradient of Buffer A (50 mM sodium phosphate, 300 mM NaCl, and 30 mM imidazole, pH 6.8) to Buffer B (50 mM sodium phosphate, 300 mM NaCl, and 400 mM imidazole, pH 6.8) over ten column volumes. AcID-containing fractions were pooled and underwent a second purification via cation exchange (HiTrap SP HP, GE Healthcare) with a gradient of Buffer A (50 mM sodium phosphate, 1 mM DTT, pH 6.8) to Buffer B (50 mM sodium phosphate, 1 mM DTT, 1M

NaCl, pH 6.8) over fifteen column volumes. Pooled fractions containing purified protein were dialyzed into storage buffer (10 mM $NaPO_4$, 50 mM NaCl, pH 6.8). Following dialysis, protein was concentrated via Amicon 5,000 Da cutoff spin concentrator and quantified via UV/Vis spectroscopy on a NanoDrop at 280 nm using an extinction coefficient, ε = 22,460 $M^{-1}cm^{-1}$. Protein identity was confirmed via mass spectrometry (Agilent Q-TOF) and purity was assessed by SDS-PAGE on a 4-12% bis-tris gel stained using Quick Coomassie (Anatrace).

For isotopically labeled protein, Rosetta pLysS cells bearing the appropriate expression plasmid were grown in LB at 37 °C for 18 hrs. The resulting cells were pelleted via centrifugation for 10 minutes at 2500 RPM and the LB was decanted. Cell pellets were resuspended in 20 mL M9 minimal media followed by centrifugation for 10 mins at 2500 RPM. M9 media was decanted and the washed pellet resuspended in 10 mL M9 media. 5 mL of suspended cells were then used to inoculate a 1 L expression flask containing M9 media supplemented with Bioexpress (6 mL/L, Cambridge Isotopes) and containing the appropriate isotopically labeled reagents (1 g/L $^{15}NH_4Cl$ or 1 g/L $^{15}NH_4Cl$ and ^{13}C D-glucose, Cambridge Isotopes). Isotopically labeled protein was isolated and purified as described above, with the exception of the storage buffer used for dialysis (20 mM $NaPO_4$, 150 mM NaCl, pH 6.5).

NMR Analyses of Activator–AcID Complexes

^1H,^{15}N HSQC experiments of activator–AcID complexes were performed on a Bruker Avance III 600 MHz spectrometer equipped with a cryogenic probe at 30 °C. Titrations were conducted with Med25 AcID (20 mM $NaPO_4$, 150 mM NaCl, pH 6.5, 5% D_2O) at 50 μM, and acetylated peptides were added at 0, 0.2, 0.5, 0.8, 1.1, 2, and 3 eq with a 2% final DMSO concentration. Control spectra were obtained with Med25 AcID and DMSO only. Tethered activator–AcID complexes were prepared as previously described.[15] Data processing and visualization was performed using NMR Pipe and Sparky.[38]

Peptide Tethering

VP16-AcID tethered complexes were prepared by incubating 10 μM Med25 AcID with 4 eq of individual disulfide-capped VP16 peptides (3.1, Entries 5-18) and 2 eq of β-mercaptoethanol (β-ME) for 4 hours at room temperature in protein storage buffer (10 mM $NaPO_4$, 100 mM NaCl, pH 6.8). Immediately following the 4 hour incubation, labeling efficiency was determined via mass spectrometric analysis of an aliquot of each reaction on an Agilent Q-TOF and comparing the spectral peak heights of unlabeled Med25 AcID, the Med25 AcID-β-ME adduct, and the tethered Med25 AcID-VP16 complex using the following equation:

$$Percent\ labeled\ = \frac{Med25\ AcID_{VP16}}{Med25\ AcID_{unlabeled} + Med25\ AcID_{\beta-ME} + Med25\ AcID_{VP16}}\ x\ 100$$

Transient kinetic experiments

Stopped-flow kinetic assays were performed on a Kintek SF-2001 stopped-flow apparatus equipped with a 100-W Xe arc lamp in two-syringe mode. The 4-DMN fluorophore was excited at 440 nm and its emission was monitored at wavelengths >510 nm, using a long-pass filter (Corion). All experiments were run at 10 °C in stopped-flow buffer (10 mM sodium phosphate, 100 mM NaCl, 1% glycerol, 0.001% NP-40, pH 6.8), and the solutions were equilibrated in the instrument for at least 5 min before experiments were performed. Concentrations reported are after mixing. All kinetic traces reported are an average of 20-80 individual traces. A series of exponential equations were fit to the transient kinetic time courses, F(t) as in the equation below, to obtain the fluorescence amplitudes (F_n) and the observed rate constants (k_{obs}) for each exponential phase, where *F(0)* is the initial fluorescence intensity and *t* is time.

$$F(t) = F(0) + \sum F_n \times (1 - e^{-k_{obs,n} \times t})$$

Association experiments were performed by mixing excess Med25 with a constant concentration of labeled activator. Dissociation experiments were performed by precomplexing Med25 with the labeled activator and mixing with an excess of the corresponding unlabeled activator. The concentrations of labeled activators for association experiments were as follows 50 nM VP16 and 250 nM ERM.

Calculation of microscopic rate constants

In previous work, a Taylor series approximation of the exact equation for a two-step association mechanism was used.[20,21] However, here the kinetics did not meet the requirements to use this approximation (i.e., sufficient time-domain separation of the two phases at the limit of zero Med25). Instead, because two phases were observed, the exact expression for a dissociation experiment (below) was utilized to calculate all first-order microscopic rate constants.

$$k_{obs,off,1,2} = \frac{k_{off} + k_f + k_r \pm \sqrt{(k_{off} + k_f + k_r)^2 - 4 \times k_{off} \times k_r}}{2}$$

Because $k_{obs,2,max}$ is the sum of k_f and k_r, k_{off} is simply the sum of the two observed off-rate constants minus $k_{obs,2,max}$:

$$k_{off} = k_{obs,off,1} + k_{obs,off,2} - k_{obs,2,max}$$

A value for k_r can also be easily resolved from the exact expression, by first subtracting the two observed off-rate constants and then substituting observable parameters and rearranging:

$$k_{obs,off,1} - k_{obs,off,2} = \sqrt{(k_{off} + k_f + k_r)^2 - 4 \times k_{off} \times k_r}$$

$$= \sqrt{(k_{obs,off,1} + k_{obs,off,2})^2 - 4 \times (k_{obs,off,1} + k_{obs,off,2} - k_{obs,2,max}) \times k_r}$$

$$k_r = \frac{(k_{obs,off,1} + k_{obs,off,2})^2 - (k_{obs,off,1} - k_{obs,off,2})^2}{4 \times (k_{obs,off,1} + k_{obs,off,2} - k_{obs,2,max})}$$

k_r is then simply obtained by subtracted the calculated value of k_r from $k_{obs,2,max}$. To validate this model, a subset of kinetic transients were globally fit using the calculated microscopic rate constants as constraints with Kintek Explorer. Here, the rate constants were constrained to the calculated values and fitted fluorescence values for each state

along with small scaling factors to account for variations in lamp intensity. Good fits were found for the data, further supporting the proposed mechanisms and values.

Direct binding and competition experiments

Direct binding and competition experiments were performed using fluorescence polarization. Low volume, non-binding black 384-well plates (Corning) were used and fluorescence polarization was measured using a PHERAStar multi-mode 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 5.0. For direct binding experiments, a binding isotherm that accounts for ligand depletion (assuming a 1:1 binding model of peptide to AcID) was fit to the observed polarization values as a function of protein concentration to obtain the apparent equilibrium dissociation, 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 of the three replicates. For competition experiments, curves were fit with a non-linear regression using the built-in equation “log(inhibitor) vs response – variable slope” from which the IC_{50} value was calculated.

High-throughput screening

Assays were performed in a final volume of 20 µL in a low volume, non-binding, black 384-well plate (Corning) and read by plate reader (Pherastar) with polarized 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 (Kd) over time, with combinations of DMSO and NP-40. 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 µL FITC-VP16(465-490). The compounds were then tested for fluorescence quenching before 10 µL 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 µM with a final DMSO concentration of 1% v/v. Data was published to and analyzed using MScreen (http://mscreen.lsi.umich.edu).

The primary screening campaign had an average Z' score of 0.87, indicating an excellent assay, and a 1.6% hit rate. For the purposes of this screen, a hit was defined as any compound that resulted in inhibition greater than three standard deviations above the negative control, which corresponded to approximately ten percent inhibition. 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.

Mass spectrometry analysis of covalent adducts

Protein (Med25 WT and mutants) was diluted to a concentration of 20 µM 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 diluted protein to a final concentration of 20 µM (1 equivalent). Samples were incubated for 30 minutes at room temperature with gentle mixing on an orbital shaker. Samples were analyzed by mass spectrometry using an Agilent QToF LC/MS equipped with a Poroshell 300SB C8 reverse-phased column using 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.

3.5 References

(1) 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.

(2) Taatjes, D. J. The Human Mediator Complex: A Versatile, Genome-Wide Regulator of Transcription. *Trends Biochem. Sci.* **2010**, *35* (6), 315–322. https://doi.org/10.1016/j.tibs.2010.02.004.

(3) Cantin, G. T.; Stevens, J. L.; Berk, A. J. Activation Domain–Mediator Interactions Promote Transcription Preinitiation Complex Assembly on Promoter DNA. *Proc. Natl. Acad. Sci.* **2003**, *100* (21), 12003–12008. https://doi.org/10.1073/pnas.2035253100.

(4) Yin, J.; Wang, G. The Mediator Complex: A Master Coordinator of Transcription and Cell Lineage Development. *Dev. Camb. Engl.* **2014**, *141* (5), 977–987. https://doi.org/10.1242/dev.098392.

(5) Ansari, S. A.; Morse, R. H. Selective Role of Mediator Tail Module in the Transcription of Highly Regulated Genes in Yeast. *Transcription* **2012**, *3* (3), 110–114. https://doi.org/10.4161/trns.19840.

(6) 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.

(7) Vojnic, E.; Mourão, A.; Seizl, M.; Simon, B.; Wenzeck, 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.

(8) 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 ATF6α *. *J. Biol. Chem.* **2013**, *288* (36), 26179–26187. https://doi.org/10.1074/jbc.M113.496968.

(9) 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.

(10) 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.

(11) 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.

(12) Eletsky, A.; Ruyechan, W. T.; Xiao, R.; Acton, T. B.; Montelione, G. T.; Szyperski, T. Solution NMR Structure of MED25(391–543) Comprising the Activator-

Interacting Domain (ACID) of Human Mediator Subunit 25. *J. Struct. Funct. Genomics* **2011**, *12* (3), 159. https://doi.org/10.1007/s10969-011-9115-1.
(13) Teilum, K.; Kunze, M. B. A.; Erlendsson, S.; Kragelund, B. B. (S)Pinning down Protein Interactions by NMR. *Protein Sci. Publ. Protein Soc.* **2017**, *26* (3), 436–451. https://doi.org/10.1002/pro.3105.
(14) Milbradt, A. G.; Kulkarni, M.; Yi, T.; Takeuchi, K.; Sun, Z.-Y. J.; Luna, R. E.; Selenko, P.; Näär, A. M.; Wagner, G. Structure of the VP16 Transactivator Target in the Mediator. *Nat. Struct. Mol. Biol.* **2011**, *18* (4), 410–415. https://doi.org/10.1038/nsmb.1999.
(15) 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.* **2011**, *108* (15), 6056–6061. https://doi.org/10.1073/pnas.1102376108.
(16) Liu, P.; Kim, B.; Friesner, R. A.; Berne, B. J. Replica Exchange with Solute Tempering: A Method for Sampling Biological Systems in Explicit Water. *Proc. Natl. Acad. Sci.* **2005**, *102* (39), 13749–13754. https://doi.org/10.1073/pnas.0506346102.
(17) Sugita, Y.; Okamoto, Y. Replica-Exchange Molecular Dynamics Method for Protein Folding. *Chem. Phys. Lett.* **1999**, *314* (1), 141–151. https://doi.org/10.1016/S0009-2614(99)01123-9.
(18) Kim, J.; Straub, J. E.; Keyes, T. Replica Exchange Statistical Temperature Molecular Dynamics Algorithm. *J. Phys. Chem. B* **2012**, *116* (29), 8646–8653. https://doi.org/10.1021/jp300366j.
(19) 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.
(20) Johnson, K. A. Rapid Kinetic Analysis of Mechanochemical Adenosinetriphosphatases. *Methods Enzymol.* **1986**, *134*, 677–705. https://doi.org/10.1016/0076-6879(86)34129-6.
(21) Wands, A. M.; Wang, N.; Lum, J. K.; Hsieh, J.; Fierke, C. A.; Mapp, A. K. Transient-State Kinetic Analysis of Transcriptional Activator·DNA Complexes Interacting with a Key Coactivator. *J. Biol. Chem.* **2011**, *286* (18), 16238–16245. https://doi.org/10.1074/jbc.M110.207589.
(22) 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.* **2014**, *111* (33), 12061–12066. https://doi.org/10.1073/pnas.1406033111.
(23) Shammas, S. L.; Travis, A. J.; Clarke, J. Remarkably Fast Coupled Folding and Binding of the Intrinsically Disordered Transactivation Domain of CMyb to CBP KIX. *J. Phys. Chem. B* **2013**, *117* (42), 13346–13356. https://doi.org/10.1021/jp404267e.
(24) Shammas, S. L.; Travis, A. J.; Clarke, J. Allostery within a Transcription Coactivator Is Predominantly Mediated through Dissociation Rate Constants. *Proc. Natl. Acad. Sci.* **2014**, *111* (33), 12055–12060. https://doi.org/10.1073/pnas.1405815111.

(25) Gianni, S.; Morrone, A.; Giri, R.; Brunori, M. A Folding-after-Binding Mechanism Describes the Recognition between the Transactivation Domain of c-Myb and the KIX Domain of the CREB-Binding Protein. *Biochem. Biophys. Res. Commun.* **2012**, *428* (2), 205–209. https://doi.org/10.1016/j.bbrc.2012.09.112.
(26) Dogan, J.; Schmidt, T.; Mu, X.; Engström, Å.; Jemth, P. Fast Association and Slow Transitions in the Interaction between Two Intrinsically Disordered Protein Domains*. *J. Biol. Chem.* **2012**, *287* (41), 34316–34324. https://doi.org/10.1074/jbc.M112.399436.
(27) Hilser, V. J.; Thompson, E. B. Intrinsic Disorder as a Mechanism to Optimize Allosteric Coupling in Proteins. *Proc. Natl. Acad. Sci.* **2007**, *104* (20), 8311–8315. https://doi.org/10.1073/pnas.0700329104.
(28) Schrank, T. P.; Bolen, D. W.; Hilser, V. J. Rational Modulation of Conformational Fluctuations in Adenylate Kinase Reveals a Local Unfolding Mechanism for Allostery and Functional Adaptation in Proteins. *Proc. Natl. Acad. Sci.* **2009**, *106* (40), 16984–16989. https://doi.org/10.1073/pnas.0906510106.
(29) 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.
(30) Tompa, P.; Fuxreiter, M. Fuzzy Complexes: Polymorphism and Structural Disorder in Protein-Protein Interactions. *Trends Biochem. Sci.* **2008**, *33* (1), 2–8. https://doi.org/10.1016/j.tibs.2007.10.003.
(31) 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.; Hodoscek, 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.
(32) MacKerell, A. D.; Bashford, D.; Bellott, M.; Dunbrack, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E.; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiórkiewicz-Kuczera, J.; Yin, D.; Karplus, M. All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins. *J. Phys. Chem. B* **1998**, *102* (18), 3586–3616. https://doi.org/10.1021/jp973084f.
(33) 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.
(34) Im, W.; Lee, M. S.; Brooks III, C. L. Generalized Born Model with a Simple Smoothing Function. *J. Comput. Chem.* **2003**, *24* (14), 1691–1702. https://doi.org/10.1002/jcc.10321.
(35) Chen, J.; Im, W.; Brooks III, C. L. Balancing Solvation and Intramolecular Interactions. *J. Am. Chem. Soc.* **2006**, *128* (11), 3728–3736. https://doi.org/10.1021/ja057216r.

(36) Karpen, M. E.; Tobias, D. J.; Brooks III, C. L. Statistical Clustering Techniques for the Analysis of Long Molecular Dynamics Trajectories: Analysis of 2.2-Ns Trajectories of YPGDV. *Biochemistry* **1993**, *32* (2), 412–420.
(37) 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.
(38) Delaglio, F.; Grzesiek, S.; Vuister, G. W.; Zhu, G.; Pfeifer, J.; Bax, A. NMRPipe: A Multidimensional Spectral Processing System Based on UNIX Pipes. *J. Biomol. NMR* **1995**, *6* (3), 277–293. https://doi.org/10.1007/BF00197809.

Chapter 4

TMPRSS2 Inhibitor Discovery Facilitated Through an *in silico* and Biochemical Screening Platform[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. Here we utilize virtual screening to curate large libraries into a focused collection of potential inhibitors. Optimization of a recombinant expression and purification protocol for the TMPRSS2 peptidase domain facilitates subsequent biochemical screening and characterization of selected compounds from the curated collection in a kinetic assay. In doing so, we demonstrate that serine protease inhibitors camostat, nafamostat, and gabexate inhibit through a covalent mechanism. We further identify new non-covalent compounds as TMPRSS2 protease inhibitors, demonstrating the utility of a combined virtual and experimental screening campaign in rapid drug discovery efforts.

4.1 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 coronoviruses, infection with SARS-CoV-2 requires host cell cooperation; the spike (S) protein protruding outside the viral coat requires priming by TMPRSS2, a human transmembrane serine protease, for viral entry via the receptor ACE2 (*Figure 4.1A*).[2–5]

[1] Portions of this chapter were adopted from: **Peiffer, A. L.**; Garlick, J. M.; Wu, Y.; Soellner, M. B.; Brooks III, C. L.; Mapp, A. K. "TMPRSS2 inhibitor discovery facilitated through an *in silico* and biochemical screening platform," *bioRxiv* doi:10.1101/2021.03.22.436465 (2021).

While many have focused 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] Along with SARS-CoV-2, TMPRSS2 has been implicated in priming other pathogenic coronaviruses such as SARS-CoV and MERS, as well as influenza.[9–12] 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.[13] 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.[14] 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. Thus, TMPRSS2 is a desirable drug target for treating SARS-CoV-2 and future coronavirus infections.

To date, there are few reported TMPRSS2 inhibitors. Camostat, a compound initially discovered as a Matriptase 2 inhibitor, also inhibits TMPRSS2.[4,15] Nafamostat and gabexate have also been reported to inhibit TMPRSS2. However, both camostat and nafamostat inhibit a wide range of serine proteases and are rapidly metabolized in mammals to structures with poorly defined activity. [16–18] It has been reported that each of these compounds all form a covalent bond with the active site serine of serine proteases via the central ester, also a site of metabolic breakdown (*Figure 4.1A*).[19,20] Additionally, molecular modeling studies on TMPRSS2 supports covalent bond formation with both camostat and nafamostat, as well as the camostat metabolite FOY 251.[21] Thus, TMPRSS2 inhibitors with less reactive architectures are highly desirable.

As a strategy for rapid TMPRSS2 inhibitor discovery, we developed a combined *in silico* and biochemical workflow. Strategic development of a TMPRSS2 expression protocol, utilizing an autocatalysis-based affinity tag removal, facilitated purification for biochemical assay development. This allowed existing TMPRSS2 inhibitors to be profiled and characterized as covalent. Our protocol for virtual screening against a TMPRSS2 homology model (*Figure 4.2*) comprises a molecular dynamics/simulated annealing-based docking that employs flexible receptor side chains to capture subtle changes, both

conformational and energetic, for improved compound scoring. Our approach curated a subset of promising TMPRSS2 ligands, which upon subsequent biochemical testing were identified as active inhibitors. In doing so, we identify new non-covalent hit compounds that can be both repurposed for SARS-CoV-2 infections as well as derivatized to yield improved TMPRSS2 inhibitors.

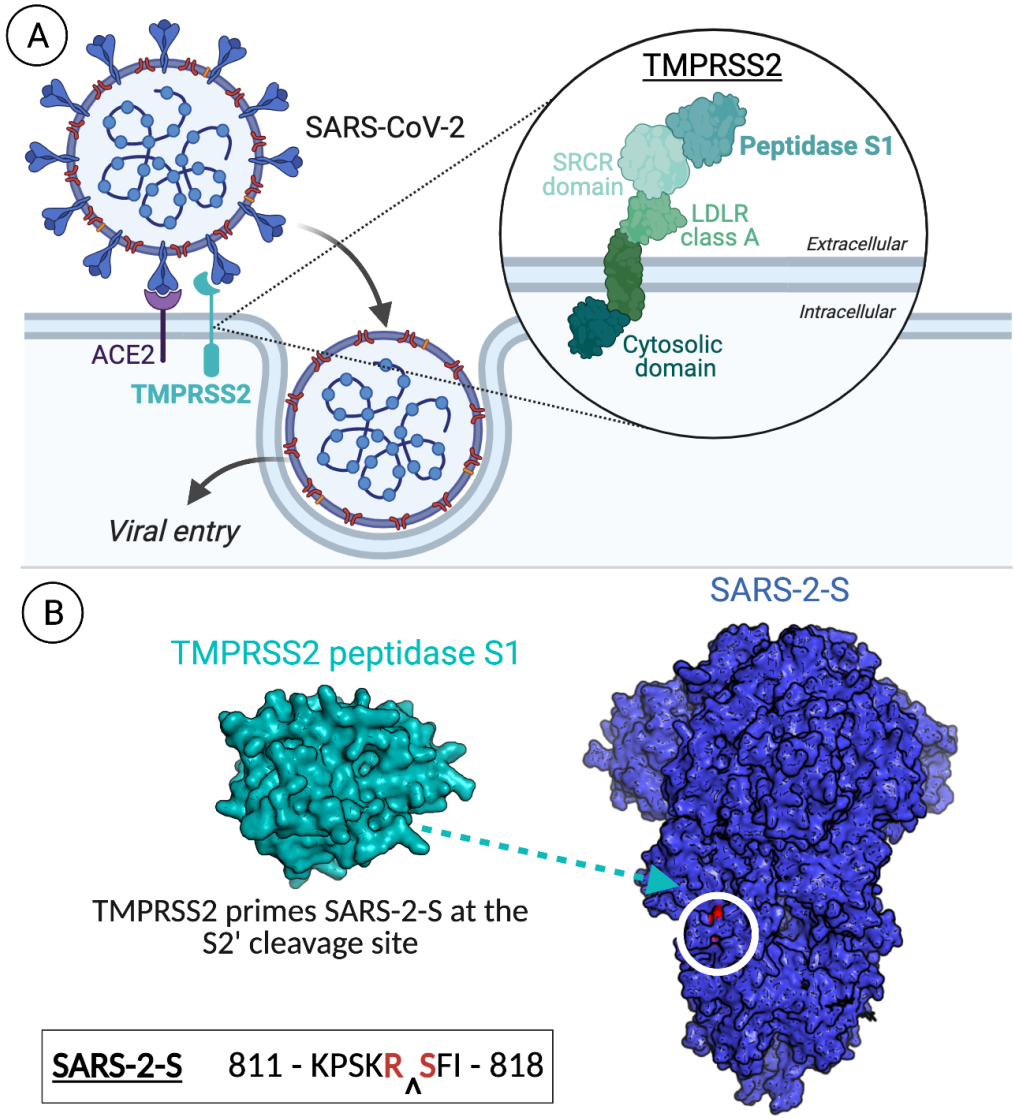


Figure 4.1. *The role of TMPRSS2 in SARS-CoV-2 infection*. (A) TMPRSS2 primes the viral S protein (SARS-2-S), which promotes membrane fusion and ultimately viral entry. TMPRSS2, part of the type II transmembrane serine protease family and hepsin/TMPRSS subfamily, is anchored at the cell membrane.[4] 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.[22] (B) The peptidase S1 domain of TMPRSS2 cleaves SARS-2-S (PDB 6VXX) at the S2' cleavage site.[23,24]

4.2 Results and Discussion

Active TMPRSS2 Peptidase S1 can be expressed recombinantly 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.[25–27] 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, called TMPRSS2(247–492), with an *N*-terminal 6xHis tag for purification (*Figure 4.2*).

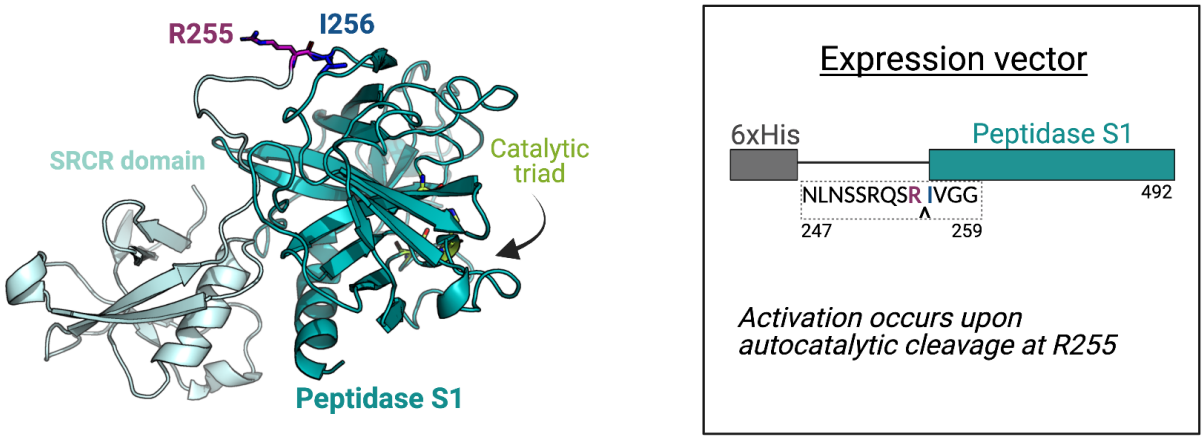


Figure 4.2 *TMPRSS2 activation occurs through autocleavage*. (Left) The homology model for TMPRSS2 contains the SRCR domain (light blue) and the protease domain (teal). Activation occurs through autocatalysis at R255 (purple), leaving I256 (dark blue) free to insert under protease domain and stabilize the active conformation. (Right) The recombinant expression vector contains an extended portion of the Peptidase S1 (247-492) of TMPRSS2 along with a 6xHis tag that allows for purification of the denatured protease and subsequent cleavage of the affinity tag after proper refolding.

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 4.3*). As seen in *Figure 4.4*, while TMPRSS2 expression yields highly concentrated protein that is denatured, the refolding procedure significantly reduces the concentration. Nonetheless, silver staining was able to detect protein at the correct mass, and a TMPRSS2 antibody showed furth validation of obtaining the correct protein.

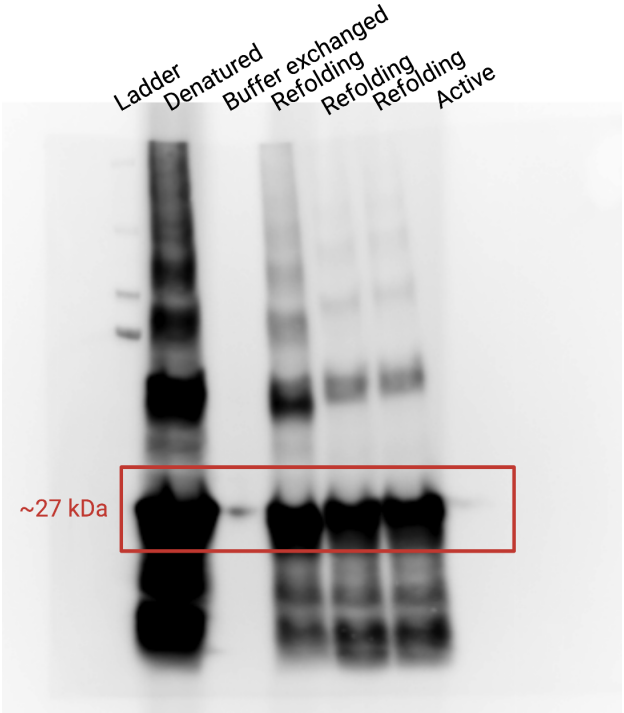


Figure 4.3. *Western blot for 6xHis tag at various stages of TMPRSS2 purification*. Significant signal is seen after denaturation and at various time points during refolding. A small amount of His-tagged protein is

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 performed with Dr. Julie Garlick.

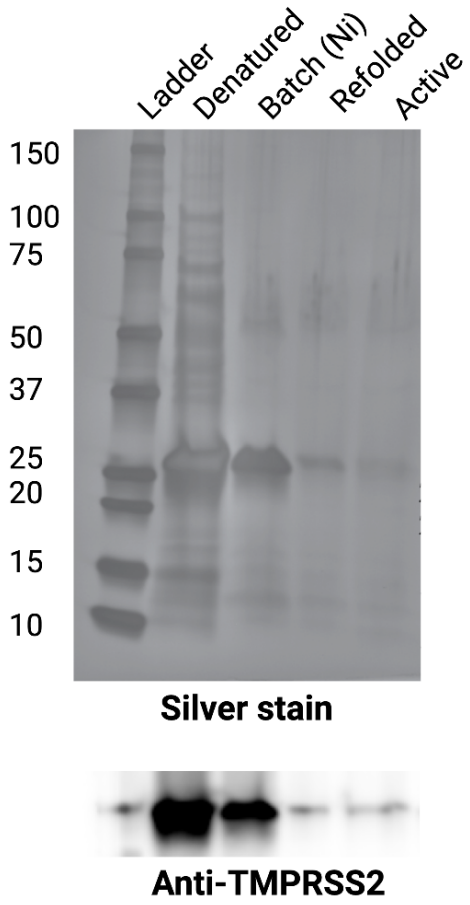


Figure 4.4. *Silver stain gel and Western blot for TMPRSS2 protease.* TMPRSS2 protein is expressed and purified by expression in inclusion bodies (labeled denatured), purified by batch binding to Ni-NTA resin, refolded, and dialyzed into assay buffer. Silver staining shows purity and protein levels at 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. Experiments performed with Dr. Julie Garlick.

Activity of TMPRSS2 peptidase domain

The classic trypsin substrate Boc-QAR-AMC has been reported as a TMPRSS2 substrate, which we used to confirm activity of the purified protein (*Figure 4.5A*).[16,29] In doing kinetic fluorescence experiments, we determined the K_M to be 5.1 ± 0.4 μM with TMPRSS2 (*Figure 4.5D* and *4.5E*), which is comparable to the K_M for trypsin with this substrate, 7.8 ± 0.7 μM (*Figure 4.5B and 4.5C*).

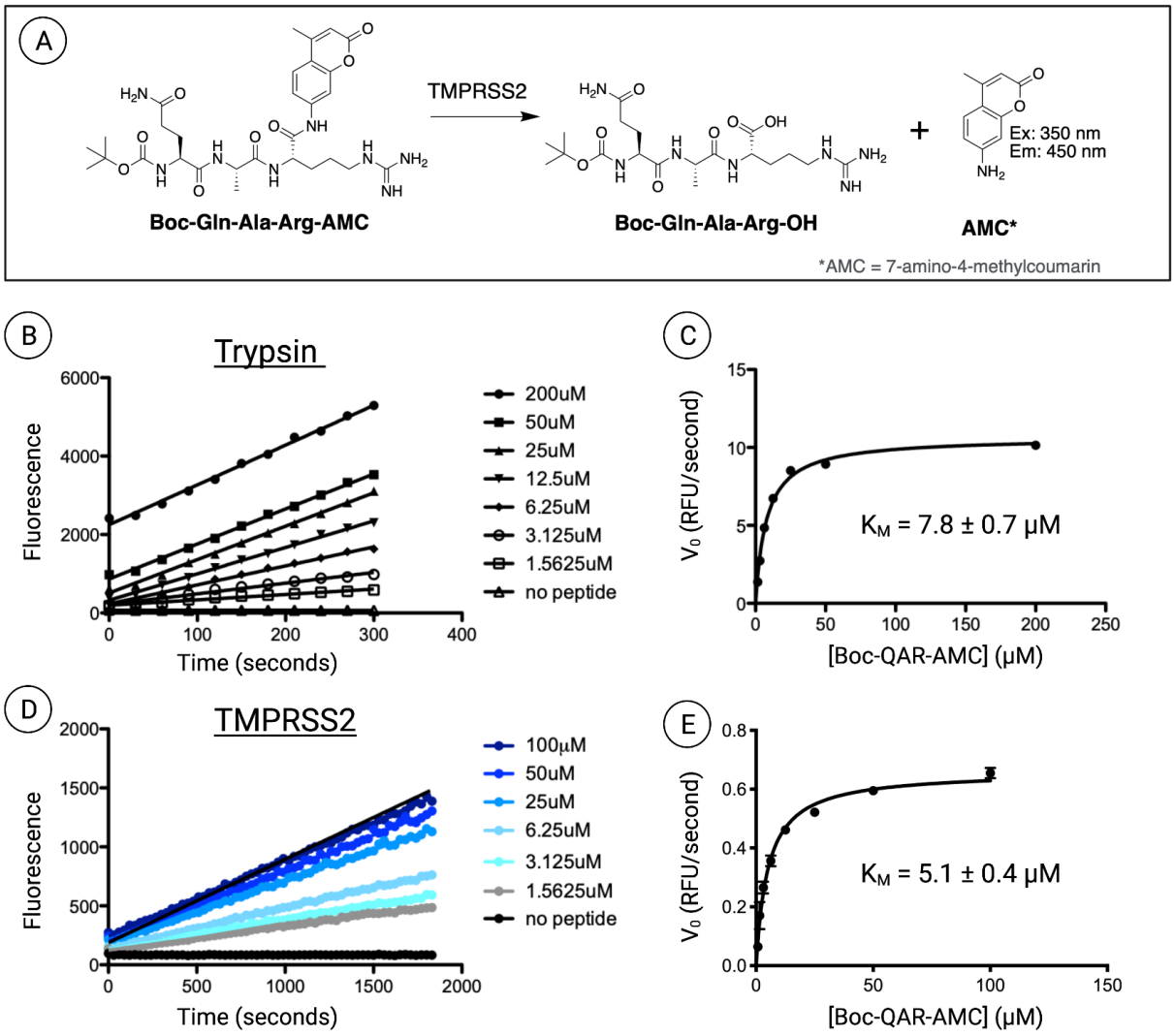


Figure 4.5. *Fluorescence-based kinetic activity assay for serine proteases.* (A) Schematic of the biochemical assay used to monitor TMPRSS2 activity. Raw fluorescence reads over time for (B) trypsin (D) TMPRSS2 using Boc-Glu-Ala-Arg-AMC (Boc-QAR-AMC) are shown. K_M is determined by plotting initial velocity, calculated as the slope of the kinetic trace at less than 10 percent substrate cleavage, vs substrate concentration, which is shown for (C) trypsin and (E) TMPRSS2. Data was fit to the Michaelis-Menten equation using GraphPad Prism. Experiments performed with Dr. Julie Garlick.

Using a NanoDrop, we calculated that we had ~1.7 μM protein in our active sample. However, we noted that does not reflect the concentration of *active* protein. Thus, to determine approximate concentration of active protein, we measured TMPRSS2 activity while titrating in the covalent protease inhibitor FPR-chloromethylketone (also known as PPACK).[30–32] 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 4.6*). In doing so, we determined that we had roughly 0.5 nM active protein in our samples.

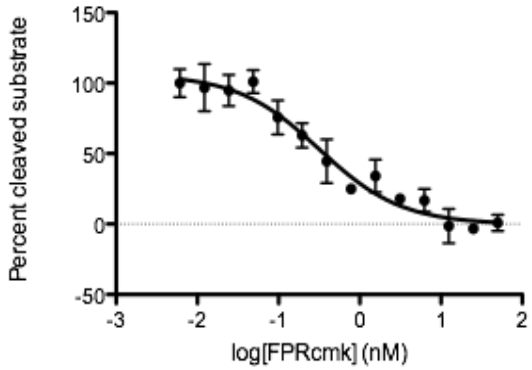


Figure 4.6. *Inhibition curve of TMPRSS2 activity obtained with increasing [FPR-cmk].* The IC_{50} obtained here is 0.32 ± 0.09 nM, allowing us to approximate 100% inhibition at about 0.64 nM. Thus, assuming 1:1 complex formation of protein and inhibitor, we can determine the protein concentration of 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. Experiments performed by Dr. Julie Garlick.

Experimental conditions for high throughput screening (HTS) in 384-well plates were optimized using 2.5 μM 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 μM 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 0.79, respectively, indicating an excellent HTS assay (*Figure 4.7*).

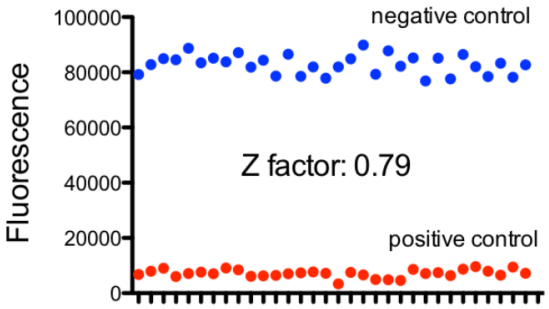


Figure 4.7. *HTS with TMPRSS2 protease.* TMPRSS2 activity assay was adapted to 384 well format with the Z factor indicating an excellent assay. Negative controls correspond to protein + substrate, while positive controls correspond to substrate only. Experiment performed by Dr. Julie Garlick.

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 4.8*). IC_{50} values for these compounds agree with previously reported values obtained using similar biochemical approaches. 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.

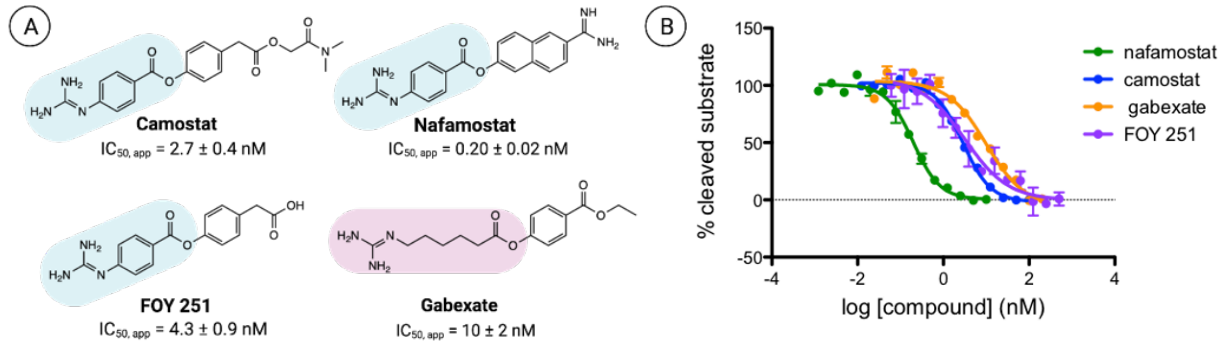


Figure 4.8. *TMPRSS2 inhibition with published molecules*. (A) The molecules that have commonly been cited as TMPRSS2 inhibitors are camostat, FOY 251, nafamostat, and gabexate. All contain a reactive ester, which can form a covalent bond with the catalytic serine, as well as a guanidium group. (B) Inhibition 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. Experiments performed with Dr. Julie Garlick.

Analysis of covalent TMPRSS2 inhibitors

Although not previously shown experimentally, others have suggested that the inhibitors camostat and nafamostat covalently modify TMPRSS2 with a similar mechanism to other proteases.[21] 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 4.9*). With equimolar protein and compound, an increase in mass of 161 Da is observed for camostat, nafamostat, and FOY 251 (*Figure 4.9*). Gabexate requires 10X compound compared to protein for adduct formation to be observed, which corresponds to a mass increase of 155 Da (*Figure 4.9*).

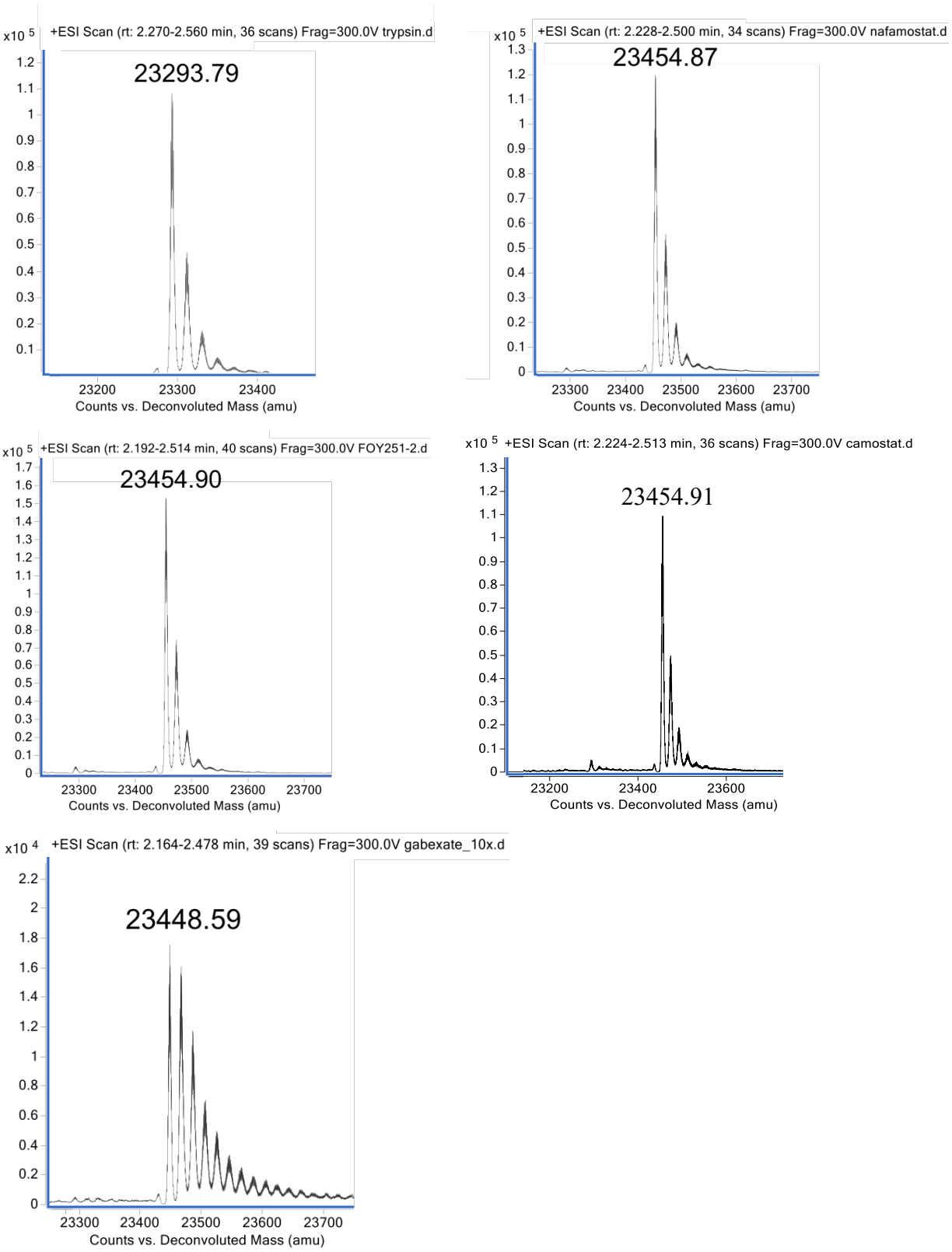


Figure 4.9. *Covalent inhibition with trypsin using mass spectrometry.* Deconvoluted LC-MS data of (Top left) apo trypsin, incubated with (top right) equimolar nafamostat, (second row, left) FOY 251, or (second row, right) camostat. After 30 minutes incubation, a mass increase of about 161 Da is observed, indicating

covalent adduct formation. (Bottom) Deconvoluted LC-MS data of trypsin incubated with 10X gabexate. After 30 minutes of 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 gabexate. Experiments performed by Dr. Julie Garlick.

Since mass spectrometry was not a viable option with TMPRSS2, covalency was tested through looking at IC_{50}'s with varying incubation times with inhibitors. A time-dependent decrease in apparent IC_{50} is observed with camostat in biochemical activity assays for both trypsin and TMPRSS2 (*Figure 4.10*). 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 4.10*). Thus, while these compounds exhibit low nM IC_{50} values *in vitro*, this is likely due, at least in part, to the covalent mechanism of action.

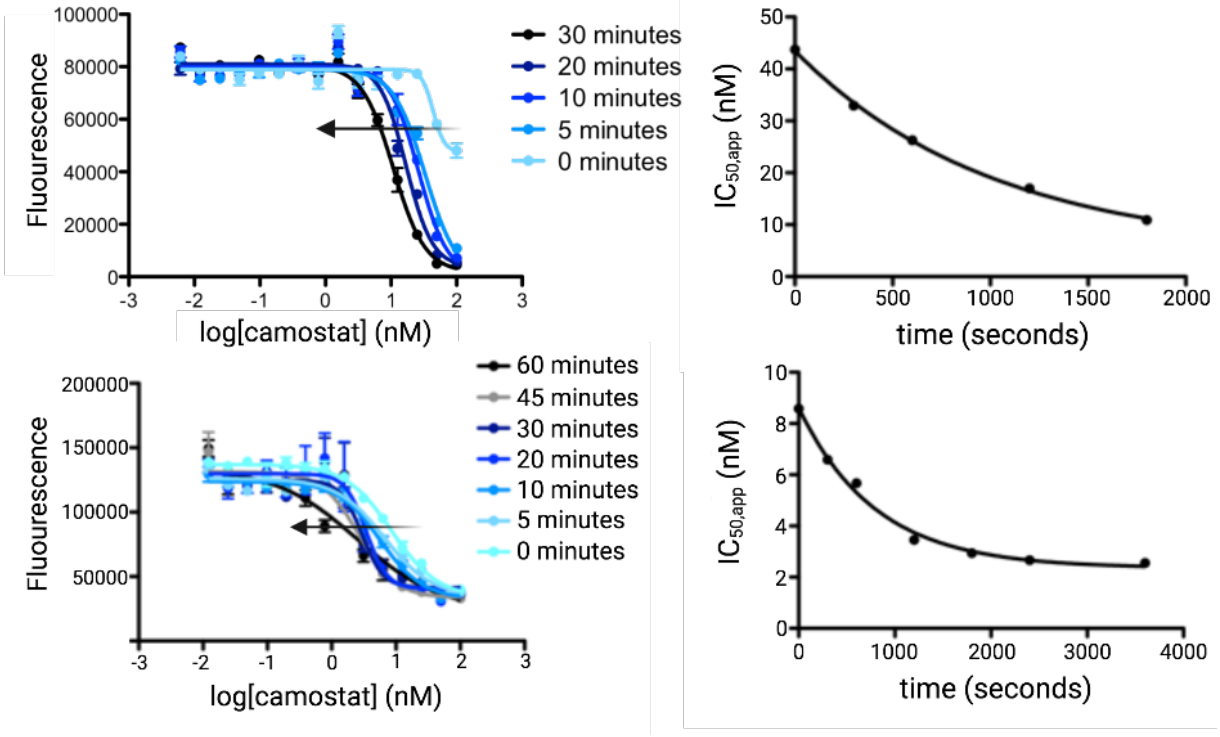


Figure 4.10. *Decreasing IC_{50} values with longer incubations suggest a covalent mechanism of inhibition.* Decreasing IC_{50} values for camostat with increased incubation time for trypsin (top) and TMPRSS2 (bottom) is consistent with covalent inhibition. Experiments performed with Dr. Julie Garlick.

It is important to note that in a cellular context, camostat has a very short half-life of <1 min.[33] Rapid hydrolysis to (4-(4-guanidinobenzoyl-oxy)phenylacetic acid), also known as the protease inhibitor FOY 251 (*Figure 4.8A*), occurs both *in vitro* in serum as well as *in vivo*.[34–36] We observe similar potencies of FOY 251 and camostat in our

biochemical assay, with apparent IC_{50} = 4.3 ± 0.9 nM and 2.7 ± 0.4 nM, respectively (*Figure 4.8*). The half-life of FOY 251 is longer than camostat, though it is metabolized to 4-guanidinobenzoic acid, which shows minimal TMPRSS2 inhibition.[34] Thus, it would be advantageous to identify novel inhibitors that have alternative molecular scaffolds.

Construction and refinement of TMPRSS2 homology model

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*. 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 4.2*).[37–41] 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 as described below. 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.[42–44]

Virtual screening yields 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 4.11*). A total of 134,109 molecules were collected from multiple databases, 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 (*Figure 4.11*). The second relied upon a novel 3D pharmacophore *fastdock* framework,

which operates by superposing pharmacophores onto compounds bound in experimentally solved structures.

From this initial stage, 4,307 Level 1 screening candidates were determined and subjected to GPU accelerated Flexible-CDOCKER methods that were recently developed as part of the CHARMM molecular modeling package.[45–48] 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.

We utilize a novel scoring methodology by performing conformational clustering of the flexible side chains and the ligand, which provides key contributions to the ligand scoring from the entropic variation of the side chains to accommodate various ligand poses. The ligands are 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.[49] The rescoring is 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.

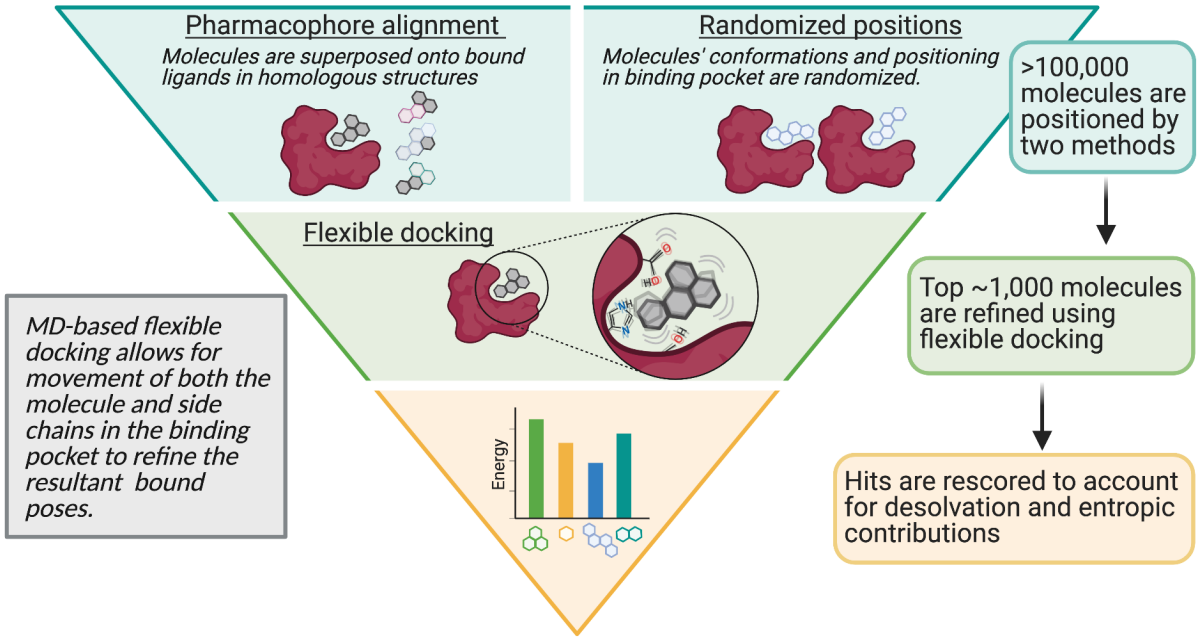


Figure 4.11. *The virtual docking workflow*. The virtual screening protocol involves multiple refinements, including using MD-based docking. Top ranked compounds were tested for *in vitro* inhibition.

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. Residues Gln-438 and Lys-342 in particular show the greatest conformational change upon ligand binding, suggesting that they participate in stabilizing bound compounds. The known inhibitors camostat, nafamostat, and gabexate all 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 positioning the guanidinium functionality to form a salt-bridge interaction with the active site Asp-435. As the mechanism of covalent inhibition involves His-296 deprotonating Ser-441, we performed a subsequent docking experiment using the three molecules with these charge changes (*Figure 4.12*). For all three molecules, the deprotonated serine positions itself into a more reactive state to attack the carbonyl carbon. 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.

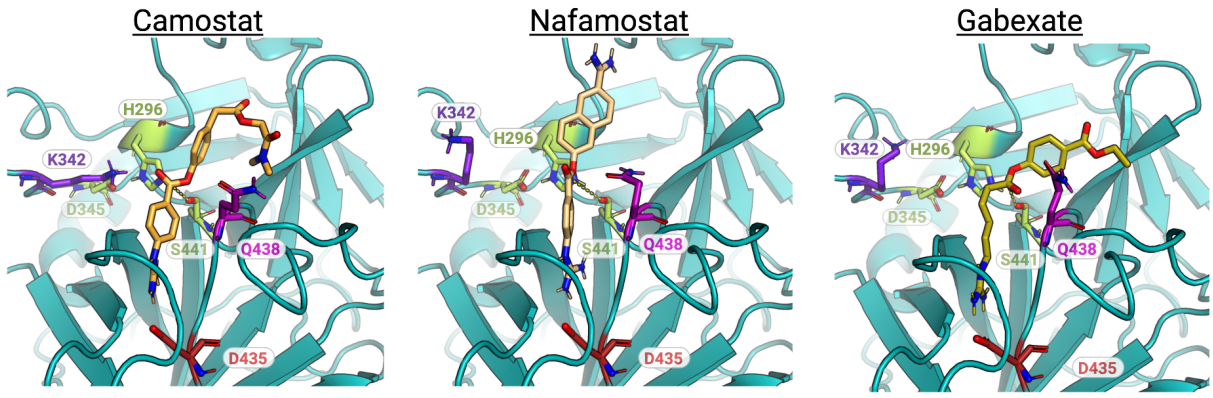


Figure 4.12. *Docked poses of known inhibitors*. Docked poses of camostat, nafamostat, and gabexate in the TMPRSS2 active site with the proton transfer from Ser-441 to His-296 that occurs after substrate binding. Shown in red is D435, which resides at the bottom of the S1 binding pocket, and lime green 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 performed by Yujin Wu.

Identification of noncovalent inhibitors

Several clinically approved drugs emerged as top ranked compounds in the virtual screen, which we selected to test in our *in vitro* assay. 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 4.13*). Biochemically, we found that all three molecules did in fact inhibit TMPRSS2 activity with varied potencies, with debrisoquine being the least potent (*Figure 4.13*). 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 (*Figure 4.13*). Pentamidine and propamidine are of similar size to camostat and nafamostat, typical of small molecule inhibitors (>350 MW), while debrisoquine is quite small, at 175.2 MW, classifying it as a fragment rather than a small molecule. However, debrisoquine has the greatest 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 4.14*). However, the fragment debrisoquine has modest selectivity for TMPRSS2 over trypsin (~3 fold). As observed with the covalent inhibitors (camostat,

nafamostat, and gabexate; *4.12*), the docking studies with the noncovalent compounds showed significant structural rearrangements for the two non-catalytic residues, Lys-342 and Gln-438 (*Figure 4.13*). 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.

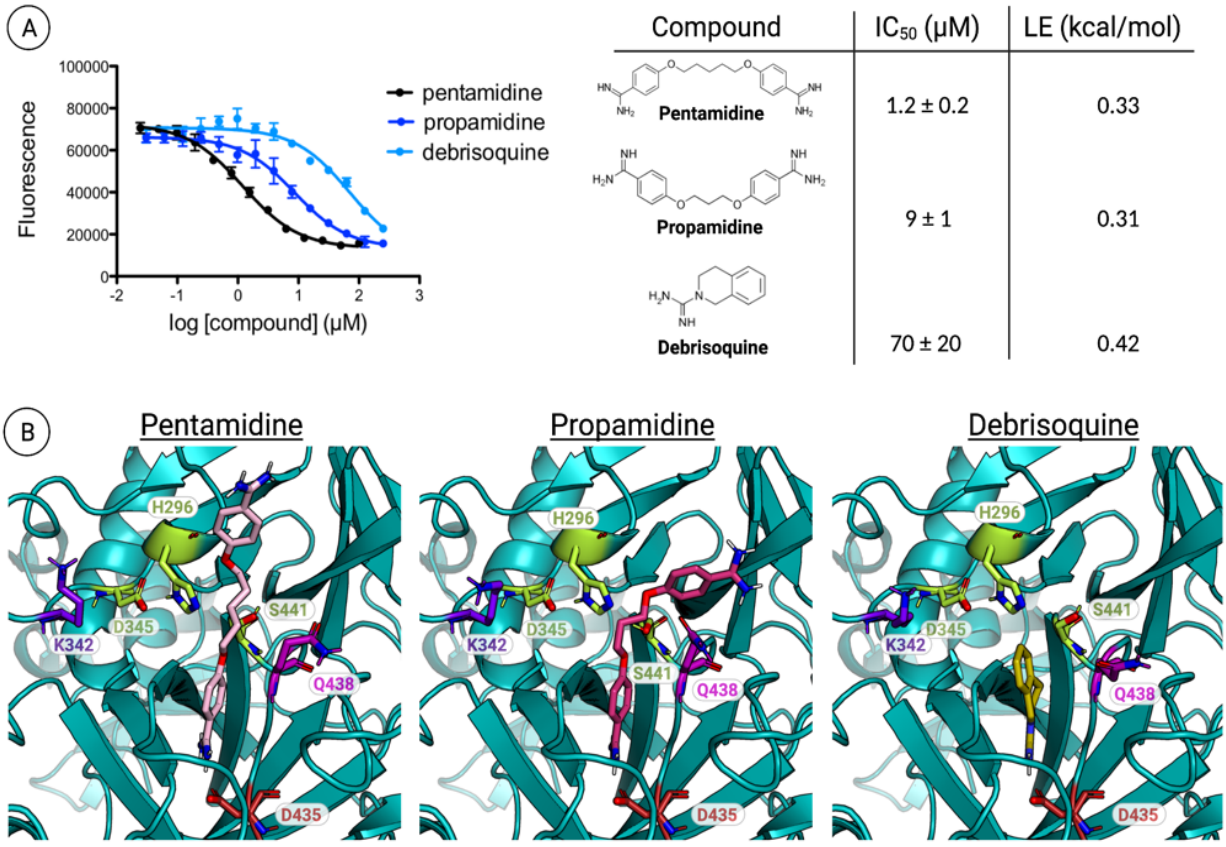
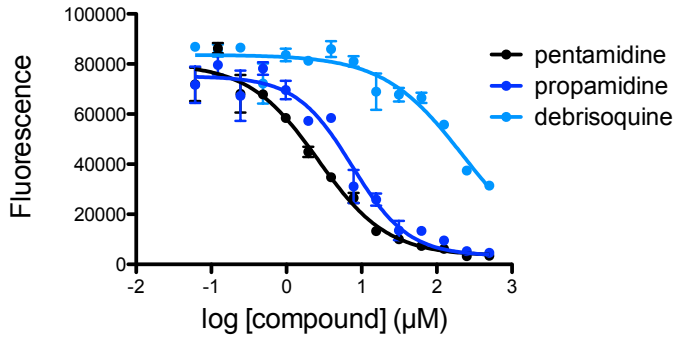


Figure 4.13. *Noncovalent hits from virtual screen.* Top ranked hits from virtual screen inhibit TMPRSS2 in biochemical activity assay. Left: Raw inhibition values used to obtain IC_{50} curves. Data is the average of duplicate experiments conducted in technical triplicate. Right: Table showing structures of hits, calculated IC_{50} values, and ligand efficiencies (LE). (B) 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 those residues. Binding experiments performed with Dr. Julie Garlick. Docking experiments performed by Yujin Wu.



	IC_{50} (μM)
debrisoquine	220 ± 30
pentamidine	2.6 ± 0.4
propamidine	7 ± 1

Figure 4.14. *Inhibition data with the three noncovalent compounds against trypsin.* Inhibition of Trypsin by computational hit compounds pentamidine, propamidine, and debrisoquine with 30 minutes preincubation of protein and inhibitor. Data points represent the average of technical triplicate with error bars indicating the standard deviation (left). Calculated IC_{50} values (right). Experiments performed with Dr. Julie Garlick.

4.3 Conclusion

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.[50,51] 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. In the present study, we demonstrate 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.

4.4 Materials and Methods

Mass spectrometric analysis of trypsin

Trypsin (Sigma, T9201), 10 μM, was incubated with compound, 10 μM, 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, 0.9% 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, 0.9% 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 µL) 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 $CaCl_2$, 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 µL 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 µL, at various concentrations) followed by addition of TMPRSS2 protein (8 µL, 0.5 nM final concentration). After 30 minute incubation (unless otherwise specified) at room temperature, substrate (2 µL, 2.5 µM 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 µM substrate.

General flexible docking setup

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 *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.[52,53] 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.[54] Open Babel was used to generate ligand random conformations.[55] MOE was used to predict the correct protonation state for the ligands at pH 7.4.[56] ParamChem was used to prepare the ligand topology and parameter files with the CGenFF force field.[57–59] 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.[48,60]

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.[45] 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 4.1.

Table 4.1. Soft-core potentials used in flexible receptor docking

name	E^*_{max}(vdw)	E^*_{max}(att)	E^*_{max}(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}(vdw), E_{max}(att)* and *E_{max}(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.[49]

4.5 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) Shulla, A.; Heald-Sargent, T.; Subramanya, G.; Zhao, J.; Perlman, S.; Gallagher, T. A Transmembrane Serine Protease Is Linked to the Severe Acute Respiratory Syndrome Coronavirus Receptor and Activates Virus Entry. *J. Virol.* **2011**, *85* (2), 873–882. https://doi.org/10.1128/JVI.02062-10.
(10) 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.
(11) 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.
(12) Shen, L. W.; Mao, H. J.; Wu, Y. L.; Tanaka, Y.; Zhang, W. TMPRSS2: A Potential Target for Treatment of Influenza Virus and Coronavirus Infections. *Biochimie* **2017**, *142*, 1–10. https://doi.org/10.1016/j.biochi.2017.07.016.
(13) 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.
(14) 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.
(15) 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.
(16) 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.
(17) 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.
(18) 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.
(19) Spraggon, G.; Hornsby, M.; Shipway, A.; Tully, D. C.; Bursulaya, B.; Danahay, H.; Harris, J. L.; Lesley, S. A. Active Site Conformational Changes of Prostasin 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.
(20) Sun, W.; Zhang, X.; Cummings, M. D.; Albarazanji, 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.
(21) 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.
(22) 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.
(23) 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.
(24) 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.
(25) 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.
(26) 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.
(27) Hedstrom, L. Serine Protease Mechanism and Specificity. *Chem. Rev.* **2002**, *102* (12), 4501–4524. https://doi.org/10.1021/cr000033x.
(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) 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.

(30) 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.

(31) 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.

(32) 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.

(33) 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.

(34) 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.

(35) 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.

(36) 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.

(37) 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.

(38) 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.

(39) 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.
(40) 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.
(41) 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.
(42) Lee, M. S.; Salsbury, F. R.; Brooks, C. L. Novel Generalized Born Methods. *J. Chem. Phys.* **2002**, *116* (24), 10606–10614. https://doi.org/10.1063/1.1480013.
(43) Lee, M. S.; Feig, M.; Salsbury, F. R.; Brooks, 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.
(44) Gong, X.; Chiricotto, M.; Liu, X.; Nordquist, E.; Feig, M.; Brooks, 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.
(45) Gagnon, J. K.; Law, S. M.; Brooks, 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.
(46) Ding, X.; Wu, Y.; Wang, Y.; Vilseck, J. Z.; Brooks, 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.
(47) Wu, G.; Robertson, D. H.; Brooks, 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.
(48) Brooks, B. R.; Brooks, 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.; Hodoscek, 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.
(49) 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.
(50) 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.

(51) 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.

(52) 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.

(53) 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.

(54) Feig, M.; Karanicolas, J.; Brooks, 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.

(55) 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.

(56) 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.

(57) 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.

(58) 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.

(59) 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.

(60) 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.

Chapter 5

Conclusions and Future Directions

5.1 Conclusions

One of the major challenges in drug design is protein target viability. The work presented in this dissertation aids in expanding the druggable proteome by establishing rules of molecular recognition in difficult targets, namely disordered/dynamic proteins. In Chapters 2 and 3, we utilize molecular dynamics simulations and biophysical *in vitro* experiments to interrogate dynamic protein-protein interactions (PPIs) involved in transcriptional regulation to establish ways in which native assembly occurs. We study two structurally distinct coactivators, CBP KIX (Chapter 2) and Med25 AcID (Chapter 3), and we find that rules of engagement with transcription factors can be generalized; ABDs utilize high conformational plasticity to engage with multiple partners, where a wide distribution of micro-states can be selected by each ligand respectively.

The utility of disorder in transcriptional regulation

Chapter 2 explores the ways in which KIX forms ternary complexes with two pairs of activator: c-Myb/MLL and pKID/MLL. Previous research on these proteins has established that while c-Myb and pKID bind to the same general area on KIX, the mechanisms by which they do this differs between the two.[1,2] Both c-Myb and pKID participate in allosteric communication through KIX, showing increased affinity for the protein when MLL binds first.[3] To interrogate the ways in which KIX utilizes conformational flexibility in complex assembly, we utilized molecular dynamics simulations.

Simulations of all systems (apo, binary, ternary) yielded atomic level resolution for binding interactions. In answering the question of how the distribution of KIX conformational states changes through activator binding, we utilized structural clustering and developed a method to project distribution changes in two dimensions. First, we utilized K-means clustering on all KIX complexes for a given system to determine the number of configurational micro-states as their relative populations.[4] In the apo distribution of states, we found that KIX occupies a wide distribution of states that are all relatively lowly populated. This is not surprising, given that the structure of apo KIX defies structural experimentation (i.e. crystallography or NMR) because of its inherent "floppiness."[1,2] Subsequent activator binding stabilizes KIX by winnowing down the number of accessible states. This, in turn, allows for allostery to occur, where the reweighted distribution of states increases or decreases the probability of the second binding event.

One finding of particular interest is that KIX can employ multiple mechanisms of molecular recognition. From our structural clustering methods, we found that when c-Myb, pKID, or MLL bind, each activator engages with a unique subset of KIX micro-states, reweighting relative populations in the distribution. However, unlike pKID or MLL, c-Myb also induces conformational changes into KIX. Thus, the benefits of flexibility inherent to the protein are two-fold. First, it allows multiple activators to bind because the apo distribution is so wide. And second, the structural plasticity of the protein allows for binding to induce conformational changes. We see similar effects with the Tethered inhibitor 1-10, which induces structural changes that disfavor pKID binding. Hence, we establish that dynamic domains like KIX could be "drugged" through allosteric small molecules that induce unfavorable structural changes.

Structurally different ABDs exhibit consistent mechanisms of activator engagement

While the ABD KIX has been studied relatively extensively, Med25 AcID has only more recently emerged as a domain of interest.[5–7] The structure of Med25 differs significantly from that of KIX, where instead having dynamic helices, it contains a stable β-barrel core.[5,6,8] While apo KIX has never been structurally elucidated via crystallography/NMR, Med25 AcID has only been solved in the apo state.[6,8] One of the

reasons for this is because of the “fuzziness” of interactions between Med25 and binding partners; activators including VP16 and ERM show preferential binding to one Med25 face over another, however, they do have propensity to bind weakly at opposing faces as well.[7,8] Further, when studying bound Med25 in complex with activators, extensive line broadening in NMR suggests that there are multiple bound conformational states.

In Chapter 3, we demonstrate how the binding interactions between activators and Med25 can be elucidated using molecular dynamics simulations. In particular, we opt to use enhanced sampling methods, namely temperature replica exchange, as PPIs could incur large energy barriers in binding that are difficult to overcome in a single unbiased MD simulation. Further, we utilize disulfide Tethering to localize the activators to pertinent areas on the protein, namely the H1 face. This was made easier since Med25 has native cysteines, and we demonstrated that only one of the three cysteines forms a disulfide bond with the peptides. Through simulations, we ultimately produce models of bound Med25 complexes, which aligns with chemical shift perturbations (CSPs) from NMR. In doing so, we find that even while Med25 is very structurally different from more typical ABDs, it engages with activators in the same way by utilizing the mobile regions on the protein to uniquely accommodate partners.

In addition to exploring cognate binding partners, we utilize high-throughput screening to identify the covalent small molecule inhibitor norstictic acid (NA). NA was found to covalently link to lysine side chain residues on the H2 face of Med25, yet the molecule can inhibit interactions with activators not only on the H2 face, but H1 face interactions as well. Using molecular dynamics simulations, we find that H1 face interactions may be inhibited through altered dynamics, as the mobile loops surrounding the H1 binding surface show decreased fluctuations. Through this work, we establish that drugging Med25 may be achieved by targeting dynamic substructures on the protein, which can inhibit both ortho- and allosterically, much like we saw with KIX.

Identification of TMPRSS2 inhibitors

The COVID-19 pandemic shifted my research focus towards identifying antiviral drugs with the goal of mitigating SARS-CoV-2 infection. The transmembrane serine protease TMPRSS2 has been implicated in coronaviral infections, including with SARS

and MERS, as well as other viral infections.[9–11] Thus, in Chapter 4, we utilize a multi-pronged strategy for TMPRSS2 inhibitor discovery, relying on *in silico* virtual screening alongside biochemical testing. We first curate a subset of small molecules that show propensity to bind in the TMPRSS2 active site through molecular docking by using a homology model for the protease domain. Biochemically, we optimize an expression and purification protocol that includes a delicate protein refolding of the peptidase domain. Subsequent kinetic experiments with hits from virtual screening show that current approved drugs have the potential to be repurposed for TMPRSS2 inhibition.

5.2 Future directions

Expanding our understanding of KIX allostery using additional activators

Our modeling generated for the KIX systems illustrates how three activators (c-Myb, pKID, and MLL) are able to differentially engage with the available distribution of KIX configurational states. In building a larger understanding of the ABD's role in activator engagement, other binding partners will be explored. Using molecular dynamics and the clustering protocols developed in Chapter 2, conformational selection by different activators including p65, FOXO3a, E2a, and HBZ will be determined (*Figure 5.1*). This will help to build our understanding the KIX conformational landscape, where we can identify common KIX micro-states that are important for engaging with multiple activators.

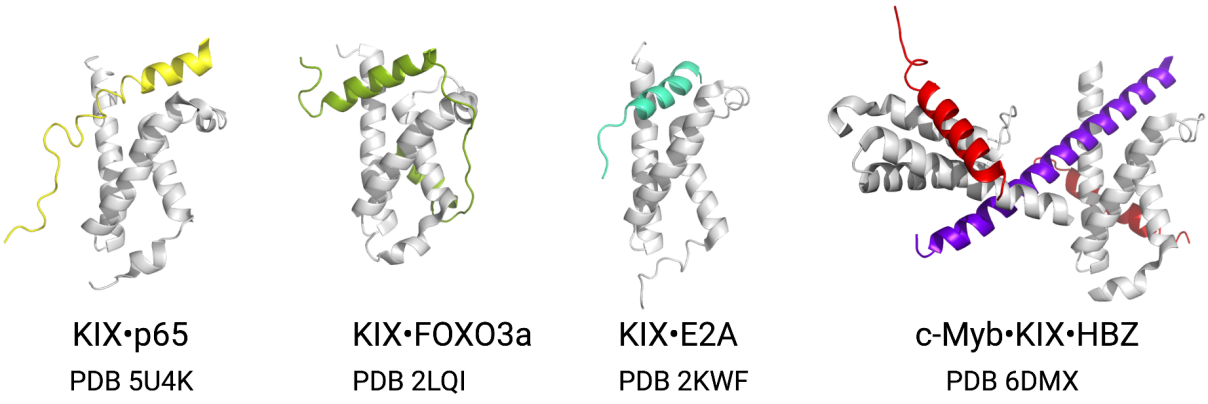


Figure 5.1. *CBP KIX binary and ternary complexes*. KIX is shown in light gray, and all of the activators are shown in color: p65 (yellow), FOXO3a (green), E2A (seafoam green), c-Myb (red), and HBZ (purple). PDB structures shown include: KIX•p65 (PDB 5U4K)[12], KIX•FOXO3a (PDB 2LQI)[13], KIX•E2A (PDB 2LQI), and c-Myb•KIX•HBZ (PDB 6DMX).[14]

Identification of Med25 orthosteric and allosteric inhibitors

With its role in various diseases, there is high interest in developing chemical probes for Med25. However, targeting dynamic PPIs is non-trivial. One molecule that arose from a disulfide Tethering screen was molecule 22, which preferentially binds to $Med25_{C506}$ (*Figure 5.2*)[15]. More interestingly, while 22 binds to the H1 face, enhances binding of activators at the H2 face. I performed explicit solvent all-atom molecular dynamics simulations on the bound complex to identify ways in which the molecule can elicit cooperative effects. Structures show that while the molecule can adopt a linear pose, it binds to Med25 predominantly in a bent conformational state, fitting between the bottom two loops on Med25. Conformational changes are observed on the H1 face, with a majority of the changes corresponding to the more dynamic regions on the protein (*Figure 5.2*).

With the identification of norstictic acid, which binds to the H2 face and allosterically inhibits, as well as molecule 22, which binds to the H1 face and allosterically enhances, we establish the importance of dynamic substructures on malleable proteins. Targeting mobile loops with small molecules allows for global conformational changes, redistributing the ensemble of accessible conformational micro-states. Thus, future research efforts can be directed towards high-throughput screens that contain reactive compounds (for example, disulfides) to look for molecules that form bound adducts. By increasing the number of both inhibitors and enhancer molecules for Med25, we can begin to better understand ways in which molecules can be strategically designed for enhancement/inhibition of particular Med25 interactions. This can be done using machine learning as well as rational design.

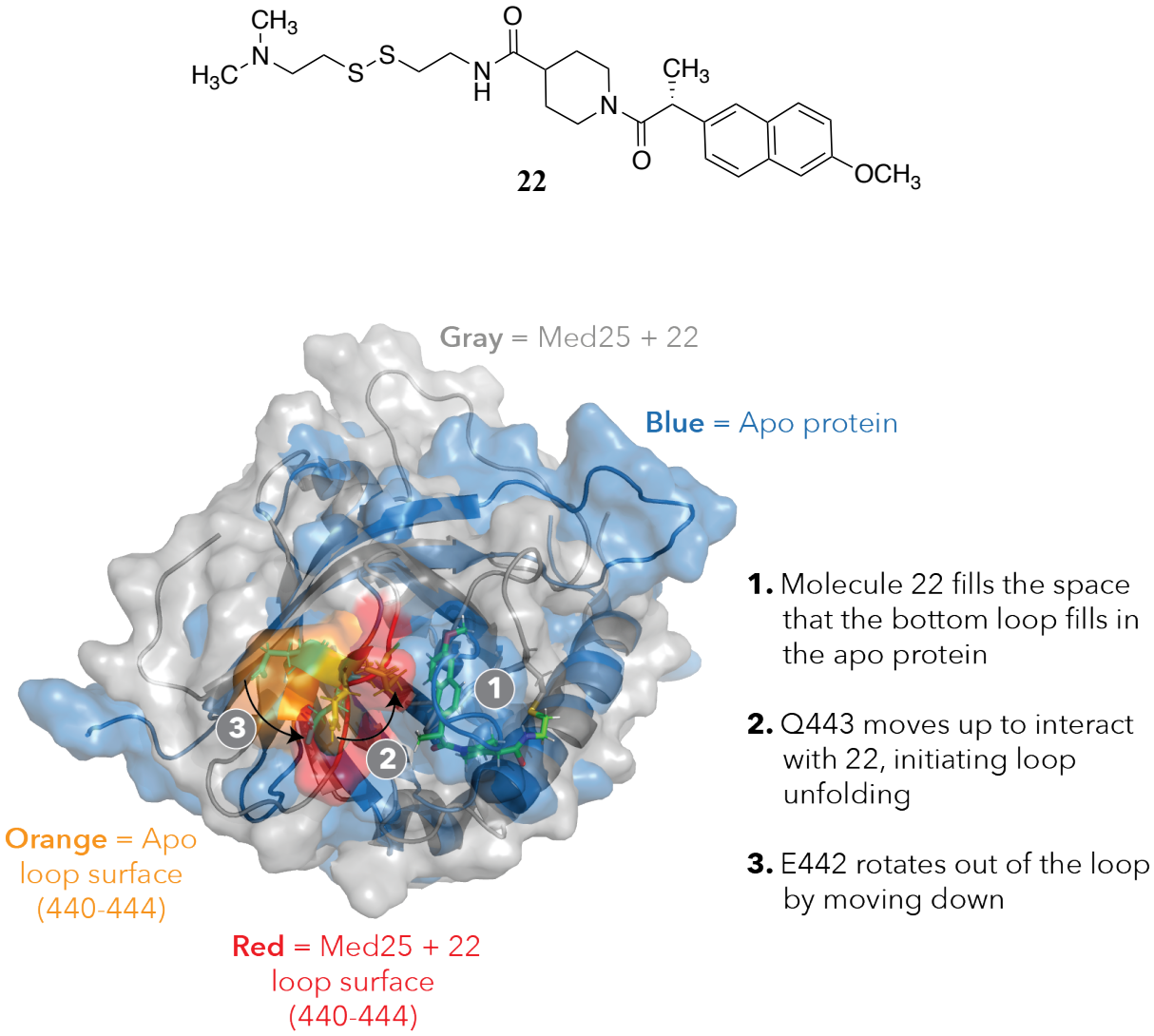


Figure 5.2. *Modeling with molecule 22 covalently bound to Med25$_{C506}$*. Molecule 22 (top) was covalently linked to Med25$_{C506}$ via a modified PATCH in CHARMM. (Bottom) Molecular dynamics simulations over the course of 100 ns show that the molecule preferentially adopts a bent shape and binds between the two bottom loops of Med25. In doing so, the bottom helix on the protein unwinds. Binding in this pose induces structural changes to the upper loop.

Derivatization of TMPRSS2 inhibitors

From the work done in Chapter 4, we identified small molecules that inhibit TMPRSS2 protease activity. Debrisoquine was the least potent inhibitor at 70 ± 20 μM, however it had the highest ligand efficiency (LE) of 0.42. Thus, the fragment provides a promising starting point for further derivatization to increase potency while maintaining key contacts formed between the molecule and the protease active site. The structural models generated from virtual screening are a crucial first step beginning to expand the

molecule. For example, we identified non-catalytic residues that show conformational flexibility between all bound molecules. In particular, residues K342 and Q438 showed large movements, interact with each bound molecule differently. Looking at K342, we see that when structurally aligning TMPRSS2 to other serine proteases, the residues in that loop are not conserved (*Figure 5.3*). Additionally, the size and conformation of the loop differs between proteases. This provides a direction for how to derivatize debrisoquine and build out towards this loop, where we can achieve high selectivity and potency for targeted inhibition.

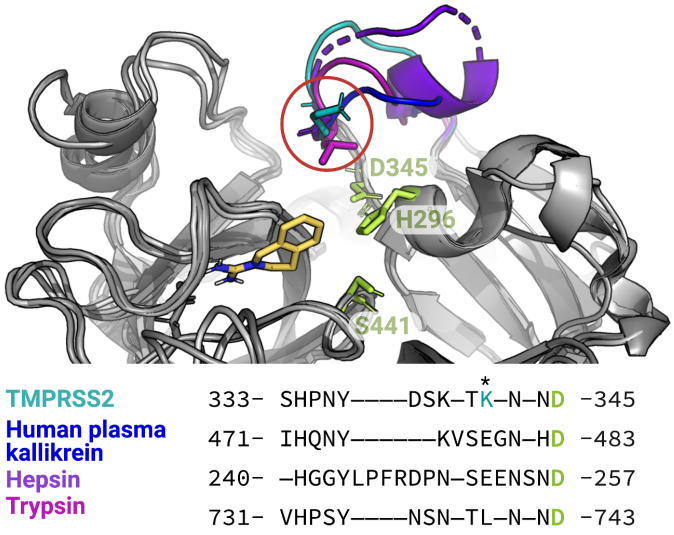


Figure 5.3. *TMPRSS2 K342 shows promise for derivatizing debrisoquine*. The proteases are all shown in gray. Debrisoquine (yellow) is docked into the active site of TMPRSS2. K342 in TMPRSS2, shown in teal, is not conserved across these other proteases.

5.3 References

(1) De Guzman, R. N.; Goto, N. K.; Dyson, H. J.; Wright, P. E. Structural Basis for Cooperative Transcription Factor Binding to the CBP Coactivator. *J. Mol. Biol.* **2006**, *355* (5), 1005–1013. https://doi.org/10.1016/j.jmb.2005.09.059.

(2) Brüschweiler, S.; Konrat, R.; Tollinger, M. Allosteric Communication in the KIX Domain Proceeds through Dynamic Repacking of the Hydrophobic Core. *ACS Chem. Biol.* **2013**, *8* (7), 1600–1610. https://doi.org/10.1021/cb4002188.

(3) 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.

(4) Karpen, M. E.; Tobias, D. J.; Brooks III, C. L. Statistical Clustering Techniques for the Analysis of Long Molecular Dynamics Trajectories: Analysis of 2.2-Ns Trajectories of YPGDV. *Biochemistry* **1993**, *32* (2), 412–420.
(5) 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.
(6) Eletsky, A.; Ruyechan, W. T.; Xiao, R.; Acton, T. B.; Montelione, G. T.; Szyperski, T. Solution NMR Structure of MED25(391–543) Comprising the Activator-Interacting Domain (ACID) of Human Mediator Subunit 25. *J. Struct. Funct. Genomics* **2011**, *12* (3), 159. https://doi.org/10.1007/s10969-011-9115-1.
(7) 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.
(8) Vojnic, E.; Mourão, A.; Seizl, M.; Simon, B.; Wenzeck, 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.
(9) 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**, S0092867420302294. https://doi.org/10.1016/j.cell.2020.02.052.
(10) Heurich, A.; Hofmann-Winkler, H.; Gierer, S.; Liepold, T.; Jahn, O.; Pöhlmann, S. TMPRSS2 and ADAM17 Cleave ACE2 Differentially and Only Proteolysis by TMPRSS2 Augments Entry Driven by the Severe Acute Respiratory Syndrome Coronavirus Spike Protein. *J. Virol.* **2014**, *88* (2), 1293–1307. https://doi.org/10.1128/JVI.02202-13.
(11) Iwata-Yoshikawa, N.; Okamura, T.; Shimizu, Y.; Hasegawa, H.; Takeda, M.; Nagata, N. TMPRSS2 Contributes to Virus Spread and Immunopathology in the Airways of Murine Models after Coronavirus Infection. *J. Virol.* **2019**, *93* (6). https://doi.org/10.1128/JVI.01815-18.
(12) Lecoq, L.; Raiola, L.; Chabot, P. R.; Cyr, N.; Arseneault, G.; Legault, P.; Omichinski, J. G. Structural Characterization of Interactions between Transactivation Domain 1 of the P65 Subunit of NF-KB and Transcription Regulatory Factors. *Nucleic Acids Res.* **2017**, *45* (9), 5564–5576. https://doi.org/10.1093/nar/gkx146.
(13) Wang, F.; Marshall, C. B.; Yamamoto, K.; Li, G.-Y.; Gasmi-Seabrook, G. M. C.; Okada, H.; Mak, T. W.; Ikura, M. Structures of KIX Domain of CBP in Complex with Two FOXO3a Transactivation Domains Reveal Promiscuity and Plasticity in Coactivator Recruitment. *Proc. Natl. Acad. Sci.* **2012**, *109* (16), 6078–6083. https://doi.org/10.1073/pnas.1119073109.
(14) Yang, K.; Stanfield, R. L.; Martinez-Yamout, M. A.; Dyson, H. J.; Wilson, I. A.; Wright, P. E. Structural Basis for Cooperative Regulation of KIX-Mediated

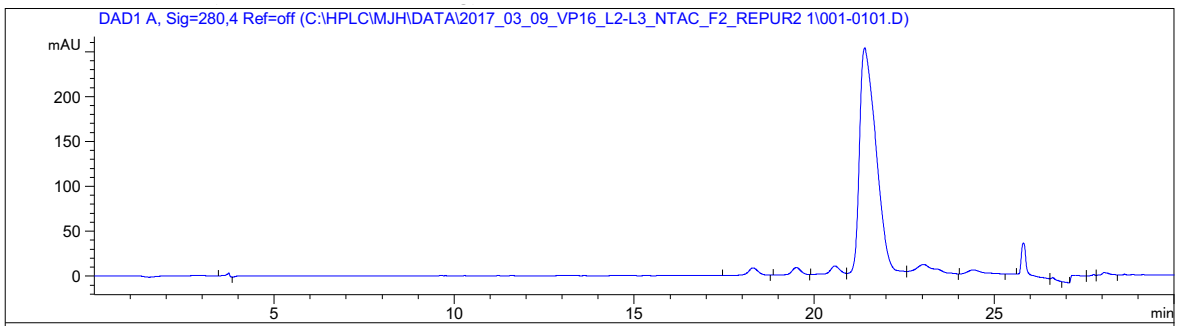
Transcription Pathways by the HTLV-1 HBZ Activation Domain. *Proc. Natl. Acad. Sci. U. S. A.* **2018**, *115* (40), 10040–10045. https://doi.org/10.1073/pnas.1810397115.

(15) 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 III, 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.

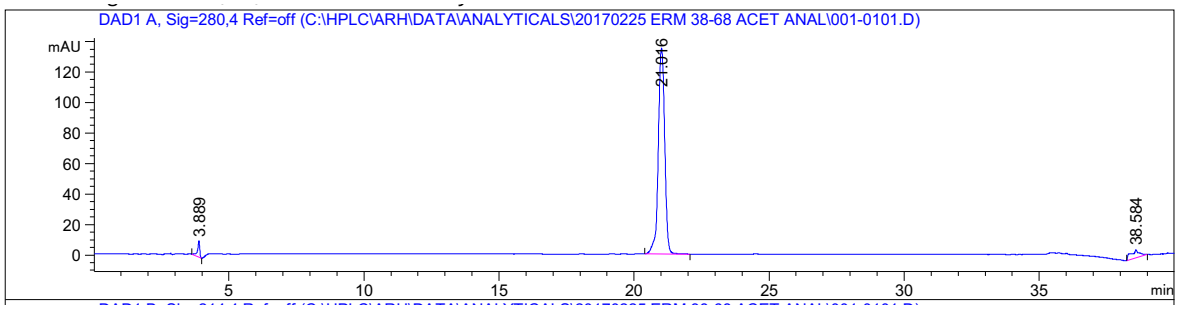
Appendix A

Characterization of Synthesized Peptides

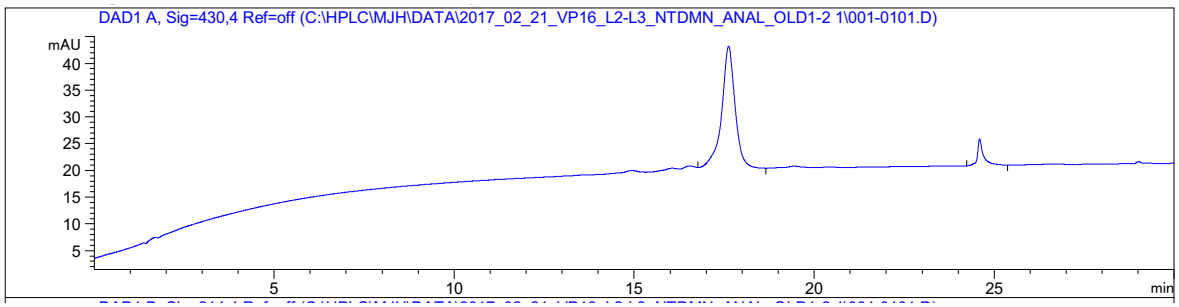
This appendix contains analytical HPLC chromatograms for the activator synthesized peptides used throughout this thesis.



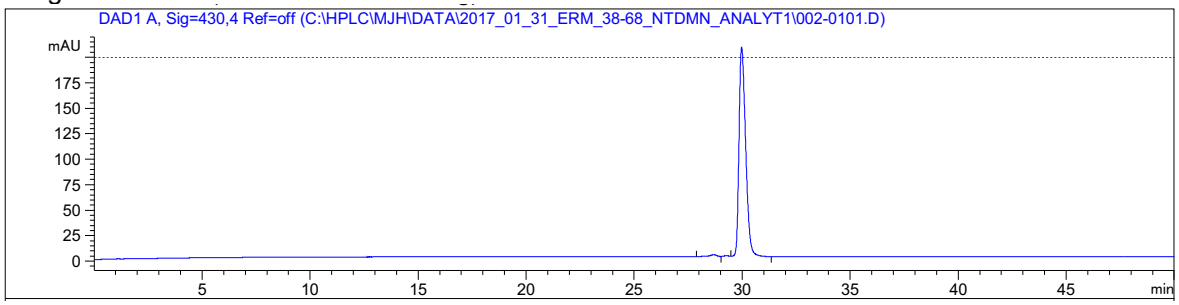
Analytical HPLC UV/Vis trace of **VP16 (438-490)**, monitored at 280 nm. Analytical sample was run in a 100 mM ammonium acetate(pH 7)/acetonitrile binary solvent system. The sample was injected with an isocratic flow of 82% 100 mM ammonium acetate and 18% acetonitrile. After 2 mins, the solvent gradient was increased from 18-28% ACN over 20 mins.



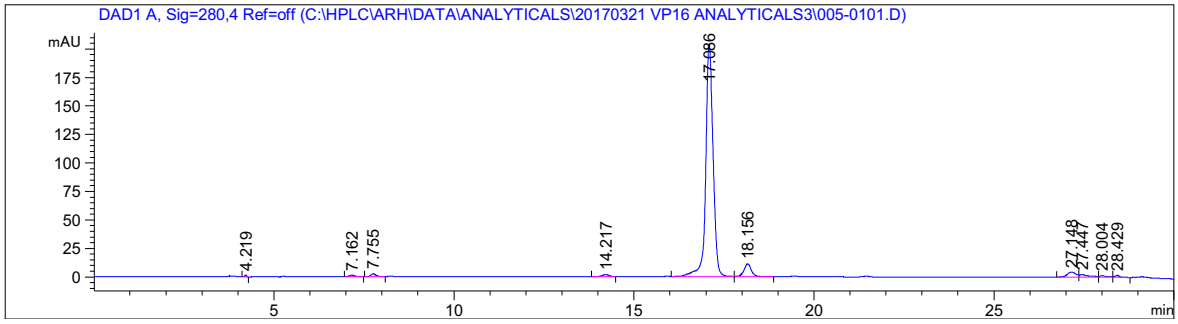
Analytical HPLC UV/Vis trace of **ERM (38-68)**, monitored at 280 nm. Analytical sample was run in a water (with 100 mM ammonium acetate)/acetonitrile system. The sample was injected with an isocratic flow of 85% water (with 100 mM ammonium acetate) and 15% acetonitriile. After 2 mins, the solvent gradient was increased from 15-30% acetonitrile over 20 mins.



Analytical HPLC UV/Vis trace of **DMN-VP16 (438-490)**, monitored at 430 nm. Analytical sample was run in a water (with 100 mM ammonium acetate)/ acetonitrile system. The sample was injected with an isocratic flow of 78% water (with 100 mM ammonium acetate) and 22% acetonitrile. After 2 mins, the solvent gradient was increased from 22-32% acetonitrile over 20 mins.

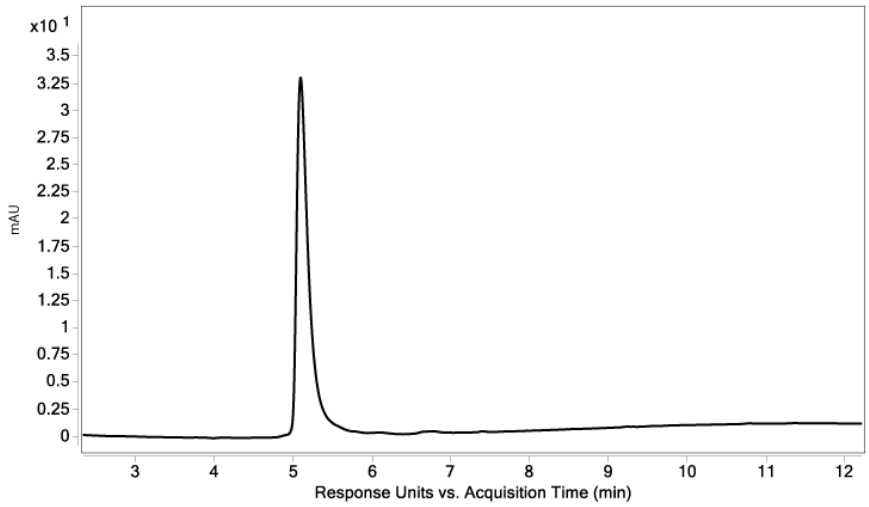


Analytical HPLC UV/Vis trace of **DMN-ERM (38-68)**, monitored at 430 nm. Analytical sample was run in a water (with 100 mM ammonium acetate)/ acetonitrile system. The sample was injected with an isocratic flow of 90% water (with 100 mM ammonium acetate) and 10% acetonitrile. After 2 mins, the solvent gradient was increased from 10-40% acetonitrile over 40 mins.

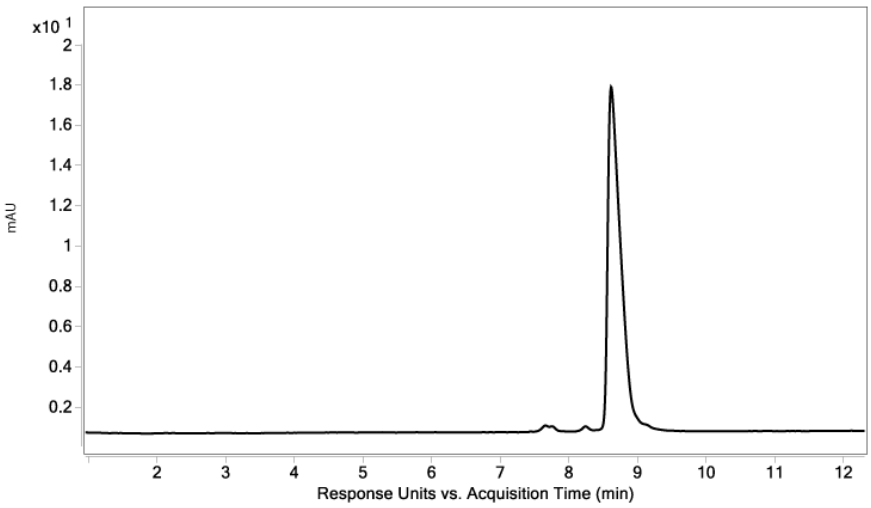


Analytical HPLC UV/Vis trace of **VP16 (438-454) G450C**, monitored at 280 nm. Analytical sample was run in a water (with 100 mM ammonium acetate)/ acetonitrile system. The sample was injected with an isocratic flow of 85% water (with 100 mM ammonium acetate) and 15% acetonitrile. After 2 mins, the solvent gradient was increased from 15-30% acetonitrile over 20 mins.

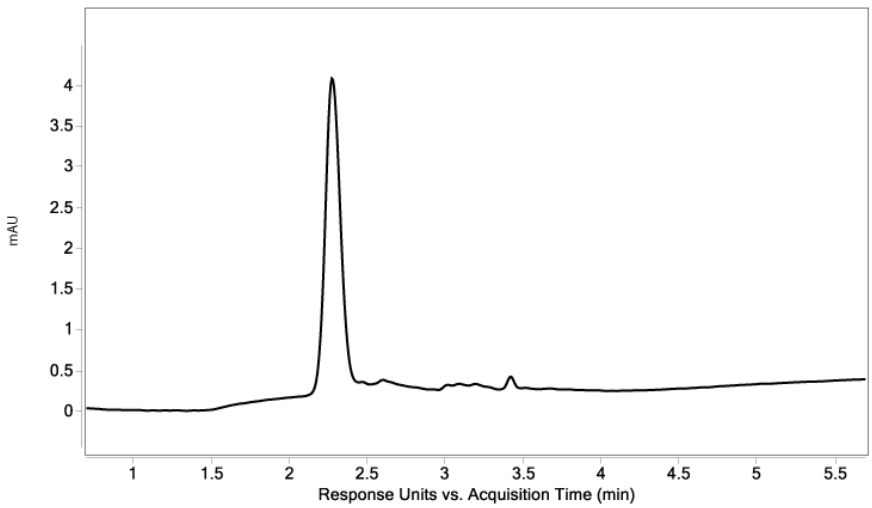
A)



B)



C)

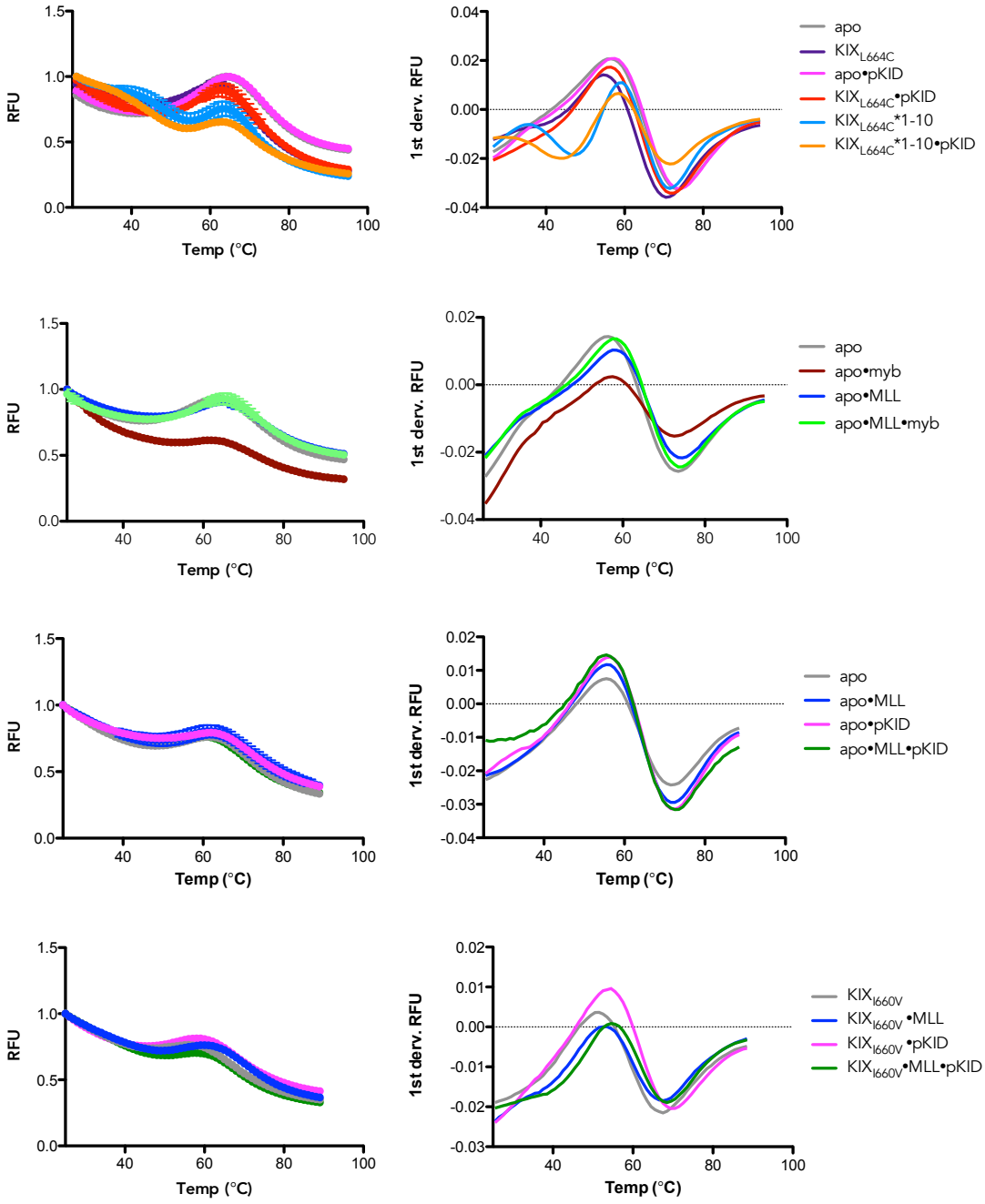


Analytical HPCL traces for KIX system peptides. Analytical traces at 280 nm of purified peptides for A) c-Myb, B) MLL, and C) pKID. Experiments performed by Dr. Stephen Joy.

Appendix B

DSF melting curves

This appendix contains melting curves for KIX complexes using SYPRO Orange dye.



DSF melting curves for KIX in various complexes. DSF melt curves (left) normalized to relative fluorescence units (RFU) and the first derivative of the curves (right). Experiments performed by Dr. Julie Garlick.