UCSF
UC San Francisco Electronic Theses and Dissertations

Title
Regulation of Transcriptional Machinery and Its Relationship to Disease

Permalink
https://escholarship.org/uc/item/5h0616jc

Author
Bruning, John Michael

Publication Date
2017

Peer reviewed|Thesis/dissertation

eScholarship.org Powered by the California Digital Library
University of California

Regulation of Transcriptional Machinery and Its Relationship to Disease

by

JOHN BRUNING

DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Pharmaceutical Sciences and Pharmacogenomics

in the

GRADUATE DIVISION

of the

UNIVERSITY OF CALIFORNIA, SAN FRANCISCO

[redacted]

Dedicated to My Dearest and Sweetest Wheezy

and to

My Fallen Heroes Who Have Passed During My Graduate Career in:

Joost Van Der Westhuizen (Motor Neuron Disease / ALS), SA #9

Jonah Lomu (Kidney Disease), NZ #11

Jerry Collins (Car Accident), NZ #6

Dedication and Acknowledgements

I think it is appropriate to begin by thanking my parents for providing me with the seemingly endless opportunities that allowed me to study here at UCSF. My predisposition for research science began at an early age when my mom would allow me to concoct different 'potions' using ingredients from the spice rack. I distinctly remember synthesizing volatile brown slurries that smelled mostly of star anise. She insisted that I used to taste my 'potions', but I never recall sampling any of them. While I think a lot of curiosity -- in particular my scientific curiosity -- is rooted in my mom, I feel that I inherited my reason and logic from my dad, the attorney.

My entire family, including my sister, have a somewhat insane work ethic. I grew up in 'the house that never sleeps'; my mom and sister often worked late (until 2am) and my dad often woke up early for work (~4:30am). I did my best to adhere to a normal circadian rhythm, but it was needless to say quite challenging. To my family -- my mom and dad, sister, and my late grandmother -- thank you for all of your support, always.

I am also grateful for my best friends from back home who often kept me in harms way. I think the lifetime of trouble they helped me accrue in high school alone allowed me to succeed in my undergraduate experience, as I had largely by then 'been there, done that.' I am especially thankful for my time with my late best friend Greg Macias -- Greg, you would have been a great surgeon and I would let you operate on me any day.

I have had many scientific mentors over the years. At UCSC, I worked in the Singaram lab with a graduate student at the time, Jinsoo Kim. I could not have asked for a more patient and

excellent teacher than Jinsoo. As a research assistant at UCSC, I worked under Jeremy Sanford. Our long days in the lab extended onto the rugby pitch as coaches together. Having the opportunity to tackle your boss without repercussions is a truly rare opportunity that I relished and doubt I will revisit.

As for my time at UCSF, I was mentored by fantastic graduate students -- in particular, Kris Kuchenbecker, Laura Lavery, Jason Fernandes, and David Booth. This list also coincides with my 'most respected biochemist/biologist' list. I am grateful for all of the members of the Fletterick and Jacobson Lab, especially CK, for their help over all of these many years.

A special extended acknowledgment to David Booth -- I love you bestie. You are the best biochemist I know but you really don't need to use an ammonium sulfate cut in every protein purification…

Thank you to Dianna Kroetz and Frank Szoka for giving me the opportunity to interview here even though I missed the formal interviews. It's a long story, but thank you. And thank you to Rebecca Brown and Debbie Acoba-Idlebi for the hard job of keeping me on track with the various administrative responsibilities that I always forgot to perform.

I was lucky enough to work very closely with some great faculty at UCSF -- namely, my advisers (Robert and Matt) a close collaborator (Pam England) and my 3rd Thesis Committee Adviser Jaime Fraser. I learned massive amounts about research science from this diverse set of thinkers. I was also lucky enough to work somewhat closely with James Holton, the beamline

scientist at ALS 8.3.1. He is, in my opinion, the pinnacle of a patient teacher and brilliant scientist and lover of Mountain Dew.

I am also immensely grateful to the members of the Agard lab. I can, with certainty say, that any successes I have had can somehow be directly attributed to them. Whether it was experimental design, access to instrumentation, or general support in the form of > 5pm Coors Lights, the members of the Agard lab were always there to support me. My best fortune in graduate school without a doubt was being embedded into the MSG next to the Agard lab.

Finally, thank you to my dearest Louisey. I am growing increasingly concerned that you will never know how I feel about you, for at times like these when I reflect on our time together, I have simply no words to describe how much I love and appreciate you. Thank you for always listening to the trials and tribulations of my many bad days in graduate school -- you have always been the unwavering support that I needed to make it through these often challenging and confusing times. I love you to the moon and back and back and back and back...

Regulation of Transcriptional Machinery and Its Relationship to Disease

John Bruning

Abstract

Transcription is the process in which our genetic information in the form of DNA is first utilized to create the machinery, structural integrity, and regulatory components that allows for cells to simply survive. While there are many, many, many regulatory modes and mechanisms that govern processes from enzyme catalysis to cell differentiation (and everything in between), the regulation of transcription is the initial decision to 'put the players on the pitch.' Transcriptional regulation is immensely complex, as it rightfully should be, as it is required for both simple procedures, for example simply increasing the mRNA copy number of a single species in response to a cellular cue, to intricately coordinated events such as decisions regarding cell fate. Transcription factors are the protein components that interplay between environmental cues and gene expression. In particular, nuclear receptors are ligand dependent transcription factors that are modulated by a small molecule ligand which provides information about a particular metabolic pathway. The focus of my research has been a biochemical and structural understanding of how this subset of transcriptional machinery is implicated in the etiology of disease.

Mutations in the cardiac transcription factor Gata4 results in heart abnormalities in the form of septal defects at birth. While the genotype-phenotype relationship was clear for this mutant, the structural and biochemical underpinnings of how genetic aberration was coupled to this birth defect was not entirely unclear. While ongoing efforts from the Srivastava lab and others demonstrated that this mutation led to ineffective communication between Gata4 and another

cardiac transcription factor Tbx5. To bridge this gap, I dedicated several years to the biochemical reconstitution of ternary complexes between Gata4 and Tbx5 on various DNA elements.

In addition to this early thesis project, I also collaborated with the HARC (HIV Accessory Related Complexes) center at UCSF and Berkeley to better understand how and when an HIV protein, TAT, commandeered host transcriptional machinery to coordinate the transcription of its own viral genome. Working closely with a structural lab at Berkeley, I provided computational support in the form of *in silico* simulations of this HIV-host transcriptional complex decorated with various post-translational modifications and binding partners. From these simulation studies I predicted how a single mutation (M62V) in a constituent called AFF4 could enhance complex stability. Additionally, I was able to use a information theory metric in the Kullback-Leibler Divergence to characterize how post-translational modifications on both the host and viral component were dynamically coupled to functional sites in the complex that were integral to viral virulence.

Finally, the heart and soul of my thesis work biochemically and structurally characterizing a novel post-translational modification in the nuclear receptor Nurr1. Nurr1 is a master regulator of the dopaminergic neuron phenotpye, and consequently, small molecule agonism of its function could in theory help prevent the loss of dopaminergic neurons as seen in Parkinson's disease. We have identified a small molecule metabolite that regulates Nurr1 function via formation of a cysteine adduct. Quantifying and characterizing this post-translational modification proved unbelievably challenging to gain biochemical traction on for various reasons. We were ultimately able to solve a 3.2 Å crystal structure of Nurr1 bound to this

dopamine metabolite. This work suggests that Nurr1 acts as a sensor of dopaminergic neuronal stasis by receiving input from a molecule that is linked simultaneously to redox stasis and neuronal health, as well as dopamine concentrations. This regulatory mechanism is exactly what you want and expect, in my opinion, for this exquisitely sensitive system -- the fact that disparate readings from this system converge onto a single molecule in DHI is to regulate cellular fate is yet another example of nature's elegant logic *par excellence.*

Table of Contents

Chapter 1: 1

A Synopsis of My Doctoral Research

Chapter 2: 15

A Novel Redox Sensitive Post-Translational Modification Regulates the Function of the Atypical Nuclear Receptor Nurr1

References 73

Chapter 3: 78

Post-Translational Modifications in the HIV Semi-Super Elongation Complex are Dynamically Coupled to Functional Sites Necessary for Virulence

References 120

Chapter 4: 122

Transcriptional Logic In Cardiac Reprogramming

References 146

Appendix: 149

Bioinformatic Identification of Tandem Response Elements

List of Tables

Table 2.1: 63

Data Collection and Refinement Statistics after 8 Rounds of Refinement

Table 4.1: 145

Mutations of Interest in Human Gata4

List of Figures

Chapter 2

Figure 2.1: 46

DHI but Not DHICA Stabilizes the Nurr1 LBD

Figure 2.2: 47

Kinetic Characterization of a DHI-Nurr1 Interaction by SPR

Figure 2.3: 51

Structural Alignment of the apo Nurr1 LBD and DHI-bound LBD

Figure 2.4: 57

Effect of Cysteine Mutations on Nurr1-DHI Interaction

Figure 2.5: 58

Cell Based Assays Reporter Assays Suggest DHI Enhances Transcriptional Activity

Figure 2.6: 59

Target Gene Transcription in Zebrafish

Figure 2.7: 60

In vivo Imaging of Neuronal Fluorescence Upon DHI Administration

Supp. Figure 2.1: 48

Redox Dependent Kinetic Effects on the DHI-Nurr1 Interaction

Supp. Figure 2.2: 49

Temperature Dependent Kinetic Effects on the DHI-Nurr1 Interaction

Supp. Figure 2.3: 50

LC-MS Characterization of a DHI Adduct

Supp. Figure 2.4: 52

Potential energy surface of thiolate attack for DHI (catechol) at each 4 different linkage sites

Supp. Figure 2.5: 53

Potential energy surface of thiolate attack for IQ (quinone) at each 4 different linkage sites

Supp. Figure 2.6: 54

Side chain prediction for both IQ and DHI at each of the proposed linkage sites.

Supp. Figure 2.7: 55

Side chain prediction of IQ at linkage site C2.

Supp. Figure 2.8: 56

Superposition of IQ linked at all 3 linkage sites (C2,C4, and C7)

Supp. Figure 2.9: 61

Neuronal Pattering and Position Measurement Upon Administration of DHI

Supp. Figure 2.10: 62

Nurr1:RXR Heterodimer Models

Chapter 3

Figure 3.1: 98

Functional Sites Critical to TAT-mediated HIV Transactivation Exhibit Sample A Different Conformational Landscape Upon Acetylation or Phosphorylation

Figure 3.2 : 99

H-Bonding of K28 with Nearby H-Bond Acceptors

Figure 3.3 : 100

TRM Conformational Sampling Upon TAT Acetylation

Figure 3.4 : 100

Acetylation of Lys28 Is Coupled to the CDK9 Substrate Binding Pocket

Figure 3.5 : 101

Phosphorylation of Ser175 Alters the Dynamics of Nearby Residues

Figure 3.6 : 101

Poisson-Boltzmann Calculations

Figure 3.7 : 109

Conformational Analysis for R254 Upon TAT Acetylation.

Figure 3.8 : 111

Conformational Analysis for I255 (~249) Upon TAT Acetylation.

Figure 3.9 : 113

Conformational Analysis for W258 (~252) Upon TAT Acetylation.

Figure 3.10 : 115

Conformational Analysis for E53 (CDK9) Upon TAT Acetylation.

Supp. Figure 3.1a : 102

Sum KL Divergence Distribution for Each System

Supp. Figure 3.1b : 102

Max KL Divergence Distribution for Each System

Supp. Figure 3.2 : 103

Average Structures

Supp. Figure 3.3 : 104

TRM Sampling Upon PTM Introduction

Supp. Figure 3.4 : 104

Acetylation Lys28 Is Coupled to the CDK9 Substrate Binding

Supp. Figure 3.5 : 105

Phosphorylation of Ser175 of CDK9 Alters Local Environment

Supp. Figure 3.6 : 106

Backbone atom RMSD (time-dependent) of the semi-SEC complex

Supp. Figure 3.7 : 106

Distribution of backbone atom RMSD of the entire semi-SEC complex

Supp. Figure 3.8 : 107

Backbone atom RMSD of important functional sites (frequency)

Supp. Figure 3.9 : 108

Backbone atom RMSD of important functional sites (average)

Supp. Figure 3.10 : 110

Conformational Analysis for R254 Upon TAT Phosporylation

Supp. Figure 3.11 : 112

Conformational Analysis for I255 (~249) Upon CDK9 Phosphorylation

Supp. Figure 3.12 : 114

Conformational Analysis for W258 (~252) Upon CDK9 Phosphorylation

Supp. Figure 3.13 : 116

Conformational Analysis for E53 (CDK9) Upon CDK9 Phosphorylation

Supp. Figure 3.14 : 117

Conformational Analysis for K143 (CDK9) Upon CDK9 Phosphorylation

Supp. Figure 3.15 : 118

Conformational Analysis for K178 (CDK9) Upon CDK9 Phosphorylation

Supp. Figure 3.16 : 119

Conformational Analysis for K12 (TAT) Upon CDK9 Phosphorylation

Chapter 4

Figure 4.1 : 143

Arrangements of tandem TRE's observed in CHIP-seq results

Figure 4.2 : 143

Tandem RE's drive the formation of unique binding surfaces

Figure 4.3 : 144

Examples of Gata4:Tbx5 gene expression dependency

Figure 4.4 : 144

Sequences of 10 test TRE's.

Figure 4.5 : 144

Analysis of ternary complex assembly.

Figure 4.6 : 145

Structural View of G296S Mutation of Gata4

Chapter 1

A Synopsis of My Doctoral Research

I have dedicated my time at UCSF, almost exclusively, to the understanding of the mechanistic underpinnings of various diseases. As a rotation student in the Frankel Lab, I attempted to elucidate interactions between the HIV RNA binding protein Rev with host RNA transcripts. The hypothesis was that Rev, in addition to interacting with its cognate viral RNA the Rev Response Element, could additionally modulate pathogenesis by interacting with host RNA transcripts to either encourage or discourage their (a) translation into protein or (b) their regulatory functions as RNA. In another rotation in the McKerrow lab, I found myself working to purify a metalloprotease responsible for virulence in schistosomiasis. This protein prep required the manipulation of live, infectious cercarie of *S. mansoni* which was slightly intimidating as this particular life cycle stage can penetrate human skin using a cocktail of proteases, include the metalloprotease I was trying to purify.

After joining the Fletterick and Jacobson lab, my research focus narrowed to regulatory mechanisms of transcriptional machinery as they pertain to disease. This dissertation will dedicate a chapter to each of my primary projects I initiated after becoming a Ph D candidate. They will be presented in reverse chronological order, which is somewhat appropriately the same order I consider them to be in on a scale of more successful to least successful. In the past years, I have dedicated my focus to understanding how a transcription factor intimately involved with dopaminergic neuron health, and consequently Parkinson's disease, is regulated. Nurr1 is a transcription factor that falls into a unique class of ligand dependent transcriptional regulators called nuclear receptors. The Fletterick lab has studied this particular family of protein from a structural and biochemical lens for many, many years, so it is somewhat appropriate that I adopted this project (although quite late in my graduate career). The project

was ironically initially conceived not by Robert, but by Matt. In group meeting, Matt presented a few slides about this atypical nuclear receptor Nurr1 and a brief introduction into its relationship with Parkinson's disease. He outlined that, unlike other nuclear receptors, Nurr1 deviates from the canonical regulatory mechanisms -- namely, it was believed at the time to be ligand-independent independent based on the fact that it adopted a constitutively folded conformation typically seen in ligand-bound nuclear receptors. Additionally, it lacked a co-regulator binding site that typically allows for input from varying signalling pathways. Because Nurr1 is a master regulator of the dopaminergic neuron phenotype, and Parkinson's disease is pathologically characterized as the loss of this particular type of neuron, it followed somewhat logically that a Nurr1 agonist could serve as a potential therapeutic avenue for Parkinson's disease. Matt then proposed that, despite Nurr1 lacking a cavity that could accept a larger, hydrophobic molecule (for example estrogen in the case of the estrogen receptor), it could potentially accommodate a small ligand if some means of enhanced side chain sampling/ induced fit sampling was allowed for a potential ligand.

I was immediately captivated, not so much by the idea of developing Nurr1 modulators, but rather trying to understand how a protein absolutely critical to cell survival could seemingly have little to no regulatory mechanisms. At UCSC I was taught my 3rd Quarter of Biochemistry (Metabolism) by Robert Ludwig. Robert had a booming voice and looked like the Santa Cruz version (read Tommy Bahama shirt) version of the Monopoly man. In this class, he always, always emphasized the concept of regulation, often simply by screaming the word "REGULATION!" I could almost hear his voice when reading the Wang (2003) paper that solved the initial structure of Nurr1 in the apo, but conformationally active form. The authors

went on to explain how Nurr1 was a pariah of sorts, in that it had been stripped of most of its regulatory mechanisms seen in nuclear receptors. No ligand binding pocket. No coactivator binding site. Fully active conformation. The ultimate conclusion was that Nurr1 was ligand-independent and received minimal input from biological macromolecules (except of course it's heterodimer partner RXR and of course its cognate DNA response element). Although more suggestions of regulatory mechanisms were published over the years, I feel that Matt and I both had the suspicion that Nurr1 could, in fact, be regulated by a metabolite. We suspected that this metabolite should, in fact, be small and somehow be able to be utilized as a sensor for the dopamine-stasis of the cell.

With that in mind, Matt suggested an oxidative by-product of dopamine in 5,6, dihydroxyindole (DHI). This molecule has proven to be the bane of my biochemical existence for a number of years now as I have explored the relationship between this metabolite and Nurr1. DHI is somewhat of a nightmare to work with biochemically:

- it can polymerize to form an insoluble polymer precursor to neuromelanin
- it spectroscopically active across a broad range of wavelengths (280-700nm)
- it oxidizes to an incredibly reactive electrophile in the form of 5,6, indolequinone

With these in mind, I would like to think of DHI as somewhat of a blessing as a student of biochemistry, for, in seemingly every possible assay or technique, there was opportunity for some nasty, insidious artifact. The aggregate work of Ito and D'Ischia over many decades on the properties of DHI have gained my utmost respect and appreciation.

In addition to my training as an experimentalist, the development of my computational skills has been arguably my most enjoyable component of my graduate training. Prior to graduate school, I had absolutely no experience with a Unix/Linux system at the level of a terminal interface. My first day as a rotation student in the Jacobson Lab, a senior scientist CK asked me what sort of expertise I had with shell scripting. My response of "Sorry, but what is shell scripting" I think provided him with all of the information he needed to know. However, with the help of CK and other members of the Jacobson Lab and Agard Lab, I now what "shell scripting" is and I can proudly say that I have written some rather challenging scripts to facilitate processing of molecular dynamics simulations, x-ray crystallography data, and day-to-day "this would only take my 2 minutes to do this manually but I am going to spend half a day writing some code in case I ever have to do this 1E9 times" tasks. I initially joined both the Fletterick lab and Jacobson lab so I could develop my skills as structural biologist as well as a computational chemist. A simple metric that I 'achieved' this goal is my recent offer (and acceptance) to the Stanford Chem-H Medicinal Chemistry group and Structural group; I was offered the position based upon my experience expressing, purifying, and crystallizing protein in conjunction with my familiarity and 'expertise' with the Schrodinger Computational Chemistry suite.

A more concise and brief summary of my research experiences at UCSF are as follows:

Institution: University of California, San Francisco

Advisor : Matt Jacobson, Ph.D

Dates: Fall Rotation, 2010

Project Title: Characterizing the function of an unknown operon in *Agrobacterium tumafaciens*

My fall rotation project in the Jacobson lab was to use homology modeling and docking to discover the function of an unknown operon in *Agrobacterium tumafaciens*. While this operon had a few known structures and functions of its components, many members neither had any solved structure or clear function. For these, I constructed homology models. With a structural representation for each of the members in the operon, I screened each structure against the Kyoto Encyclopedia of Genes and Genomes metabolite library. After analyzing these results, it became clear that this operon was used to import and metabolize polysaccharides. I worked closely under a Senior Scientist in the lab, Chakrapani Kalyanaraman, who has extensive experience in computational approaches to orphan enzyme function identification.

Institution: University of California, San Francisco

Advisor : Alan Frankel, Ph.D

Dates: Winter Rotation, 2011

Project Title: Characterizing host-viral interaction between HIV Rev and host RNA's

Rev is a protein in HIV responsible for viral RNA transport across the nucleus. While it's role in the viral life cycle is becoming more and more clear, it is still not well understood if and how Rev interacts with host RNA's. In order to elucidate any of these viral-host interactions, I utilized the CLIP assay which I had learned in the Sanford lab at the University of California, Santa Cruz. In this rotation, I gained further experience with Western blotting, cell culture, and the CLIP assay. Very preliminary results suggested that Rev was in fact interacting with human RNA's.

Institution: University of California, San Francisco

Advisor : James McKerrow, Ph.D

Dates: Spring Rotation, 2011

Project Title: Native purification of the metalloprotease M8 from *Schistosoma mansoni* cercarie

Schistosomiasis is a disease caused by an infection from the fresh water dwelling parasite *Schistosoma mansoni.* While there is currently an effective treatment to kill the parasite once a patient is infected, vaccination would be the best route to decreasing multiple infections from this parasite. Under the guidance of UCSF parasitologist Judy Sakanari Ph.D, I focused on purifying the metalloprotease M8. This protease is partly responsible for *S. mansoni's* ability to penetrate directly through undamaged skin. While there were other efforts in the lab to recombinantly purify this protein, my goal was to purify natively from infectious *S. mansoni* cercarie. I was able to purify this protein natively, but in quantities unsuitable for biochemistry or crystallography.

Institution: University of California, San Francisco

Advisors : Robert Fletterick, Ph. D; Leor Weinberger Ph. D

Dates: Fall 2012 - present

Project Title: Structural elucidation of the higher order IE2-CRS protein:DNA complex

Recently, I have been mentoring a research technician, Brian Linhares, from the Weinberger group at the Gladstone Institute at UCSF. Brian and I are working to assemble a higher order complex consisting of a protein from cytomegalovirus called IE2 and its cognate DNA binding site, the cis-repressive sequence (CRS). I have introduced Brian to a number of basic biochemical techniques necessary to macromolecular assembly. While our goal is to characterize this complex by cryo-EM and crystallography, we have recently characterized a solution structure of this complex by small-angle X-ray scattering (SAXS). The scattering data reveals that ~ 7 IE2 monomers (fused to maltose binding protein) oligomerize on the CRS. I will dedicate time to this project in the form of research and mentoring for about 2 more years.

Dissertation Research

Institution: University of California, San Francisco

Advisor : Robert Fletterick, Ph. D; Matt Jacobson, Ph. D

Dates: Summer 2011 - present

Project Title: The Role of Gata4 & Tbx5 Tandem Response Elements in Cardiac Reprogramming

Proposal Abstract: My dissertation research is a joint project between the labs of Matt Jacobson and Robert Fletterick. The ultimate goal of this research is to understand how two biomedically relevant cardiac transcription factors communicate with each other in different DNA contexts. These two factors have proven to be instrumental to the proper development of the human heart. In fact, mutations in either transcription factor result in congenital heart defects, the most common defect in live births. Additionally, these two factors in conjunction with just a few others are able to dictate cellular fate. More specifically, Gata4, Tbx5, and Mef2C alone can sufficiently reprogram fibroblasts into functional cardiomyocytes. Our goal is to fully understood how Tbx5 and Gata4 cooperate at a number of different levels. In brief, we plan to (i) learn the kinetic details of ternary complex assembly in various regulatory contexts (ii) characterize any protein-protein interactions between these two factors and at the same time understand how DNA mediates cross-talk to encourage this interaction (iii) reveal at a functional level how the composition of these tandem response elements dictates transcriptional activity.

Dissertation Research

Institution: University of California, San Francisco

Advisor : Matt Jacobson, Ph. D

Dates: 2012 - 2015

Project Title: Post-Translational Modifications in the HIV Semi-Super Elongation Complex are Allosterically Coupled to Functional Sites Necessary for Virulence

Abstract: The HIV Transactivator of Transcription (TAT) is critical to the successful transcription of HIV genes and the HIV genome. Once bound to pTEFB and other important host transcriptional machinery, for example the scaffolding protein AFF4, TAT then docks this complex to the short, emerging transcribed Transactivation Response Element (TAR). Finally, phosphorylation of both a paused RNA Polymerase II and other transcriptional repressors advances HIV transcription through promoter escape. While it has been well demonstrated that post-translational modifications (PTM's) throughout this complex can either drive an infection towards latency or activation, the molecular underpinnings of these events are largely unclear. Through the use of molecular dynamics simulations of the TAT-AFF4-pTEFB complex modeled with a range of PTM's, we have explored at the atomic level just how these perturbations are coupled to the biophysical and functional observations seen experimentally. Our approach to studying these perturbations and PTM's includes a.) using a priori knowledge of the system to identify differences across simulations and b.) utilizing non-biased approaches to uncover less obvious modes of communication between PTM's and both cryptic and established functional sites. Upon analysis, each system exhibits a unique conformational ensemble in the T-Loop of Cdk9. In particular when compared to the TAT-free complex, the presence of TAT — and to a

greater degree the presence of TAT acetyl K28 — restricts the conformational freedom of the T-Loop. We hypothesize that constraint of the T-Loop around some active conformer is coupled to HIV transactivation, possibly by increasing Cdk9 kinase activity or stabilizing interactions with the Cdk9 substrate, the C-terminal domain of RNA Polymerase II. Importantly, experimental observations characterizing transactivation and kinase activity for these systems corroborate this proposed mechanism.

Institution: University of California, San Francisco

Advisor : Matt Jacobson, Ph. D

Dates: 2012 - 2015

Project Title: A Redox Sensitive Post-Translational Modification Regulates the Atypical Nuclear Receptor Nurr1

Abstract: Parkinson's disease (PD) is the 2nd most prevalent neurodegenerative disease worldwide. PD is characterized by the loss of dopaminergic (DA) neurons of the *substantia nigra* with concomitant movement dysfunction often in the form of tremor, bradykinesia, and postural instability. The transcription factor Nurr1 is intimately tied to PD, as it is critical to both the development and maintenance of DA neurons. Therefore, the development of a Nurr1 agonist could revert PD symptoms and potentially halt disease progression. Although Nurr1 is a nuclear receptor, it lacks many of the key regulatory mechanisms typical to this class of ligand-dependent transcription factors. In particular, Nurr1 lacks a canonical ligand binding pocket yet adopts a fully active conformation as observed in other ligand-bound nuclear receptor structures. However, we have identified a small dopamine metabolite which binds to Nurr1 in a seemingly reversible covalent manner via a cysteine adduct. Kinetic studies demonstrate that this molecule binds to Nurr1 with an extremely long residency time provided that the protein remains folded. A crystal structure (~3Å) of this complex localizes this interaction to a cysteine adjacent to an important regulatory helix. Incidentally, this would serve as the first ligand-bound crystal structure of Nurr1. Because Nurr1 appears to be constitutively active, we hypothesized that this non-enzymatic post-translational modification would inhibit Nurr1 function. *In vivo* Nurr1 reporter assays performed in zebrafish demonstrate that

Supplementary treatment with the metabolite does in fact decrease Nurr1 transcriptional activity, yet does not eradicate dopaminergic neurons as compared to a neurotoxic control.

CHAPTER 2

A Novel Redox Sensitive Post-Translational Modification Regulates the Function of the Atypical Nuclear Receptor Nurr1

Abstract

Parkinson's disease (PD) is the 2nd most prevalent neurodegenerative disease worldwide. PD is characterized by the loss of dopaminergic (DA) neurons of the *substantia nigra* with concomitant movement dysfunction often in the form of tremor, bradykinesia, and postural instability. The transcription factor Nurr1 is intimately tied to PD, as it is critical to both the development and maintenance of DA neurons. Therefore, the development of a Nurr1 agonist could revert PD symptoms and potentially halt disease progression. Although Nurr1 is a nuclear receptor, it lacks many of the key regulatory mechanisms typical to this class of ligand-dependent transcription factors. In particular, Nurr1 lacks a canonical ligand binding pocket yet adopts a fully active conformation as observed in other ligand-bound nuclear receptor structures. However, we have identified a small dopamine metabolite which binds to Nurr1 in a seemingly reversible covalent manner via a cysteine adduct. Kinetic studies demonstrate that this molecule binds to Nurr1 with an extremely long residency time provided that the protein remains folded. A crystal structure (~3.2Å) of this complex localizes this interaction to a cysteine adjacent to an important regulatory helix. Incidentally, this would serve as the first ligand-bound crystal structure of Nurr1. Because Nurr1 appears to be constitutively active, we hypothesized that this non-enzymatic post-translational modification would inhibit Nurr1 function. *In vivo* Nurr1 reporter assays performed in zebrafish demonstrate that Supplementary treatment with the metabolite does in fact decrease Nurr1 transcriptional activity, yet does not eradicate dopaminergic neurons as compared to a neurotoxic control.

An early proposal for this project to the Michael J. Fox Foundation went unrequited…

Characterization of Small Molecule Binding Modes through Experimental and Computational Approaches

Structural characterization of any of the Nurr1 domains is sparse. In fact, there exists only a single crystal structure of Nurr1-- that of the ligand binding domain (Wang, 2003). This crystal structure, solved in the apo form, classified Nurr1 as a novel ligand-independent, constitutively active nuclear receptor. It is interesting to note that while molecules that directly bind to Nurr1 have been identified (Poppe et al, 2007; Kim et al, 2015), no co-crystal structures have been solved to date. We have had initial success with solving a co-crystal structure of Nurr1 with a dopamine metabolite, dihydroxyindole. While our crystallographic datasets have not exceeded a ~3Å resolution, even at this resolution we can see unaccounted for electron density at a single cysteine residue (with partial occupancy) that is consistent with size of dihydroxyindole. Furthermore, this apparent dihydroxyindole-cysteine adduct is consistent with the incredibly long off-rates of the Nurr1-DHI interaction seen by SPR. While reproducing crystal conditions was itself challenging, solving this low resolution structure was only achievable after diagnosing an undocumented, insidious crystal pathology that has likely thwarted other efforts in structural determination of the Nurr1 LBD. Higher resolution information of the binding mode of this natural scaffold is essential for any further medicinal chemistry efforts to synthesize a potent Nurr1 agonist, especially in the development of reversible and irreversible covalent agonists. Analogously, we will incorporate disulfide-reactive hits from our tethering screen into our crystallization efforts to provide potency for structure-based drug design.

Alongside our experimental characterization of small molecule-bound Nurr1 complexes, we are utilizing our computational resources to predict binding modes of both DHI-adducts and disulfide-mediated tethering hits. It was no surprise that our initial attempts at employing covalent docking to generate binding modes for these covalent molecules was unsuccessful, resulting in only binding modes that protruded out into solvent. This inability to generate enthalpically favorable binding modes in silico is inconsistent with our ability to tether disulfide-reactive compounds resistant to excess thiol competition. In other words, it appears that our tethering hits interact with the Nurr1 topology as opposed to simply through formation of a disulfide bond. This can easily be explained by remembering that the crystal structure describes a single low energy conformer on the Nurr1 energy landscape. In other words, our tethering screen allows molecules to interact with Nurr1 and its associated energetic landscape, while our preliminary in silico approaches force these molecules to engage with a single, low energy conformer.

Based on the apo crystal structure, it is widely believed that Nurr1 lacks a canonical ligand binding pocket. While it is undeniable that the canonical NR binding pocket is occluded in the case of Nurr1, we also recognize that proteins are inherently dynamic molecules. With this in mind, we will utilize induced fit covalent docking to generate binding modes of these small molecules. In this version of docking, a select set of residues within an appropriate distance of the proposed binding site (e.g. cysteine in this case) are allowed to sample rotameric space, as opposed to the frozen rotamers observed in the crystal structure. This enhanced sampling may

allow us to generate novel binding modes that are unimaginable with the currently limited structural information about this ligand binding domain.

Direct Binding Measurements of Disulfide-Free Tethering Hits with Surface Plasmon Resonance

Our initial tethering screen has identified a number of molecules that interact with Nurr1 supported by a disulfide bond. While this approach is certainly valuable, especially in fragment-based early drug discovery, these molecules must tested as non-covalent potential binders to truly assess potency. We will use surface plasmon resonance to classify tethering hits that have an intrinsic affinity for Nurr1 in the absence of a disulfide. The number of small molecules that bind to Nurr1 are scarce limiting the number of positive controls equally scarce.

However, even with a limited set of tools to explore direct binding, we were still able to validate this approach specifically to Nurr1. Namely, we were able to arrive at a comparable affinity to Poppe et al. for their lead molecule, although we too were unable to collect a full binding isotherm due to restraints in solubity. We also have preliminary direct binding measurements for the Nurr1 agonist proposed by Smith et al., although our estimated affinity (~200μM) deviates significantly from the reported EC50 (2.5nM). While we plan to further explore this remarkably low affinity for Nurr1, we are intrigues that this Nurr1 agonist may be eliciting effects indirectly through an undetermined Nurr1 mechanism. Finally, this method was successful in identifying dihydroxyindole as a Nurr1 binder, and fortuitously, access to the kinetics of the interaction suggested this interaction is covalent. Our confidence in our ability to characterize Nurr1-small molecule interactions will be indispensable for these screens. The ~ ##

ligands identified from the tethering screen is easily manageable number to screen at multiple concentrations against both Nurr1 and RXR (cross-reactivity) and in replicate. Furthermore, access to the Biacore T100 in the Biosensor Facility run by our collaborators in the Fletterick lab will ensure the timely identification of binding-capable, non-covalent scaffolds to insert into our medicinal chemistry pipeline. These screens will also be cross-validated with differential scanning fluorimetry, another well-accepted methodology in ligand validation.

Small Molecule Mediated Cooperation of Nurr1 Dimerization at Nurr1 cis-Regulatory Sites

A number of small molecules that modulate Nurr1-dependent RXR function have been identified. For instance, XCT0135908 (Wallen-Mackenzie et al, 2003) specifically enhances RXR activity in the presence of Nurr1 but not other nuclear receptors. Importantly, the Nurr1 P560A mutant that cripples Nurr1-RXR communication is unable to enhance RXR activity even in the presence of this potent agonist. While the further pursuit of Nurr1-dependent RXR modulators is being fervently explored by many other labs, our pursuit of RXR-dependent and RXR-independent Nurr1 agonists is still largely uncharted. It is clear that we can confidently detect direct Nurr1 binders through our aforementioned biophysical assays. We now plan to dissect mechanistically if these compounds alter communication Nurr1-DNA binding in the context and absence of RXR and its cognate response elements. Using biolayer interferometry, we can monitor both kinetics and thermodynamics of DNA binding domains interacting with their cognate response elements.

For these experiments, we plan to probe both monomeric and dimeric Nurr1 DBD-LBD interactions on the well-defined nerve growth factor inducible-β response elements (NBRE) in

the presence of both our endogenous scaffold, DHI, and our disulfide-free compounds that exhibited direct binding by SPR. Additionally, assess the capability of these molecules to enhance cooperativity of Nurr1 and RXR on a RARβ2 tandem response element (Castro et al, 1999). We will capture biotinylated RE's on a streptavidin coated sensor tip, and sequentially expose these cis-elements to Nurr1 or Nurr1 and RXR in the presence and absence of our identified Nurr1 direct binders. While we are hoping to further identify Nurr1 specific scaffolds from our tethering screens, we already have, in hand, at least two molecules that we can observe direct Nurr1 binding: namely, DHI and the Sanofi compound. Since the Sanofi compound is a clear and potent Nurr1 agonist, one hypothesis for its mechanism would be an enhanced the residency time of RXR and Nurr1 at critical neurotrophic and neuroprotective enhancer regions. Because of its selectivity for Nurr1:RXR heterodimers, XCT0135908 can be probed as well as a somewhat positive control. The constructs necessary for these experiments, namely the DBD-LBD domains of both RXR and Nurr1, have already been generated by our lab with additional success in the purification of these constructs.

Cell-Based Assessment of Small Molecule Modulators of the Nurr1 LBD

The NR4A nuclear receptor family are truly pariahs of the nuclear receptor family. In addition to adopting an active fold in the absence of both ligand and even a canonical ligand binding pocket, Nurr1 lacks the hallmarks of a traditional NR co-regulator interaction surface. However, somewhat in spite of this, Nurr1 still exhibits cell-type specific activity suggesting that other perhaps protein co-regulators can still direct Nurr1 function. For example, the Nurr1-LBD is profoundly more active in 293 cells compared to a JEG3 cell line in a Gal4-mediated transcriptional activity assays (Castro et al, 1999; Wang et al., 2003). With this in mind, we plan

to assess how our small molecules effect Nurr1 transcriptional activity in vivo. While our examination of Nurr1 homodimerization and heterodimerization on NB and RARβ2 response elements tests the hypothesis that these molecules modulate residency at these cis-elements, our cell-based assays investigate if these molecules alter the recruitment of Nurr1 co-regulators. While regulation by LxxLL-type co-regulators is a primary mode of regulation for almost all other nuclear receptors, the degree of involvement of co-regulator recruitment in the case of Nurr1 is unclear (Codina et al, 2004). To investigate this, we will adopt the well-established Gal4-UAS transactivation assay using a Gal4-fused Nurr1-LBD utilizing a luciferase reporter. (Wallen-Mackenzie et al, 2003, Castro et al, 1999). Importantly, we expose the Nurr1 LBD to variious cellular consituents in the form of 3 different cell lines including 293 (human embryonic kidney), JEG3 (placenta-derived human chariocarcinoma), and PC12 (rat-adrenal medulla). These three cell lines are all amenable to transient transfection and cover both neuronal and non-neuronal cell type.

Using these broadly disparate cell types can aid in suggesting if these Nurr1 agonists enhance interactions with the more ubiquitous transcriptional co-regulators or with neuron-specific co-regulators. This experiment utilizes a minimalist approach, as the Nurr1 LBD is uncoupled from its true DNA binding domain. While this simplified approach should inform us in a straightforward manner if the Nurr1 LBD modulators are influencing transcriptional activity. However, Nurr1 co-regulator recruitment might only be possible in the context of RXR. To address this, we will extend our transactivation assays to a luciferase experiment under the regulator of a RARβ2 cis-element using transfected Nurr1 DBD-LBD. In this way we can monitor the induction of Nurr1-transactivation as a heterodimer with its well established

binding partner. Importantly, each molecule identified as a direct binder from the tethering screen will also be checked for cross-reactivity against RXR. In this way, we can determine whether the origin of increased transcriptional activity is from a true Nurr1 agonist or a RXR-agonists that enhance transcription at Nurr1 specific sites.

What follows is a manuscript to be submitted for review for publication in Cell Chemical Biology.

John Bruning, Ph. D Candidate

UCSF, Genentech Hall

600 16th Street, S416

San Francisco, CA 94158

Dear Editorial Staff of Cell Chemical Biology,

We are excited to share with you a fascinating story entitled, "A Redox-Sensitive Post-Translational Modification Regulates Transcriptional Activity of the Atypical Nuclear Receptor Nurr1." As described below, this work involves the discovery of a novel mode of regulation of a nuclear receptor critical to dopaminergic neuron survival and consequently Parkinson's Disease (PD). We feel that this manuscript resonates with Cell Chemical Biology's dedication to publishing innovative research focusing on the interface between important biological systems and novel chemical processes.

Nurr1 is a nuclear receptor expressed almost exclusively in dopaminergic neurons of the substantia nigra pars compacta. With this in mind, it is not surprising that Nurr1 is indispensable to both the development and maintenance of the dopaminergic neuron phenotype. As PD is characterized by the loss of this subset of neurons, Nurr1 has been pursued as a target for agonist development with the hope that increasing Nurr1 activity could delay PD progression and mitigate Parkinsonism. Our driving motivation for this research was to understand mechanisms of Nurr1 regulation and its relationship to PD development. Since the structural elucidation of the apo Nurr1 LBD, it has been widely accepted that Nurr1 is a constitutively active, ligand independent nuclear receptor that, unlike other nuclear receptors, does not accept input from co-activators or co-repressors in a traditional way. The incongruence

between the critical role of Nurr1 to dopaminergic neuron survival yet its paucity of regulation motivated us to consider unconventional regulatory switches that might be unique to dopamine metabolism.

We have identified a small (150Da) dopamine metabolite, 5,6 dihydroxyindole/ indolequinone (DHI/IQ), which binds covalently to Nurr1 and consequently acts as an agonist in both cell based and in vivo transcriptional assays. Two orthogonal biophysical assays can detect direct binding of Nurr1 and DHI. Furthermore, a crystal structure of the Nurr1 LBD covalently bound to Cys566 confirms a ligand binding pocket previously previously identified as a result of a high-throughput drug screen. The combination of a lower resolution (3.3 Å) structure, a fragment lacking heavy atoms, and the possibility of labeling heterogeneity drove further computational studies to predict both a binding mode and electrophilic site on DHI/IQ. Quantum mechanical studies suggest that, in addition to the well known C4 and C7 Michael Addition sites on DHI/IQ, there is a kinetically more accessible linkage possible at the C2 position. Furthermore, QM studies with Cys566 and IQ within in the ligand binding pocket indicate an even lower energy barrier, implicating both a 'catalytic' specificity and reversibility of this covalent modification.

Both cell-based reporter assays and in vivo target gene transcriptional studies indicate that a single exposure to DHI can further enhance Nurr1 function. Interestingly, the target genes influenced by DHI are involved in managing excess cytosolic dopamine, not dopamine biosynthesis. This is consistent with the hypothesis that this Nurr1-DHI interaction should only occur under conditions of oxidative stress and mismanaged cytosolic dopamine. Essentially, this interaction communicates to Nurr1 that returning to a healthier homeostasis should be prioritized over neurotransmitter synthesis.

Our work suggests that dopamine and redox stasis is communicated to a master regulator of the dopaminergic neuron phenotype via a post-translational modification. In addition to contributing to the understanding of Nurr1 and its relationship to dopaminergic neuron health, this work also provides the foundation for Nurr1 agonist development. The identification of an agonist binding pocket adjacent to a reactive cysteine as the foundation for covalent drug design.

Normally we would suggest Kevan Shokat as an editor, but because because of our inter-institutional relationship, we have decided to suggest Craig Crews instead. Additionally, we feel that Eranthie Weerapana (redox biology), Tom Scanlan (experience with nuclear receptors), and Dennis Dougherty (biophysical expertise) would serve as an appropriate review board. Thank you for your consideration.

In Brief

The majority of Parkinson's disease cases cannot be easily referenced to a genetic perturbation. While there is clearly an unmet need for drugs that either halt disease progression or activate neuroregeneration, ignorance of the basic biochemical underpinnings of PD further complicate our ability to rationally develop chemotherapeutics. This work describes how a protein drug target critical to dopaminergic neuron health is regulated by a dopamine metabolite under cellular distress.

Highlights

- DHI, a dopamine metabolite, binds directly to the nuclear receptor Nurr1 via a cysteine adduct

- The DHI binding pocket lies outside of the canonical nuclear receptor ligand binding pocket
- QM calculations predict a novel linkage for this post-translational modification
- DHI increases Nurr1 transcriptional activity for genes involved in managing excess cytosolic dopamine
- Over-exposure to DHI decreases transcriptional activity at a TH enhancer without causing neuronal ablation

Summary

Nurr1 is a nuclear receptor that is critical to both the development and maintenance of the dopaminergic neuron phenotype. Because of its intimate connection to dopaminergic neuron health, Nurr1 agonists have been pursued as a chemotherapeutic treatment for Parkinson's disease. However, Nurr1 is structurally divergent from other readily-druggable nuclear receptors, and consequently does not adhere to canonical regulatory mechanisms observed in other nuclear receptors. We have identified a direct interaction between Nurr1 and an oxidative by-product of dopamine, 5,6, dihydroxyindole. This interaction proceeds via a covalent cysteine modification at a site adjacent to, but outside, of the typical nuclear receptor ligand binding pocket. This binding event relays information about redox- and dopamine-stasis simultaneously. In response to high concentrations of cytosolic dopamine and appreciable oxidative stress, DHI signals to Nurr1 to prioritize transcription of machinery that can manage excess dopamine. The identification of an endogenous Nurr1 agonist that binds covalently at a non-canonical NR binding pocket suggests opportunities for covalent drug development for Parkinson's Disease.

Significance

Only ~10% of Parkinson's disease cases can be attributed to a genetic or an environmental risk factor. Implicit in the notion that the majority of PD cases are sporadic is a general lack of understanding about the underpinning biochemical processes underlying the disease. This study provides insight into how a master regulator of the dopaminergic neuron phenotype responds to cellular distress in the form of oxidative stress. Furthermore, the novel interaction between this transcription factor and a dopamine-related metabolite suggests covalent drug development as a possible avenue for PD treatment.

A Redox-Sensitive Post-Translational Modification Regulates Function of the Atypical Nuclear Receptor Nurr1

John Bruning, Yan Wang, Harrison Liu, Francesca Otrabella, Boxue Tian, Paulomi Bhattacharya, So Guo, James Holton Robert Fletterick, Pamela England, Matt Jacobson.

Abstract

Parkinson's disease is the 2nd most common neurodegenerative disorder, affecting ~1% of the population older than 65 years of age. While genetic and environmental contributors to PD have been identified, the etiology of ~90% of PD cases is still outstanding and considered sporadic. At the core of dopaminergic neuronal health, and therefore PD, is the nuclear receptor Nurr1. Unlike many other nuclear receptors, Nurr1 adopts a constitutively active conformation and lacks an appreciable hydrophobic cavity capable for endogenous metabolite binding. Additionally, Nurr1 lacks a canonical co-regulator binding site inherent to nuclear receptors. In

the absence of the key regulatory mechanisms, we were interested in exploring novel modes of regulation specific to DA health. We postulated that an ideal signaling molecule should be able to report on both dopamine stasis and redox stasis to help determine if a neuron is healthy enough to continue dopamine biosynthesis. We identified 5,6, dihydroxyindole (DHI), an oxidative by-product of dopamine metabolism, as an endogenous small molecule Nurr1 modulator. This interaction occurs at low μM concentrations of DHI and seemingly proceeds via a covalent adduct as determined by surface plasmon resonance. We structurally characterized this novel post-translational modification by x-ray crystallography revealing a covalent adduct at Cys566 of H11 of the Nurr1 ligand binding domain. Interestingly, this position coincides with a 'regulatory hotspot' for Nurr1 as it packs directly against the regulatory helix, H12, and additionally lies in the RXR heterodimerization interface. Cell-based transcriptional activity assays with a a Nurr1-LBD Gal4-DBD fusion protein determined DHI to be a modest Nurr1 activator. Furthermore, DHI could increase transcription of a subset of Nurr1-dependent DA specific genes in zebrafish, which was further exacerbated by concomitant administration of a tyrosinase inhibitor, PTU. However, repeated exposure of DHI, mimetic of a brain under uncorrected oxidative stress as is the case in PD, decreased expression of a fluorescent reporter under a TH promoter in a chemo-genetic zebrafish model of PD. This work suggests that redox homeostasis is communicated to Nurr1 via a redox-sensitive, dopamine metabolite via post-translational modification of a cysteine.

Introduction

Parkinson's disease (PD) is a neurogenerative disorder that impacts approximately 1% of the aging population above 60 years old (MJFF). PD is pathologically characterized as the loss of

dopaminergic neurons (DA) of the *substantia nigra pars compacta.* The concomitant motor impairment associated with the specific form of neurodegeneration includes bradykinesia, tremor, postural instability, and rigidity (Connolly et al., 2014). While careful clinical monitoring alongside both chemical and physical therapeutic intervention can minimize some symptoms of Parkinsonism, there still exists the unmet need for a drug that can safely halt disease progression or ideally induce neuroregeneration.

While both environmental and genetic factors have been clearly associated with PD risk and development, the majority (~90%) of PD instances are of unknown etiology and classified as sporadic. However, the identification of both genetic and environmental factors have ultimately uncovered malfunctioning maintenance processes that result in DA loss. For example, genetic aberrations in the PINK1/Parkin relationship have clearly elucidated the connection between the inability to recycle damaged mitochondria (mitophagy) and PD development (Pickrell et al., 2015). In addition to the already identified genes underlying familial PD cases (Klein et al., 2012), Nurr1 has been extensively studied to better understand its role in DA health and decay.

Nurr1 is a nuclear receptor that has been clearly identified as a master regulator of the dopaminergic neuron phenotype (Saucedo-Cardenas et al., 1998; Zetterstrom et al., 1998). Nurr1 is required for both DA precursors to differentiate into TH+ neurons and fully differentiated DA neurons to maintain the dopaminergic neuron phenotype (Kadkhodaei B, 2009). The ability of Nurr1 to drive DA neuronal cell fate stems from regulating genes involved in dopamine biosynthetic enzymes (tyrosine hydroxylase (TH), aromatic amino acid carboxylase (AADC), dopamine packaging and transporting transporters (DAT, VMAT2), and neuronal health

machinery (SOD2, RET). Clearly the ability to stimulate Nurr1 function is appealing from a chemotherapeutic perspective, particularly because Nurr1 falls into a class of protein that has been historically very druggable -- the nuclear receptors (NR's).

However, unlike other NR's, Nurr1 has seemingly been stripped of the ligand-dependent regulation that largely define the nuclear receptor family. Nurr1 adopts an active conformation as characterized by a folded H12 pinned to Helix 10/11 even in the absence of a ligand (Wang et al., 2003). Additionally, the initial structural elucidation of Nurr1 suggested that it lacked both a ligand binding pocket and co-regulator surface (AF2), the primary constituents of ligand-dependent, co-regulator recruitment observed in many NR's. However, the overarching theme that unites nuclear receptors is the ability to selectively drive transcription of genes in response to a cellular cue, for example presence of a hormone. We postulated that while Nurr1 is structurally divergent from other NR's, its biological ability to modify gene expression in response to a signalling event should be conserved.

Because Nurr1 plays such a pivotal role in dopamine-related processes, we postulated that direct sensing of dopamine-stasis through the Nurr1 LBD would allow a dopamine-Nurr1 feedback mechanism. As a catechol, dopamine is susceptible to oxidation to the highly reactive dopa-quinone. Quinones such as these ultimately exhibit their cytotoxicity through conjugation with protein thiols if not preemptively detoxified through redox buffering (as glutathione) or enzymatic metabolism (catecholamine methyltransfersase, for example) (Segura-Aguilar et al., 2014). Under conditions of oxidative stress, excess dopamine can be further shunted into melanogenesis pathway with enzymatic oxidation from tyrosinase (Zecca et al., 2001; Ito, 2003).

We posited that if Nurr1 were to be regulated by an endogenous metabolite that the ligand should be able to communicate information about dopamine stasis as well as neuronal health, particularly in regards to oxidative stress. The dopamine metabolite 5,6 dihydroxyindole (DHI) and its further oxidized form 5,6 indolequinone (IQ) serves this purpose as its presence conveys both an abundance of unsequestered dopamine as well as a state of oxidative stress required for its formation.Two orthogonal biophysical assays can detect direct binding of Nurr1 and DHI/IQ. Furthermore, a crystal structure of the Nurr1 LBD covalently bound to Cys566 confirms a ligand binding pocket previously identified in a high-throughput drug screen (Poppe et al., 2007). The combination of a lower resolution (3.3 Å) structure, a fragment lacking heavy atoms, and the possibility of labeling heterogeneity drove further computational studies to predict both a binding mode and electrophilic site on DHI/IQ. Quantum mechanical studies suggest that, in addition to the well known C4 and C7 Michael Addition sites on IQ, there is a kinetically more accessible linkage possible at the C2 position. Furthermore, QM studies with Cys566 and IQ within in the ligand binding pocket indicate an even lower energy barrier, implicating both a 'catalytic' specificity and reversibility of this covalent modification.

Functionally, both cell-based reporter assays and *in vivo* target gene transcriptional studies indicate that a single exposure to DHI/IQ can further enhance Nurr1 function. Interestingly, the target genes influenced by DHI/IQ are involved in managing excess cytosolic dopamine, not dopamine biosynthesis. This is consistent with the hypothesis that this Nurr1-DHI/IQ interaction should only occur under conditions of oxidative stress and mismanaged cytosolic dopamine. Essentially, this signalling communicates to Nurr1 that returning to a healthier

homeostasis is prioritized over neurotransmitter synthesis. Our work suggests that dopamine and redox stasis is communicated to a master regulator of the dopaminergic neuron phenotype via a post-translational modification. In addition to contributing to the understanding of Nurr1 and its relationship to dopaminergic neuron health, this work also provides insights into Nurr1 agonist development. Namely, the identification of an agonist binding pocket adjacent to a reactive cysteine as the foundation for covalent drug design.

Results

DHI but Not DHICA Stabilizes the Nurr1 LBD

To initially investigate a possible interaction between Nurr1 and DHI, we utilized differential scanning fluorimetry to determine if the presence of DHI can stabilize the Nurr1 LBD when subject to a thermal gradient. We compared the ability of DHI and the chemically similar oxidative by-product of L-Dopa, 5,6 dihydroxyindole carboxylic acid (DHICA), to shift the melting temperature of Nurr1. DHI, but not DHICA, was able to increase the melting temperature of Nurr1 by ~1ºC at maximal concentration (50μM) with a subsequent dose-dependent response. **(Figure 2.1)** While higher concentrations of DHI continued to stabilize the Nurr1 LBD to a further extent, achieving saturation was hindered by a concomitant decrease in fluorescence as DHI concentration increased. This fluorescence quenching is likely the result of absorbance of DHI/DHI polymer from either the incident excitation light or the emitted light from the dye. While DHI exhibits an absorbance max at ~295nm with little absorbance with red-shifted light, DHI polymers exhibit absorbance across the entire UV-visible region, characteristic of melanins. DHI was also able to induce a dose-dependent change in fluorescence at 25ºC prior to the thermal gradient. Two possible explanations for this change in initial fluorescence are (i)

DHI-dependent conformational change that results in decreased hydrophobic exposure, or (ii) displacement of the hydrophobic-sensing dye from a region on the Nurr1 LBD. These preliminary direct binding studies of DHI and the Nurr1 LBD warranted further biochemical investigation, ideally with a spectroscopically-independent assay.

5,6 Dihydroxyindole Exhibits An Unusually Long Residence Time on Nurr1

Initial kinetic studies (data not shown) using surface plasmon resonance warranted a more detailed kinetic analysis of this interaction between Nurr1 and 5,6 dihydroxyindole. An experiment in which 3 concentrations of DHI were injected for 3 different contact times over freshly generated Nurr1 surfaces was performed. The regeneration of the Nurr1 surface was achieved using the CAP system provided by GE in conjunction with *in vivo* biotinylated, recombinantly purified Nurr1 LBD. From these experiments it is clear that DHI binds to Nurr1 with a considerable residency time, or possibly even in a covalent manner. For comparison, low molecular weight small molecules of comparable size to DHI used in fragment screening by SPR will typically exhibit reversible, step-like 'fast-on, fast-off' kinetics. After the longest association period of 320 seconds and 2.5μM DHI, it can be seen that this interaction is converging towards steady state. Furthermore based on the accrued RU, these responses do not suggest any super stoichiometric binding at these concentrations, although the most extreme concentration and contact time (2.5μM, 320s) suggests a 2nd binding event **(Figure 2.2).**

Parameter estimation for kinetic constants (namely, kon, koff, and kchemistry) was complicated by the > 1 stoichiometry inferred from the theoretical Rmax of each reaction. To this end, we utilized a geometric simplex minimization to fit our data to the numerical solutions of a system

of biochemical equations which included two independent covalent labeling steps. A single labeling event, as predicted, was incapable of achieving the observed Rmax.

While this experiment supports the notion of a direct interaction between the Nurr1 LBD and DHI, the inability to achieve saturation at much higher concentrations (> 2.5μM) indicates a non-specific binding/labeling event of unknown origin. Extensive efforts were made to identify the origin of this non-specific binding. Pretreatment of DHI with 5mM glutathione was able to attenuate a portion of these high responses when compared to an experiment lacking glutathione. **(Supplementary Figure 2.1).** We postulate that this ~1000-fold excess of glutathione was able to neutralize any monomeric DHI, but would have no effect on the presence of any already polymerized species. Additionally and somewhat counterintuitively, higher resonance units and faster associations were observed at 10ºC compared to 25ºC **(Supplementary Figure 2.2).** This observation is consistent again with the insoluble higher-order polymeric species accumulating on the chip due to decreased solubility at lower temperatures.

While there are numerous reports of reactive dopamine species covalently modifying proteins, quantitatively characterizing these adducts by mass-spectrometry has largely proven elusive. Consistent with this, we were unable to detect adduct formation by LC-MS. **(Supplementary Figure 2.3)**.

5,6, Dihydroxyindole Modifies a Cysteine in a Regulatory Hotspot of the Nurr1 LBD

We turned to protein crystallography to experimentally characterize at the atomic level the binding mode of DHI on the Nurr1 LBD. We were able to solve a co-crystal to 3.3 Å of the Nurr1

LBD that revealed chemical modification of Cys566 with DHI (**Table 2.1)** Cys566 lies on H11, and depending on the rotamer, can either be pointing towards the traditional ligand binding pocket of Nurr1, or towards a more superficial pocket formed by H4,H11 and H12. In the apo structure solved by Wang et al. only the rotamer facing towards the ligand binding pocket was observed. An alignment between the published apo Nurr1 LBD and the DHI-bound Nurr1 LBD resulted in a C-alpha RMSD of 0.52 Å **(Figure 2.3)**

Unfortunately due to the lower resolution and potentially partial labeling, assigning a binding mode to DHI is difficult. Furthermore, the redox state of this metabolite, catechol (DHI) or ortho-quinone (IQ) is ambiguous. Furthermore, binding mode heterogeneity (i.e. interactions at multiple linkages) will further diffuse the corresponding ligand density. To enhance the signal of our ligand density, we generated Polder difference maps which take into consideration bulk-solvent flattening effects that traditional OMIT maps do not.

In addition to the well-described alpha-beta unsaturated ketone electrophilic sites at C4 and C7 of IQ, we propose that the C2 position is also capable of reacting with a thiolate nucleophile. According to our quantum mechanics calculations, DHI is unlikely to react with thiolate, as the reactions are endothermic and no transition state exists. However, IQ reacts with thiolate at positon 2, 4 and 7 in a reversible manner. The energy barriers for Position 2, 4 and 7 are 6.3, 14.7 and 16.7 kcal/mol, respectively, implying that position 2 is kinetically the most favorable position for thiolate attack. Position 3 is not reactive because the energies of the transition state and the corresponding product are high (Ea > 30.0 kcal/mol, . **(Supplementary Figure 2.4).** Computational side chain prediction for both redox states at each of the proposed linkages sites

(Supplementary Figure 2.5) suggests that IQ linked at the C2 is poised for hydrogen bonding with key residues that define this pocket -- E415 (H4), R445 (H8-H9 loop), and R563 (H12) **(Supplementary Figure 2.6).** For these reasons, we will focus on the C2 linkage of Cys566 to IQ.

The interactions that would stabilize this interaction include i.) a hydrogen bond the guanidinium of R515 and the carbonyl off of C6 of DHI, ii.) a hydrogen bond between the carboxylate of E415 and the indole NH from DHI iii.) cation-pi interactions between both/either R515 and R563. Interestingly, in the apo structure, this pocket is occupied by 2-4 ordered waters (depending on the chain). Therefore, DHI binding could drive receptor desolvation, and in liberating these ordered waters, help to drive the initial reversible binding event. These predicted binding poses help partially explain the preference of DHI over DHICA, as it would be difficult to accommodate a carboxylate in this binding pocket. It is unclear how the presence of the carboxylate at the C2 position would influence adduct formation.

A C566T Mutant can Attenuate Nurr1 Binding

In addition to the C566 site, we mutated the the surface cysteines C534 and C505. We speculated one potential Nurr1 source of non-specific binding could be at C534. In small molecule tethering screens, we observed C534 to be promiscuously labeled by the majority of test compounds (unpublished). We probed a series of mutants (C566T, C505A/C534T, and C505A/C534T/C566T) against a series of DHI concentrations **(Figure 2.4)**. All mutants were able to attenuate responses at higher DHI concentrations compared to the wildtype. However, only the C566T was able to attenuate responses at the lowest concentration of 625nM. Responses did not decrease monotonically with the number of cysteines removed. In fact, the triple mutant

C505A/C534T/C566T exhibited the highest responses of the cysteine mutants. We have observed that removal of multiple cysteines jeopardizes Nurr1 stability as assessed by DSF. These mutants can still be expressed and purified, but can exhibit a Δ 5-10ºC decrease in melting temperature. Again, DHI polymer is suspected as a contaminant contributing to non-specific binding. Regardless, a simple comparison of the C566T and the wildtype clearly shows that the removal of Cys566 eliminates the highest affinity, and therefore most specific, Nurr1-DHI interaction.

DHI Increases Nurr1 Transactivation in a Non-Neuronal Cell Line

In order to deconstruct the relationship between DHI and Nurr1 transcriptional activity, we utilized a well-characterized Nurr1-LBD Gal4-DBD fusion protein that can selectively drive expression of Firefly luciferase through an upstream UAS cis-element (Castro et al., 1999). Luciferase activity was measured after a 6h incubation with DHI at 24h post-transfection. Co-transfection with a constitutively expressed (Nurr1 independent) Renilla Luciferase allowed for the normalization of transfection efficiency. Normalized firefly luciferase expression was enhanced according to a DHI dose-response with a maximal change in luciferase activity at 100μM DHI only in the presence of the Nurr1 LBD **(Figure 2.5).** While a saturable dose-response curve is obviously desirable, this was unachievable due to DHI cellular toxicity at > 100μM. This functional data corroborates with our direct binding data, suggesting that DHI but not DHICA, can bind to and active Nurr1 via the Nurr1 LBD. Furthermore, this emphasizes that the presence of the Nurr1 LBD is required for a DHI response, at least in this transcriptional context.

DHI Enhances Transcription for a Subset of Nurr1-Dependent Genes *in vivo*

To further validate the relationship between DHI and Nurr1 transcriptional activity, we performed target gene transcription in zebrafish treated with a single dose of DHI. Zebrafish contain two isoforms of NR4A2: NR4A2A, predominantly expressed in DA, and NR4A2B, predominantly expressed in the eye. Importantly, the sequence identity between the LBD's of zebrafish NR4A2A/B and human Nurr1 is remarkably high (~95%, ~76%, respectively), which includes conservation of Cys566.

Zebrafish embryos (3 days post-fertilization, 3dpf) were exposed to various doses of DHI for either 6h or 24h, at which time embryos were collected for RNA extraction. qPCR was then utilized to assess the degree of transcription of well-established Nurr1-dependent genes to establish an *in vivo* relationship between DHI dose and Nurr1 activity. At 6h post treatment, DHI was able to induce transcription in the dopamine transport related genes (VMAT2, DAT), but not dopamine biosynthetic enzymes. Biologically, this is consistent with the notion that the presence of DHI suggests both excess cytosolic dopamine and oxidative stress. Therefore, in this case, DHI would be communicating to Nurr1 to prioritize handling excess dopamine over synthesizing more dopamine. At 24h post-treatment, in addition to VMAT2, D2RA and TH are also upregulated by DHI. At this time point and again with only a single exposure to DHI, it would suggest these neurons have partially rebounded from the initial oxidative insult and are ready to re-engage with dopamine biosynthesis and dopamine signaling. The work of Smith et al., 2015 in the identification of SA00025 as a potent Nurr1 agonist highlights the the complexity of Nurr1 regulation temporally. TH was upregulated 3-fold after 1h post-treatment, and subsided to basal levels as seen in each time point thereafter. C-ret, on the other hand, was downregulated at 4h post-treatment, followed by upregulation at 24h post-treatment.

Repeated Exposure to DHI Causes Neurodegeneration and Decreased Expression at a TH Enhancer-Driven Fluorescent Reporter

We hypothesize that a single exposure to DHI should invoke a cytosolic clean-up through Nurr1 by expressing genes that can manage excess cytosolic dopamine. However, how would a system handle chronic exposure to these dopamine metabolites, specifically DHI? To this end, we monitored a fluorescence reporter, mCherry, under the regulation of a TH enhancer in zebrafish frequently exposed to DHI. Both 1dpf and 3dpf were exposed to fresh DHI at various concentrations every 12h until 5dpf. Fish were then imaged to assess total neuronal fluorescence as described in Liu et al., 2016.

1dpf fish exposed to high concentrations (200uM DHI) exhibited significant tail deformities (upward curvature) that ultimately complicated imaging. Even though 6/14 survived, we were only able to capture 1 fish image due to these deformities. While developmental complications due to DHI exposure are interesting, we will focus on the 3dpf fish because the survival rate was 100%, therefore facilitating statistical analysis. At 200μM DHI 3dpf showed a statistically significant ($p = 0.0003$) decrease in total neuronal fluorescence when compared to a DMSO control. The 100μM concentration point was considered significant at the $p < 0.1$ threshold ($p = 0.053$).

The TH-driven fluorescent reporter mCherry is also conjugated to nitroreducatese (NTR) an enzyme that can metabolize the protoxin MTZ into a neurotoxic compound. MTZ

administration phenocopied the decrease in neuronal fluorescence 200μM DHI treatment. **(Figure 2.6).** Interestingly, while high concentrations of DHI are sufficient to attenuate fluorescence similarly to MTZ, the neuronal architecture of DHI treated zebrafish is preserved, unlike those treated with MTZ **(Supplementary Figure 2.9)**. This suggests that DHI-dependent fluorescence loss is not due to simple neuronal destruction, but rather a more specific loss of activity at a TH cis element.

Discussion

It is somewhat surprising that vertebrates have evolved to use potentially toxic catecholamines, such as dopamine, epinephrine, and norepinephrine, as the primary signalling molecules for neural communication. One possible explanation for this is that these molecules, in addition to communicating at the synapse, can also relay information about the overall stasis of the cell. We propose that in addition to its clear role in neuronal communication, dopamine can communicate with the transcriptional regulator Nurr1 in a redox dependent manner through its oxidized and cyclized state. Namely, under oxidative stress and a depleted redox defense system (e.g. depleted cytosolic [glutathione]), dopamine can oxidize and cyclize into the highly reactive metabolite DHI/IQ. Our work suggests that oxidized IQ can selectively increase Nurr1 transcriptional activity via a reversible cysteine adduct at Cys566.

There are numerous instances that suggest dopamine can covalently modify proteins intimately involved in PD generation (SOD2: Beluzzi et al., 2012; tyrosine hydroxylase: Kuhn et al., 1999; DJ1: Girotto et al., 2012; alpha-syn: Bisaglia et al., 2007). For example, Parkin solubility and consequently activity were shown to be inhibited by the formation of a dopamine metabolite

adduct (LaVoie et al., 2005). It is important to note that, in many of these cases, the source of quinone reactive species was tyrosinase-treated DA. This approach to quinone generation will not only produce dopa-quinone (DQ), but additional downstream oxidized species such as DHI/IQ, especially in the absence of high reducing reagent conditions. Furthermore, DA concentrations, in some cases, exceed > 200μM. At these concentrations of DA in the presence of tyrosinase under non-reducing conditions, even small amounts of generation of DHI/IQ could possibly be serving as the electrophilic species in adduct formation Therefore, the true covalent species in some of these cases is ambiguous.

In many of these cases, dopamine-derived quinone modification leads to the formation of higher order species and consequently aggregation, insolubility, and loss of function when tested. For example, DA labeling of DJ-1 led to both higher species, and either a -1ºC or -5ºC decrease melting temperature. In our study, we have described a specific reaction between Nurr1 and DHI that is detectable in the low μM range that stabilizes the Nurr1 LBD. There were no indications that this interaction induced aggregation or crosslinking. Because of the PAINs (Baell, J. et al, 2015) nature of this molecule, it is important to note the the Nurr1 transcriptional activity was primarily increased, not decreased, in the presence of DHI. While there are many ways to imagine how DHI or any higher order polymers could be inhibiting Nurr function (e.g. partial unfolding, crosslinking), non-specific agonism of a transcriptional regulator is largely inconceivable.

These dopamine adducts are largely undetectable by mass-spectrometry. Somewhat counterintuitively, these dopamine covalent species can be observed by incorporation of 14C

dopamine followed by SDS-page and autoradiography. Because both methods cause denaturation and are inherently non-equilibrium methods, it appears that these modifications are somewhat resistant to protein unfolding and dilution unlike other reversible covalent modifications (Bradshaw et al., 2015). Therefore, it is possible that either the drop in pH or the ionization during LC-MS could be sufficient to cause loss of these adducts.

NMR studies by Poppe et al. initially suggested both the nucleophilicity of Cys566 as well as a potential ligand binding pocket formed by H4, H11, and H12 of the Nurr1 LBD. In these experiments, they were able to readily alkylate Cys566 in addition to two other cysteines with 13C beta mercapoethanol. With this site -specific labeling approach, they could ultimately approximate a binding site for their lead molecule. This previously identified binding pocket on Nurr1 coincides with that of DHI. In line with these studies, our own disulfide-mediated small molecule tethering screen again identified 3 reactive cysteine, and further mutagenesis experiments determined that ~95% of hits localized to Cys566 (data not published). Additionally, the identification of 3 reactive cysteines in the Nurr1 LBD may partially explain the source of non-specific binding observed in the direct binding assays.

Based on this, we suggest that DHI specifically interacts with Nurr1 at Cys566 in the very low μM range, but may further react with DHI/IQ at much higher concentrations at other cysteines. This idea is consistent with our *in vivo* experiments in which single exposure to DHI could invoke transcriptional activity at VMAT, DAT, and DR2A, suggestive of the need to manage excess cytosolic dopamine. However, repeated exposure to DHI was sufficient to decrease transcriptional activity at a Nurr1-dependent promoter. While a single exposure to DHI might

mimic a brief, manageable degree of oxidative stress, the recurrent exposure to DHI might be more similar to an irrecoverable degree of oxidative distress as is the case in PD.

The underlying mechanism for exactly how this post-translational modification alters Nurr1 activity is unclear, especially in light of the fact the DHI/IQ binding does not drive any large conformational changes in the Nurr1 LBD. Nurr1 is capable of interactions with DNA as a monomer, a homodimer, or a heterodimer with RXR (Zetterstrom et al., 1996; Perlmann et al., 1995). Furthermore, a model of the RXR LBD-Nurr1 LBD heterodimer places Cys566 in the middle of the dimerization interface **(Supplementary Figure 2.10)**. Because Nurr1 transcriptional activity was only observed at a subset of Nurr1-dependent genes, it is possible that this modification somehow alters Nurr1 dimerization preference.

The identification of a endogenous Nurr1 covalent-agonist is potentially inspiring for Nurr1 agonist development. From a structure based drug design perspective, this work suggests Nurr1 is potentially druggable through reversibly covalent or covalent compounds. Furthermore, it seems that the natural biological response to high cytosolic dopamine concentrations and a low redox buffering capacity is to invigorate expression of genes to help rectify this cellular distress. Therefore, development of Nurr1 agonists that can preferentially express genes responsible for neuronal health (e.g. Nurr1, C-ret, SOD2, VMAT) instead of driving the synthesis of dopamine biosynthetic machinery (e.g. TH, AADC) may be an alternative approach to PD chemotherapeutic intervention.

Significance

Nurr1 is a transcriptional regulator that directs expression of genes that define dopaminergic neurons such dopamine biosynthetic enzymes, dopamine transporters, and dopamine signaling receptors. While Nurr1 is a nuclear receptor, the structure of its ligand binding domain suggested that, if Nurr1 were to bind an endogenous metabolite like other members of its protein family, it would require some novel binding mode. We have elucidated an interaction of Nurr1 with an oxidative by-product of dopamine, DHI. This interaction occurs via a cysteine adduct in a non-canonical NR binding pocket. Functionally, this interaction drives transcription of only a subset of Nurr1 dependent genes related to managing cellular distress specific to dopaminergic neurons.

Figures

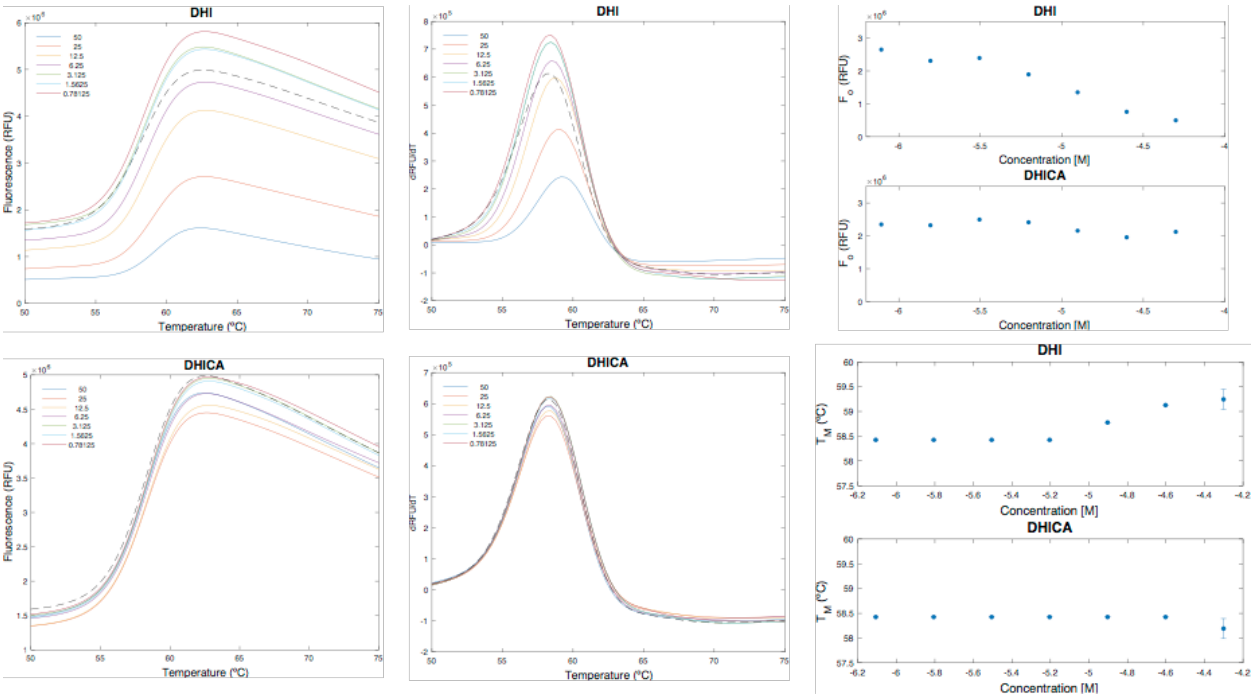


Figure 2.1 - Thermal shift assays monitoring Nurr1 LBD melting temperature in the presence of either DHI or DHICA (concentrations in μM). Panels A,B: Temperature vs Raw Fluorescence; Panels C,D: Temperature vs dRFU/dT; Panel E: Concentration (M) vs Initial Fluorescence at 25ºC for DHI and DHICA; Panel F: Temperature vs. Melting Temperature (Tm; ºC) for DHI and DHICA.

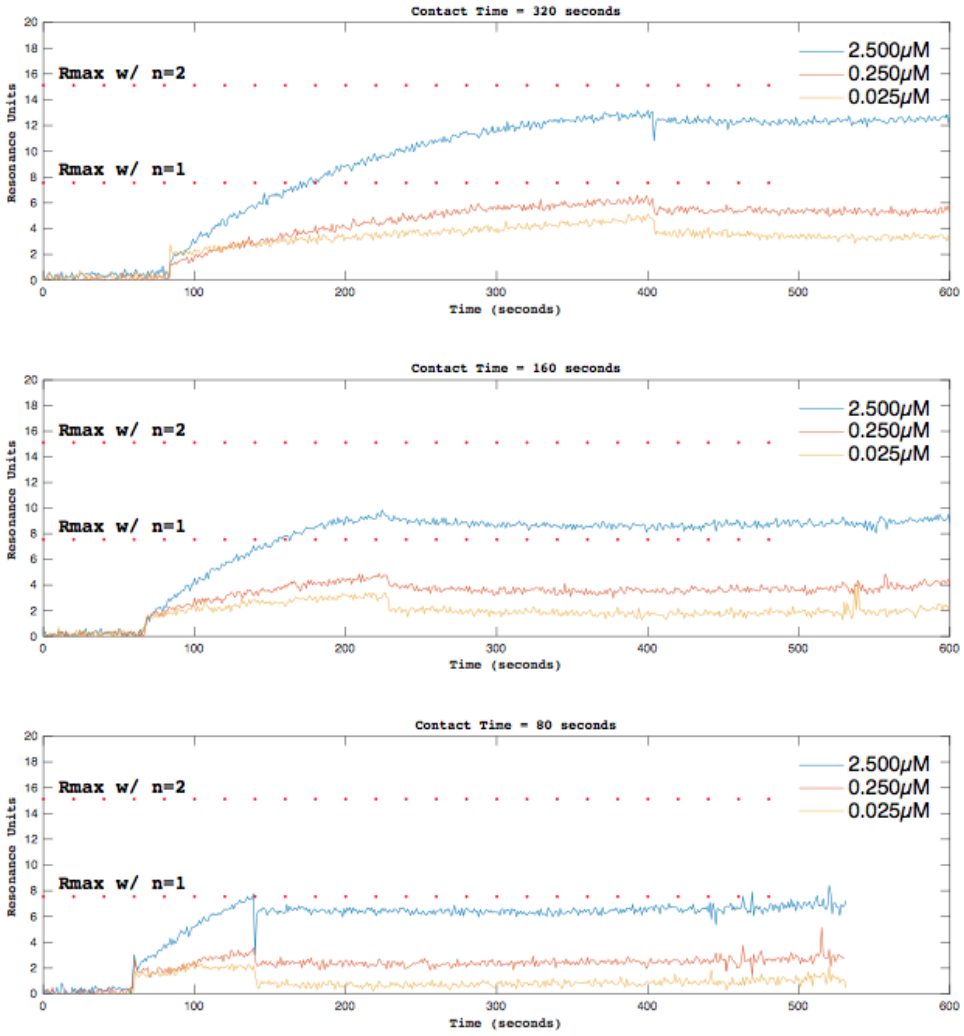
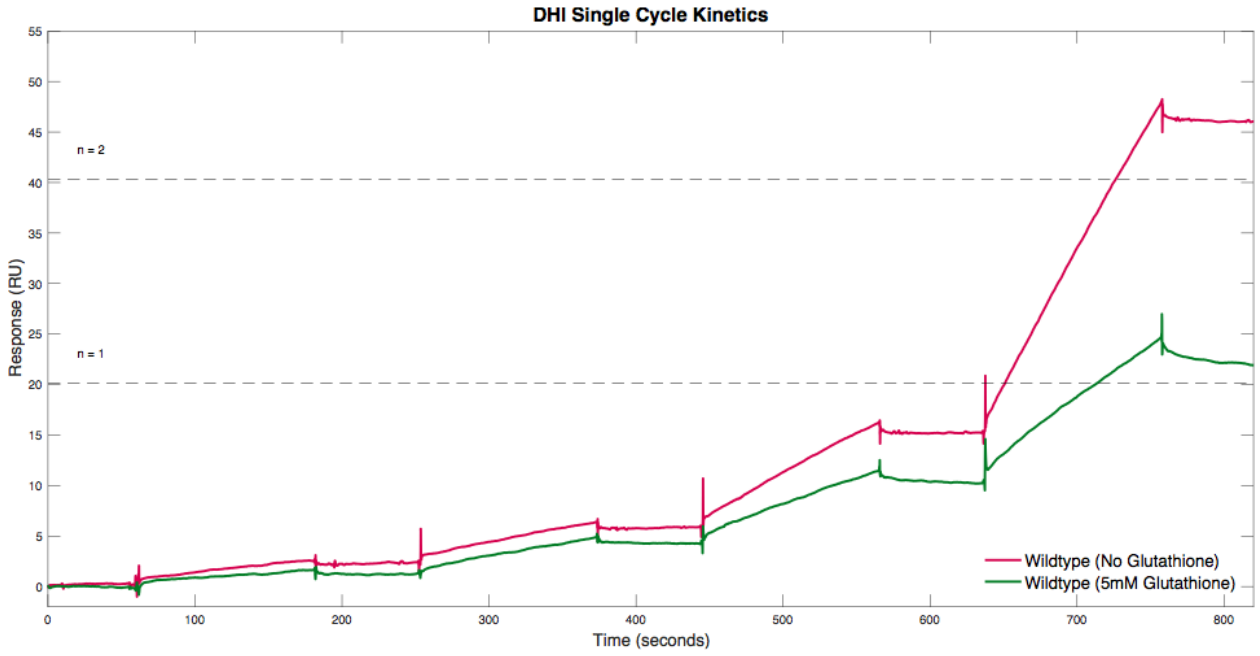
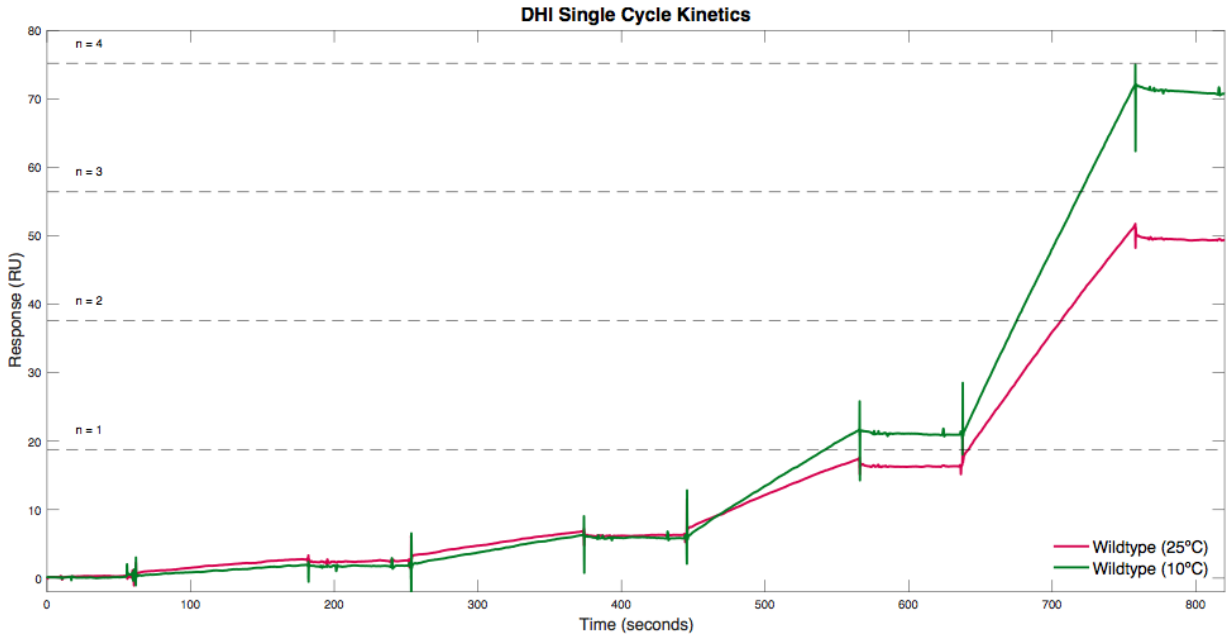


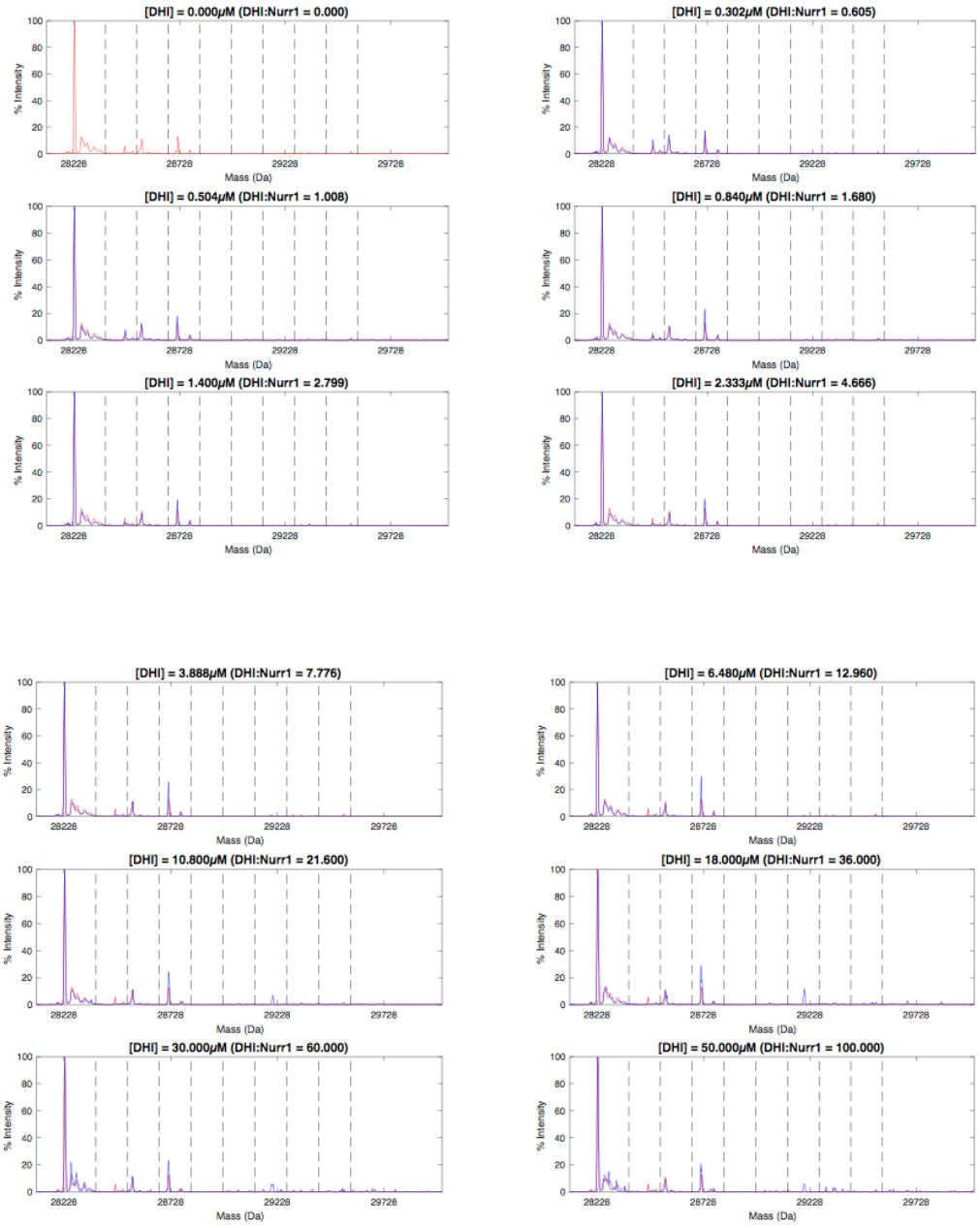
Figure 2.2 - Time dependent association of DHI with Nurr1 LBD by surface plasmon resonance. All panels contain 3 DHI concentrations (2.5μM, 0.25μM, 0.025μM) injected for various lengths (top panel: 320s, middle panel: 160s, bottom panel: 80s). The surface was regenerated with the CAP regeneration solution followed by capture of avi-tagged Nurr1 LBD prior to each of the injections. Data was referenced to both the reference flow and a buffer subtraction of equivalent contact time.



Supplementary Figure 2.1 - Time dependent association of DHI with Nurr1 LBD by surface plasmon resonance. Each experiment was performed as single cycle kinetic runs with no surface regeneration in between concentrations. Responses were normalized according to initial protein captured for better comparison. DHI was injected in increasing concentrations for 4 concentrations (0.6125, 1.25, 2.5, 5μM). Contact times were 120s with an intermittent 60s dissociation time. Dashed black lines indicate the expected response from *n* additions of DHI. Running buffers were equivalent with the exception of +/- glutathione. To minimize bulk shift responses due to buffer differences, DHI samples were pre-incubated with the +/- glutathione buffer as well.



Supplementary Figure 2.2 - Time dependent association of DHI with Nurr1 LBD by surface plasmon resonance. Each experiment was performed as single cycle kinetic runs with no surface regeneration in between concentrations. Responses were normalized according to initial protein captured for better comparison. DHI was injected in increasing concentrations for 4 concentrations (0.6125, 1.25, 2.5, 5μM). Contact times were 120s with an intermittent 60s dissociation time. Dashed black lines indicate the expected response from *n* additions of DHI. Experiments were identical excluding the instrument temperatures (25ºC and 10ºC)



Supplementary Figure 2.3 - 500nM Nurr1 LBD was incubated with various concentrations of DHI (ranging from sub-stoichiometric to 10-fold excess DHI) or DMSO. Red traces denote DMSO control for comparison. Dashed black lines designate the expected mass shifts corresponding to sequential additions of DHI adducts. No enrichment of peaks at these markers was observed after treatment with DHI.

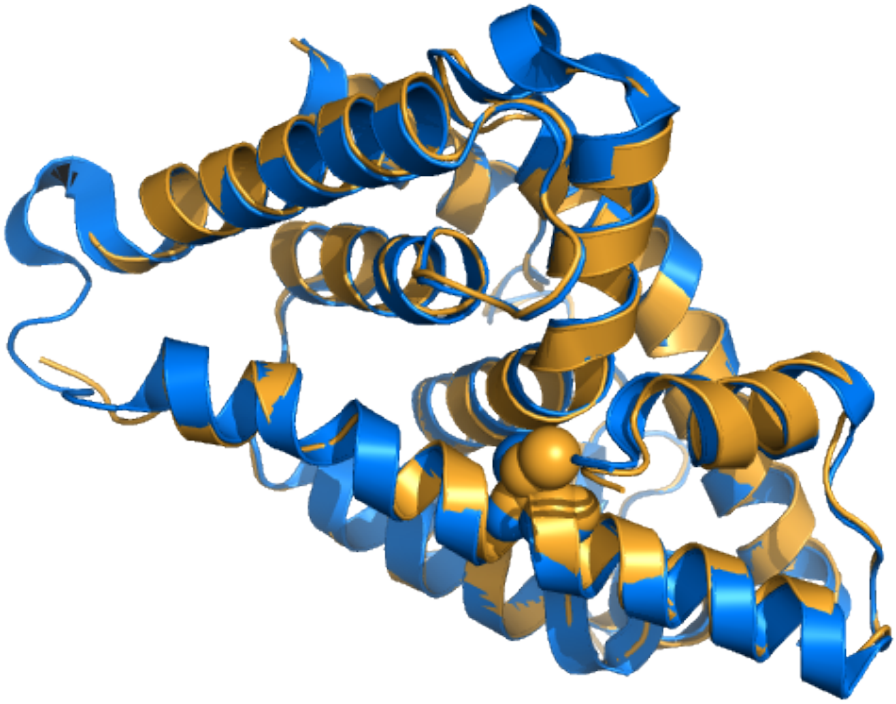
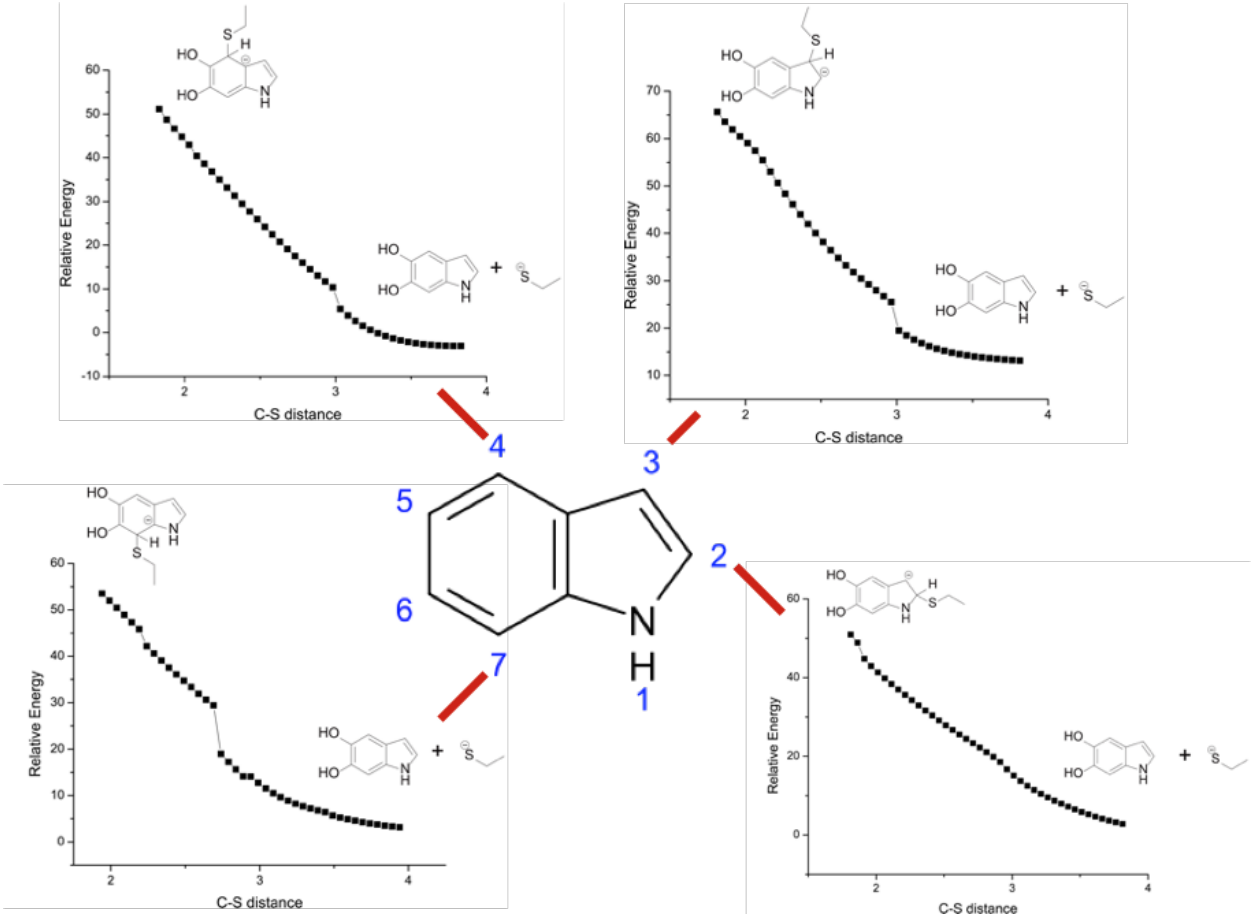
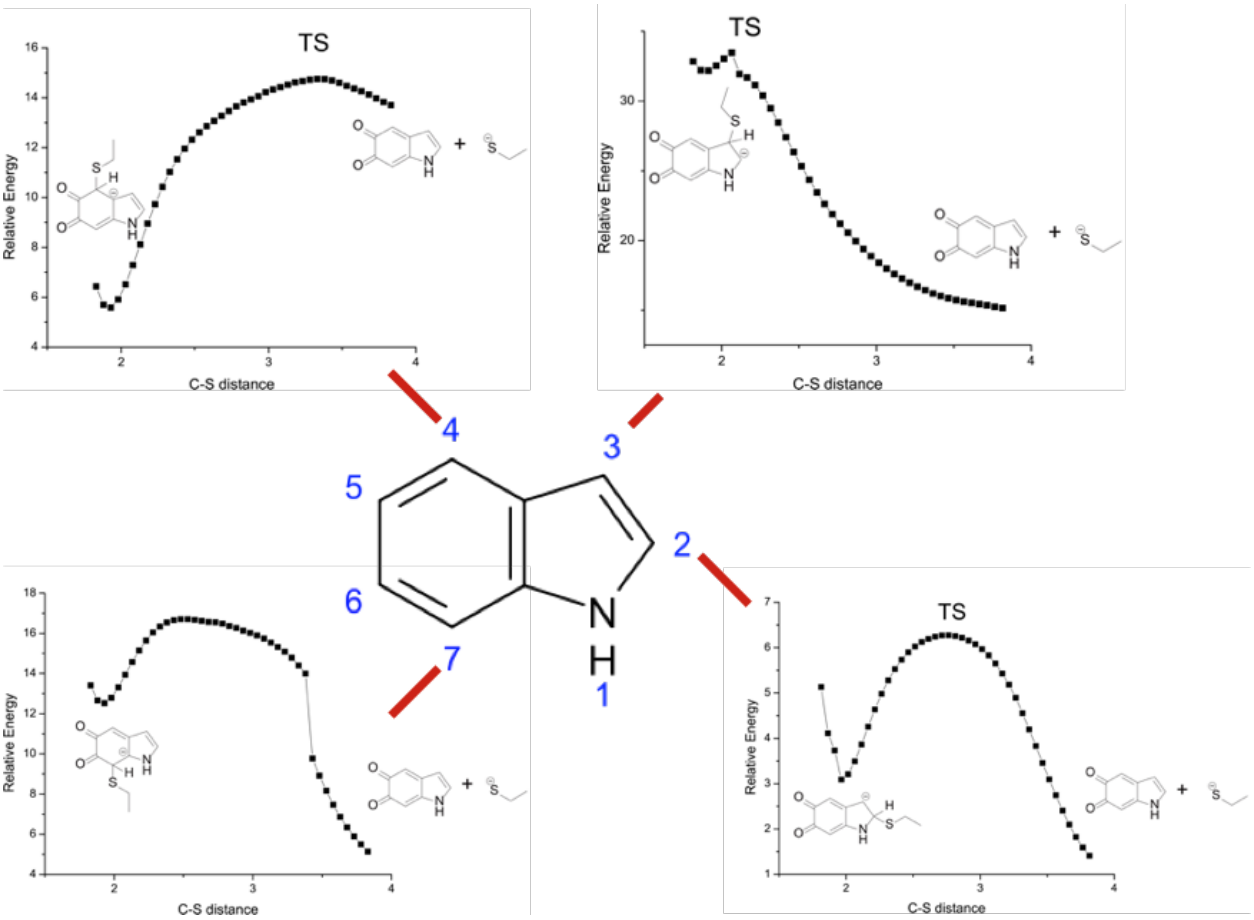


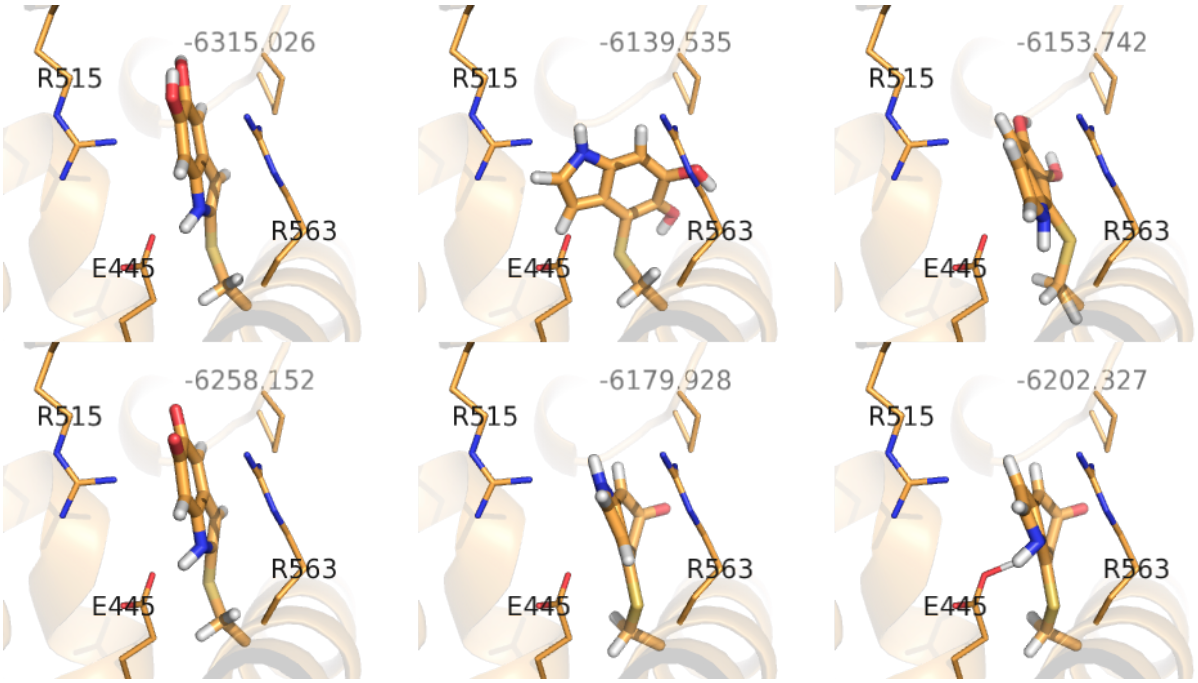
Figure 2.3 - Alignment between the apo (1ovl; chain A; marine) and DHI bound (chain A; brightorange) Nurr1 LBD with a C-alpha RMSD of 0.52Å.



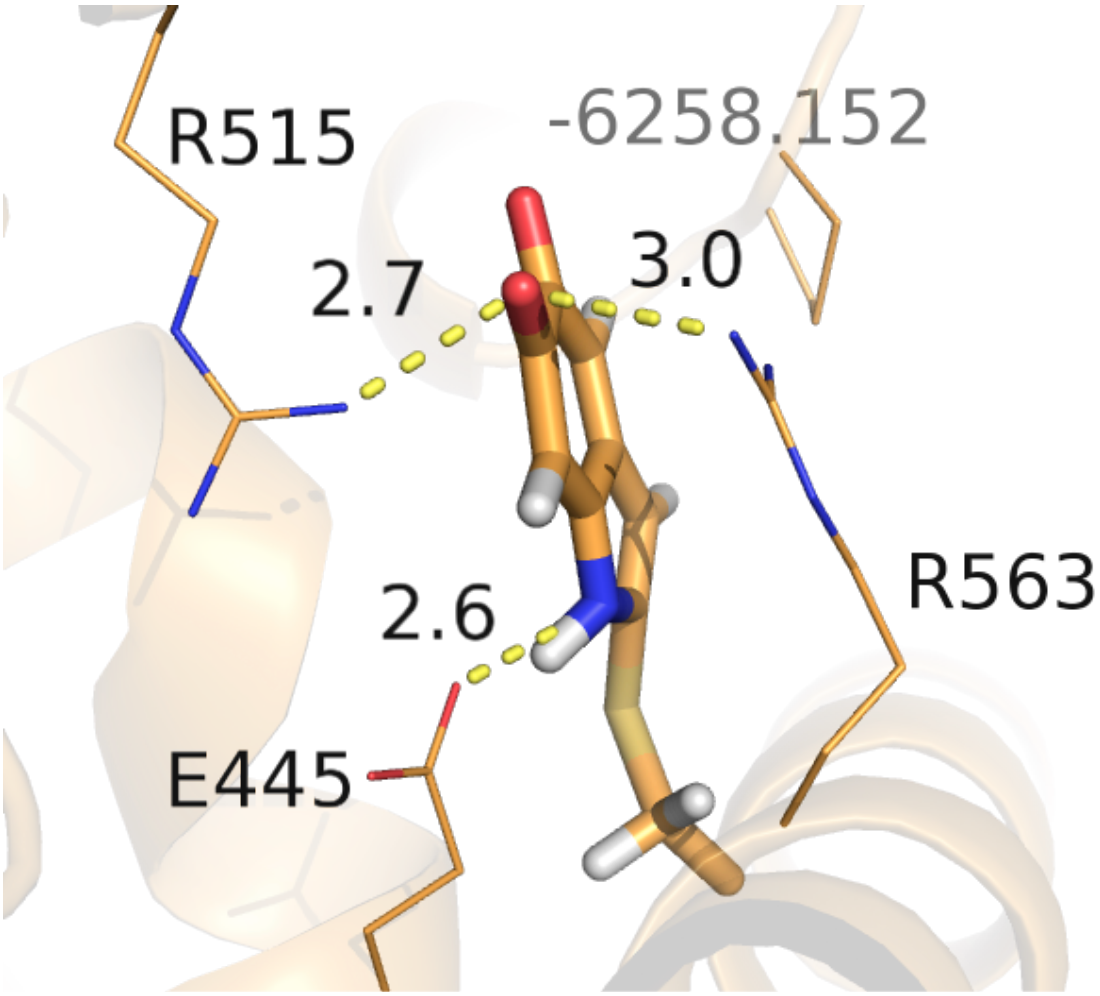
Supplementary Figure 2.4 - Potential energy surface of thiolate attack for DHI (catechol) at each 4 different linkage sites. None of these reactions exhibit an energy barrier that would indicate formation of a C-S bond at any position.



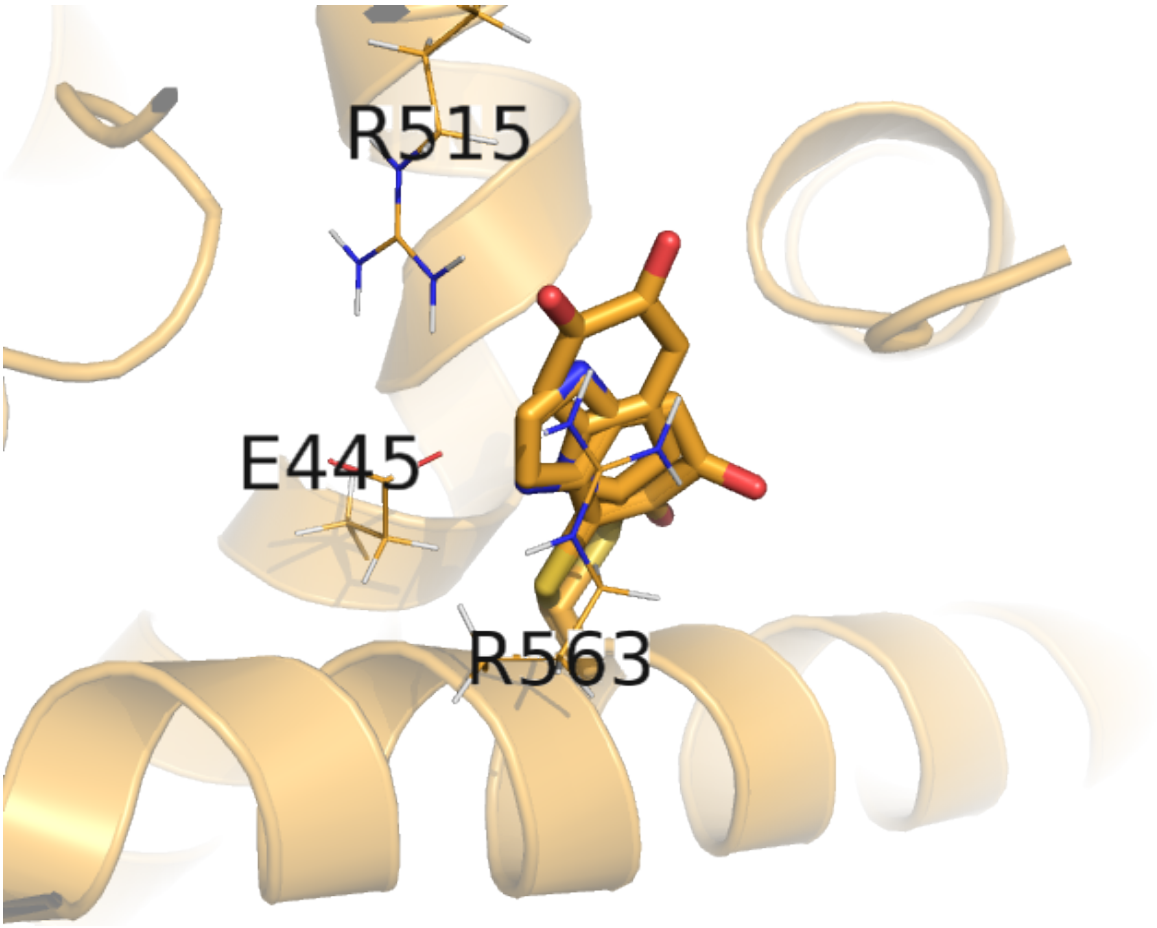
Supplementary Figure 2.5 - Potential energy surface of thiolate attack for IQ (quinone) at each 4 different linkage sites. Each reaction (except C3) exhibit an observable energy barrier leading to a stable C-S linked molecule. The energy barriers for Position 2, 4 and 7 are 6.3, 14.7 and 16.7 kcal/mol, respectively. In particular, the low reaction energy barrier observed in the S-C2 reaction suggest a reversibility to the reaction.



Supplementary Figure 2.6 - Side chain prediction for both IQ and DHI at each of the proposed linkage sites. Top Row: DHI, Bottom Row: IQ. Columns 1-3 correspond to C2, C4, and C7 linkages, respectively. Only Cys566-DHI was optimized; all other residues that constitute the ligand binding pocket were not optimized or minimized in order to preserve the conformations from the crystal structure. PLOP energies are located in the top right corner of each panel.



Supplementary Figure 2.7 - Side chain prediction of IQ at linkage site C2. This redox/linkage combination allows for hydrogen bonding with 3 residues in the binding pocket and possible cation-pi interactions with the displayed arginines. Distance between atoms are in Å.



Supplementary Figure 2.8 - Superposition of IQ linked at all 3 linkage sites (C2,C4, and C7) to help visualize how labeling heterogeneity would be reflected in electron density

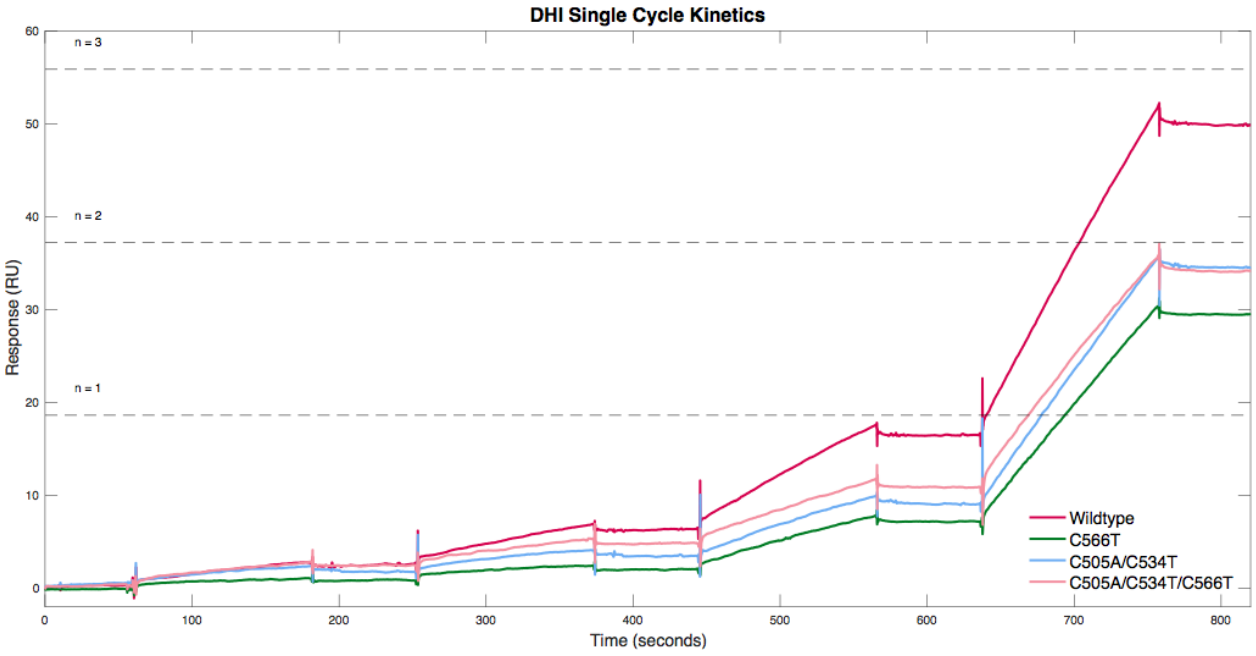


Figure 2.4 - Time dependent association of DHI with Nurr1 LBD by surface plasmon resonance. Each experiment was performed as single cycle kinetic runs with no surface regeneration in between concentrations. Responses were normalized according to initial protein captured for better comparison. DHI was injected in increasing concentrations for 4 concentrations (0.6125, 1.25, 2.5, 5μM). Contact times were 120s with an intermittent 60s dissociation time. Dashed black lines indicate the expected response from *n* additions of DHI.

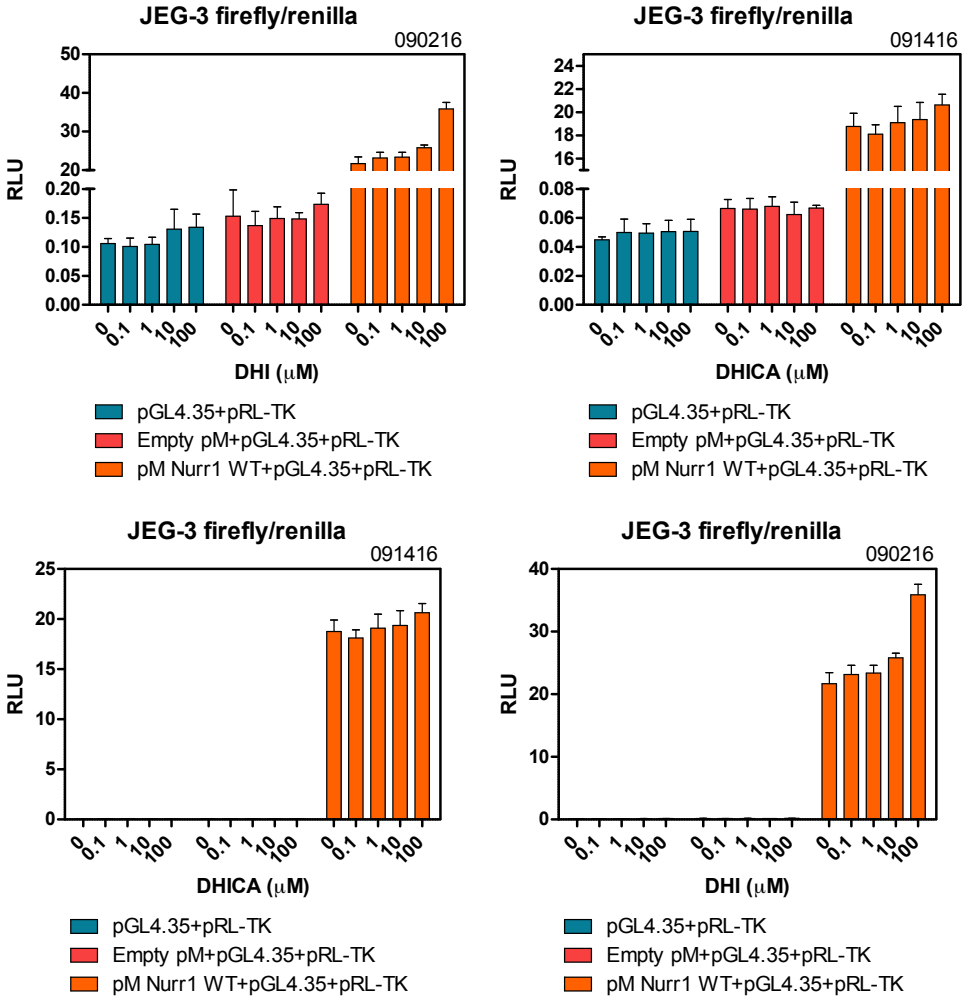


Figure 2.5 - JEG3 cells were co-transfected with a reporter plasmid encoding the firefly luciferase gene driven by an upstream 9x Gal4 UAS (pGL4.35) and a plasmid encoding a Gal4 DBD (Empty pM) by itself or fused to the Nurr1 LBD (pM Nurr1 WT). Every well was also transfected with a plasmid encoding the Renilla luciferase gene constitutively expressed under a thymidine kinase promoter. Cells were incubated with DHI/DHICA or DMSO for 6h before luciferase signal was measured. Top row and bottom row differ only in the y-axis (break plot and no break plot, respectively). One-way Analysis of Variance (ANOVA), in comparison with the data of 0 μM, **$p < 0.01$, ***$p < 0.001$

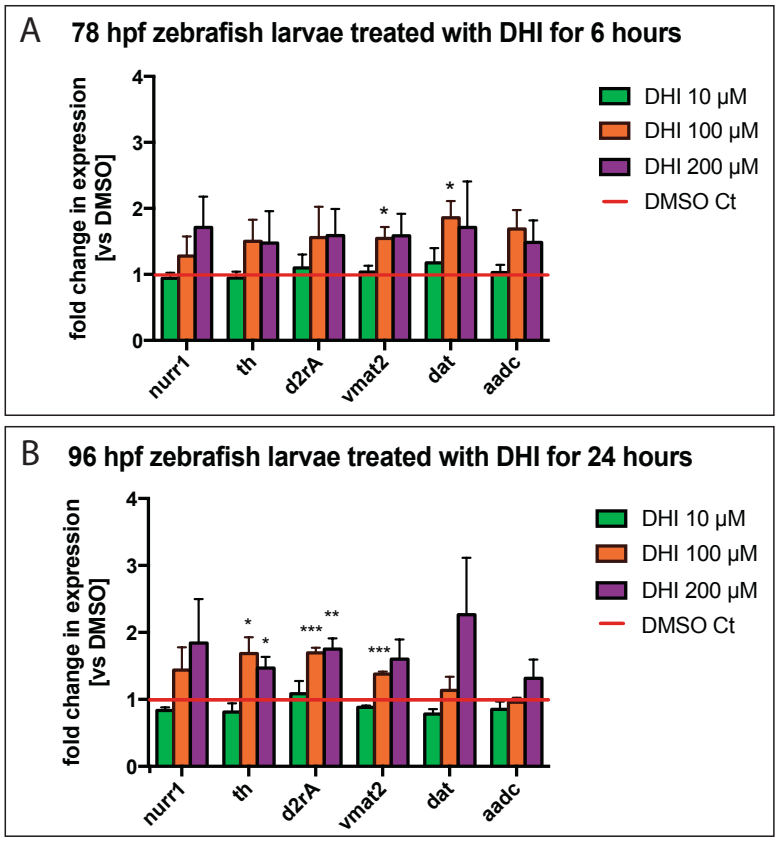


Figure 2.6 - Treatment with DHI causes increased gene expression in genes involved in dopamine metabolism downstream *nurr1*. X axis: *nurr1, th, d2rA, vmat2, dat, aadc* A) qPCR analysis of mRNA of 78 hpf wild type zebrafish larvae prepared after 6 hours of DHI treatment shows statistically significant increased gene expression in *vmat2* and *dat* at 100 µM concentration compared to DMSO Ct (red bar). B) qPCR analysis of mRNA of 96 hpf wild type zebrafish larvae prepared after 24 hours of DHI treatment shows statistically significant increased gene expression in *th* and *d2rA* at 100 and 200 µM concentrations and in *vmat2* gene at 100 µM concentration compared to DMSO Ct (red bar). *elf1a* gene was used as internal control to determine the relative mRNA expression. Results are from four independent experiments. Relative average expression ± SEM; $P<0.05$ by Student t test. GraphPad Prism 7 software was used for statistical analysis

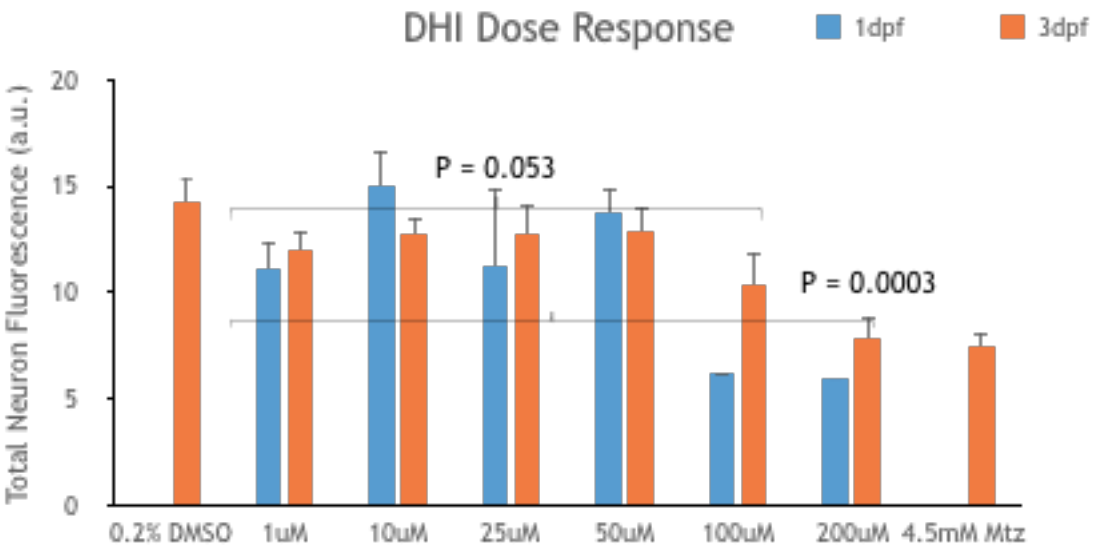
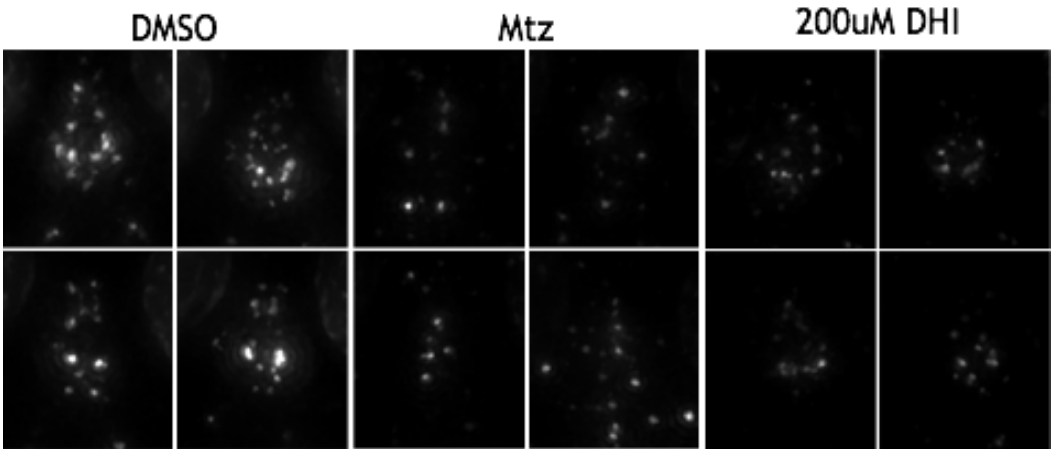
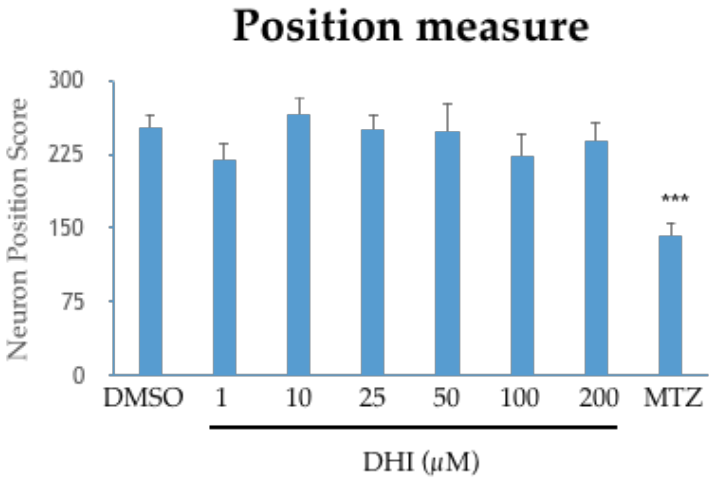
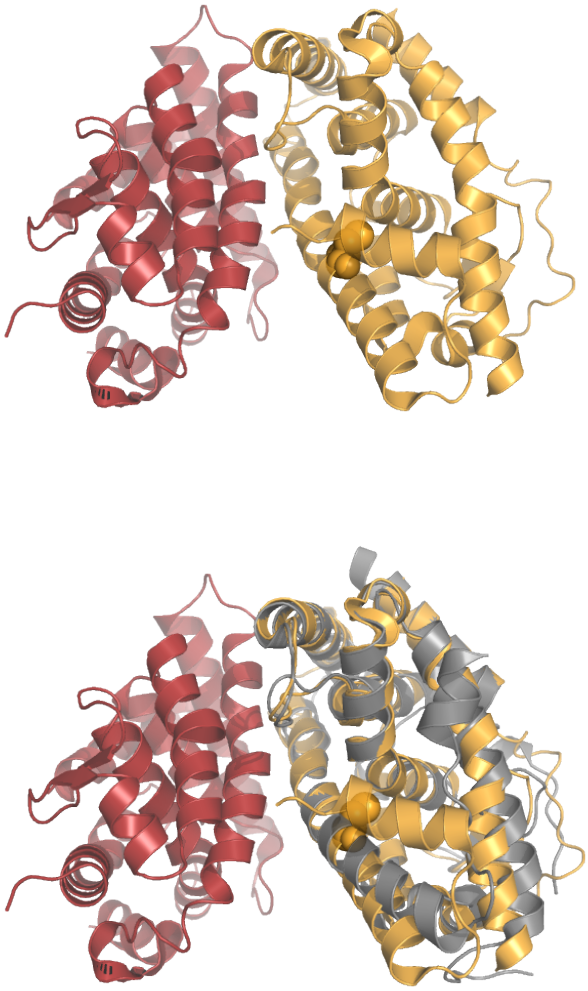


Figure 2.7 - 1dpf and 3dpf transgenic zebrafish harboring a fluorescent reporter (mCherry-nitroreductase fusion protein) driven by a TH promoter were exposed to fresh DHI every 12h until 5dpf. Total neuronal fluorescence was measured after the last DHI treatment. DHI was toxic to 1dpf fish at 200μM (6/14 survival rate); therefore only data from 3dpf fish (100% survival rate) was analyzed. Both the 100μM and 200μM DHI treatments exhibited a statistically significant decrease in neuronal fluorescence relative to the DMSO treated fish. 4.5mM MTZ was administered as a pro-toxin that is metabolized by nitroreductase to serve as a postive control for neuronal ablation.





Supplementary Figure 2.9 - Quantification of neuronal pattern comparing DHI treatment with a positive control for neuronal ablation (MTZ). Top panel: Raw fluorescence images from 3dpf zebrafish treated with either DMSO, 4.5mM MTZ, or 200μM DHI; Bottom Panel: Quantification of neuron position after treatment with DMSO, DHI, or 4.5mM MTZ. Only MTZ exhibits a disruption in neuronal patterning relative to the DMSO control ($p << 0.01$ ***)



Supplementary Figure 2.10 - Model of RXR LBD:Nurr1 LBD heterodimer based off of the RXR LBD: RAR LBD heterodimer structure (PDB entry: 1DKF). The RMSD between the RAR LBD and Nurr1 LBD was 1.437 Å. Top panel: Nurr1 (brightorange), RXR (red); Bottom Panel: Nurr1 (brightorange), RXR (red), RAR (gray). Cys566 of Nurr1 is shown in spheres. The RXR:RAR heterodimer and RXR:Nurr1 heterodimer share a common DR5 spacing at the cis-element level.

Table 2.1: Data Collection and Refinement Statistics after 8 Rounds of Refinement

	DQ2*	DQ4*	DQ7*
Wavelength	11111eV	11111eV	11111eV
Resolution range	40.29 - 3.3 (3.418 - 3.3)	40.29 - 3.3 (3.418 - 3.3)	40.29 - 3.3 (3.418 - 3.3)
Space group	P 31 2 1	P 31 2 1	P 31 2 1
Unit cell	80.586 80.586 225.805 90 90 120	80.586 80.586 225.805 90 90 120	80.586 80.586 225.805 90 90 120
Total reflections	155425 (15669)	155425 (15669)	155425 (15669)
Unique reflections	13457 (1318)	13457 (1318)	13457 (1318)
Multiplicity	11.5 (11.9)	11.5 (11.9)	11.5 (11.9)
Completeness (%)	99.58 (99.92)	99.58 (99.92)	99.58 (99.92)
Mean I/sigma(I)	14.90 (1.63)	14.90 (1.63)	14.90 (1.63)
Wilson B-factor	100.37	100.43	100.40
R-merge	0.183 (2.044)	0.183 (2.044)	0.183 (2.044)
R-meas	0.1915 (2.135)	0.1915 (2.135)	0.1915 (2.135)
R-pim	0.0558 (0.6156)	0.0558 (0.6156)	0.0558 (0.6156)
CC1/2	0.999 (0.653)	0.999 (0.653)	0.999 (0.653)
CC*	1 (0.889)	1 (0.889)	1 (0.889)
Reflections used in refinement	13411 (1317)	13410 (1317)	13411 (1317)
Reflections used for R-free	1323 (126)	1323 (126)	1323 (126)
R-work	0.2627 (0.3864)	0.2562 (0.3865)	0.2659 (0.3953)
R-free	0.3097 (0.3927)	0.3012 (0.3960)	0.3223 (0.4088)
CC(work)	0.960 (0.664)	0.959 (0.661)	0.954 (0.669)
CC(free)	0.903 (0.713)	0.898 (0.700)	0.887 (0.738)
Number of non-hydrogen atoms	5500	5496	5502
macromolecules	5426	5426	5433
ligands	20	22	23

solvent	54	48	46
Protein residues	674	674	674
RMS(bonds)	0.004	0.015	0.006
RMS(angles)	0.95	1.15	0.99
Ramachandran favored (%)	94.12	95.67	93.34
Ramachandran allowed (%)	4.49	3.10	5.11
Ramachandran outliers (%)	1.39	1.24	1.55
Rotamer outliers (%)	3.03	3.19	3.19
Clashscore	11.67	13.23	10.45
Average B-factor	120.94	121.43	121.72
macromolecules	121.59	121.98	122.27
ligands	151.80	153.20	149.06
solvent	43.47	43.99	43.27
Number of TLS groups	14	14	14
Statistics for the highest-resolution shell are shown in parentheses.			

Methods

Differential Scanning Fluorimetry

Nurr1 LBD was buffer exchanged from Base Buffer into 25mM HEPES, pH 7.4, 150mM NaCl. (For initial DHI vs DHICA comparison, pNurr1 LBD was used; for cysteine mutants, avi-tagged Nurr1 LBD was used). The final reaction mixture was composed of 4μM protein, 2.5x SYPRO Orange (ThermoFisher/Life Technologies, from 5000x stock), and specified concentration of ligand. The reaction buffer was 1x HBS, pH 7.4 + 2% DMSO for each reaction at a volume of 50μL. Samples were allowed to incubate at 25ºC, for 30', in the dark prior to exposure to thermal gradient. Fluorescence was monitored using the ViiA 7 Real Time PCR System (ThermoFisher) in a 96 well polypropylene plate (Agilent Technologies Cat #410088) using a 2 MicroAmp Optical Adhesive Film, PCR compatible (P/N 4311971). The filter used for fluorescent monitoring corresponded to an excitation/emission of x1(470 ± 15)nm and m2(558±11)nm, respectively. The thermal gradient was performed from 25ºC to 95ºC at a rate of 0.05 ºC/s. Melting temperature determination corresponded to the temperature at the maximum of the derivative (temperature vs dRFU/dt) data.

Surface Plasmon Resonance

CAP Regeneration Approach

Detailed kinetic analysis (3 concentrations of DHI, 3 contact times) was performed on a CAP chip (GE Health Care sciences). Running buffer for each experiment was 25mM HEPES, pH 7.4, 150mM NaCl, 0.05% P20, and 2% DMSO. All reactions were monitored at 25ºC. Avi-tagged Nurr1 LBD was captured to ~2500 RU (maximum capacity for CAP chip) prior to each injection, and surface was regenerated after sample injection using the defined regeneration

protocol by manufacturers (6M Guanidium HCl + 0.25M NaOH regeneration solution). The flow rate for each experiment was 30μL/min. DHI samples were prepared in running buffer w/ out DMSO but resulted in a final of 2% DMSO to match running buffer. Data processing included double referencing (reference flow cell and buffer subtracted using a buffer injection of appropriate contact time for the given injection) and no solvent correction was deemed necessary. Because of the suspected covalent nature of the interaction, kinetic fits were performed in MATLAB using only the curves from the longest (320s) injection. Briefly, ODE's were solved according to a simple covalent model of : A + B $\Leftrightarrow$ C -> C* (w/ A = Nurr1, B = DHI, C = reversible complex, C* = covalent complex) and parameter estimation was performed using a geometric simplex method. Parameters estimation included kon, koff, kchemistry (bind formation), a bulk shift offset for each curve, and a background drift correction.

Single Cycle Kinetics Approach

Single cycle kinetic experiments were performed on a Biacore T200 instrument (GE). Running buffer for each experiment was 10mM HEPES, pH 7.4, 150mM NaCl, 3mM EDTA, 0.05% P20, and 2% DMSO (diluted from 10x HBS-EP+; GE Healthcare Life Sciences); for one experiment, running buffer was supplemented with 5mM glutathione. All reactions were monitored at 25ºC unless explicitly stated otherwise. Each experiment utilized a new SA (streptavidin) chip (GE Healthcare Life Sciences) which was prepared and conditioned according to standard protocol. ~4000RU of avi-tagged protein was captured for each experiment. Each experiment was run in dual mode (2-1, or 4-3) and never in tandem with another experiment. The flow rate for each experiment was 30μL/min. DHI samples were prepared in running buffer w/out DMSO but resulted in a final of 2% DMSO to match running buffer. Data processing included double

referencing (reference flow cell and buffer subtracted to a single cycle kinetic run of just buffer injections) and no solvent correction was deemed necessary.

Crystallography

Crystallization

The Nurr1 LBD was buffer exchanged into 150mM NaCl, 25mM HEPES, pH 7.4 using a 7kD cutoff Zeba Desalting Column. DHI was then added to a final concentration of 200μM (2% DMSO) from a 10mM stock to a protein solution of 228μM. The reaction was allowed to proceed on ice for 15'. Excess DHI and DMSO were removed via another round of buffer exchange. The labeling procedure was then repeated, and excess DHI and DMSO were again removed via buffer exchange. In our experience, the presence of DMSO seemed to hinder ordered crystal growth. Crystals were then screened against a custom screen (pH 6.5-8.0 as a mixture of various volumes of HEPES 6.5 and HEPES 8.0) based off of the initial reported crystallization condition 18% PEG3350, 200mM KBr and 100 mM HEPES, pH 6.5 at a concentration of 4-7mg/mL. Crystals grew in a 96 well hanging drop format as 500nL drops consisting of 1:1 protein:condition. Crystals were transferred briefly into the crystallization condition supplemented with 20% ethylene glycol to serve as a cryoprotectant. The best diffracting crystal described in this paper was ~200μm x 50μm x 50μm.

Data Collection

Data was collected as the Advanced Light Source, Beamline 8.3.1 at an energy of 11111eV on a ADSC Quantum 315r S/N 926 detector under a stream of liquid nitrogen.

Structure Determination

Data was integrated and scaled using XDS. Although these crystals grew isomorphous to the deposited apo structure (1ovl), this crystal form can reportedly grow in either the P31 or P3121 space group. Initial attempts at reaching a molecular replacement solution was unsuccessful, regardless of which space group the data was processed in. We were initially able to refine this model as a hexamer in P31 against 1ovl after applying the appropriate P31 reindexing operator as -h,-k,l or -k,-h,-l. However, the Rwork-Rfree gap was ~10% after many rounds of refinement (and could not be closed), indicating that the higher symmetry P3121 might be the appropriate space group. We utilized Zanuda to help explore P3121 as a potential space group. Zanuda successfully identified one potential trimer that could serve as an initial model for refinement. The unmerged data was the merged as P3121 using this trimer as a reference with the program pointless. Iterative rounds of phenix.refine and model building with COOT were able to refine the model to within reasonable structure validation metrics. After several rounds of refinement and rebuilding, Cys-IQ adducts were introduced at Cys566 for chain A,B, and C using computational modeling (side chain prediction) described below. Ligand restraints utilizing AM1 geometry optimisation were generated using phenix.elbow for later rounds of refinement. Polder maps were generated using phenix.polder to help guide ligand placement.

Computational Side Chain Prediction for DHI and IQ adducts

Chain A (clearest density for both ligand and local side chains) from an almost fully refined stlsructure (prior to DHI/IQ introduction) was used as starting models for side chain prediction. The unnatural amino acid (Cys-DHI or Cys-IQ w/ C2,C4, and C7 linkages each) was initially built using Maestro from the Schrodinger Suite. Hetgrp_ffgen (Schrodinger utilities)

was then used to generate parameters. The program PLOP was then used to perform side chain prediction for each protein chain at each potential Cys-DHI linkage (C2,C4,C7) over 500,000 iterations followed by energy minimization of the introduced adduct. No other residues were optimized/minimized as to not deviate from the experimental ligand density

Positional Scanning Simulations

To explore the positional preference for thiolate attack, potential energy surfaces for the catechol and the quinone forms of DHI were scanned All quantum mechanics calculations were carried out by using the Gaussian 09 software [ref 1]. Geometries were optimized at the B3LYP/ 6-311+G(d) level.

Luciferase Assays

Human choriocarcinoma JEG-3 cells were maintained in Eagle's MEM with Earle's BSS (MEM) supplemented with 10% FBS, 1% NEAA, 1% L-Glutamine, 1% Sodium Pyruvate and 1% Penicillin/Streptomycin. For transfection, cells were seeded in 96-well plate in antibiotic free medium one day before transfection. Transfection were carried out with Lipofectamine 2000 (Invitrogen) according to manufacture's protocol. Lipofectamine/DNA complexes prepared in Opti-Mem medium and incubated with cells overnight. The Nurr1 LBD fragment was subcloned into pM plasmid (Addgene) containing GAL4 DNA binding domain. The reporter plasmid, pGL4.35[luc2P/9XGAL4 UAS/Hygro] Vector (Promega), contains 9 repeats of GAL4 UAS (Upstream Activator Sequence), which drives transcription of the luciferase reporter gene luc2P in response to binding of Gal4 DNA binding domain fused protein. pRL-TK (Promega) was used as an internal control. The amounts of pM Nurr1 LBD, pGL4.35 and pRL-TK were 50

ng, 50 ng and 5 ng per well respectively. Six hours before luciferase measurement, different doses of DHI were added and incubated. Cells from each well were lysed with 100 μl Passive Lysis Buffer (Promega). The luciferase activities of firefly and renilla were measured using Dual-Luciferase® Reporter Assay (Promega) according to manufacturer's protocol. The results were expressed as average of firefly/renilla luciferase activity ± SD (n=4).

Target Gene Transcription in Zebrafish

DHI Administration

50 larvae/concentration/experiment for a total of 200 larvae/concentration (fn=4) were treated in a 10-cm Petri dish at 72 hours post fertilization with DHI (10 μM, 100 μM, 200 μM) diluted in 40 ml of blue egg water (lab stock) (0.2% DMSO final concentration). DHI was diluted to reach the desired concentration in blue egg water from a 100 mM stock in DMSO. Control larvae were exposed to equal amounts of DMSO (0.2%). Samples were collected for further analysis 6 and 24 hours post treatment.

RNA Isolation and Transcript Quantification

Total RNA was extracted from zebrafish larvae (collection post-treatment at 78 hpf after 6 hours DHI exposure and 96 hpf after 24 hours DHI exposure) using TRIzol reagent (Invitrogen) by homogenization and purified using RNeasy Mini Kit (Qiagen). cDNA was synthesized from 160 ng of purified RNA using qScript cDNA SuperMix (Quanta Biosciences) and used as template. qPCR was performed using Applied Biosystems SYBR Green PCR Master Mix and the ABI7900HT machine. Forward and Reverse primers, listed in Supp. Info, were designed using

NCBI/primer-BLAST software with exon-exon junction parameters and Danio rerio RefSeq for off targets. elf1a primers were used as standard CT to generate Ct values.

Significance Testing:

*Selected statistical significance

P > 0.05 ns; P ≤ 0.05 *; P ≤ 0.01 **; P ≤ 0.001 ***; P ≤ 0.0001 ****

Significant results (P values):

6 hours DHI treatment

100 uM: Vmat2 * 0.018969254170036; Dat * 0.015055297540142

24 hours DHI treatment

100 uM: th * 0.030026918076409; d2rA *** 0.000106487447673; vmat2 *** 0.000078305313436

200 uM: th * 0.031174373114953 ; d2rA ** 0.003182492284801

In vivo Imaging

For total neuron fluorescence, images were analyzed as described in the "Image Analysis" in (Liu et al, 2016). Total neuron fluorescence is simply the summation of all the identified neuron objects within the image of a single fish. "Determination of Brain Health Score" section immediately following was not performed for any of the data generated above. For the position measure, a binary image of the identified neuron object was first created. A binary image of the brain template image was generated using the Otsu method to calculate an appropriate threshold. The position measure was the summation of the product of these two binary images.

The position measure quantifies how many "neuron pixels" overlap with "brain pixels," giving us an intensity-independent measure of how closely the neuronal pattern mimics that of a healthy brain.

References

Baell JB, Holloway GA (2010) New substructure filters for removal of pan assay interference compounds (PAINS) from screening libraries and for their exclusion in bioassays. J Med Chem 53: 2719–2740.

Belluzzi E, Bisaglia M, Lazzarini E, Tabares LC, Beltramini M, et al. (2012) Human SOD2 modification by dopamine quinones affects enzymatic activity by promoting its aggregation: Possible implications for Parkinson's disease. PLoS One 7: 1–9.

Bisaglia M, Mammi S, Bubacco L (2007) Kinetic and structural analysis of the early oxidation products of dopamine: Analysis of the interactions with alpha-synuclein. J Biol Chem 282: 15597–15605.

Bradshaw JM, Jesse M McFarland, Ville O Paavilainen, Phan1 VT, et al. (2015) Prolonged and tunable residence time using reversible covalent kinase inhibitors. Nat Chem Biol 11: 525–531.

Castro DS, Arvidsson M, Bolin MB, Perlmann T (1999) Activity of the Nurr1 carboxyl-terminal domain depends on cell type and integrity of the activation function 2. J Biol Chem 274: 37483–37490.

Connolly BS, Lang AE (2014) Pharmacological Treatment of Parkinson Disease. J Neurol 311: 442–449.

Gaussian 09, Revision E.01, M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, G. A. Petersson, H. Nakatsuji, X. Li, M. Caricato, A. Marenich, J. Bloino, B. G. Janesko, R. Gomperts, B. Mennucci, H. P. Hratchian, J. V. Ortiz, A. F. Izmaylov, J. L. Sonnenberg, D. Williams-Young, F. Ding, F. Lipparini, F. Egidi, J. Goings, B. Peng, A. Petrone, T. Henderson, D. Ranasinghe, V. G. Zakrzewski, J. Gao, N. Rega, G. Zheng, W. Liang, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, K. Throssell, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, T. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, J. M. Millam, M. Klene, C. Adamo, R. Cammi, J. W. Ochterski, R. L. Martin, K. Morokuma, O. Farkas, J. B. Foresman, and D. J. Fox, Gaussian, Inc., Wallingford CT, 2016.

Girotto S, Sturlese M, Bellanda M, Tessari I, Cappellini R, et al. (2012) Dopamine-derived quinones affect the structure of the redox sensor DJ-1 through modifications at Cys-106 and Cys-53. J Biol Chem 287: 18738–18749.

Ito S (2003) A Chemist's View of Melanogenesis. Pigment Cell Res 16: 230–236.

Kadkhodaei B, Ito T, Joodmardi E, Mattsson B, Rouillard C, et al. (2009) Nurr1 Is Required for Maintenance of Maturing and Adult Midbrain Dopamine Neurons. J Neurosci 29: 15923–15932.

Kim C-H, Han B-S, Moon J, Kim D-J, Shin J, et al. (2015) Nuclear receptor Nurr1 agonists enhance its dual functions and improve behavioral deficits in an animal model of Parkinson's disease. Proc Natl Acad Sci 112: 8756–8761.

Klein C, Westenberger, A (2012) Genetics of Parkinson's Disease. Cold Spring Harb Perspect Med 2:a008888

Kuhn DM, Arthur RE, Thomas DM, Elferink LA (1999) Tyrosine hydroxylase is inactivated by catechol-quinones and converted to a redox-cycling quinoprotein: Possible relevance to Parkinson's disease. J Neurochem 73: 1309–1317.

LaVoie MJ, Ostaszewski BL, Weihofen A, Schlossmacher MG, Selkoe DJ (2005) Dopamine covalently modifies and functionally inactivates parkin. Nat Med 11: 1214–1221.

Perlmann T, Jansson L (1995) A novel pathway for vitamin A signaling mediated by RXR heterodimerization with NGFI-B and NURR1. Genes Dev 9: 769–782.

Pickrell AM, Youle RJ (2015) The roles of PINK1, Parkin, and mitochondrial fidelity in parkinson's disease. Neuron 85: 257–273.

Poppe L, Harvey TS, Mohr C, Zondlo J, Tegley CM, et al. (2007) Discovery of Ligands for Nurr1 by Combined Use of NMR Screening with Different Isotopic and Spin-Labeling Strategies. J Biomol Screen 12: 301–311.

Saucedo-Cardenas O, Quintana-Hau JD, Le WD, Smidt MP, Cox JJ, et al. (1998) Nurr1 is essential for the induction of the dopaminergic phenotype and the survival of ventral mesencephalic late dopaminergic precursor neurons. Proc Natl Acad Sci U S A 95: 4013–4018.

Segura-Aguilar J, Paris I, Muñoz P, Ferrari E, Zecca L, et al. (2014) Protective and toxic roles of dopamine in Parkinson's disease. J Neurochem 129: 898–915. doi:10.1111/jnc.12686.

Smith GA, Rocha EM, Rooney T, Barneoud P, McLean JR, et al. (2015) A Nurr1 agonist causes neuroprotection in a parkinson's disease lesion model primed with the toll-like receptor 3 dsRNA inflammatory stimulant poly(I:C). PLoS One 10: 1–14.

Volakakis N, Tiklova K, Decressac M, Papathanou M, Mattsson B, et al. (2015) Nurr1 and Retinoid X Receptor Ligands Stimulate Ret Signaling in Dopamine Neurons and Can Alleviate -Synuclein Disrupted Gene Expression. J Neurosci 35: 14370–14385.

Wang Z, Benoit G, Liu J, Prasad S, Aarnisalo P, et al. (2003) Structure and function of Nurr1 identifies a class of ligand-independent nuclear receptors. Nature 423: 555–560. Available:

Zecca L, Tampellini D, Gerlach M, Riederer P, Fariello RG, et al. (2001) Substantia nigra neuromelanin: structure, synthesis, and molecular behaviour. Mol Pathol 54: 414–418.

Zetterstrom R, Solomin L, Jansson L, Hoffer B, Olson L, Perlmann T (1997) Dopamine Neuron Agenesis in Nurr1-Deficient Mice. Science (80-) 276: 248–250.

Zetterstrom RH, Solomin L, Mitsiadis T, Olson L, Perlmann T (1996) Retinoid X receptor heterodimerization and developmental expression distinguish the orphan nuclear receptors NGFI-B, Nurr1, and Nor1. Mol Endocrinol 10: 1656–1666.

https://www.michaeljfox.org/understanding-parkinsons/living-with-pd/topic.php?causes&navid=causes

CHAPTER 3

Post-Translational Modifications in the HIV Semi-Super Elongation Complex are Dynamically Coupled to Functional Sites Necessary for Virulence

This chapter contains unpublished work that was conducted between by 3rd and 5th year in collaboration with the Zhou and Hurley labs. The collaboration was initiated by U. Schulze-Gahmen, who requested computational support to help explain how perturbations in the HIV Semi-Super Elongation Complex might support stable complexes for structural elucidation. Over the course of ~2 years, we held regularly meetings with these collaborators, mutually providing updates to help support and validate the others findings. Unfortunately, my work on this research culminated culminated on a publication in which I was not formally acknowledged as an author. Instead, my findings were, without permission, published as "previously unpublished; personal communications" without my consent. I stumbled upon this publication while preparing for a HARC update talk. After confronting the authors, we determined that this lack of acknowledgement wasn't malicious or intentional — just incredibly thoughtless and careless. Regardless, I learned a lot about both (i) executing and analyzing biomolecular simulations as well (ii) the role of post-translational modifications in this fascinating piece of HIV transcriptional machinery. What follows is a manuscript that was never submitted for review.

Below is the excerpt in question:

“Molecular dynamics simulations of the Tat-AFF4-PTEFb complex show the flexible Tat K28 side-chain interacting with the backbone of AFF4 as well as the Tat/TAR recognition motif (TRM) of CycT1 (M. Jacobson, personal communications). The M62V mutation in AFF4 changes the equilibrium of these interactions. In particular, K28 hydrogen bonds more frequently with the carbonyl oxygen of AFF4 E61 and has fewer interactions with the CycT1 TRM, allowing the TRM and especially the tryptophan side-chain in the TRM to adopt different conformations. Several residues in the TRM, including W258, are known to be critical for TAR interactions (37). This change in the interaction network of K28 is likely the cause for increased affinity of Tat for AFF4 M62V-P-TEFb and enhanced Tat-transactivation by AFF4 M62V”

Lu H, Li Z, Zhang W, Schulze-Gahmen U, Xue Y, et al. (2015) Gene target specificity of the Super Elongation Complex (SEC) family: How HIV-1 Tat employs selected SEC members to activate viral transcription. Nucleic Acids Res 43: 5868–5879.

Take Home Message

Viral life cycle regulation is heavily governed by post-translational modifications that modulate P-TEFb-related functions through non-redundant mechanisms.

Title

Post-Translational Modifications in the P-TEFb:TAT:AFF4 Complex Are Allosterically Coupled to Functional Sites Critical to HIV Transactivation

Authors and Affiliations

John Bruning [1], Ursula Schulze-Gahmen [2], and Matt Jacobson [1,3]

[1] Department of Pharmaceutical Sciences and Pharmacogenomics, University of California San Francisco, San Francisco, CA, 94158 , [2] Department of Molecular and Cell Biology, University of California Berkeley, Berkeley, CA, 94720 , [3] Department of Biophysics, University of California San Francisco, San Francisco, CA, 94158.

Abstract

The HIV Transactivator of Transcription (TAT) is critical to the successful transcription of the HIV genome. In the absence of TAT, RNA Polymerase II becomes transcriptionally arrested at the 5′ viral LTR resulting in the synthesis of only short viral transcripts. However upon expression of TAT, the semi-super elongation complex (semi-SEC) consisting of HIV TAT:P-TEFb:AFF4 is assembled and ultimately anchored to the viral genome via the viral RNA hairpin TAR. This TAT mediated recruitment of P-TEFb to the site of transcriptional arrest promotes transcriptional elongation via the phosphorylation of the C-terminal domain of the stalled RNA

polymerase II. In addition to TAT expression levels, HIV transactivation is regulated through a range of post-translational modifications throughout each component of the TAT:P-TEFb:AFF4 complex. In particular, TAT Lys28 acetylation stabilizes the semi-SEC and consequently enhances HIV transactivation. Furthermore, TAT Lys28 is of the most highly conserved positions in a genome notorious for its genetic drift. In addition to HIV transactivation, post-translational modifications in the semi-SEC can re-awaken HIV out of latency. Interestingly, phosphorylation on the non-viral human CDK9 —part of the P-TEFb heterodimer along with Cyclin T1 — drives the formation of the semi-SEC even at the vanishingly low levels of TAT observed in viral latency. While these and a few other PTM's have been partially characterized experimentally, there remains a gap in our structural understanding of how these perturbations exhibit their functional effects. Using *in silico* modeling and molecular dynamics simulations, we have been able to elucidate the structural mechanisms that couple these PTM's to functional sites of the semi-SEC that are responsible for i.) TAR RNA binding, ii.) phosphorylation, and iii.) complex stability.

Author Summary

Viruses are often labeled as molecular pirates as they require their own viral proteins to hijack host machinery to be viable. In archetype of this behavior is HIV TAT which commandeers the host kinase P-TEFb in order to preferentially synthesize its own viral genes. Since HIV TAT expression levels and activity serve as an important decider of the viral infectivity, carefully regulating TAT function is paramount to viral success. In order to fine-tune TAT activity, HIV capitalizes on post-translational modifications (PTM's) to both viral and host proteins in order to modify TAT activity. Biophysical experimentation with these modifications is difficult, as

purifying homogenous populations of PT-modified proteins is challenging, if not intractable. Therefore, we turned to computational approaches in order to explore how these PTM's influence the molecular motions in the P-TEFb:TAT:AFF4 complex. We report that these somewhat subtle modifications to TAT and P-TEFb are dynamically linked to both nearby and distant functional sites responsible for viral RNA binding, substrate binding, and protein-protein interactions. Our findings not only corroborate with experimental support, but also propose additional explanations for how these PTM's might drastically steer the direction of the viral lifecycle.

Introduction

The life cycle of the human immunodeficiency virus (HIV) is exquisitely regulated at the level of post-translational modification (PTM). The consequences of these covalent linkages are most readily observable for the viral transactivator of transcription (TAT) protein. The PTM's in HIV TAT can be either activating, driving the virus towards proliferation and increased infection, or deactivating, rendering the virus in a latent, non-replicative state (Huo et al, 2011; D' Orso et al, 2009) Furthermore, the chemical diversity of these modifications includes acetylation, mono- di- and tri-methylation, phosphorylation, and ubiquitination. This necessity for such complex regulation at the level of PTM's can be rationalized by the fact that TAT expression levels and activity dictate the trajectory of viral infection. High levels of TAT promote the transcription of longer viral transcripts necessary for viral translation and viral genome packaging, while low levels of TAT only result in short transcripts that are unable to yield sufficiently long transcripts suitable for replication. It is this regulation by PTM's in TAT that determines whether an

infection should lie dormant or proceed to replication, and, ultimately the degree in which to do so.

As with many other viral proteins, HIV TAT redirects the function of a host protein to benefit viral survival. The positive transcription elongation factor B (P-TEFb) is a heterodimer of cyclin-dependent kinase 9 (CDK9) and its cognate cyclin, Cyclin T1. P-TEFb normally relieves transcriptionally arrest via the phosphorylation of the C-terminal domain (CTD) of a paused RNA Polymerase II on a host RNA. In the viral context, P-TEFb is recruited to nascently transcribed HIV RNA transactivation response element (TAR) where RNA Polymerase II is transcriptionally stalled. Upon phoshorylation by P-TEFb, RNA Polymerase II can then synthesize longer viral transcripts required for HIV proliferation. While it is unsurprising that P-TEFb itself is carefully regulated by PTM's, the fact the HIV can capitalize on changes in an accessory protein like P-TEFb is quite remarkable.

Recently, two crystal structures of P-TEFb and TAT along with the scaffolding protein AFF4 were solved independently (Gu et al 2014; Schulze-Gahmen et al 2013). This P-TEFb:TAT:AFF4 model serves as an excellent foundation for exploring how PTM's in TAT and CDK9 influence the molecular motions of a complex critical to HIV proliferation. Through the use of molecular modeling and molecular dynamics simulations, we investigated how two PTM's — Ser175 phosphorylation in CDK9 and Lys28 acetylation in TAT — altered the dynamics of the P-TEFb:TAT:AFF4 complex. While biophysical experiments utilizing homogenous populations of purified proteins is incredibly challenging, a computational approach to studying these PTM's is far more feasible. Of the many known PTM's available to model in this complex, these two

perturbations have experimental support suggesting the viral benefits of these PTM's. Namely, CDK9 Ser175 phosphorylation appears to have multifold consequences including: i.) stabilization of the P-TEFb:TAT complex, and 2.) modulation in the kinase activity of P-TEFb for its natural substrate, the CTD of RNA Pol II. TAT Lys28 acetylation has been determined to stabilize the TAT:TAR:P-TEFb complex and consequently increase viral transactivation.

In this study, we report mechanistic explanations of how these PTM's provide an advantage to viral transactivation over basal host transcription. Namely, CDK9 Ser175 phosphorylation and TAT Lys28 acetylation alter the molecular motions of key functional sites related to the P-TEFb:TAT:AFF4 complex. While these differences are not immediately obvious when comparing standard metrics of dynamics (e.g. RMSD, RMSF, etc), the implementation of the non-biased Kullback-Leibler divergence analysis revealed quantitatively distinct changes of the conformational landscape at the per residue level. By using this statistically rigorous analysis, we determined that TAT Lys28 acetylation is directly coupled to the mobility of the TAT-recognition motif (TRM) of Cylin T1, and allosterically coupled to the more distant (~40Å) CDK9 substrate binding pocket. On the other hand, CDK9 S175 phosphorylation stabilizes the P-TEFB:TAT complex by acting as a cationic sinkhole for nearby arginines and lysine's on both TAT and CDK9. In addition to providing mechanistic insight into the already experimentally characterized functions of these PTM's, we also propose that these PTM's may influence CDK9 substrate specificity and substrate affinity.

Results:

KL Divergence Analysis Identifies Differences in Molecular Motions of Important Functional Sites Upon PTM Incorporation

A non-biased statistical metric in the Kullback-Leibler (KL) divergence was used as an initial first pass analysis to understand how these PTM's altered the conformational landscape of the P-TEFb:TAT:AFF4 complex (McClendon et al, 2012). In brief, the KL divergence compares the relative entropies associated with a given residue between a reference ensemble (wildtype) and a perturbed ensemble (PTM). Because these relative entropies are calculated in torsional space using backbone and side chain dihedrals, the KL divergence captures meaningful differences in molecular motions between the two ensembles of interest. The distribution of the sum KL divergences for both phospho-Ser175 and acetyl-Lys28 demonstrate that the statically rigorous KL divergence analysis used here identifies the majority (> ~85%) of residues as having no significant deviation in the conformational landscape between the reference and the PTM ensemble. The distribution of sum and max KL divergence values are shown in **Supplementary Fig. 3.1a and 3.1b,** respectively**.**

Because the KL divergence is a per-residue metric like the more well-known atomic fluctuation or B-factor, KL divergence values can be mapped on structures to better visualize which areas of a structure are most greatly influenced by the perturbation of interest (in this case, a PTM). The sum KL divergence values were mapped onto the average structure across all 60ns of simulation time for each PTM **(Figure 3.1)**. Average structures for all 3 systems are shown in **Supplementary Fig. 3.2.** Interestingly, Lys28 acetylation primarily effects the nearby (~8Å) TAT-TAR recognition motif (TRM) of Cyclin T1 (Box I). Additionally, acetylation of TAT is also allosterically coupled to the substrate binding pocket of CDK9 (Box II). It is remarkable that

these two sites are not only dynamically coupled over a distance of ~40Å, but also that this signal is propagated from TAT through Cyclin T1 to CDK9. While Lys28 acetylation exhibits both local and long-range changes in conformational dynamics, Ser175 phosphorylation on CDK9 only alters the local conformational landscape of nearby residues on the activation loop of CDK9 (Box III). This non-biased approach identified 3 primary regions that were effected by PTM introduction, all of which serve as important functional sites for successful HIV transactivation.

TAT Acetylation Alters the Conformational Landscape of TAR-Interacting Residues in the Cyclin T1 TRM

The KL divergence analysis identified the Cyclin T1 TRM as a hot spot for differences in conformational sampling upon TAT acetylation. Interestingly, phosphorylation of Ser175 of CDK9 had no appreciable influence on TRM dynamics **(Figure 3.5)**. It is interesting to note that KL divergence values do not necessarily correlate with RMSF values, as residues with high KL divergence values (K253, R254, I255, W256) do not have outstandingly large changes in thermal fluctuation when compared to the wildtype. On the other hand, some residues with a lower but still significant KL divergence values (W258, R259, A260, C261) have ~ 2-fold higher B-factors compared to the wildtype. Overall, the TRM up until residue L252 shares the same restricted motion in the wild type and perturbed systems. However, only upon acetylation does the rest of the TRM K253 to C261) gain an increase in molecular motion as determined by the B-factor calculations. While RMSD is a common metric for dynamics, the backbone RMSD of the TRM in both the acetylated and phosphorylated systems fails to capture any deviation from the wild

type dynamics as highlighted by the KL divergence or even thermal fluctuation calculations (Supp. Figure #)

In addition to stabilizing the P-TEFb:TAT complex, the TRM of Cyclin T1 mediates interactions with HIV TAR. The residues necessary for forming these higher order P-TEFB:TAT:TAR complexes have been identified experimentally by alanine scanning followed by TAR gel shift assays (Garber et al, 1998). In particular, R254, I255, and W258 promote higher order P-TEFb:TAT:TAR complexation, but are dispensable for TAT binding. When comparing the conformational ensemble between the wild type and acetylated systems, R254 inherits a disparate backbone and side chain dihedral distributions. The phi sampling in the acetylated case shifts towards 0º, while the psi distribution is also translated towards 0º. In regards to side chain sampling, acetylation simply redistributes the dihedral populations at $\chi 1$ and $\chi 2$ without accessing any novel rotamers. **(Figure 3.7; Supp Figure 3.10)** Conversely, phosphorylation of CDK9 does not have any significant effect on R254 conformational sampling.

I255 has the largest KL divergence value in the TRM upon acetylation. There is an appreciable loss in entropy for both backbone dihedral angles as characterized by the narrowing of each distribution. Conversely, there is an apparent gain in entropy in both $\chi 1$ and $\chi 2$ as evident by the shift towards multi modality upon TAT acetylation. Isoleucine along with other amino acids with γ-heavy atoms have a pronounced relationship between their position in the Ramachandran plot and accessible rotamers (Dunbrack et al 1994). For I255, it appears that changes in backbone sampling accommodate a more uniform distribution of natural rotamers at $\chi 1$, around 60º, 180º, and -60º (**Figure 3.8; Supp Figure 3.11)**.

While the KL divergence value of residue W258 is only ~0.5, it too exhibits both changes in the thermal motion and torsional sampling. The phi distribution in the wildtype ensemble is both narrow and normal but slightly non-uniform, but upon TAT acetylation both broadens and deviates from normality to a larger degree The psi sampling for both the acetylated and wild type systems is bi-modal with means of ~0º and 150º, although TAT acetylation drives a population inversion shifting the preference towards the 0º peak. The $\chi 1$ distribution is multimodal in the presence and absence of TAT although there is a slight redistribution in the conformational landscape upon acetylation. In summary, TAT acetylation, but not CDK9 phosphorylation alters both the conformational landscape and overall molecular motion of residues in the TRM critical to TAR binding.

TAT Acetylation is Allosterically Coupled to the Substrate Binding Pocket of CDK9

Upon binding Cyclin T1 and phosphorylation of Thr186 in the activation loop, CDK9 relieves transcriptional pausing via the phosphorylation of both the CTD of RNA Polymerase II and transcriptional repressors bound to the RNA Pol II complex. Exactly which positions CDK9 phosphorylates and the associated combinatorial outcomes is still a point of contention. Interestingly, the KL divergence identified that substrate binding pocket of CDK9 to exhibit altered molecular motions upon TAT acetylation **(Supp Figure 3.4).** In contrast, phosphorylation of the nearby T-Loop of CDK9 does not regulate the dynamics of key residues in the substrate binding pocket. Alongside any redistribution in the conformational landscape of CDK9 residues 49-59, an overall increase in entropy is observed upon acetylation as well as evident through the backbone B-factors of these residues. Within this 11 residue stretch that feeds directly into the C-

helix of CDK9, there is a disproportionately high amount of charged residues (2 lysines, 3 glutamates). Because these residues may be involved in interacting or repulsing the heavily phosphorylated CTD, we first performed conformational analysis to better understand where these large KL divergence values originate. For example, E53 exhibits radically different sampling in phi sampling, shifting from trimodality to a single peak upon TAT acetylation. Phi sampling on the other hand is primarily translated towards 0º with a slight loss in entropy (**Supp Figure 3.13)**. χ1 and χ2 share the same expected profile with sampling primarily around 60, 180, and -60, however there is a redistribution in the populations upon TAT acetylation. The CDK9 substrate binding pocket interacts with the C-terminal domain of a transcriptionally arrested RNA polymerase II. Upon phosphorylation at some combination of Ser2, Ser5, and/or Ser7 of the repeated heptapeptide YSPTSPS, RNA Polymerase II

Ser175 Phosphorylation Nucleates an Electrostatic Network Between CDK9 and TAT

Ser175 of CDK9 becomes phosphorylated in T-Cells at least partially by the MAPK/ERK pathway upon stimulation with either Protein Kinase C agonist PMA (phorbol 12-myristate 13-acetate) or antibody-specific stimulation of CD3 or CD28. Along with phosphorylation of the nearby Thr186 also on the T-Loop, phosphorylation at Ser175 excludes interactions with the inhibitory 7SK snRNP -HEXIM1 complex, and thus acts as a marker for kinase competent P-TEFb in the cell. Additional biochemical studies with the phosphomimetic S175D suggest that phosphorylation changes binding partner specificity from BRD4 to TAT and enhances kinase activity with the CTD (Mbonye et al, 2013). Because this phosphorylation event has been experimentally characterized to have these profound influences on HIV transactivation, we

were curious how this PTM influenced the dynamics in the context of the P-TEFb:TAT:AFF4 complex.

In contrast to the TAT acetylation, the effects of phosphorylation of Ser175 do not propagate throughout the entire complex into TAT, CyclinT1, or even outside of the T-Loop **(Figure 3.5).** A closer look at the KL divergence values within the T-Loop indicate that K178 - with the largest KL divergence value of ~6) samples dihedral space dramatically different than the wild type ensemble. Upon inspection of the trajectories, it is clear that Ser175 phosphorylation nucleates an elaborate electrostatic network consisting of primarily of residues Lys144 and Lys 178 of CDK9 and Glu11 and Lys12 of TAT **(Supp Figure 3.15).** In agreement with this observation, the interatomic distance between the γO of Ser175 in CDK9 and the ζN of TAT Lys12 is considerably closer upon phosphorylation. **(Supp Figure 3.16).**

A closer look at the conformational analysis reveals that Lys143 of CDK9 only needs slight rotameric adjustments at χ3 and χ4 to accommodate the nearby phosphate, with no changes in backbone sampling **(Supp Figure 3.14).** Conversely, Lys178 located on the CDK9 T-Loop requires drastic alterations of to join this electrostatic network. Notably, each residues that deviates from wild type sampling loses appreciable entropy in order reap this enthalpic reward in the form of a strong electrostatic interaction. In regards to TAT's contributions to this network, Lys12 appears to undergo torsional strain in its backbone dihedrals as a consequence of interacting with the nearby phosphpate. Again, a loss of entropy at χ2 and χ4 is consistent with this electrostatically-driven conformational rearrangement.

Discussion: Post-translational modifications influence a vast number of cellular functions, such as protein-protein interactions, protein function, and protein stability, through disparate mechanisms. For example, a post-translational modification can induce drastic large-scale conformational change, more subtle redistributions of the energy landscape locally and globally, and any combination of the two. In this study, we explored how post-translational modifications critical to the HIV life cycle influence the dynamics of a large, multicomponent system containing both host and viral factors. Using molecular dynamics simulations and a non-biased statistical metric to compare the dihedral sampling of every ϕ, ψ, and χ in the complex, we identified three primary regions that were affected either by CDK9 phosphorylation or TAT acetylation. Importantly, these three regions correspond to critical functional sites that are critical to HIV transactivation. Namely, TAT acetylation altered the dynamics of both the nearby (~7Å) Cyclin T1 TRM and distant (~40Å) CDK9 Substrate Binding Pocket, while CDK9 phosphorylation redistributed the conformational landscape of nearby residues locally on the T-loop.

Lys28 acetylation by the acetyltransferase PCAF has been demonstrated to strengthen higher order TAT:TAR:Cyclin T1 complexes and, consequently, increase HIV transactivation (D' Orso et al, 2009). Interestingly, the acetyl-mimetic Lys28Gln is unable to enhance this TAT-driven HIV transactivation. Furthemore, a mutation with charge conservation (i.e. Lys28Arg) exhibits an equal magnitude of transactivation to the acetyl-mimetic. This suggests that in addition to a change in chemical properties (i.e. charged to uncharged residue), sequential rounds of acetylation and deacetylation might be necessary to achieve high levels of transactivation. Our simulation studies reveal that acetylation is dynamically coupled to the Cyclin T1 TRM and the

CDK9 substrate binding pocket. in regards to the Cyclin T1 TRM, we propose that acetylation liberates a previously conformationally restrained TRM. In conjunction with this increase in atomic fluctuation, residues critical to TAR binding sample drastically different conformers and rotamers, specifically, Arg254 and Trp256. While it is not trivial to actually elucidate the mechanism describing the physical coupling of two sites, in this instance it appears that W258 is at least partially involved in communicating this post-translational modification to the rest of the TRM. More specifically, in the wild type case, Trp258 intercalates between Lys28 andLys29. Energetically, this positioning is quite sensical, as it minimizes the loss of entropy from the surrounding solvent required to solvate a solvent exposed tryptophan. Additionally, this positioning of Trp258 between two lysine affords cation-π interactions and insulates the two adjacent positive charges emanating from the neighboring lysine residues. Perhaps to conserve this mode of regulation, Lys28 is invariable across sequence and species while the neighboring Lys29 position can accommodate an Arg, or a Trp, which could be stabilized by analogous cation-π interactions or π-π interactions, respectively. However, upon acetylation, Trp258 is ejected from its previous binding pocket to accommodate the arrival of the bulky acetylated-Lysine. After acetylation, Trp258 becomes solvent exposed and would desperately seek out a more hydrophobic hiding spot, such as between two nucleotides in HIV TAR.

In addition to protein dynamics, the TAR RNA landscape is populated by a major and a minor conformer. Interestingly, the ground state conformer of TAR (87% population) is characterized by more flexibility it U31 and A35, where in the ground state these bases are involved in base pairing. More specifically, the solvent exposed residues U31 and the adjacent G32 could serve as an intercalation site for W256 supporting π-π stacking of W258 with either U31 or G32. This

acetylation-driven interaction could then be further enhanced by through backbone electrostatic or base-specific interactions with TRM residues R251, K253, R254, and R259. This model is an agreement with the experimental observation that W256 is not critical to forming higher TAT:TAR:Cyclin T1 interactions in a non-acetylated context, but could facilitate this interaction in a acetylation-dependent manner.

Why would HIV transactivation would be mediated in large part by acetylation as opposed to another mode of PTM? Acetylation is sensitive it relative concentrations of Acetyl-CoA. Namely, in a nutrient depleted state, Acetyl-CoA concentrations drop while NAD+ concentrations increase, ultimately shifting the equilibrium towards deacetylation. Upon return to a healthier state marked by basal level or increased Acetyl-CoA, the equilibrium shifts back in favor of acetylation. This global shift in acetylation-deacetylation equilibrium potentially adds another layer of complexity to this mode of regulation. It follows that viral and host transcription — which is strongly regulated via histone acetylation — would want to proceed during a resource-rich state characterized high levels of Acetyl-CoA. (Verdin, 2014)

While most well-known in chromatic remodeling via histone acetylation, the underlying theme of regulation by post-translational acetylation is also evident in other non-histone transcriptional machinery for example in the transcription factors p53 and NF-κB. (Verdin, 2014)

Extended Discussion Thoughts:

- the introduction of this diatonic phosphate group acts as an electropositive sink for nearby cationic arginines and lysines.

- mechanism of acetylation in histone tail
- combinatorial outcomes of CTD phosphorylation (PTM complexity propogating further PTM complexity)
- combinatorial outcomes of multiple PTM's in the semi-SEC; both acetylation and methylation can occur on the same histone tails.
- introduction of ε-N-acetylation creates a neutral side chain from a cationic side chain and affords the binding of bromodomains
- While most well-known in chromatic remodeling via histone acetylation, the underlying theme of regulation by post-translational acetylation is also evident in other non-histone transcriptional machinery for example in the transcription factors p53 and NF-κB. (Verdin, 2014)
- An additional layer of complexity can be imagined when studying the PTM's of the entire PTEF-b:TAT:AFF4 complex combinatorially when considering that a particular PTM function can change depending on the relative PTM environment. For example, hetero-PTM markers have been observed on a given histone tail. (Walsh 2005)
- Global acetylation: coincidental that both HIV transactivation and heterochromatic remodeling occur within the same signal?
- "enhance HIV transcription"
- "proviral transcriptional elongation"
- many potential binding partners of the CTD including Mediator complex, histone methyltransferases, mRNA capping enzymes, and polyadenylation factors (RNA processing enzymes) (Buratowski et al 2003)

- TAT acetylation, which neutralizes a positive charge on lysine, has a pronounced effect on the CDK9 substrate binding pocket. Even though CDK9 phosphorylation is closer and of a greater magnitude, changing a neutral residue to a diatonic residues, TAT acetylation is the only PTM to change the dynamics of the substrate binding pocket. While this is intuitively perplexing at first, this can be rationalized by the fact the the electrostatic effect of the phosphorylation is neutralized immediately by the surrounding lysine and arginine residues of TAT and CDK9.
- it is interesting to note that this electrostatic network is mediated by lysine's and not arginines.
- TCR stimulation triggers a number of events necessary for HIV rebound such as:
  - expression of hCycT1 (which is normally kept low in resting CD4+ cells by miRNA)
  - phosphorylation of Ser175 of CDK9
  - translocation of NFAT and NF-KB to the nucleus to encourage remodeling of the HIV LTR
  - promotes dissociation of CDK9 from 7SK snRNP complex (D' Orso et al, 2010)

Methods: Parameter files for post-translational modifications were generated by first building the non-natural amino acid in Maestro (Schrodinger Suite), followed by parameter generation using the Schrodinger utility hetgrp_ffgen. Post-translational modifications were then built into the starting model (provided by Ursula Schulze-Gahmen prior to deposition into the PDB) using the modeling program PLOP. If parameter files for non-protein components were available (e.g. Zn2+, *http://www.mayo.edu/research/labs/computer-aided-molecular-design/projects/zinc-protein-simulations-using-cationic-dummy-atom-cada-approach)* these were used. The AMBER simulations package (Case et al, 2013) was used for simulations. After solvent and system

minimization, systems were allowed to equilibrate for 300ps. These equilibrated systems were then used for 6 x 10ns independent simulations. Simulation data was extracted using Amber Tools and data was processed in MATLAB. KL Divergence calculations were performed with as instructed in McClendon et al, 2015.

Other global metrics of overall dynamics can be found in the figures below. Many of these calculations serve as important controls but did warrant acknowledgement in the this manuscript.

Acknowledgements: Thank you to K. Lexa for help setting up simulations, K. Ball for help with dihedral angle extraction scripts, C. McClendon for troubleshooting KL-Divergence software.

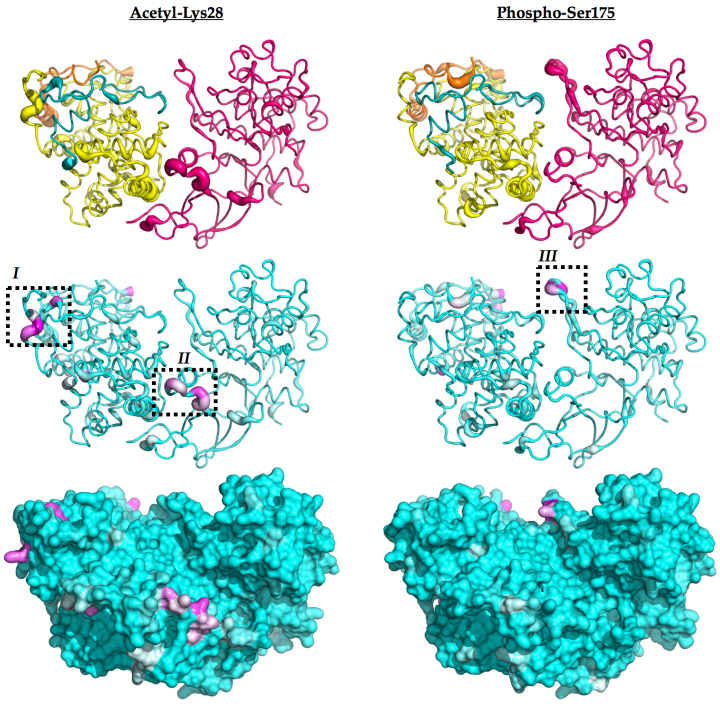


Figure 3.1. Functional Sites Critical to TAT-mediated HIV Transactivation Exhibit Sample A Different Conformational Landscape Upon Acetylation or Phosphorylation. The Kullback-Leibler divergence values were mapped on to the average structure for each of the models. In this case, KL Divergence values for each residue are equal to the sum KL value of each dihedral (ϕ, ψ, χ) within the residue. KL Divergence values increase with the thickness and according to the color spectrum (below). The wildtype complex was used as the reference in the KL divergence calculations. Acetyl-Lys28 (Left), Phospho-Ser175 (right). Cdk9, *pink;* Cyclin T1, *yellow;* HIV TAT, *teal;* AFF4, *orange* (top). Putty Representation with KL Divergence spectrum coloring (middle); Surface Representation with KL Divergence spectrum coloring (middle). Functional Sites: TRM of CyclinT1 (*Box I*), Substrate Binding Pocket of CDK9 *(Box II)*, T-Loop of CDK9 *(Box III)*.

Figure 3.2.

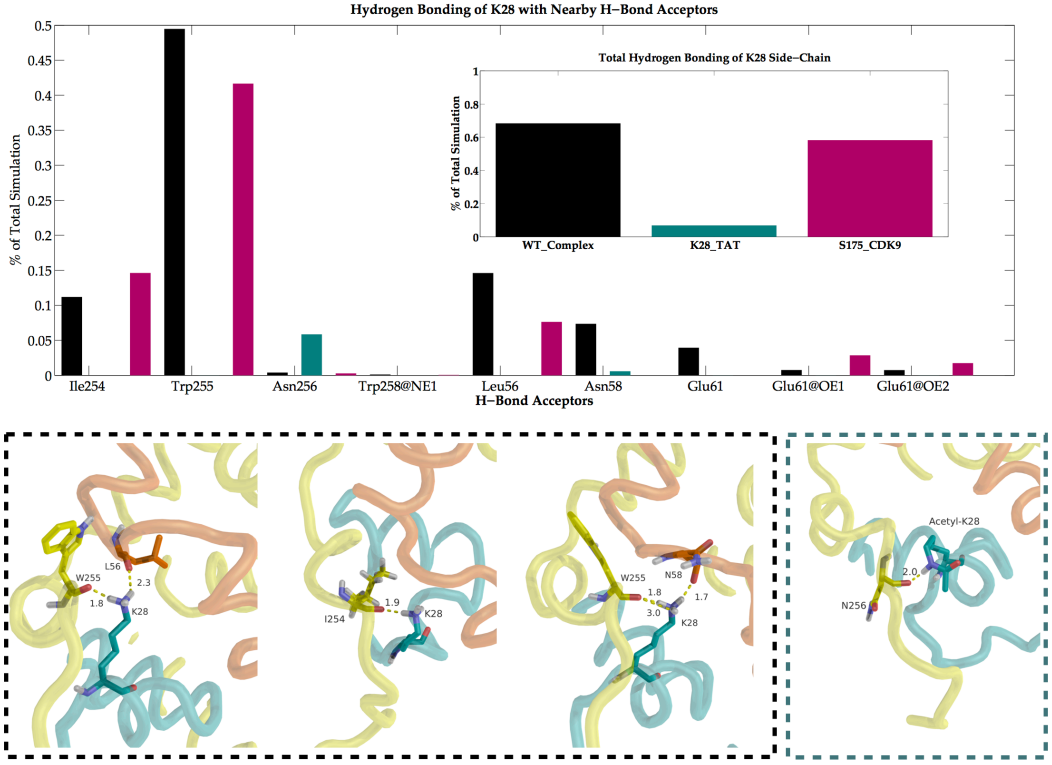

Hydrogen Bonding of K28 with Nearby H−Bond Acceptors
% of Total Simulation
H−Bond Acceptors
Ile254
Trp255
Asn256
Trp258@NE1
Leu56
Asn58
Glu61
Glu61@OE1
Glu61@OE2
Total Hydrogen Bonding of K28 Side−Chain
WT_Complex
K28_TAT
S175_CDK9
L56
W255
K28
I254
N58
Acetyl-K28
N256

Figure 3.3 The TRM of Cyclin T1 Samples A Drastically Different Conformational Landscape Upon TAT Acetylation but Not CDK9 Phosphorylation. Sum KL values for both post-translationally modified semi-SEC models for the TAT-TAR Recognition Motif (Cyclin T1 residues 245-264).

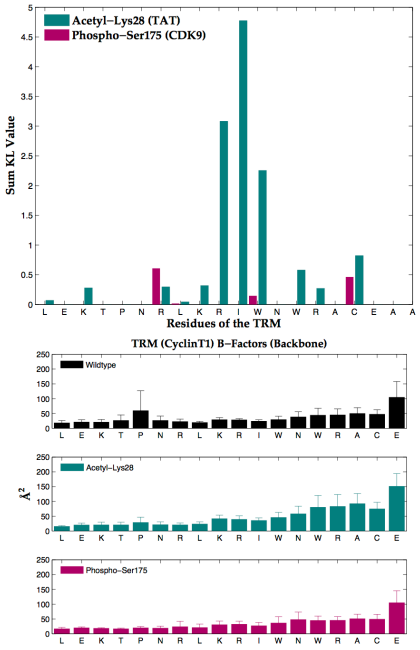


Figure 3.4. Acetylation of TAT Lys28 Is Allosterically Coupled to the CDK9 Substrate Binding Pocket. Sum KL Values corresponding to the CDK9 substrate binding pocket (residues 49 - 59) for both post-translationally modified systems.

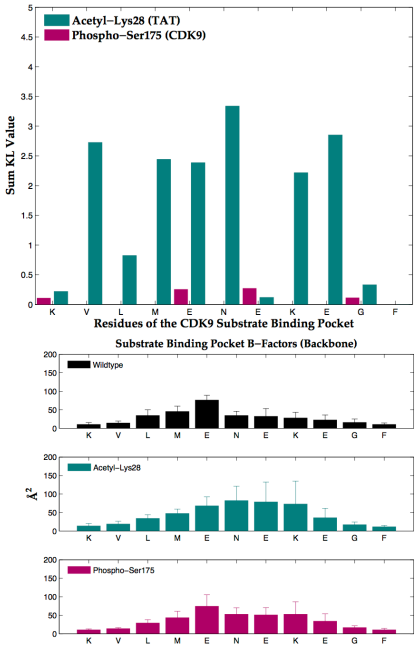


Figure 3.5. Phosphorylation of Ser175 of CDK9 Alters the Conformational Dynamics of Nearby Residues but Does Not Exert Any Appreciable Influence Over Long-Ranges. Sum KL values for both post-translationally modified semi-SEC models for the CDK9 Activation Loop (CDK9 residues 167-198).

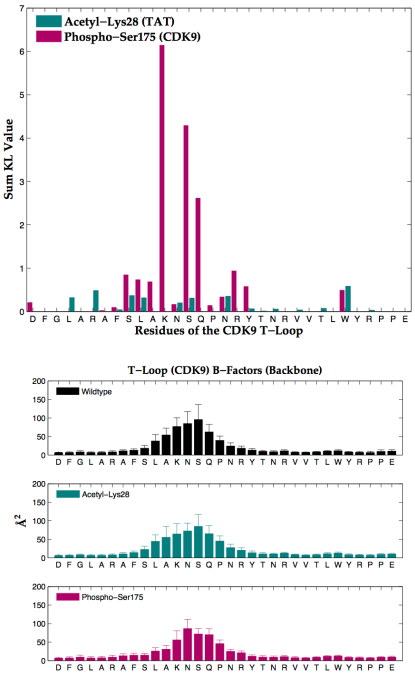
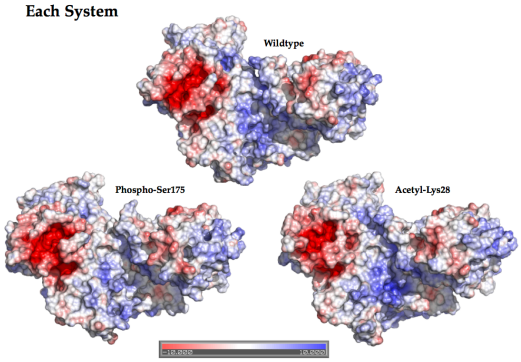
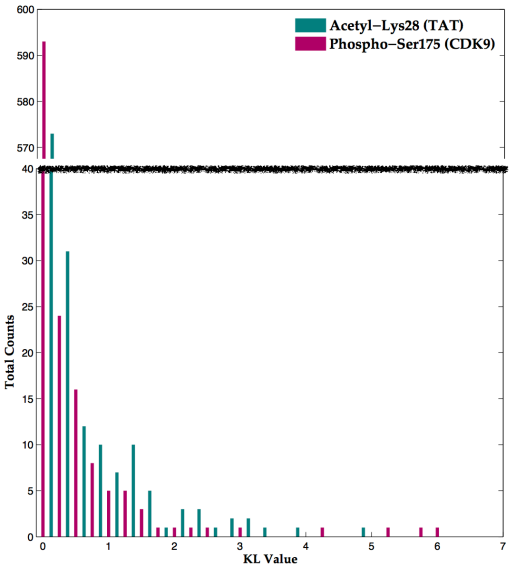


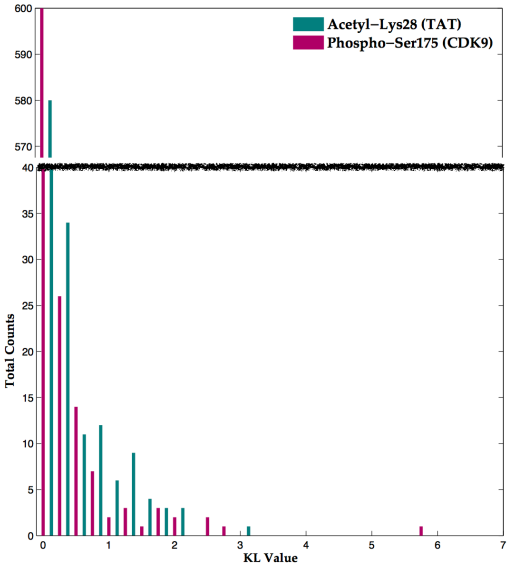
Figure 3.6. Poisson-Boltzmann Calculations Reveal Differences in Electrostatic Potential in the Substrate Binding Pocket in Each System



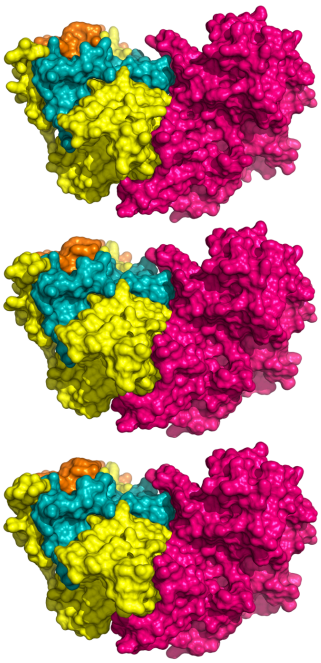
Supplementary Figure 3.1a. Sum KL Divergence Distribution for Each System. A histogram (bin-width = 0.25) of the sum KL Divergence by residue (sum value for all ϕ, ψ, and χ of a given residue). Note the majority of residues exhibit no statistically significant difference in KL values between the wildtype semi-SEC and the post-translationally modified semi-SEC (i.e. KL divergence = 0).



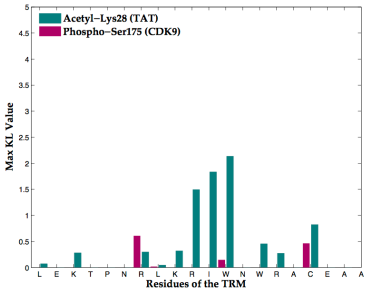
Supplementary Figure 3.1b. Max KL Divergence Distribution for Each System. A histogram (bin-width = 0.25) of the max KL Divergence by residue (max value for any ϕ, ψ, or χ of a given residue). Note the majority of residues exhibit no statistically significant difference in KL values between the wildtype semi-SEC and the post-translationally modified semi-SEC (i.e. KL divergence = 0).



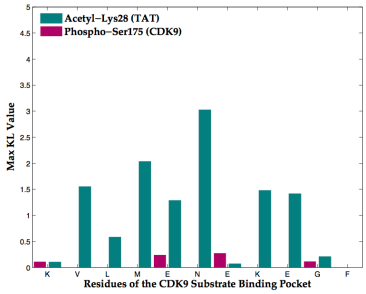
Supplementary Figure 3.2. Average Structures of the Wildtype and Post-translationally Modified Semi-SEC. Average structures for each of the models were generated across all 60ns of simulation time. Cdk9, *pink;* Cyclin T1, *yellow;* HIV TAT, *teal;* AFF4, *orange*. Wildtype (top), Acetyl-Lys28 (middle), Phospho-Ser175 (bottom)



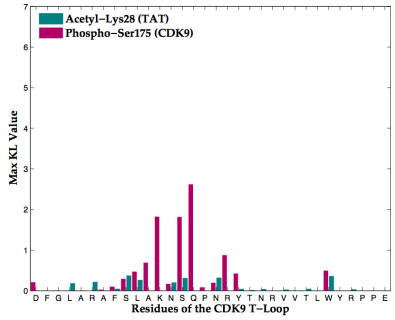
Supplementary Figure 3.3 The TRM of Cyclin T1 Samples A Drastically Different Conformational Landscape Upon TAT Acetylation but Not CDK9 Phosphorylation. Max KL values for both post-translationally modified semi-SEC models for the TAT-TAR Recognition Motif (Cyclin T1 residues 245-264).



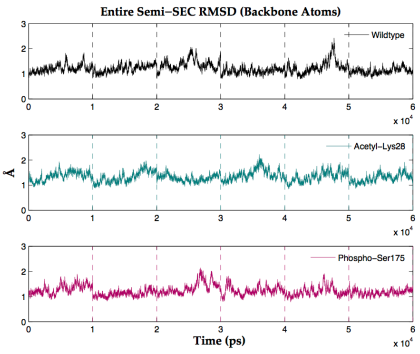
Supplementary Figure 3.4. Acetylation of TAT Lys28 Is Allosterically Coupled to the CDK9 Substrate Binding Pocket. Max KL Values corresponding to the CDK9 substrate binding pocket (residues 49 - 59) for both post-translationally modified systems.



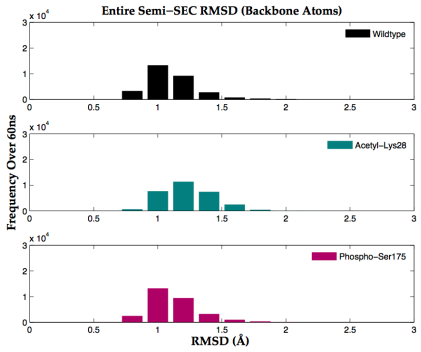
Supplementary Figure 3.5. Phosphorylation of Ser175 of CDK9 Alters the Conformational Dynamics of Nearby Residues but Does Not Exert Any Influence Over Long-Ranges. Max KL values for both post-translationally modified semi-SEC models for the CDK9 Activation Loop (CDK9 residues 167-198).



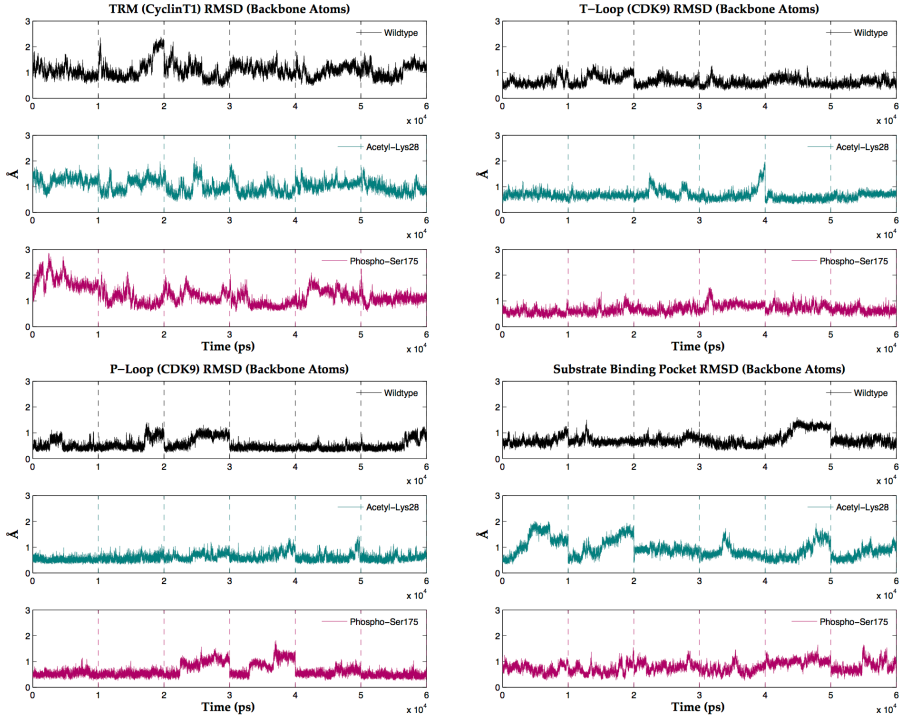
Supplementary Figure 3.6. Backbone atom RMSD (time-dependent) of the semi-SEC complex. RMSD's were calculated using the average structure from all 60ns of simulation time as a reference. Each 10000ps (delineated by dashed vertical line) indicates a single 10ns simulation (6 in total for each system).



Supplementary Figure 3.7. Distribution of backbone atom RMSD of the entire semi-SEC complex. RMSD's were calculated using the average structure from all 60ns of simulation time as a reference. (bin size = 0.2Å)



Supplementary Figure 3.8. Backbone atom RMSD of important functional sites in the semi-SEC. RMSD's were calculated using the average structure from all 60ns of simulation time as a reference. Each 10000ps (delineated by dashed vertical line) indicates a single 10ns simulation (6 in total for each system).



Supplementary Figure 3.9. Distribution of backbone atom RMSD of important functional sites in the semi-SEC. RMSD's were calculated using the average structure from all 60ns of simulation time as a reference. (bin size = 0.2Å)

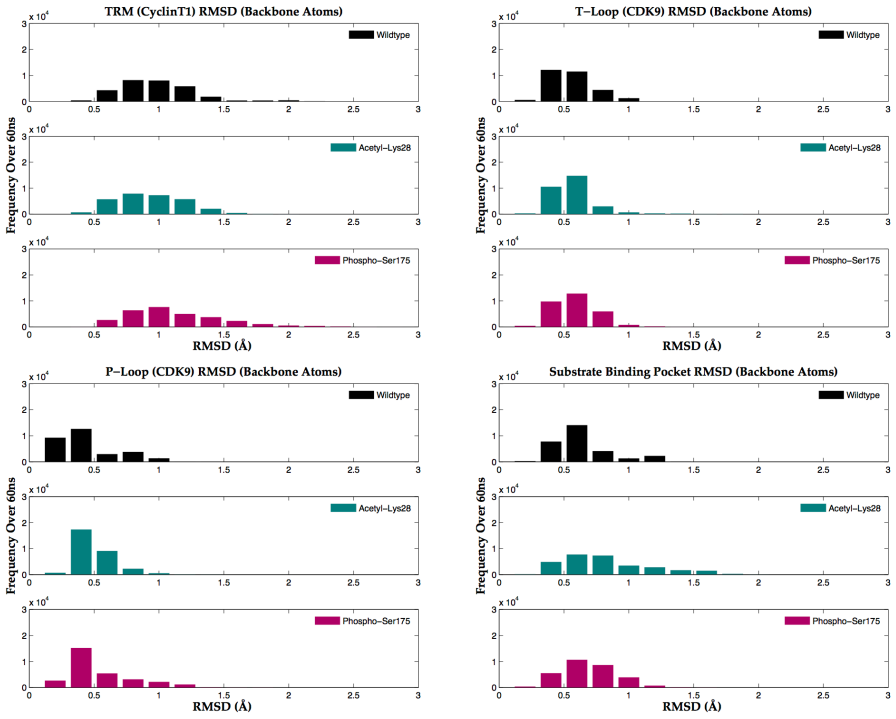
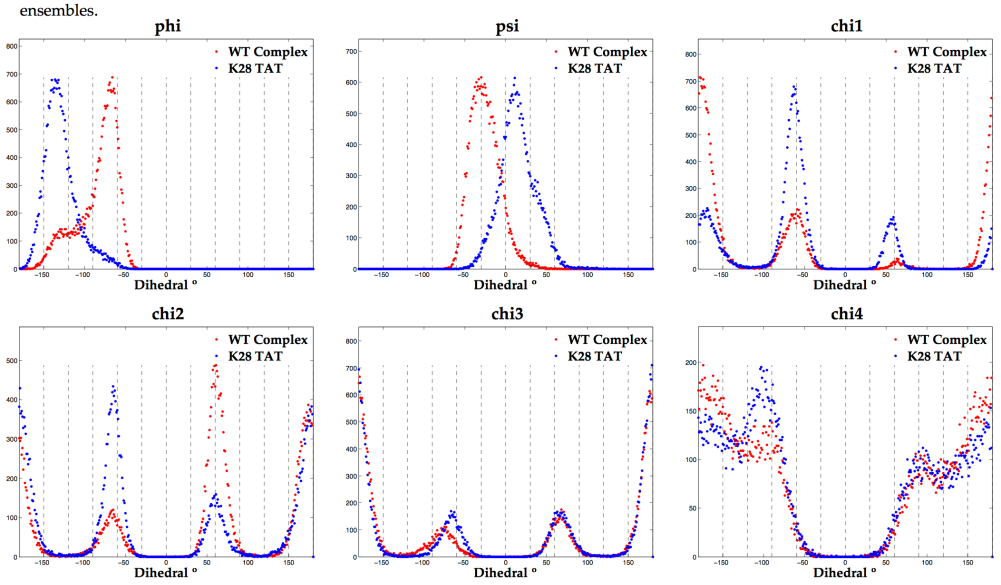


Figure 3.7. Conformational Analysis for R254 (~248) for Backbone and Side Chain Dihedral Angles Upon TAT Acetylation. Each dihedral angle was calculated across all 60ns of simulation and binned every 1°. Dashed vertical lines show the bin width (30°) used in calculating the KL divergence to help better visualize the relative entropies between the reference (wildtype) and the perturbed (PTM) ensembles.



Supplementary Figure 3.10. Conformational Analysis for R254 (~248) for Backbone and Side Chain Dihedral Angles Upon CDK9 Phosphorylation. Each dihedral angle was calculated across all 60ns of simulation and binned every 1°. Dashed vertical lines show the bin width (30°) used in calculating the KL divergence to help better visualize the relative entropies between the reference (wildtype) and the perturbed (PTM) ensembles.

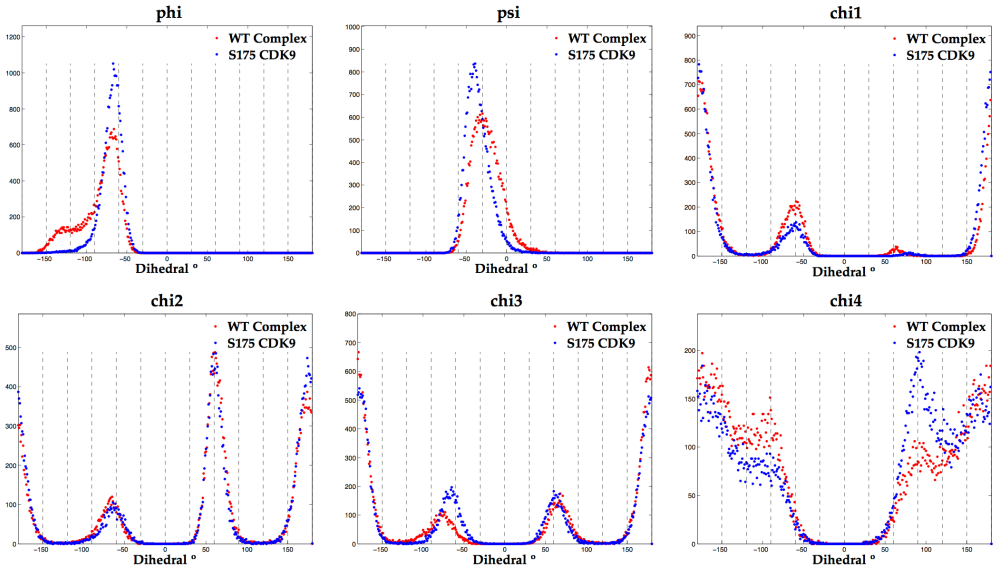
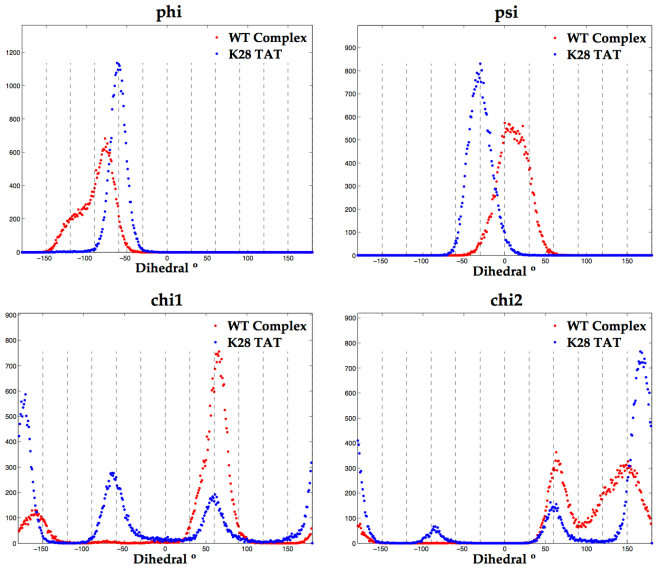


Figure 3.8. Conformational Analysis for I255 (~249) for Backbone and Side Chain Dihedral Angles Upon TAT Acetylation. Each dihedral angle was calculated across all 60ns of simulation and binned every 1°. Dashed vertical lines show the bin width (30°) used in calculating the KL divergence to help better visualize the relative entropies between the reference (wildtype) and the perturbed (PTM) ensembles.



Supplementary Figure 3.11. Conformational Analysis for I255 (~249) for Backbone and Side Chain Dihedral Angles Upon CDK9 Phosphorylation. Each dihedral angle was calculated across all 60ns of simulation and binned every 1°. Dashed vertical lines show the bin width (30°) used in calculating the KL divergence to help better visualize the relative entropies between the reference (wildtype) and the perturbed (PTM) ensembles.

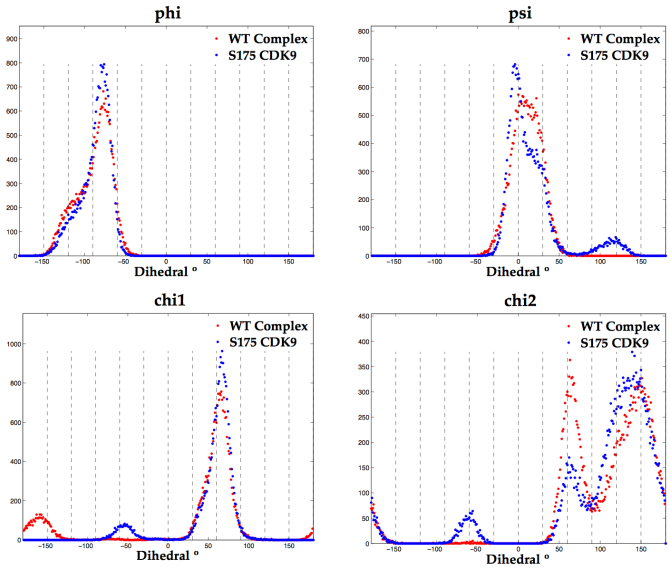
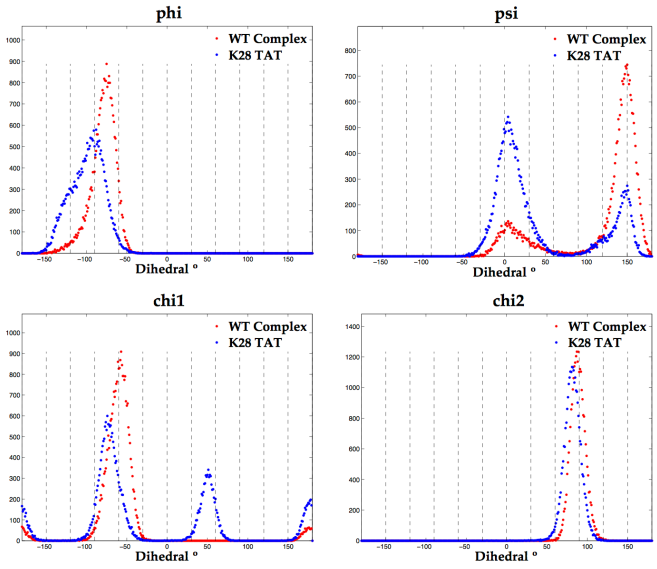


Figure 3.9. Conformational Analysis for W258 (~252) for Backbone and Side Chain Dihedral Angles Upon TAT Acetylation. Each dihedral angle was calculated across all 60ns of simulation and binned every 1°. Dashed vertical lines show the bin width (30°) used in calculating the KL divergence to help better visualize the relative entropies between the reference (wildtype) and the perturbed (PTM) ensembles.



Supplementary Figure 3.12. Conformational Analysis for W258 (~252) for Backbone and Side Chain Dihedral Angles Upon CDK9 Phosphorylation. Each dihedral angle was calculated across all 60ns of simulation and binned every 1°. Dashed vertical lines show the bin width (30°) used in calculating the KL divergence to help better visualize the relative entropies between the reference (wildtype) and the perturbed (PTM) ensembles.

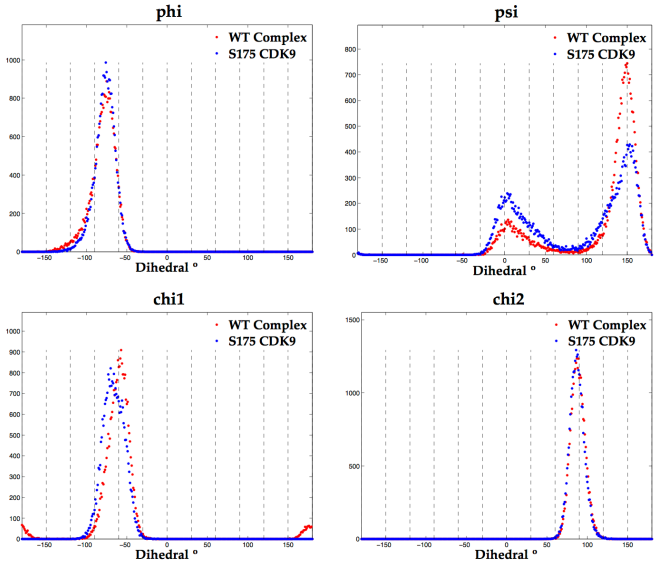
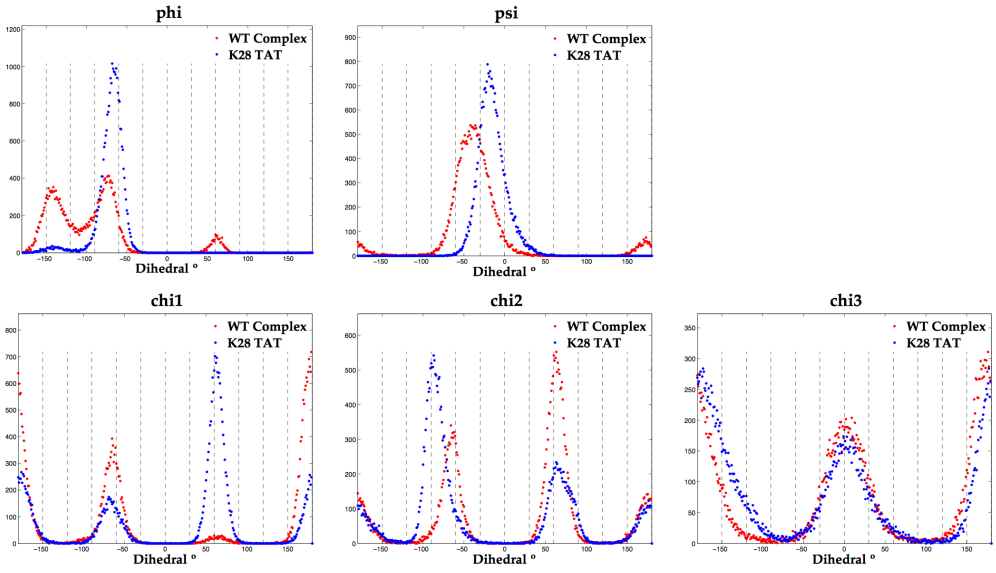
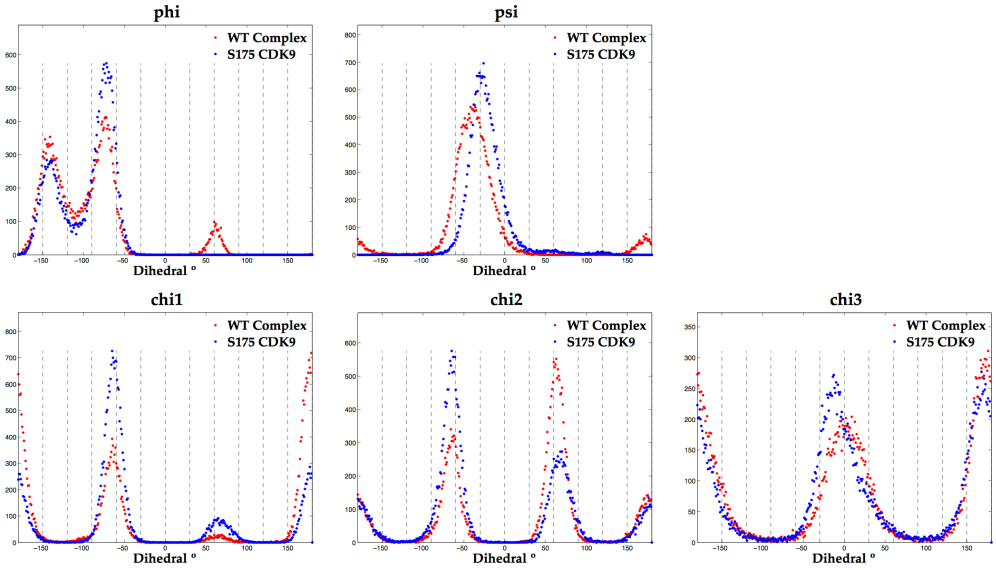


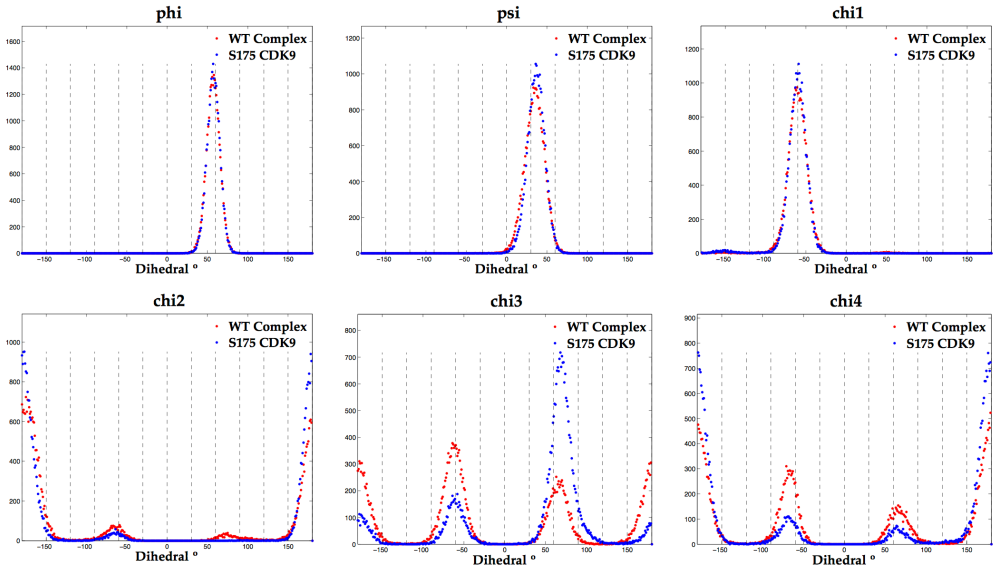
Figure 3.10. Conformational Analysis for E53 (CDK9) for Backbone and Side Chain Dihedral Angles Upon TAT Acetylation. Each dihedral angle was calculated across all 60ns of simulation and binned every 1°. Dashed vertical lines show the bin width (30°) used in calculating the KL divergence to help better visualize the relative entropies between the reference (wildtype) and the perturbed (PTM) ensembles.



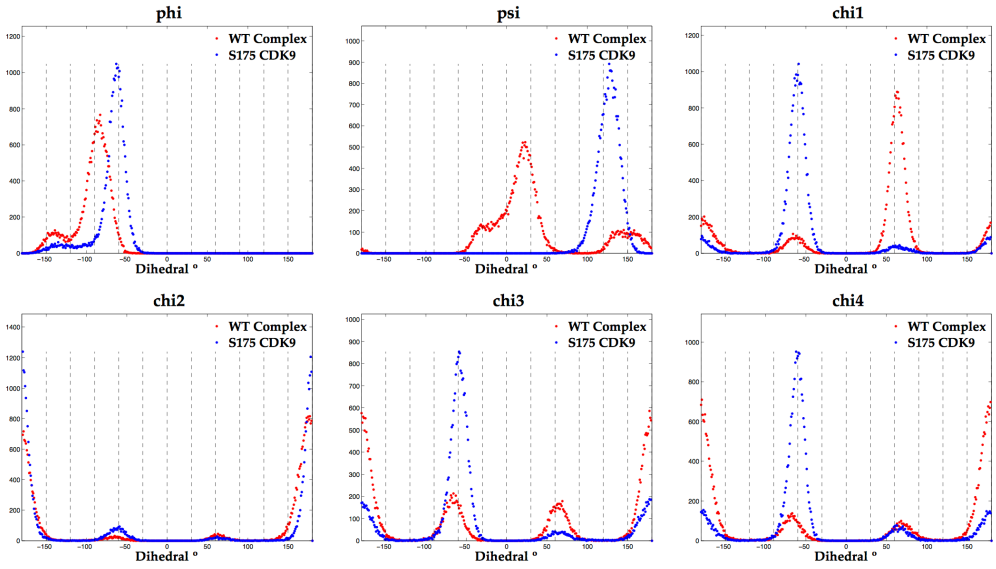
Supplementary Figure 3.13. Conformational Analysis for E53 (CDK9) for Backbone and Side Chain Dihedral Angles Upon CDK9 Phosphorylation. Each dihedral angle was calculated across all 60ns of simulation and binned every 1°. Dashed vertical lines show the bin width (30°) used in calculating the KL divergence to help better visualize the relative entropies between the reference (wildtype) and the perturbed (PTM) ensembles.



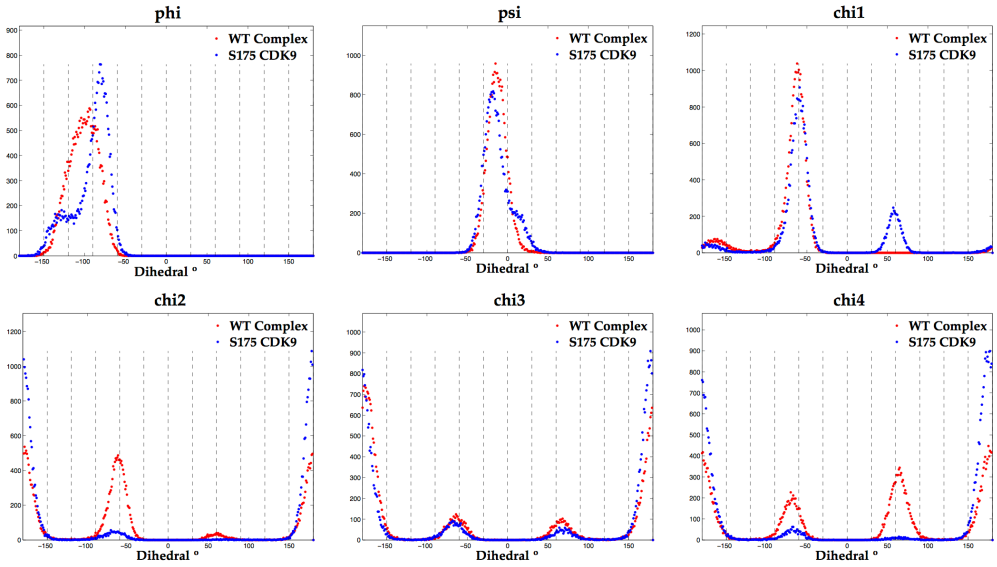
Supplementary Figure 3.14. Conformational Analysis for K143 (CDK9) for Backbone and Side Chain Dihedral Angles Upon CDK9 Phosphorylation. Each dihedral angle was calculated across all 60ns of simulation and binned every 1°. Dashed vertical lines show the bin width (30°) used in calculating the KL divergence to help better visualize the relative entropies between the reference (wildtype) and the perturbed (PTM) ensembles.



Supplementary Figure 3.15. Conformational Analysis for K178 (CDK9) for Backbone and Side Chain Dihedral Angles Upon CDK9 Phosphorylation. Each dihedral angle was calculated across all 60ns of simulation and binned every 1°. Dashed vertical lines show the bin width (30°) used in calculating the KL divergence to help better visualize the relative entropies between the reference (wildtype) and the perturbed (PTM) ensembles.



Supplementary Figure 3.16. Conformational Analysis for K12 (TAT) for Backbone and Side Chain Dihedral Angles Upon CDK9 Phosphorylation. Each dihedral angle was calculated across all 60ns of simulation and binned every 1°. Dashed vertical lines show the bin width (30°) used in calculating the KL divergence to help better visualize the relative entropies between the reference (wildtype) and the perturbed (PTM) ensembles.



References:

Buratowski S (2003) The CTD code. Nat Struct Biol 10: 679–680.

Case, D.A, D.S. Cerutti, T.E. Cheatham, III, T.A. Darden, R.E. Duke, T.J. Giese, H. Gohlke, A.W. Goetz, D. Greene, N. Homeyer, S. Izadi, A. Kovalenko, T.S. Lee, S. LeGrand, P. Li, C. Lin, J. Liu, T. Luchko, R. Luo, D. Mermelstein, K.M. Merz, G. Monard, H. Nguyen, I. Omelyan, A. Onufriev, F. Pan, R. Qi, D.R. Roe, A. Roitberg, C. Sagui, C.L. Simmerling, W.M. Botello-Smith, J. Swails, R.C. Walker, J. Wang, R.M. Wolf, X. Wu, L. Xiao, D.M. York and P.A. Kollman (2017), AMBER 2017, University of California, San Francisco.

Dunbrack RL, Karplus M (1994) Conformational analysis of the backbone-dependent rotamer preferences of protein sidechains. Nat Struct Biol 1: 334–340.

D'Orso I, Frankel AD (2009) Tat acetylation modulates assembly of a viral-host RNA-protein transcription complex. Proc Natl Acad Sci U S A 106: 3101–3106.

D'Orso I, Frankel AD (2010) RNA-mediated displacement of an inhibitory snRNP complex activates transcription elongation. Nat Struct Mol Biol 17: 815–821.

Garber ME, Wei P, KewalRamani VN, Mayall TP, Herrmann CH, et al. (1998) The interaction between HIV-1 Tat and human cyclin T1 requires zinc and a critical cysteine residue that is not conserved in the murine CycT1 protein. Genes Dev 12: 3512–3527.

Gu J, Babayeva ND, Suwa Y, Baranovskiy AG, Price DH, et al. (2014) Crystal structure of HIV-1 Tat complexed with human P-TEFb and AFF4. Cell Cycle 13: 1788–1797.

Huo L, Li D, Sun X, Shi X, Karna P, et al. (2011) Regulation of Tat acetylation and transactivation activity by the microtubule-associated deacetylase HDAC6. J Biol Chem 286: 9280–9286.

Mbonye UR, Gokulrangan G, Datt M, Dobrowolski C, Cooper M, et al. (2013) Phosphorylation of CDK9 at Ser175 Enhances HIV Transcription and Is a Marker of Activated P-TEFb in CD4+ T Lymphocytes. PLoS Pathog 9.

McClendon CL, Hua L, Barreiro A, Jacobson MP (2012) Comparing Conformational Ensembles Using the Kullback-Leibler Divergence Expansion. J Cher Theory Comput 8: 2115–2126.

Schulze-Gahmen U, Upton H, Birnberg A, Bao K, Chou S, et al. (2013) The AFF4 scaffold binds human P-TEFb adjacent to HIV Tat. Elife 2013: 1–14.

Verdin E, Ott M (2014) 50 Years of Protein Acetylation: From Gene Regulation To Epigenetics, Metabolism and Beyond. Nat Rev Mol Cell Biol: 1–7.

Walsh CT, Garneau-Tsodikova S, Gatto GJ (2005) Protein Posttranslational Modifications: The Chemistry of Proteome Diversifications. Angew Chemie Int Ed 44: 7342–7372.

CHAPTER 4

Transcriptional Logic In Cardiac Reprogramming

This chapter describes early graduate work on understanding the relationship of two transcription factors that, when in conjunction with a 3rd, are entirely capable of reprogramming heart scar tissues (cardiac fibroblasts) into functional beating heart cell cells (cardiomyocytes). This work was largely pioneered by a researcher at the Gladstone Institute at UCSF, Deepak Srivastava. During my time, I met with Deepak to discuss my research goals and experimental design. He also served as a member of my qualifying exam committee. This conversations were invaluable to, at least, understanding the deeper biology of this reprogramming phenomenon. This project was ultimately abandoned after initial successes in crystallizing Nurr1 and after years of struggling to biochemically reconstitute the ternary complexes of Gata4/Tbx5 and a tandem response element in the form of DNA.

What Follows is a Proposal to the American Heart Association Regarding this Project and its Implications to Regenerative Medicine

Project Narrative

As the human heart has little to no regenerative capabilities on its own, therapeutic intervention to help repair a damaged heart --as in the case of a person afflicted with cardiovascular disease-- is both appealing and necessary. One promising approach that essentially rejuvenates a compromised heart back into a viable, healthy heart is the minimally invasive administration of a few critical proteins, called cardiac transcription factors, which regulate cardiac gene expression. Our goal is to reveal the molecular underpinnings of just how these cardiac transcription factors communicate with each other in different DNA environments in order to provide the insight fundamental to further advancing this therapy into humans.

Project Summary

Gata4 and Tbx5 are two cardiac transcription factors that, when in conjunction with a third transcription factor Mef2c, are capable of directly reprogramming cardiac fibroblasts into functional cardiomyocytes. A mechanistic understanding of how these proteins cooperate with each other will be pivotal to further advancing this potential therapy for rejuvenating a damaged heart. Despite the increasing number of instances in which just a few key transcription factors have been shown to initiate cellular reprogramming, the molecular underpinnings of this particular event are largely unknown. At the moment, the biochemical and functional relationship between Tbx5 and Gata4 is best understood (when compared to Mef2c & Tbx5/ Gata4). In addition to co-localizing regulatory regions in the genome, Tbx5 and Gata4 have

shown to achieve transcriptional synergy in the context of a number of important cardiac specific genes. Additionally, a single point mutation in Gata4 is enough to abrogate a specific protein-protein interaction with Tbx5. Moreover, the disruption of this interaction results in abnormalities in heart development, particularly congenital septal defects. We hypothesize that these disparate arrangements of Tbx5 and Gata4 tandem response elements (TRE's) allow for different routes of ternary complex assembly on DNA. Furthermore, the binding of these two transcription factors on disparate DNA's will generate a range of unique 3-dimensional topologies. Finally, it is these DNA dependent topologies that serve as binding surfaces for co-regulators from various signaling pathways. It is through this mechanism that a number of complex signaling pathways can integrate information through merely three proteins in order to undergo drastic cellular metamorphosis.

Specific Aims

Myocardial infarction often results in the transformation of once functional heart tissue into fibrotic scar tissue. This scar tissue, composed of cardiac fibroblasts, decreases overall heart function while increasing the susceptibility for further detrimental cardiac events. It was demonstrated that the administration of key cardiac transcription factors Mef2c, Tbx5, and Gata4 are capable of directly reprogramming fibroblasts into functional cardiomyocytes. We will investigate how two of these three factors --namely, Tbx5 and Gata4 -- coordinate with each other in solution and in different DNA environments to promote this cellular repair process. There are numerous instances in which cardiac transcription factors interact with each other in order to modulate transcriptional activity (e.g. Gata4 and Tbx5 on both the *NPPA* and *GJA5*

promoter). As well, Tbx5 and Gata4 appear to be involved in a physical protein-protein interaction. The abrogation of this interaction by a single point mutation in Gata4 (G296S) results in the most common observed developmental defect in live births -- congenital heart defects. While these two proteins have been, at least to some degree, structurally characterized individually as crystal or solution structures, there is still minimal insight into how these factors engage with each other at the biochemical level. We will be focusing on understanding the relationship between Gata4 and Tbx5; this pair was chosen based on previous studies demonstrating (i) a protein-protein interaction between the two (ii) transcriptional synergy exhibited between the two factors on important cardiac specific genes (iii) an altered gene expression pattern when these two proteins do not interact properly. We plan to uncover the biochemical details that allow two transcription factors to share information with each other in order to initiate cardiac reprogramming and proper heart development.

Specific Aim 1: Biophysically characterize the formation of Gata4:Tbx5 heterodimers and Gata4:Tbx5:DNA ternary complexes.

Rationale

Tbx5 and Gata4 have been shown to co-occupy regulatory regions in the genome in numerous instances. These tandem response elements (TRE's) are diverse in sequence, spacing, orientation, and polarity. The goal of this aim is to biophysically characterize how Tbx5 and Gata4 assemble on disparate DNA contexts. We will utilize the Biacore T100 in our lab to study the kinetics of both heterodimerization between Gata4 and Tbx5 and ternary complex assembly between Gata4:Tbx5:DNA. A number of single point mutants of Gata4 will be screened against

wildtype Tbx5 to identify and characterize mutations that disrupt this protein-protein interaction. Additionally, We will investigate 10 different oligos derived from regulatory regions 10kb upstream of genes effected by the disruption of a Gata:Tbx5 interaction. This observation was revealed in the analysis of gene expression data in knock-in mice that were heterozygous $^{G295S/+}$. Each of these TRE's contain cognate RE's for Tbx5 and Gata4 within 10bp of each other-- this criterion will make these probes amenable to SPR and for structural characterization (Aim2).

Specific Aim 2: Reveal at the atomic level (i) how Tbx5 and Gata4 engage with each other on disparate tandem response elements (ii) how a subtle mutation in Gata4 disrupts interactions with Tbx5.

Rationale

To investigate how different TRE's guide the formation of unique Tbx5 and Gata4 ternary complexes, we will analyze be size-exclusion chromatography the regulatory regions (from Aim 1) that promote stable ternary complex formation. Data collected from Aim 1 would also help to identify regulatory regions that promote this macromolecular assembly. Our hypothesis is that each ternary complex forms a unique binding surface for a cognate co-regulator. In this manner, a range of signaling pathways could integrate information into transcriptional regulation through binding a co-regulator to a particular topology. Small-angle X-ray scattering (SAXS) could also be employed to reveal these DNA driven topologies. Molecular dynamics will be used to examine how the protein motions of single point mutants vary from the dynamics of wildtype Tbx5 and Gata4. As there is no solved structure for Gata4, a model will be built off of

the highly homologous mouse Gata3 structure. These computational studies could reveal how these mutations disrupt the protein:protein interactions necessary for the transcription of cardiac genes.

Specific Aim 3: Assess Tbx5 and Gata4 synergy in regulatory regions of cardiac specific genes through luciferase assays

Rationale

Tbx5 and Gata4 have been previously shown to cooperate to regulate cardiac specific genes (e.g. *Nppa* or *Gja5*) at the transcriptional level. Our goal is to expand upon these previous studies by using the 10 regulatory regions examined in Aim 1 in luciferase assays. By using a number of different minimal response elements, it can be further revealed the extent to which these factors regulate important genes. Of interest is comparing the functional output from Aim 3 in the form of transcriptional activation to the biophysical output from Aim 1 to examine how the mechanistic details of complex assembly factor are coupled to transcriptional regulation.

Research Strategy

Background and Significance

The overarching theme of this proposal is understanding the relationship between two cardiac transcription factors (TF's)--Tbx5 and Gata4. We hope to learn how these factors interact with each other in solution and in different DNA contexts in order to regulate cardiac related gene expression. These transcription factors are critical to study from a biomedical prospective for three reasons:

1. Tbx5 and Gata4 are critical to proper heart development. Non-synonymous mutations in either gene lead to congenital heart defects, the most common congenital abnormalities in live births.
2. alongside Mef2C, Tbx5 and Gata4 can reprogram cardiac fibroblasts into cardiomyocytes. The administration of these factors is in development for the treatment of cardiovascular disease.
3. Tbx5 and Gata4 serve as a model system for understanding the transcriptional complexity in eukaryotic gene expression.

A Second Chance at a Healthy Heart

Cardiovascular disease is a the leading cause of mortality world wide. In addition to a number of other contributing factors to this statistic, the adult human heart has little to no independent regenerative capabilities.. In other words, a heart damaged by, for example a myocardial infarction, will not only not get better with time, but will likely get much worse. One consequence of infarction is the formation of fibrotic scar tissue on the heart. This scar tissue, in the form of cardiac fibroblasts, sequesters tissue normally occupied by functional cardiomyocytes; this replacement leads to decreased cardiac function and potentially death. Recently, it has been discovered that a few key regulators in the form of TF's can reprogram one cell type into another, . Particularly, the retroviral administration of Gata4, Mef2c, and Tbx5 -- 3 cardiac TF's important in heart development -- can successfully transform cardiac fibroblasts into cardiomyocytes. The induced cardiomyocytes (iCM's) are not only comparable to native cardiomyocytes in gene expression profile and morphology, but are even capable of beating spontaneously in culture. This potential treatment for cardiovascular disease has a number of

benefits, including (i) induction of a patient's own fibroblasts (as opposed to transplantation from a donor) minimizes the risk of rejection (ii) the direct reprogramming of fibroblasts into cardiomyocytes bypasses a stem-cell progenitor, decreasing the likelihood of tumor formation associated with stem cell based treatments. While the proof-of-concept for this potential therapy has proven to be both successful in culture and in mice, the transformation efficiency of this process is incredibly low (0.03% in culture with Gata4, Tbx5, and Mef2C alone). Understanding just how these three TF's communicate with each other in different genomic contexts will be necessary to safely and effectively carry this treatment over to humans.

The Incestuous Life of the Cardiac TF:

The fact that merely 3,000 protein regulators are responsible in controlling the expression of ~25,000 genes in a number of cell types is quite remarkable. Most researchers believe that key protein regulators from a range of signaling pathways interact in a combinatorial fashion to regulate expression of genes at a given loci at a specific time . Recently, ChIP-seq in cardiomyocytes with a subset of key cardiac TF's has demonstrated that 5 TF's -- including Tbx5 and Gata4-- co-occupy (within 500bp) transcriptionally active segments of DNA (He, 2011). In addition to numerous instances in which at least two TF's could be mapped in close proximity to each other, certain genes contained regulatory regions composed of response elements (RE's) for all five cardiac TF's **(Figure 4.1)**. In line with this, cardiac TF's have also demonstrated to tightly regulate cardiac related genes. For example, the transcriptional cooperativity of Gata4, Nkx2-5, Sall4, and Tbx5 has been characterized on both the *Gja5* and *Nppa* promoter regions. From this study, it is evident that the combination of TF's alone does not dictate whether or not transcription will be activated or repressed. Rather, it is both the combination of the regulating

factors and the particular DNA context that will dictate transcriptional outcome. For example, while Tbx5 and Gata4 alone activate transcription for both *Gja5* and *Nppa,* the addition of Sall4 can cause either activation or repression. In other words, Tbx5, Gata4, and a particular DNA context allow Sall4 the flexibility of Sall4 to either act as a co-activator or co-repressor. The extent to which complex transcriptional logic extends itself to other cardiac specific genes, let alone the entire genome, remains to be seen.

Recently, it has been determined that mutations in hGata4 are responsible for congenital cardiac abnormalities, particularly atrial-, ventricular-, and atrialventricular septal defects[1,]. Most notably, a heterozygous G296S single nucleotide polymorphism (SNP) that results in atrial septal defects decreases both DNA binding ability and transcriptional activity. Unlike the wildtype, this G296S mutant is incapable of interacting with Tbx5 as determined by co-immunoprecipitation assays. The abrogation of this protein-protein interaction appears specific, as both the G296S mutant and the wild-type proteins were still able to interact with another cardiac TF, Nkx2-5[1]. Exactly how this subtle SNP alters the expression of genes regulated by both Tbx5 and Gata4 is still not understood.

Gata4 and Tbx5 Work Together to Regulate Cardiac Gene Expression:

It is thought that the specific relationship between a pair of given TF binding sites, or response elements (RE's)-- namely polarity, spacing, and strand location -- will result in the recruitment of cognate TF's that form a unique stereo-specific binding scaffold. It is these topologies that support further context-specific co-regulator interactions and, ultimately, differential expression **(Figure 4.2)**. This notion rationalizes how a limited number of TF's can fine-tune expression to accommodate the responses to intracellular and extracellular cues. For example in the case of

the nuclear hormone receptor VDR-RXR heterodimer, binding to two different RE's results in the presentation of two disparate co-activator binding surfaces as determined by hydrogen-deuterium exchange followed by mass-spectrometry. While this provides structural clues as to how differential gene expression can be executed, studies with NF-kB dimers demonstrate functionally that response element-directed dimer binding can recruit co-activators from different signaling pathways. Tbx5 and Gata4 also collaborate to control gene expression of cardiac related genes. For example, Tbx5 and Gata4 show transcriptional synergy in the expression of both NPPA and GJA5. Interestingly, the addition of Sall4 in the NPPA context causes transcriptional repression, in contrast to the GJA5 context, in which supplementation with Sall4 further enhances transcriptional activity[7]. This is a situation in which the presentation of different binding surfaces by a Tbx5:Gata4 complex could explain how Sall4 acts as either a transcriptional co-activator or co-repressor in a DNA context dependent manner. The atomic details of multiple, disparate ternary complexes of Gata4 and Tbx5 on varying RE's acquired through X-ray crystallography and SAXS can elucidate how RE arrangement mediates the succession of TF topology presentation, co-factor recruitment, and ultimately differential gene expression.

How can just a handful of TF's coordinate such a drastic cellular change?

This question is at the heart of this proposal. It has been observed that cardiac TF's, particularly Tbx5 and Gata4, co-localize in close proximity at various regulatory regions of cardiac related genes. We hypothesize that the response elements for the DNA binding proteins drives the formation of binding scaffolds for various co-regulators. It is in this manner that just a few factors could integrate information from a variety of signaling pathways for the purpose of

drastic cellular reprogramming. The experiments proposed here would (i) identify how the co-localization of these Gata4 and Tbx5 differs depending on the regulatory region (ii) provide structural insight as to how Gata4 and Tbx5 work together with several DNA compositions (iii) examine how Tbx5 and Gata4 functionally regulate various genes in a cellular context. Through these experiments, we will better understand the molecular underpinnings that drive the transformation of fibroblasts into cardiomyocytes. It is our hope that insights from these experiments will serve to rationally guide the advancement of this heart regeneration therapy.

Research Design

Specific Aim 1: Biophysically characterize the communication between Gata4 and Tbx5 in the presence and absence of DNA using surface plasmon resonance

Significance

Our goal is to quantitatively characterize the interactions of Tbx5 and Gata4 in solution, and in the different environments. We hypothesize that the kinetics governing the formation of each distinct ternary complex at least partially explains how these two factors can regulate an array of cardiac related genes. Additionally, we hypothesize that, if a protein-protein interaction between Tbx5 and Gata4 is required for transcriptional activation, only a subset of genes would be impacted by a single point mutation (e.g. G296S) **(Figure 4.3)**. Evidence of Gata4:Tbx5 dependent expression is observed when heterozygous G295S mice displayed altered gene expression for only a select set of genes. In contrast, other cardiac related genes were unaffected by this mutation . Recently, it was determined that protein:DNA interactions are not only governed by the DNA sequence, but also the position of other nearby DNA-bound factors. This

through-DNA communication is still not well understood. Testing this phenomenon in biologically significant TF's could provide insight into how these transcriptional partners regulate a wide array of genes with different regulatory regions.

Technical Details

All measurements will be performed on our in-house Biacore T100. All experiments will include periodic buffer injections for double referencing. Collected data will be fit using a global-fitting analysis. These assays will be performed using recombinantly purified DNA-binding domains of Gata4 and Tbx5 (tandem, N-terminal, C-terminal Zinc Finger for Gata4, and the T-Box domain for Tbx5). Biotinylated protein (required for 1.1) will be acquired through *in vivo* biotinylation followed by recombinant purification. Biotinylated DNA (required for 1.2) will simply be purchased from a local supplier. 3 dimensional binding isotherms will be generated by evaluating ternary complex assembly with simultaneous titrations of Gata4 and Tbx5 (Aim 1.2).

1.1 Quantitatively compare the kinetics and binding affinities of the wild-type Tbx5 and Gata4 protein-protein interaction to various mutant:wildtype interactions

Approach

The protein-protein interaction with Gata4 and Tbx5 was revealed through the discovery that a single point mutation in Gata4 (G296S) can interact with other cardiac TF's but not specifically with Tbx5. In addition to a decreased ability to activate transcription at the *Nppa* promoter, this mutation also abrogates a physical interaction with Tbx5. A similar phenomenon was observed

with mutations in *Sall4:* more specifically, the inability to neither physically interact with Tbx nor synergize with Tbx5 at the *Fgf10* promoter. Clearly, physical interactions between cardiac TF's, at least in these instances, influence whether transcriptional synergy can also occur. Known mutations of Gata4 (D210N, I250N, V267M [8], T280M , G296S) will be kinetically characterized with wild-type Tbx5 by surface plasmon resonance. Below describes what is currently known about each of the mutations **(Table 4.1)**. I*n vivo* biotinylated Tbx5 will be captured onto streptavidin coupled biosensor chips. The various mutants of Gata4 would then be flowed over the Tbx5 conjugated flow cells to reveal which mutations abrogate heterodimerization. While the heart defects caused by the the inability of cardiac TF's to interact have been documented, the biochemical details that govern these interactions are still ambiguous. From these experiments, we can learn the extent to which these mutations interrupt the Gata4:Tbx5 interaction, and at the same time, identify mutations that may effect Tbx5:Gata4 dependent gene expression.

1.2 Examine how different DNA contexts (i.e. tandem Gata4/Tbx5 response elements) drive the formation of Gata4:Tbx5:DNA ternary complexes

Approach

Gata4 and Tbx5 have demonstrated to show transcriptional synergy for at least two regulatory regions. Our goal is to evaluate how different TRE's direct the assembly of Gata4:Tbx5:DNA ternary complexes. The 10 TRE's used in these experiments are shown in **Figure 4.4**. Although there were numerous instances of TRE's in 10kb upstream for each of these genes, these were

selected to represent the diverse arrangements and orientations of the Gata4 and Tbx5 response elements. The criteria for these test regulatory regions are as follows:

- no more than 10kb from a cardiac related gene
- no more than 7bp spacer to (more feasible for crystallography)
- only strict Tbx5 (CCACA), C-terminal Zinc Finger (GATA), or N-terminal ZF (GATG/C/T) recognized,
- minimal redundancy in TRE composition to represent maximal arrangement and orientation space.

Of interest, 4 out of these 5 genes demonstrate altered expression in heterozygous G295S knock in mice. This suggests that, while other genes are unaffected, these genes are dependent at least in part by effective communication between Tbx5 and Gata4. These experiments reveal the biophysical nature of Gata4:Tbx5 communication. Additionally, a comparison of ΔG between binary and ternary complexes how TRE arrangement affords cooperativity (or anti-cooperativity) **(Figure 4.5).** From a biomedical prospective, this knowledge could hopefully be used to accelerate the development of cardiac reprogramming therapy. From a basic science view, a mechanistic understanding of how these two important TF's behave will provide further insight into the complexities of eukaryotic gene expression.

Specific Aim 2: Reveal at the atomic level (i) how Tbx5 and Gata4 engage with each other on disparate TRE's (ii) how a subtle mutation in Gata4 disrupts interactions with Tbx5.

Significance

While it is clearly evident that transcriptional regulators, particularly cardiac TF's, synergize to alter gene expression, there are still only a few well understood instances where the molecular underpinnings of these protein DNA complexes have been elucidated. In particular, the structural details of the Oct and Sox proteins have been characterized by both crystallography and NMR. Biochemically, it has been shown that this pair of TF's can occupy TRE's of diverse sequence and arrangement. Oct4 and Sox2, for example, interact differently depending on the composition of the TRE (e.g. FGF4 vs. UTF1 regulatory regions). This Oct:Sox 'partner code' directs cells to varying degrees of differentiation and pluripotency. In a similar fashion, understanding the 'partner code', between cardiac TF's (e.g. Tbx5 and Gata4) will provide insight into the role of Gata4 and Tbx5-dependent gene expression in cardiac reprogramming. Crystal structures of these ternary complexes will reveal any (i) protein-protein interactions between Tbx5 and Gata4 (ii) DNA-mediated communication between the two factors (iii) unique conformations driven by specific TRE composition. Molecular dynamics simulations will describe the motions of both the wildtype and mutant cardiac TF's-- this will possibly elucidate how a subtle mutation (e.g. Gata4 G296S) can disrupt protein-protein interactions through a particular mechanism (e.g. instability, redistribution of conformational ensemble, re-shaping of a dimerization interface, etc.)

Technical Details

Protein complexes will be assembled for crystallography by size-exclusion chromatography. Our lab has already had success crystallizing a ternary complex of Tbx5:Nkx2-5:DNA with this approach (unpublished). Diffraction (X-ray crystallography) and scattering (SAXS) data will be collected at the Advanced Light Source (Beamline 8.3.1), located at the Lawrence Berkeley

National Laboratory. Homology models needed for computational studies will be generated using the PRIME modeling tool in the Schrodinger suite. Molecular dynamics simulations will be run on using the GPU-modified AMBER force field-- this allows for nanoseconds of simulation time per day (depending on system size).

2.1 Solve crystal structures of Tbx5 and Gata4 on multiple TRE's

Approach

Crystal structures have been previously solved for both the T-Box family of TF's (including apo and DNA bound Tbx5[14] and the GATA family (various zinc fingers species of mouse Gata3 bound to DNA [15,16]. While this is not a predictor of whether structures of Gata4:Tbx5:DNA ternary complexes can crystallized and solved, it at the very least indicates that both of these families of proteins can be solubly expressed in significant amounts, purified sufficiently for crystallography, and crystallized under the appropriate conditions. Recombinantly purified Gata4 and Tbx5 will be incubated with each of the 10 test regulatory regions from Aim 1. These mixtures will then be subjected to size exclusion chromatography (SEC) to identify which of these oligos supports stable ternary complexes. Data gathered in Aim 1.2 could also be used to help select suitable TRE for crystallography trials. Solving these these various structures could help to answer the following questions:

1. To what extend does DNA act as an allosteric modulator of the TF's?
2. What is the structural basis for any heterodimerization between these two proteins?
3. How do the overall topologies of these structures compare depending on TRE composition?

While all of these questions could be answered by solving crystal structures of multiple ternary complexes, a single structure would provide key insight as well. If, however, protein crystals cannot be grown for any of these ternary complexes, small angle X-ray scattering (SAXS) would also reveal the unique topologies of Tbx5 and Gata4 ternary complexes. We hypothesize that these potential co-regulator binding surfaces, driven by DNA composition, allow for the integration of information from a variety of signaling pathways.

2.2 Perform molecular dynamics simulations to compare protein motions between single point mutants of Tbx5 and Gata4 with their respective wild-type proteins

Approach

Mutations in Gata4 and Tbx5 have already demonstrated to have effects on DNA binding, protein-protein interactions, and transcriptional activity. While Aim 1 will quantitatively evaluate some of these deleterious effects in Gata4 mutants, they provide no mechanistic description of how these mutations hinder basic TF functions. To this end, a comparison of the wildtype vs mutant protein motions can provide the missing information necessary to understanding the biochemical basis for these disease causing mutations. Preliminary 10ns simulations of a homology model of DNA-bound human Gata4 and human Gata4 G296S have already been performed. This homology model was built off of the highly homologous mouse Gata3 structure[16] (~83% sequence identity for tandem zinc finger+C terminal basic tail). Essentially, this mutation introduces a hydrogen bond donor in the form of a hydroxyl group of serine that appears to be within hydrogen bonding distance of a nearby backbone carbonyl **(Figure 4.6).** This could be introducing rigidity into this molecule where flexibility is required

for either protein-protein interactions or DNA binding. Further analysis and longer simulations would be required to support this hypothesis. In addition to more routine trajectory analysis, Kullback–Leibler divergence could be used to compare the conformational distributions between the wildtype and the mutant for both Gata4 and Tbx5 species. Mutations for Gata4 are described in Aim1; mutations for Tbx5 can be found here[14]. This approach for comparing conformational ensembles was developed in the Jacobson lab, and it has been used to characterize how subtle perturbations (e.g. mutations) redistribute the conformational ensemble. These studies would act as basis for further experimental validation. For example, could a second compensatory mutation revive the wildtype function? We hope that the these simulations will be helpful in understanding the transition from genotype to phenotype for congenital heart defects.

Specific Aim 3: Quantitatively assess how Tbx5 and Gata4 cooperate to regulate cardiac specific genes through luciferase assays.

Significance

Tbx5 and Gata4, in conjunction with Mef2c, were the first factors to be recognized that can reprogram fibroblasts into functional cardiomyocytes. Recently, Hand2 has been identified to further activate this reprogramming process. In this manner, the presence of the MyoD transactivation domain was able to further accelerate the iCM formation. While Tbx5, Gata4, and Mef2c are both necessary and sufficient to cardiac reprogramming, it is evident that additional factors greatly improve the overall efficiency of this process. While our lab focuses structurally and biophysically understanding of the underlying mechanisms of cell fate, we are

also interested in learning if and how this structural knowledge of transcriptional regulators is coupled to functional performance. Again, the knowledge is not only fascinating from a basic science prospective, but is also obviously pertinent to the cellular reprogramming therapies that are emerging.

3.1 Compare by transcriptional activity assays how Tbx5 and Gata4 cooperate to regulate a range of important cardiac genes

Approach

While Aim 1 identifies if and how TRE's drive the assembly of Gata4:Tbx5:DNA ternary complexes, Aim 2 reveals the structure of these ternary complexes. Aim 3 plans to characterize how these ternary complexes actually influence transcriptional activity at regulatory regions in important cardiac related genes.. As an example, suppose at least two oligos allow for ternary complex assembly (as determined in Aim 1). While mechanistically these two complexes could assemble in a similar fashion, one may result in activation with the other resulting in repression. Supported by previous work, we hypothesize that Gata4 and Tbx5 will behave in different ways depending on the arrangement of their TRE's. We feel that these experiments are required to comprehend how the core reprogramming machinery (i.e. Tbx5 and Gata4) utilize response element diversity to control gene expression. This basic knowledge could begin to partially describe the transcriptional logical, or gene expression blueprint, for this reprogramming process.

Conclusion

Elucidating the behavior of Tbx5 and Gata4 in a number of biologically relevant DNA contexts will provide insight as to how these factors coordinate drastic cellular events such as cardiac reprogramming and heart development. This proposal aims to truly understand this relationship at the biophysical, structural, and functional level. We hope this knowledge helps (i) cardiologists fully understand the etiology of congenital heart defects (ii) cardiovascular researchers realize cardiac reprogramming as a therapy for cardiovascular disease (iii) basic scientists better explain the intricacies intertwined into eukaryotic gene expression.

Figures

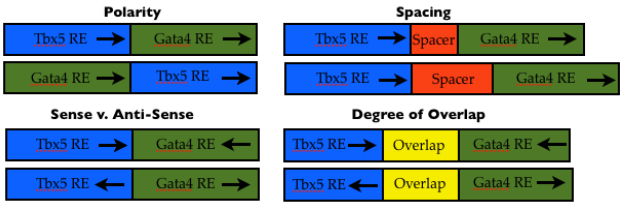


Figure 4.1. Different possible arrangements of tandem TRE's observed in CHIP-seq results (He, 2011).

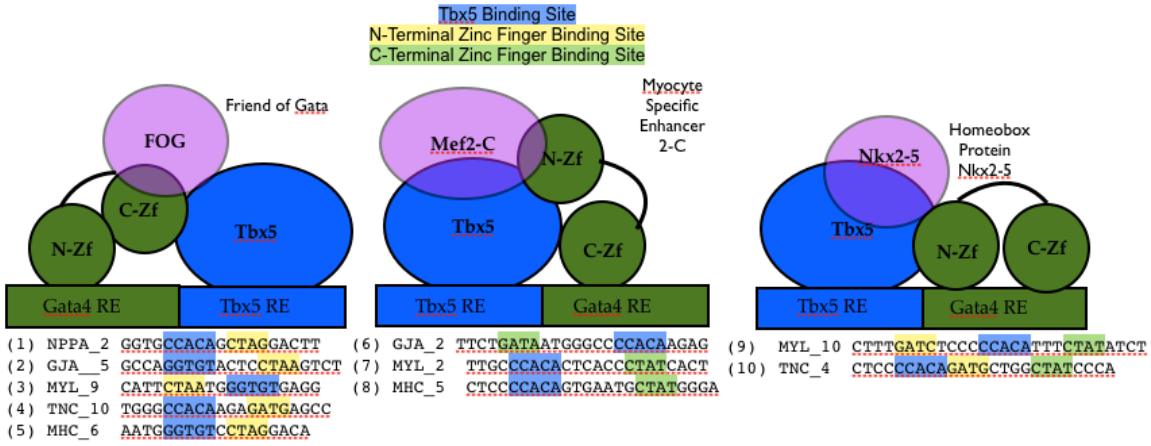


Figure 4.2. Tandem RE's drive the formation of unique binding surfaces for co-regulators. In this way, just two TF's can delicately regulate a vast array of gene in a way that would be necessary for a process like cellular reprogramming. The DNA sequences

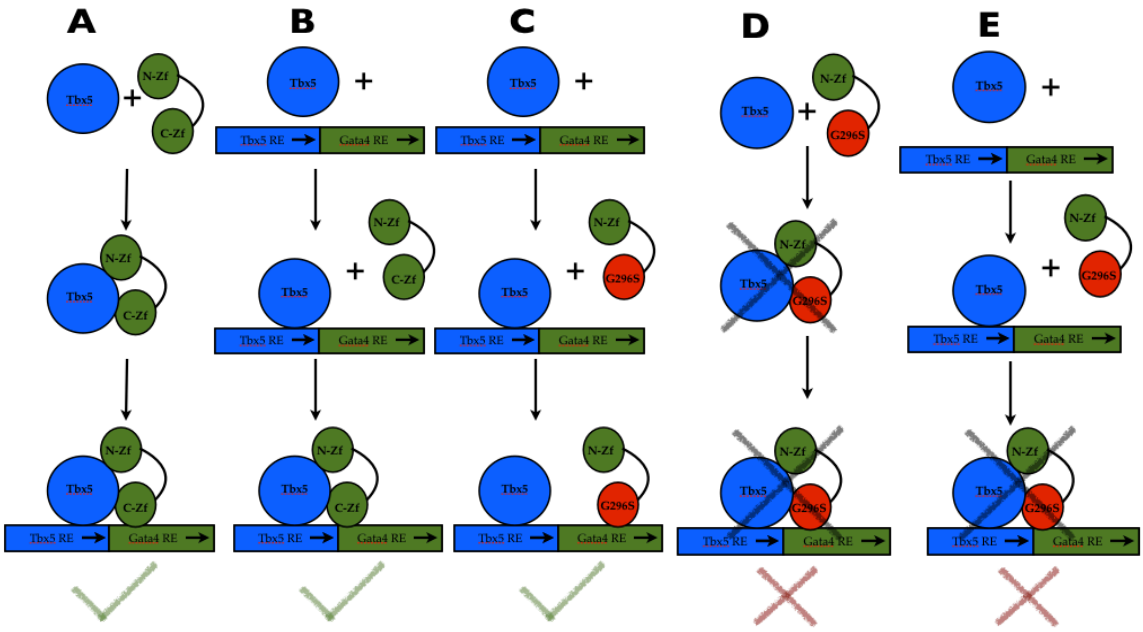


Figure 4.3. Examples (non-exhaustive) of Gata4:Tbx5 gene expression dependency. A green 'check' denotes functional regulation; a red 'X' denotes misregulation
a.) Heterodimzerization in solution prior to DNA binding (Functional)
b.) Sequential binding on the TRE resulting in a protein:protein interaction (Functional)
c.) Independent binding with no physical interaction between Tbx5 and Gata4 (therefore, unaffected by mutation) (Functional)
d.) Disruption of DNA-independent heterodimerization (Non-Functional)
e.) Inhibited ternary complex assembly on the TRE (Non-functional)

Tbx5 Binding Site
N-Terminal Zinc Finger Binding Site
C-Terminal Zinc Finger Binding Site

```
(1) NPPA_2  GGTGCCACAGCTAGGACTT
(2) GJA_5   GCCAGGTGTACTCCTAAGTCT
(3) MYL_9   CATTCTAATGGGTGTGAGG
(4) TNC_10  TGGGCCACAAGAGATGAGCC
(5) MHC_6   AATGGGTGTCCTAGGACA
(6) GJA_2   TTCTGATAATGGGCCCCACAAGAG
(7) MYL_2   TTGCCCACACTCACCCTATCACT
(8) MHC_5   CTCCCCACAGTGAATGCTATGGGA
(9) MYL_10  CTTTGATCTCCCCCACATTTCTATATCT
(10)TNC_4   CTCCCCACAGATGCTGGCTATCCCA
```

Figure 4.4. Sequences of 10 test TRE's. NPPA: atrial natriuretic factor; GJA: Connexin40/Gap junction alpha-5 protein; MHC: Myosin Heavy Chain 6; MYL: Myosin Light Chain 3; TNC: Cardiac Troponin C

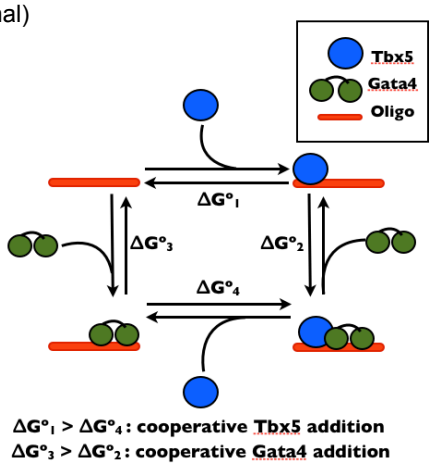


Figure 4.5. Analysis of ternary complex assembly. This depiction of a thermodynamic cycle demonstrates how complex assembly at a given TRE can be classified by its cooperativity

Mutation in Gata4	CHD Causing?	Decreased DNA binding?	Decreased transcriptional activity?	Decreased Binding to Tbx5?
D210N	Yes	No	No	No
I250N	Yes	No	No	No
V267M	No	No	Not tested	No
T280M	Yes	Yes	Yes	No
G296S	Yes	Yes	Yes	Yes

Table 4.1. Mutations of interest in human Gata4

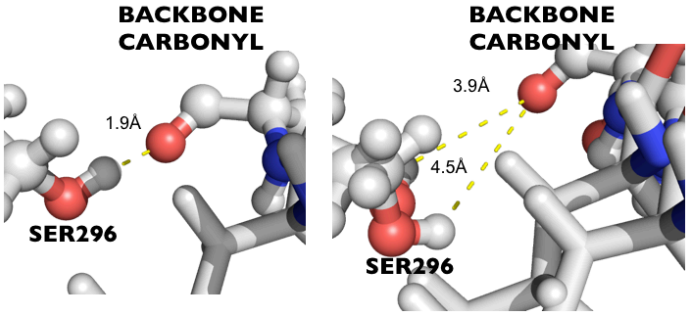


Figure 4.6. The hydroxyl group of the G296S mutant was observed to be within hydrogen bonding distance of a nearby backbone carbonyl in a multiple clusters after k-means clustering analysis. One cluster, left image; two clusters, right image. Longer simulations (~50ns) will be necessary to gain confidence in this observation.

References

Bates DL, Chen Y, Kim G, Guo L, Chen L. Crystal structures of multiple GATA zinc fingers bound to DNA reveal new insights into DNA recognition and self-association by GATA. J Mol Biol. 2008;381(5):1292-306.

Chen Y, Bates DL, Dey R, et al. DNA binding by GATA transcription factor suggests mechanisms of DNA looping and long-range gene regulation. Cell Rep. 2012;2(5):1197-206.

Chen Y, Mao J, Sun Y, et al. A novel mutation of GATA4 in a familial atrial septal defect. Clin Chim Acta. 2010;411(21-22):1741-5.

Hirai H, Katoku-kikyo N, Karian P, Firpo M, Kikyo N. Efficient iPS cell production with the MyoD transactivation domain in serum-free culture. PLoS ONE. 2012;7(3):e34149.

Ieda M, Fu JD, Delgado-Olguin P, Vedantham V, Hayashi Y, Bruneau BG, Srivastava D. Direct reprogramming of fibroblasts into functional cardiomyocytes by defined factors. Cell. 2010;142(3):375-86.

Kim S, Broströmer E, Xing D, et al. Probing allostery through DNA. Science. 2013;339(6121): 816-9.

Koshiba-takeuchi K, Takeuchi JK, Arruda EP, et al. Cooperative and antagonistic interactions between Sall4 and Tbx5 pattern the mouse limb and heart. Nat Genet. 2006;38(2):175-83.

Leung TH, Hoffmann A, Baltimore D. One nucleotide in a kappaB site can determine cofactor specificity for NF-kappaB dimers. Cell. 2004; 118(4):453-64.

Mcclendon CL, Hua L, Barreiro A, Jacobson MP. Comparing Conformational Ensembles Using the Kullback-Leibler Divergence Expansion. J Chem Theory Comput. 2012;8(6):2115-2126.

Misra C, Sachan N, Mcnally CR, et al. Congenital heart disease-causing Gata4 mutation displays functional deficits in vivo. PLoS Genet. 2012;8(5):e1002690.

Pang ZP, Yang N, Vierbuchen T, et al. Induction of human neuronal cells by defined transcription factors. Nature. 2011;476(7359):220-3.

Remenyi A, Schö|ler H, Wilmanns M. Combinatorial control of gene expression. Nature Structural & Molecular Biology. 2004;11(9):812.

Reményi A, Lins K, Nissen LJ, Reinbold R, Schöler HR, Wilmanns M. Crystal structure of a POU/HMG/DNA ternary complex suggests differential assembly of Oct4 and Sox2 on two enhancers. Genes Dev. 2003;17(16):2048-59.

Song K, Nam YJ, Luo X, et al. Heart repair by reprogramming non-myocytes with cardiac transcription factors. Nature. 2012;485(7400):599-604.

Srivastava D, Heidersbach AJ. Small solutions to big problems: microRNAs for cardiac regeneration. Circ Res. 2013;112(11):1412-4.

Stirnimann CU, Ptchelkine D, Grimm C, Müller CW. Structural basis of TBX5-DNA recognition: the T-box domain in its DNA-bound and -unbound form. J Mol Biol. 2010;400(1):71-81.

Takahashi K, Yamanaka S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. Cell. 2006;126(4):663-76.

Wang E, Sun S, Qiao B, et al. Identification of functional mutations in GATA4 in patients with congenital heart disease. PLoS ONE. 2013;8(4):e62138.

Williams DC, Cai M, Clore GM. Molecular basis for synergistic transcriptional activation by Oct1 and Sox2 revealed from the solution structure of the 42-kDa Oct1.Sox2.Hoxb1-DNA ternary transcription factor complex. J Biol Chem. 2004;279(2):1449-57.

Zhang J, Chalmers MJ, Stayrook KR, et al. DNA binding alters coactivator interaction surfaces of the intact VDR-RXR complex. Nat Struct Mol Biol. 2011;18(5):556-63.

Garg V, Kathiriya IS, Barnes R, et al. GATA4 mutations cause human congenital heart defects and reveal an interaction with TBX5. Nature. 2003;424(6947):443-7.

APPENDIX

Determination of Suitable Tandem Response Elements for Ternary Complex Assembly

Identifying a suitable segment of DNA that was contained both a Tbx5 response element and either a response element from (i) N-terminal zinc finger (N-Zf) of Gata4, (ii) C-terminal zinc finger, or (iii) both zinc fingers was challenging but possible with the help of A. Holloway from the Gladstone institute. In addition to containing the tandem response elements, the enhancer region should ideally potentially regulate a gene with a known biological function as it pertains to cardiac development/health. Below are the identified regions. 10 tandem response elements were ultimately examined for the ability to form ternary complexes with Gata4 and Tbx5.

TANDEM RESPONSE ELEMENTS WITHIN GENE

hg19_ct_UserTrack_3545_0 range=chr1:11905768-11907840 5'pad=0 3'pad=0 strand=+NPPA

NPPA_w_1: CGCCACCACACCCAGCTAATTTTTG

hg19_ct_UserTrack_3545_1 range=chr1:147228333-147232714 5'pad=0 3'pad=0 strand=+ repeatMasking=non (GJA5)

GJA_w_1: TCAAAGATTGATCCCACAGAATTTCTAATTCAA

GJA_w_2: TAATGCTAGGCAATAGATTAAAAGGTGTCCCCT

GJA_w_3: TCTGGACTATGCCCACAGAGGG

GJA_w_4: AGGAAGATTCCGTAGATGAAGTACTGGCCCACAATGAA

GJA_w_5: TGGCTCCCACACTTTAACGGATCTAGAA

GJA_w_6: CTGAGGATTAAATGATATGAGGTGTGTGATATTCTG

GJA_w_7: GGTTCCCACACACAGATCTGCCC

ihg19_ct_UserTrack_3545_0 range=chr3:46899358-46904973 5'pad=0 3'pad=0 strand=+

repeatMasking=none (MYL3):

MYL_w_1: GGGCTCCACAGAGGTCAGGATCCCCCCA

MYL_w_2: TACAGTGGTGTGATTATAGC

MYL_w_3: CCATAGGTGTCAGCGCCTACCACCA

MYL_w_4: GGACAGATGTACCTTAAGCCACAAGCCC

MYL_w_5: TGCTGCTATGCAGATACCACAGGTGG

hg19_ct_UserTrack_3545_1 range=chr3:52485108-52488057 5'pad=0 3'pad=0 strand=+

repeatMasking= (TNNC1)

TNC_w_1: CAGCATCTACTCCACACCCTTC

TNC_w_2: TCTCCCCTATCAGGCAGAGGCCACAGGGTCCCTAGGCCTG

TNC_w_3: GCAGGCTATTTCCAGGCCACAGAGTGG

TNC_w_4: GGACTCCCACAGGATAAGCAG

hg19_ct_UserTrack_3545_0 range=chr14:23851200-23877486 5'pad=0 3'pad=0 strand=+

repeatMaskin (MYH6)

MHC_w_1: TAGTAGGTGTGTTGATGAGAAA

MHC_w_2: GCCAGGATGCCATTTCTACTTCTGTCCACAAATCA

MHC_w_3: AACTTCCACATCTACTCCTGGTGTCTGG

MHC_w_4: TGGTTGATGAGGCTGGTGTTCTAGACATG

MHC_w_5: GGTGGGATTGCAGAAAATGGTGTTCTAGCCAGC

MHC_w_6: AGCCACCACAGATGAGAGTTGGCCTAAGAGAG

MHC_w_7: AGCTGGATCTGGGTGTCCTGA

MHC_w_8: GGGAGGATGAGAAGATACACGGTGTCAGCCAGGACTATCGCCC

MHC_w_9: GAACCCCACACTCCTACCCACTGTCTAAGTGGT

MHC_w_10: TGGAGGGTGTGTGTGTGACTCTACTGGGC

MHC_w_11: ACCCCCCACACAGGCTGATCGACGG

MHC_w_12: GGAGACTACCTCTTTCTACCAGCCACAGTCTG

MHC_w_13: GGCCACCACAGTCTCCTACTTCTT

MHC_w_14: TGCCACCACACCCGGCTAATTTTT

MHC_w_15: TCTGACCCACACTAGTTGAC

MHC_w_16: GGGCTCTAGACTCACCTTGGTGTGGCCA

MHC_w_17: TCTTTCCACATTCTAGTTCTC

MHC_w_18: GCCCAGGATGTTGTAGTCCACAGTGCC

MHC_w_19: GCACTCCTAATCACCCCCCACAAGGCCCACAGTGCC

MHC_w_20: CCTGGGATTCTTGGGACTCTAGTTTCTTGGGTGTAGAAGG

MHC_w_21: GGGCACCCACAGCTTGATGAAGAG

MHC_w_22: AGGAACTAGGGCTCAGAGAGGTGTAAGGA

MHC_w_23: TACTGGATGACACGCTTGGTGTTCACA

MHC_w_24: CTGAGGATCTGGGTGGGTGTCTGGG

MHC_w_25: CCCATCTACTCCACACTCCC

MHC_w_26: CTGTTGGTGTGGTTTGATTGGCAG

MHC_w_27: GGGGTGATTCTCTTGGCTGGTGTGAGCA

TBX5:N-ZF Only (Within Gene)

NPPA_w_1: CGCCACCACACCCAGCTAATTTTTG

GJA_w_1: TCAAAGATTGATCCCACAGAATTTCTAATTCAA

GJA_w_2: TAATGCTAGGCAATAGATTAAAAGGTGTCCCCT

GJA_w_4: AGGAAGATTCCGTAGATGAAGTACTGGCCCACAATGAA

GJA_w_5: TGGCTCCCACACTTTAACGGATCTAGAA

GJA_w_6: CTGAGGATTAAATGATATGAGGTGTGTGATATTCTG

GJA_w_7: GGTTCCCACACACAGATCTGCCC

MYL_w_1: GGGCTCCACAGAGGTCAGGATCCCCCCA

MYL_w_2: TACAGTGGTGTGATTATAGC

MYL_w_3: CCATAGGTGTCAGCGCCTACCACCA

MYL_w_4: GGACAGATGTACCTTAAGCCACAAGCCC

TNC_w_1: CAGCATCTACTCCACACCCTTC

TNC_w_2: TCTCCCCTATCAGGCAGAGGCCACAGGGTCCCTAGGCCTG

MHC_w_1: TAGTAGGTGTGTTGATGAGAAA

MHC_w_2: GCCAGGATGCCATTTCTACTTCTGTCCACAAATCA

MHC_w_3: AACTTCCACATCTACTCCTGGTGTCTGG

MHC_w_4: TGGTTGATGAGGCTGGTGTTCTAGACATG

MHC_w_5: GGTGGGATTGCAGAAAATGGTGTTCTAGCCAGC

MHC_w_6: AGCCACCACAGATGAGAGTTGGCCTAAGAGAG

MHC_w_7: AGCTGGATCTGGGTGTCCTGA

MHC_w_8: GGGAGGATGAGAAGATACACGGTGTCAGCCAGGACTATCGCCC

MHC_w_9: GAACCCCACACTCCTACCCACTGTCTAAGTGGT

MHC_w_10: TGGAGGGTGTGTGTGTGACTCTACTGGGC

MHC_w_11: ACCCCCCACACAGGCTGATCGACGG

MHC_w_12: GGAGACTACCTCTTTCTACCAGCCACAGTCTG

MHC_w_13: GGCCACCACAGTCTCCTACTTCTT

MHC_w_14: TGCCACCACACCCGGCTAATTTTT

MHC_w_15: TCTGACCCACACTAGTTGAC

MHC_w_16: GGGCTCTAGACTCACCTTGGTGTGGCCA

MHC_w_17: TCTTTCCACATTCTAGTTCTC

MHC_w_18: GCCCAGGATGTTGTAGTCCACAGTGCC

MHC_w_19: GCACTCCTAATCACCCCCCACAAGGCCCACAGTGCC

MHC_w_20: CCTGGGATTCTTGGGACTCTAGTTTCTTGGGTGTAGAAGG

MHC_w_21: GGGCACCCACAGCTTGATGAAGAG

MHC_w_22: AGGAACTAGGGCTCAGAGAGGTGTAAGGA

MHC_w_23: TACTGGATGACACGCTTGGTGTTCACA

MHC_w_24: CTGAGGATCTGGGTGGGTGTCTGGG

MHC_w_25: CCCATCTACTCCACACTCCC

MHC_w_26: CTGTTGGTGTGGTTTGATTGGCAG

MHC_w_27: GGGGTGATTCTCTTGGCTGGTGTGAGCA\

<u>Tbx5:C-ZF Only</u>

GJA_w_3: TCTGGACTATGCCCACAGAGGG

GJA_w_6: CTGAGGATTAAATGATATGAGGTGTGTGATATTCTG

MYL_w_5: TGCTGCTATGCAGATACCACAGGTGG

TNC_w_2: TCTCCCCTATCAGGCAGAGGCCACAGGGTCCCTAGGCCTG

TNC_w_3: GCAGGCTATTTCCAGGCCACAGAGTGG

TNC_w_4: GGACTCCCACAGGATAAGCAG

MHC_w_8: GGGAGGATGAGAAGATACACGGTGTCAGCCAGGACTATCGCCC

Tbx5:N-ZF:C-Zf

GJA_w_6: CTGAGGATTAAATGATATGAGGTGTGTGATATTCTG

TNC_w_2: TCTCCCCTATCAGGCAGAGGCCACAGGGTCCCTAGGCCTG

MHC_w_8: GGGAGGATGAGAAGATACACGGTGTCAGCCAGGACTATCGCCC

TANDEM RESPONSE ELEMENTS (10kb UPSTREAM)

hg19_ct_UserTrack_3545_0 range=chr1:11895768-11905767 5'pad=10000 3'pad=0 strand=+

(NPPA_)

NPPA_1.) TCCCAGATTAAAGGAAACCACAGAAGC

NPPA_2.) GGGTGCCACAGCTAGGACTT

NPPA_3.) TGGGACCACAGGCATGCGCCACTATGCCCAGCTAGTTTTT

NPPA_4.) TGTGCACTATCCAGGCCTCCACATTCTGTGACAAACTAACAGGG

NPPA_5.) TATGTGCTATGATAAAGCAGATGGCCTATTTGAGGAAGAGGGTGTCTGCC

NPPA_6.) AGAATGGTGTGATCACCTG

NPPA_7.) CAGCTTGGTGTTCCATGGCCCCTAGGCGAGG

NPPA_8.) AAAGCAGATCTGTGGATGGGGAGCCTACGGGTGGTAAGAAGTGGTGTTTTGT

NPPA_9.) GTATTGATGTGCAGTATTGCACCACAGCTCT

NPPA_10.)TTGCCCCACACACAGATTGCCCCGATGTCCAA

hg19_ct_UserTrack_3545_1 range=chr1:147220271-147230270 5'pad=10000 3'pad=0 strand=+

(GJA_5)

GJA__1.) ATTGAAGGTGTATTTCATTATTACTATTAAAA

GJA__2.) ATTCTGATAATGGGCCCCACAAGAGT

GJA__3.) CTCTACTAAAAATGGTGTTGCATG

GJA__4.) TTCCTGATTTCCACATCATGA

GJA__5.) GGCCAGGTGTACTCCTAAGTCTC

GJA__6.) TCAAAGATTGATCCCACAGAATTTCTAATTCAA

GJA__7.) TAATGCTAGGCAATAGATTAAAAGGTGTCCCCT

GJA__8.) TTGCCCCACACACAGATTGCCCCGATGTCCAA

hg19_ct_UserTrack_3545_0 range=chr3:46889358-46899357 5'pad=10000 3'pad=0 strand=+

(MYL3)

MYL_1.) TCAACCACACCTTCTCCTATTTCTG

MYL_2.) CTTGCCCACACTCACCCTATCACTG

MYL_3.) GGTATGATGGCAGAGGTGTTGAGG

MYL_4.) TGACCCCACAACTCTCACTACTACCC

MYL_5.) GGAGGCCACAGCCTGCCTAGGAGTG

MYL_6.) ATTCTGATGTCTGCACCCACAGCCAG

MYL_7.) CCTCAGATGCCCACACAGCC

MYL_8.) TTGTGGATAGACCACATTTGC

MYL_9.) CCATTCTAATGGGTGTGAGGT

MYL_10.) CCTTTGATCTCCCCCACATTTCTATATCTGG

MYL_11.) CGGGGCTATCCCCACAAAGCCTCTACCTCTGACCTAGAGCCACACTGCTG

hg19_ct_UserTrack_3545_1 range=chr3:52475108-52485107 5'pad=10000 3'pad=0 strand=+

(TNC)

TNC_1.) AGGCTGGTGTCCTAGCTGGG

TNC_2.) CAGTGCCACAGGCTAGCAGGT

TNC_3.) AGGCAGATCAGGGCCACAGCTCAGCCTAGTTCCC

TNC_4.) TCTCCCCACAGATGCTGGCTATCCCAC

TNC_5.) ACAGAGATGTCCACAAATGC

TNC_6.) ATGTCCTAGACAGGTGTTTGAG

TNC_7.) CTCCCCACATCTACATGC

TNC_8.) GGGAAGGGTGTGGGGGCAGCCTACTCCACACTGCG

TNC_9.) CCCCACCACAGCTGATGAAAAG

TNC_10.) TGGGCCACAAGAGATGAGCCT

TNC_11.) GACCCCTAGGCTGCCACAGTGGT

TNC_12.) GGGTTCTAGGCCCCCACAGACTCCAGATGCTGAATCCCTAGAGATGGTGCCT

hg19_ct_UserTrack_3545_0 range=chr14:23841200-23851199 5'pad=10000 3'pad=0 strand=+

(MHC6)

MHC_1.) GGGAGGCCACAAGACATTTCCTAATTTTGGCTACCAAACC

MHC_2.) GGAACCCACACCTACAGCCA

MHC_3.) GGGCACCCACAAGGGCTACTGCCT

MHC_4.) GGAGGACCACATCCGCTAGATCTACTCT

MHC_5.) GCTCCCCACAGTGAATGCTATGGGAA

MHC_6.) GAATGGGTGTCCTAGGACACC

Tbx5:N-ZF Only (10kb Upstream)

NPPA_1.) TCCCAGATTAAAGGAAACCACAGAAGC

NPPA_2.) GGGTGCCACAGCTAGGACTT

NPPA_3.) TGGGACCACAGGCATGCGCCACTATGCCCAGCTAGTTTTT

NPPA_4.) TGTGCACTATCCAGGCCTCCACATTCTGTGACAAACTAACAGGG

NPPA_5.) TATGTGCTATGATAAAGCAGATGGCCTATTTGAGGAAGAGGGTGTCTGCC

NPPA_6.) AGAATGGTGTGATCACCTG

NPPA_7.) CAGCTTGGTGTTCCATGGCCCCTAGGCGAGG

NPPA_8.) AAAGCAGATCTGTGGATGGGGAGCCTACGGGTGGTAAGAAGTGGTGTTTTGT

NPPA_9.) GTATTGATGTGCAGTATTGCACCACAGCTCT

NPPA_10.)TTGCCCCACACACAGATTGCCCCGATGTCCAA

GJA__3.) CTCTACTAAAAATGGTGTTGCATG

GJA__4.) TTCCTGATTTCCACATCATGA

GJA__5.) GGCCAGGTGTACTCCTAAGTCTC

GJA__6.) TCAAAGATTGATCCCACAGAATTTCTAATTCAA

GJA__7.) TAATGCTAGGCAATAGATTAAAAGGTGTCCCCT

GJA__8.) TTGCCCCACACACAGATTGCCCCGATGTCCAA

MYL_3.) GGTATGATGGCAGAGGTGTTGAGG

MYL_4.) TGACCCCACAACTCTCACTACTACCC

MYL_5.) GGAGGCCACAGCCTGCCTAGGAGTG

MYL_6.) ATTCTGATGTCTGCACCCACAGCCAG

MYL_7.) CCTCAGATGCCCACACAGCC

MYL_9.) CCATTCTAATGGGTGTGAGGT

MYL_10.) CCTTTGATCTCCCCCACATTTCTATATCTGG

MYL_11.) CGGGGCTATCCCCACAAAGCCTCTACCTCTGACCTAGAGCCACACTGCTG

TNC_1.) AGGCTGGTGTCCTAGCTGGG

TNC_2.) CAGTGCCACAGGCTAGCAGGT

TNC_3.) AGGCAGATCAGGGCCACAGCTCAGCCTAGTTCCC

TNC_4.) TCTCCCCACAGATGCTGGCTATCCCAC

TNC_5.) ACAGAGATGTCCACAAATGC

TNC_6.) ATGTCCTAGACAGGTGTTTGAG

TNC_7.) CTCCCCACATCTACATGC

TNC_8.) GGGAAGGGTGTGGGGGCAGCCTACTCCACACTGCG

TNC_9.) CCCCACCACAGCTGATGAAAAG

TNC_10.) TGGGCCACAAGAGATGAGCCT

TNC_11.) GACCCCTAGGCTGCCACAGTGGT

TNC_12.) GGGTTCTAGGCCCCCACAGACTCCAGATGCTGAATCCCTAGAGATGGTGCCT

MHC_1.) GGGAGGCCACAAGACATTTCCTAATTTTGGCTACCAAACC

MHC_2.) GGAACCCACACCTACAGCCA

MHC_3.) GGGCACCCACAAGGGCTACTGCCT

MHC_4.) GGAGGACCACATCCGCTAGATCTACTCT

MHC_6.) GAATGGGTGTCCTAGGACACC

Tbx5:C-ZF Only (10kb Upstream)

NPPA_3.) TGGGACCACAGGCATGCGCCACTATGCCCAGCTAGTTTTT

NPPA_4.) TGTGCACTATCCAGGCCTCCACATTCTGTGACAAACTAACAGGG

NPPA_5.) TATGTGCTATGATAAAGCAGATGGCCTATTTGAGGAAGAGGGTGTCTGCC

GJA__1.) ATTGAAGGTGTATTTCATTATTACTATTAAAA

GJA__2.) ATTCTGATAATGGGCCCCACAAGAGT

MYL_1.) TCAACCACACCTTCTCCTATTTCTG

MYL_2.) CTTGCCCACACTCACCCTATCACTG

MYL_8.) TTGTGGATAGACCACATTTGC

MYL_10.) CCTTTGATCTCCCCCACATTTCTATATCTGG

MYL_11.) CGGGGCTATCCCCACAAAGCCTCTACCTCTGACCTAGAGCCACACTGCTG

TNC_4.) TCTCCCCACAGATGCTGGCTATCCCAC

MHC_5.) GCTCCCCACAGTGAATGCTATGGGAA

Tbx5:N-ZF:C-ZF (10kb Upstream)

NPPA_3.) TGGGACCACAGGCATGCGCCACTATGCCCAGCTAGTTTTT

NPPA_4.) TGTGCACTATCCAGGCCTCCACATTCTGTGACAAACTAACAGGG

NPPA_5.) TATGTGCTATGATAAAGCAGATGGCCTATTTGAGGAAGAGGGTGTCTGCC

MYL_10.) CCTTTGATCTCCCCCACATTTCTATATCTGG

MYL_11.) CGGGGCTATCCCCACAAAGCCTCTACCTCTGACCTAGAGCCACACTGCTG

TNC_4.) TCTCCCCACAGATGCTGGCTATCCCAC

TEST SET: ALL FROM 10kb UPSTREAM

NPPA_2.) GGGTGCCACAGCTAGGACTT

GJA__5.) GGCCAGGTGTACTCCTAAGTCTC

MYL_9.) CCATTCTAATGGGTGTGAGGT

TNC_10.) TGGGCCACAAGAGATGAGCCT

MHC_6.) GAATGGGTGTCCTAGGACACC

GJA__2.) ATTCTGATAATGGGCCCCACAAGAGT

MYL_2.) CTTGCCCACACTCACCCTATCACTG

MHC_5.) GCTCCCCACAGTGAATGCTATGGGAA

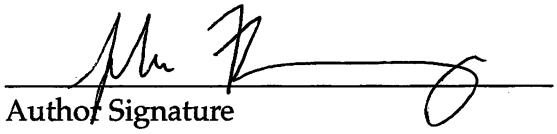
MYL_10.) CCTTTGATCTCCCCCACATTTCTATATCTGG

TNC_4.) TCTCCCCACAGATGCTGGCTATCCCAC

Publishing Agreement

It is the policy of the University to encourage the distribution of all theses, dissertations, and manuscripts. Copies of all UCSF theses, dissertations, and manuscripts will be routed to the library via the Graduate Division. The library will make all theses, dissertations, and manuscripts accessible to the public and will preserve these to the best of their abilities, in perpetuity.

I hereby grant permission to the Graduate Division of the University of California, San Francisco to release copies of my thesis, dissertation, or manuscript to the Campus Library to provide access and preservation, in whole or in part, in perpetuity.



Author Signature

6/2/2017

Date