NORTHWESTERN UNIVERSITY

Mechanistic Insights into Coactivator Recruitment by Nuclear Receptors

A DISSERTATION

SUBMITTED TO THE GRADUATE SCHOOL

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

for the degree

DOCTOR OF PHILOSOPHY

Field of Interdisciplinary Biological Sciences

By

Nicolas Duane Daffern

EVANSTON, ILLINOIS

September 2022

ABSTRACT

Nuclear receptors (NRs) are an important family of transcription factors that often regulate genes in response to ligands and by way of direct interactions with coactivator proteins. Many NR-coactivator pairs have been identified that cooperate to regulate transcription but fully understanding how NRs recruit specific coactivators involves learning which of their domains interact and how the respective structural and biochemical features govern their associations. Here, I investigated the conformation, dynamics, and interactions of NR and coactivator domains, with the goal of gaining deeper insights into molecular functions.

Initially, I characterized the ligand-binding domain (LBD) of *Drosophila* Ftz-F1 receptor, which is hypothesized to recruit coactivators in a ligand-independent manner. My studies showed that the canonical ligand-binding pocket, which appeared occupied by a short segment of the protein in the crystal structure, is dynamic, and likely samples an open conformation in solution, suggesting that Ftz-F1's recruitment of coactivators could be modulated by ligand binding. I then investigated how another atypical NR, Nurr1, which was also thought to recruit coactivators in a ligand-independent manner, can recruit the coactivator SRC1 in an LBD-independent manner. Specifically, my studies show that the Nurr1 activation function-1 (AF1) domain is unstructured in solution and contains a ~20-residue segment that binds to the SRC1 PAS-B domain, which represents a novel mechanism for coactivator recruitment by NRs. Additionally, structure-function analysis of the interaction revealed that AF1 likely forms a helix upon binding and engages with the same surface on PAS-B as the transcription factor STAT6, suggesting that binding this surface represents a conserved mechanism for transcription factors to recruit SRC1. Since PAS domains are known to bind small molecules, I then investigated whether the SRC1 PAS-B domain could bind small molecule ligands. My studies showed that a conserved, but previously uncharacterized,

pocket can bind small molecules such as prostaglandins, suggesting a potentially novel mechanism for regulating interactions between PAS-B and transcription factors. Collectively, my results have broadened our understanding of how dynamics, ligand-binding, and structure affect NR-coactivator interactions and, in the process, have uncovered new mechanisms of transcriptional activation.

ACKNOWLEDGEMENTS

Firstly, I would like to thank Dr. Ishwar Radhakrishnan for his mentorship during the PhD. He provided continual guidance with new techniques, experimental design, presentations and writing and was always available to talk about difficult projects. I would also like to thank current and past members of the Radhakrishnan lab for their support with this work. Previous to my arrival, the post-doc Zhonglei Chen laid the groundwork for the Ftz-F1 project. An undergraduate, Kelly Zhu, completed much of the cloning and protein purifications necessary for the Tau-1 project. Another undergraduate, Kade Kelly, assisted with computational modeling for the PAS-B ligand project. Finally, I'd like to thank my lab mate and fellow IBiS student, Ryan Marcum, who taught me most of the experimental techniques I know in molecular biology and helped me troubleshoot my many mistakes.

I would also like to thank the members of the core facilities who assisted with my experiments. Dr. Yongbo Zhang helped me optimize and troubleshoot many difficult NMR experiments. Saman Shafie taught me how to set up ESI and MALDI mass spectrometry experiments. Dr. Arabela Grigorescu and James Casey assisted me with ITC and SEC MALS experiments in the Keck Biophysics Facility.

I would also like to thank people who supported my work indirectly. First, and most importantly, I want to thank my family. My parents, Jeff and Nancy, were my biggest supporters during the difficulties of grad school and my wonderful sisters, Margo and Kyra, provided me with lots of laughter and joy. I'd like to thank my close friends from undergrad, Hunter and Greg, for many fun and enlightening conversations. I'd also like to thank my roommate Josh for all the lively discussion and adventures we had. A big thank you is owed to all the members of EBF and GCF who provided me a kind and supportive community through many ups and downs. Finally, I'd like

ABBREVIATIONS

AF1: activation function-1

AR: androgen receptor

bHLH: basic helix-loop-helix

CASTp: computed atlas of surface topography of proteins

CBP: CREB-binding protein

CD: circular dichroism

CPMG: Carr-Purcell-Meiboom-Gill

CSP: chemical shift deviation

DBD: DNA-binding domain

DMSO: dimethyl sulfoxide

ER: estrogen receptor

ESI: electrospray ionization

FRET: fluorescence resonance energy transfer

FTS: fluorescence thermal shift

Ftz: *Fushi Tarazu* homeodomain protein

FXR: farnesoid X receptor

Gdn-HCl: guanidine hydrochloride

GPCR: G protein-coupled receptors

GR: glucocorticoid receptor

GST: glutathione-sepharose transferase

HDX: hydrogen deuterium exchange

HIF-2: hypoxia-inducible factor 2

HPLC: high-performance liquid chromatography

HSQC: heteronuclear single quantum correlation

IDP: intrinsically disordered protein

ITC: isothermal titration calorimetry

LBD: ligand-binding domain

LBP: ligand-binding pocket

LC: liquid chromatography

LRH-1: liver receptor homolog 1

MALDI: matrix-assisted laser desorption ionization

MD: molecular dynamics

MR: mineralocorticoid receptor

MS: mass spectrometry

NF-κB: nuclear factor-κB

NMR: nuclear magnetic resonance

NOE: nuclear Overhauser effect

NR: nuclear receptor

NTD: N-terminal domain

NTR: N-terminal region

PAGE: poly-acrylamide gel electrophoresis

PAS: Per-ARNT-Sim

PCR: polymerase chain reaction

PDB: protein data bank

PG: prostaglandin

PPAR: peroxisome proliferator-activated receptor

PR: progesterone receptor

PTM: post-translational modifications

RAR: retinoic acid receptor

SAXS: small-angle x-ray scattering

SDS: sodium dodecyl sulfate

SEC: size-exclusion chromatography

SF-1: steroidogenic factor 1

SRC: steroid receptor coactivator

STAT: signal transducers and activators of transcription

TF: transcription factor

TR4: testicular receptor 4

TABLE OF CONTENTS

LIST OF FIGURES AND TABLES

CHAPTER 1: Introduction

Overview of coactivator recruitment by transcription factors

Transcription factors in biology

Sequence-specific DNA-binding transcription factors (TFs) are the key regulators of genes, determining when transcription is activated or repressed. In eukaryotes they often accomplish this by binding to DNA and assembling large protein complexes, which primarily act to modify chromatin or initiate transcription[1]. Promoter- or enhancer-binding by TFs act as nucleation sites for protein-protein interaction networks leading to the assembly of the transcription pre-initiation complex[2]. A key player in this process is a type of TF called a transcriptional coactivator. These factors lack sequence-specific DNA-binding activity but play an important role in transcriptional activation by harboring or recruiting chromatin-modifying activities as well as serving as molecular adapters linking DNA-bound TFs with the RNA polymerase. The work in this thesis seeks to further our understanding of an important class of TFs known as nuclear receptors (NRs) and their interactions with SRC coactivators[3].

TFs are often composed of multiple domains connected by flexible linker regions that range in size from tens to hundreds of residues and serve unique functions. Sequence-specific TFs have DNA-binding domains that localize them to target genes and activation domains that engage in protein-protein interactions with coactivators[4]. Likewise, coactivators are multi-domain proteins that bind DNA-bound TFs on the one hand and other coactivators including subunits of the Mediator complex makes contacts with the transcription machinery. A primary goal of this thesis was to characterize domain-level interactions between known transcriptional activator-coactivator pairs. Furthermore, these protein-protein interactions are often modulated by structural dynamics,

intrinsic disorder, and ligand binding. Thus, another major goal was to characterize how these interactions occur and are regulated at the molecular level.

Dynamics, disorder, and ligand-binding in protein-protein interactions

In the context of this thesis, dynamics will refer to the motion of atoms in space in the context of proteins. These motions can occur on a wide range of timescales and amplitudes. Smaller amplitude motions such as bond vibrations and sidechain rotations occur in the timeframe of picoseconds to nanoseconds while larger motions such as rearrangements of protein secondary or tertiary structure can occur in the timeframe of microseconds all the way up to seconds[5]. Protein crystallography was the first technique to show that proteins adopt well organized structures that are the basis of most of their function. However, techniques such as NMR, hydrogen deuterium exchange mass spectrometry (HDX-MS), and fluorescence resonance energy transfer (FRET) spectroscopy, later revealed that these structures can be flexible in solution[6].

Proteins often exist as an ensemble of closely related structures in solution with different regions varying in their stability based on protein sequence, secondary structure, and a variety of inter- and intra-molecular interactions. They can switch between different conformational states stochastically or in response to interactions with another protein or small molecule. These motions are an important aspect of protein function[7]. This includes the ability of enzymes to catalyze reactions, the selectivity of ion channels, and the recognition of ligands by numerous classes of receptors[8,9]. As will be described below, dynamics are especially important for NR LBDs when binding to ligands and recruiting coactivators.

An important sub-class of dynamics is intrinsic disorder. Proteins or domains that are intrinsically disordered exchange between many conformations that exist on a flattened energy

landscape[10]. However, intrinsically disordered proteins (IDPs) often feature short, conserved segments that can form stable secondary structures and play a direct role in binding[11]. Significant computational and experimental work has been done to try and uncover sequence determinants of disorder and IDP function, which represents one of the most challenging areas of IDP research.

The study of disorder and its role in protein function has gained momentum as the biological significance of IDPs has become increasingly clear. Indeed, it is estimated that anywhere from 33% to 50% of the human genome codes for IDPs or proteins containing disordered regions[12]. Especially relevant to the following work is the role of IDPs in protein-protein interactions. IDPs are important components of a variety of molecular processes due to their unique protein binding attributes. The energy needed to fold IDPs upon binding is taken away from their interaction energy which results in high specificity but low affinity interactions[12]. Additionally, their flexible structure can accommodate the shape of multiple binding partners and they often contain multiple interaction motifs. These traits make IDPs important parts of protein complexes, where they can act as scaffolds for the assembly of various subunits[13].

Another important effector of protein-protein interactions that is a focus of this thesis is ligand-binding. Ligands are a primary way for organisms to transmit signals or detect changes in their environment. Because of this, ligand-binding plays important role in gene regulation by binding to upstream signaling molecules, such as GPCRs, or by binding transcription factors themselves, as in the case of NRs[14]. Ligand binding usually causes downstream effects by disrupting or enhancing an interaction between two proteins in a regulatory pathway, such as for NRs where ligand binding changes their affinity for coactivators, resulting in changes in gene expression[15].

Structure and function of nuclear receptors

Historical perspective

Nuclear receptors (NRs) are an important class of transcription factors that regulate a wide variety of physiological processes in metazoans. They were first discovered as the direct molecular link between endocrine hormones that function as signaling molecules and altered patterns of gene transcription[16]. Since then, NRs have been described in species ranging from nematodes to humans that play crucial roles in virtually all major physiological processes.

NRs are soluble, intracellular proteins that employ multiple domains with different functions to transduce environmental signals and regulate gene expression. At the molecular level, NRs generally share a conserved architecture that consists of an unstructured activation function-1 domain (AF1), a small DNA-binding domain (DBD), and the C-terminal ligand-binding domain (LBD) (Fig 1.1).Transcriptional regulation is facilitated by the recruitment of transcriptional coactivators that, in turn, alter DNA accessibility through the actions of intrinsic or associated chromatin-modifying activities and by directly or indirectly or recruiting the RNA polymeras.
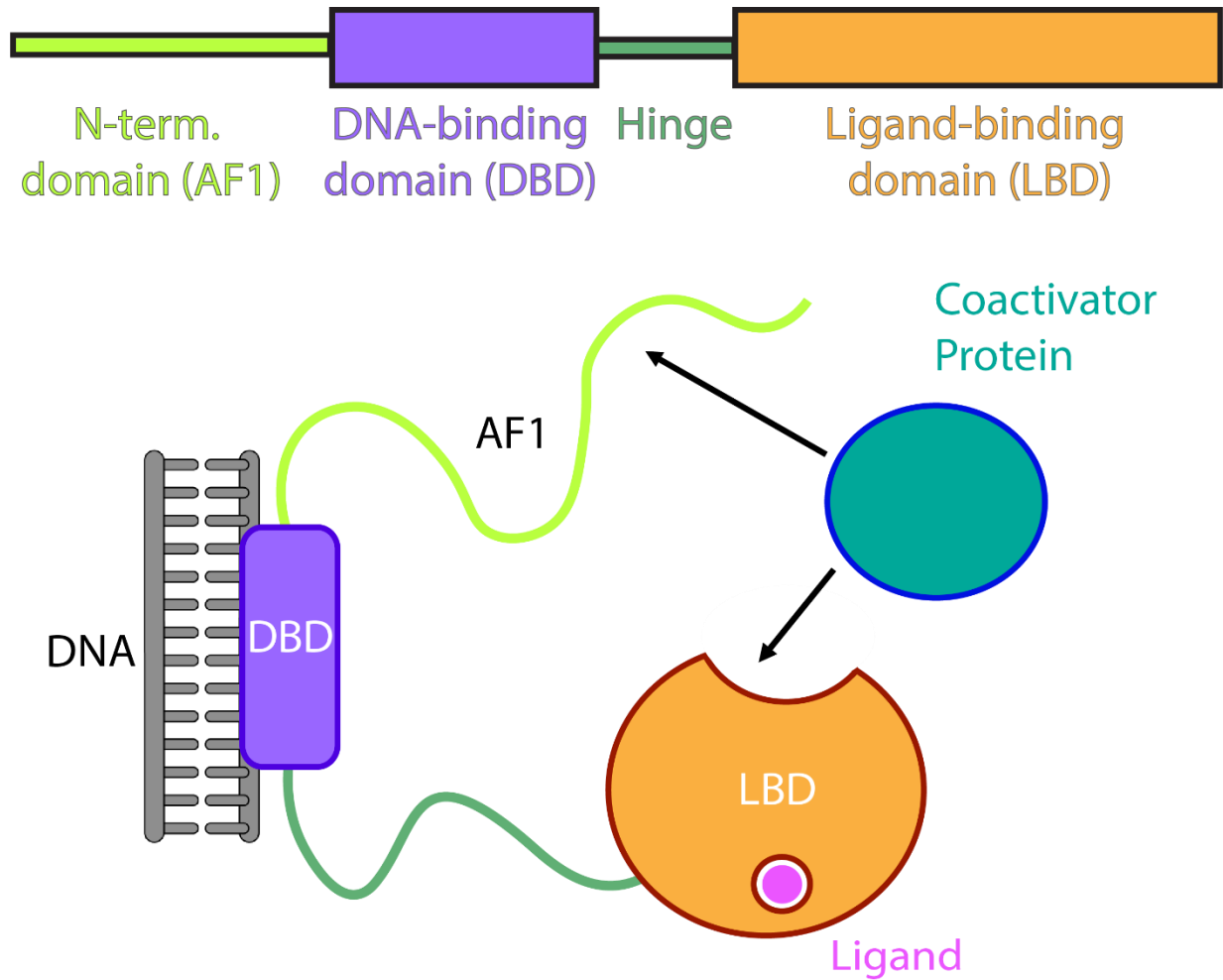
Figure 1.1 Nuclear receptor domain architecture.

A cartoon schematic highlighting the three conserved domains in NRs. NR is shown bound to DNA (grey) via the DBD and a ligand (pink) via the LBD. NRs can bind DNA as monomers, homodimers, or heterodimers, which plays a role in determining which coactivators or corepressors are bound. Recruitment of coactivators (blue) can occur by way of either the LBD or AF1 or a combination of both.

All metazoans utilize NRs and most have a complement of tens to hundreds of unique receptors that regulate different physiological processes. Interestingly, plants, fungi, archaea, and prokaryotes do not contain NR genes, suggesting evolutionarily these receptors were a recent metazoan innovation[17]. The three primary NR domains have varying levels of sequence conservation with the DBD and LBD showing high levels of sequence similarity and the AF-1 domain showing much lower conservation. The high level of conservation of the DBD and LBD has allowed for robust reconstructions of NR evolutionary history which showed that NRs fall broadly into five sub-families[18] (Fig. 1.2). Interestingly, all five families contain receptors that are ligand-dependent as well as those that are ligand-independent. Paradoxically, receptors belonging to different families have similar ligands, whereas receptors in the same family have disparate ligands, raising questions of how this superfamily evolved redundant and distinct ligand-binding functions. An important question relates to the ligand-binding status of the ancestral NR. Until recently, the ancestral NR was widely believed to lack ligand-binding activity, but new evidence suggests otherwise, with simple fatty acids functioning as potential ligands[19].
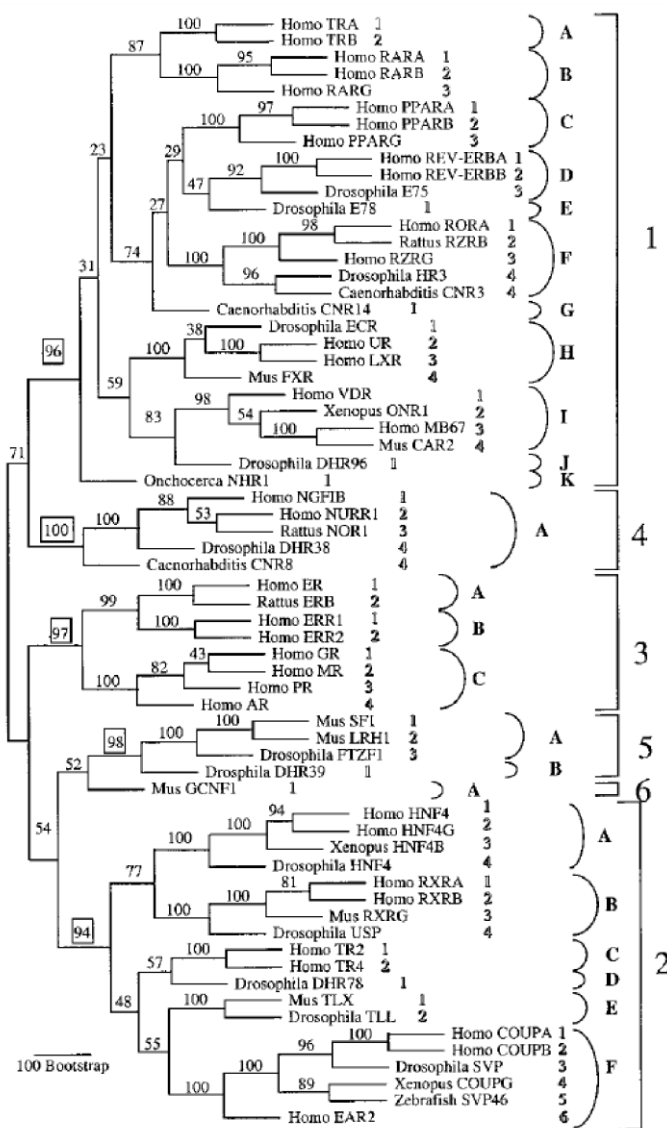
Figure 1.2 Nuclear receptor evolutionary relationships and sub-families.
A phylogenetic tree showing the relationships between the 48 known nuclear receptors in humans and a small number of their close homologs that are known in other model organisms. Sub-families are indicated by numbers at the far right and groups are indicated by uppercase letters. The tree was created based on parsimony analysis of an alignment of the nuclear receptors' sequences. Branch lengths are proportional to the values determined by a bootstrap analysis of robustness. This figure was made by the Nuclear Receptors Nomenclature Committee[20].

In humans there are 48 unique receptors from across all five families. The first receptors to be identified were those that bound to steroid hormones[16]. Because of the highly conserved nature of the LBD/DBD, once the sequences of a handful of NRs were known, researchers were able to identify by homology the other receptors in the genome. The development of co-transfection assays then allowed for the identification of physiologically relevant ligands for many orphan receptors as well as the pathways in which they took part in a relatively short period of time[16].

NRs are best known for their ability to recruit coactivators based on the presence or absence of small molecule ligands, which bind to their LBDs. However, some NRs, known as orphan receptors, have no known ligands or are thought to function in the absence of ligand[21]. Additionally, the AF1 is known to play an important role in regulation but many of the underlying molecular mechanisms are not fully understood[22]. Discovering endogenous, cognate ligands for orphan receptors and characterizing ligand-independent mechanisms of transcriptional regulation are some of the most urgent and difficult areas in contemporary NR research, due to their potential as drug targets despite their occasional lack of obvious binding pockets, the absence of a direct biological connection to a known ligand, or their interactions with low affinity ligands.

Physiological significance of nuclear receptors

Following the cloning of the first NR genes, signaling molecules were found that bound to specific NRs, often revealing the physiological processes that the NR regulated[16]. Specific receptors can have varying effects in different cellular contexts, bind to multiple ligands or no ligands, and regulate multiple unrelated processes. Many of the physiological effects of NRs have yet to be explained by a relevant pathway or molecular mechanism[23]. However, what has been learned about NRs underscores the fact that they are central regulators of much of metazoan physiology.

The steroid receptors (ER, AR, GR, MR, PR) were the first NRs to be studied extensively and their effects on human physiology remain some of the best characterized. These receptors are the primary transducers of steroid hormone signals in cells. For example, it is now well established that GR and MR, which bind glucocorticoids and mineralocorticoids, are critical for neuronal development and regulation of salt concentration[24,25]. Another important group of receptors that were discovered early on were those involved in regulating metabolism. For example, farnesoid-X-receptor (FXR) is important for the production and excretion of bile acid and peroxisome proliferator-activated receptor (PPAR) is known to regulate lipid metabolism[16,26].

NRs play an especially important role in development. It has long been known that an important regulator of development is retinoic acid which is a derivative of vitamin A. Retinoic acid has been shown to play a crucial part in early development and controls the patterning needed for structures such as the brain and lungs to be properly formed in embryos. Its cognate receptor, RAR, has accordingly been shown to regulate *hox* genes which are critical for cell patterning[27]. Two members of the NR5A sub-family are also critical for development. SF-1 is known to be necessary for the development of adrenal glands and gonads in both males and females. Its close relative, LRH-1, is found to be expressed in high levels in embryonic stem cells and it is well known that it is important for ovulation and pregnancy[28].

These roles in critical aspects of human physiology make NRs appealing drug targets. Over the years, small molecules were developed that alter NR function and NRs soon became prime drug targets for many important human diseases. For example, thiazolidinediones are a class of small molecules that bind to PPARγ which have become important drugs used for the treatment of type-2 diabetes[29]. Tamoxifen is another such drug that is used to treat breast cancer and was one of the first selective NR modulators. It was able to target only specific pathways by uniquely acting

on ERα at specific promoters and forming a unique structure when bound to the ER LBD. Drug development for NRs continues to flourish, and currently, about 16% of all small molecule drugs approved by the FDA target NRs[30].

Structure and function of the ligand-binding domain

The work in this thesis looking at NRs focuses on their two activation domains, the LBD and AF1. Work in Chapter 2 looks specifically at dynamics in the LBD, which is well conserved in both sequence and structure. It is composed of 11 to 13 helices which form a globular structure akin to a 3-layered sandwich[31] (Fig 1.3). The helices on either side extend down lower than those in the center, creating a pocket which can accommodate small molecule ligands. NRs bind lipophilic molecules and as such these pockets are primarily filled with hydrophobic residues[32]. Opposite the entrance to the ligand-binding pocket on the LBD is the primary location of coactivator binding as well as the most C-terminal helix known as the activation function 2 (AF2). This AF2 helix can be found in several different conformations (including an unstructured state) depending on the presence or absence of different ligands[33]. When the LBD is bound to an activating ligand, the AF2 helix is primarily positioned such that it forms a hydrophobic cleft. This cleft allows LxxLL motifs found in coactivators to bind to the LBD[34]. However, there are many variations on this theme, especially with regards to the absence of ligand and constitutive activity.
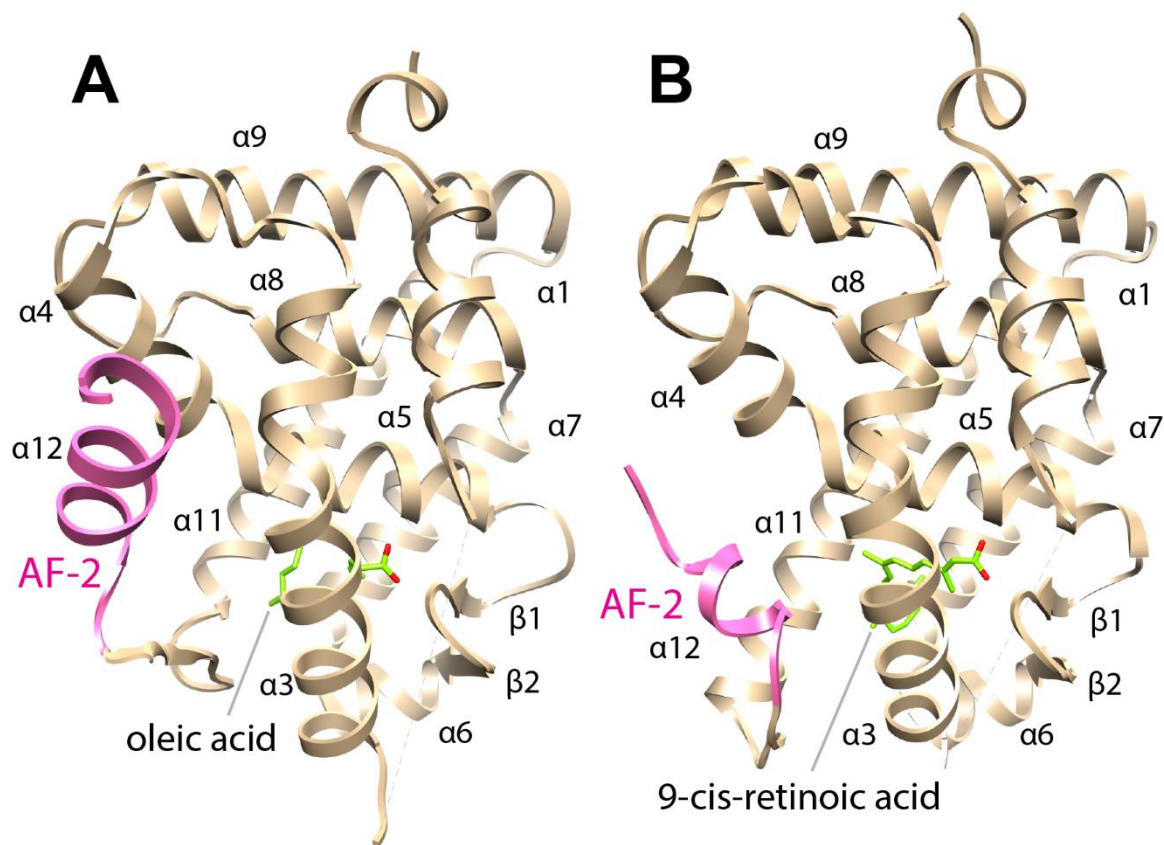
Figure 1.3 Structure of nuclear receptor ligand-binding domain.
Structures highlighting a nuclear receptor LBD in both an active and inactive conformation. The critical activation function-2 helix (AF-2) is shown in magenta. Small molecule ligands are shown in green, and are located in the hydrophobic pocket found in between the two outer layers of helices of the three-layer helical bundle that comprise the LBD. (A) Structure of the RXR LBD bound to the antagonist oleic acid with AF-2 in the inactive conformation (PDB 1DKF)[35]. (B) Structure of the RXR LBD bound to the agonist 9-cis-retinoic acid with AF-2 in the active conformation (PDB 1FBY)[36].

Protein dynamics have emerged as an important regulator of the LBD's function. The first crystal structures of NR LBDs, such as those for ER or RXR, seemed to offer a simple description of coactivator recruitment in which the AF2 helix could adopt several different ordered arrangements, which would alternatively promote or disrupt binding of the LxxLL motif[15,34]. While in some NRs this description is probably accurate, studies using NMR and HDX-MS, for example with PPAR and ER, showed that the AF2 helix may exist in a dynamic state apart from ligands and that stabilization is an important part of ligand-based activation[33]. This hypothesis is especially appealing because it answers previously unexplained phenomena, such as the structural basis for graded agonism and why LBDs are often difficult to crystallize in the absence of a ligand.

Dynamics are also critical for NR activation in the form of conformational changes and allostery. For example, crystal structures of the apo- and ligand-bound forms of RXR revealed that significant structural rearrangements occur upon binding of 9-*cis* retinoic acid that allow for formation of the coactivator binding surface[37,38]. Allostery, mediated through helix 3 and 5, has been shown to be a conserved mechanism of activation in steroid receptors and subtle modulation of that allostery probably played an important role in their evolutionary divergence[39]. Many unique examples of allostery have also arisen from the extensive efforts to develop small molecules that alter NR function. Often these small molecules can bind at locations far from the AF2 or ligand-binding pocket (LBP) and utilize novel allosteric networks to alter coactivator binding[40].

Finally, dynamics are important when considering the potential ability of orphan receptors to bind novel ligands. Many orphan receptors are considered as such because of crystal structures in which their LBP appears to have too little space to accommodate a ligand[41]. However, more recent studies have revealed that these LBPs can be more flexible than is apparent based on the crystal structure and several novel ligands have been suggested for previously reported orphan

receptors. For example, Nur77 was crystallized in the apo form with bulky hydrophobic residues filling its LBP[42]. However, NMR and HDX studies recently reported significant dynamics in the LBP and claimed unsaturated fatty acids as a potential ligand[43]. While the biological significance of these ligands has not been rigorously tested, they show that crystal structures may give incomplete pictures of the accessibility of the interior of the domain and provide additional potential regulatory mechanisms to explain Nur77 function. Likewise, several other traditional orphan receptors, such as COUP-TF2 and TR4, have been shown to bind retinoids with μM affinity and have their activity modulated in the presence of retinoids. Notably, they have helices that normally help to form the LBP that are disordered in the apo structures, further suggesting dynamics as a mechanism for accommodating novel ligands[15].

Structure and function of the activation function-1 domain

The other NR domain that is a key focus of this thesis is AF1. This is the least conserved of the NR domains in terms of both sequence and length and is considered to be mostly disordered[44]. AF1s can range from tens of residues to over 600[45] (Fig. 1.4). For some receptors specific regions of AF1s have been mapped out and shown to be critical for NR mediated gene activation or repression. Additionally, they undergo post-translation modifications (PTMs), such as phosphorylation and methylation that affect NR activity[46]. However, AF1s are the most poorly understood regions in NRs due to the lack of biochemical and structural data and much work is left to be done to better understand the molecular mechanisms that allow AF1s to regulate NR function[45].
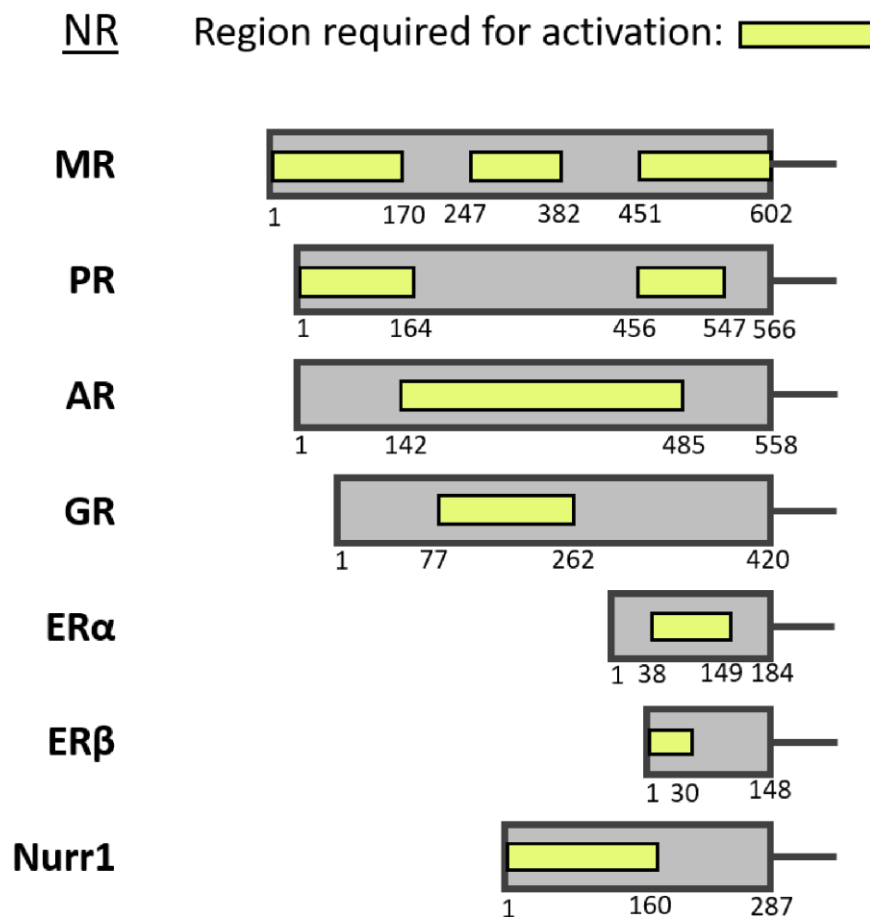
Figure 1.4 Diversity in nuclear receptor activation function-1 domains.
Schematic highlighting the lack of conservation in length and functionality of nuclear receptor AF1 domains. Grey bars show the lengths of AF1 domains for several well characterized nuclear receptors (labeled left), with residue numbers shown below. Regions experimentally shown to be important for activation are highlighted with yellow boxes. This figure was adapted from a figure made by Lavery and Mcewan[22].

Intrinsic disorder is of primary importance when considering NR AF1s. Studies employing HDX and NMR, along with secondary structure prediction software, show that the vast majority of AF1s studied are unstructured in solution[46–49]. This means that, as opposed to the classic globular LBD, the AF1 must be studied as an IDP. The structural biology mainstay of crystallography is ineffective at describing their interactions, as evidenced by the fact that almost all full-length NR structures lack density for residues in the AF1[35,50–52] However, there are still many structural techniques, such as NMR, small-angle x-ray scattering (SAXS), or circular dichroism (CD) that can be used to characterize IDPs, but at vastly different levels of resolution.

For some of the coactivators that mediate AF1-dependent gene regulation, it is not known whether that activity is by way of direct recruitment or via indirect mechanisms. Additionally, there is little data to explain how unstructured AF1 domains form stable interactions with their binding partners[53]. My thesis work addresses some of these questions with regards to specific AF1 domains.

SRCs as nuclear receptor coactivators

General role of SRCs in assembly of transcriptional machinery

Proteins that mediated the function of NRs were proposed to exist shortly after the discovery of NRs based on several unexplained experimental results. First, it was observed that overexpression of one NR would deleteriously affect the function of other NRs in cell-based assays. Second, extracts containing only purified NRs and transcriptional machinery were unable to activate transcription[55]. These data suggested that there were additional factors beyond the transcriptional machinery that worked in conjuction with, and could be shared between, different nuclear receptors. In 1995, this hypothesis was confirmed with the cloning of the gene for SRC1 (also known as NCOA1)[56]. This was the first characterized nuclear receptor coactivator and it was shown that SRC1 could bind to NRs in a ligand-dependent manner and mediate their regulatory abilities[3].

It is now known that there are three members of the SRC family of proteins: SRC1, SRC2, and SRC3[57]. Studies of these coactivators have focused primarily on their role in NR-mediated gene regulation but they can also be recruited by other transcription factors such as nuclear factor-κB (NF- κB)[58], p53[59], and signal transducers and activators of transcription (STATs)[60]. There are only three members of the SRC family, but they regulate a wide variety of physiological processes and have transcriptional effects that differ greatly. SRCs are differentially recruited to genes based on cellular context and the transcription factors recruiting them[57,61]. The SRCs, in turn, recruit additional complexes and coregulators depending on post-translational modifications and the presence of other cofactors[61].

Basic structural architecture of SRCs

Like NRs, SRC proteins have a conserved domain architecture that can engage in many different protein-protein interactions. SRCs can broadly be broadly divided into two regions: a structured region at the N-terminus while the rest of the protein is largely unstructured (Fig. 1.5). The C-terminus contains several motifs and small domains that are important for SRC function. There are three LxxLL motifs that form amphipathic helices that are well known to bind the AF2 region of NR LBDs in the presence of an activating ligand. Additionally, there are two activation domains in the C-terminus, one of which binds the paralogous coactivators CREB-binding protein (CBP) and p300 while the other has been proposed to bind proteins with methyltransferase activity[61]. Binding of CBP/p300 is especially important to SRC function because of the critical role they play in assembling transcriptional complexes and for their ability alter the acetylation state of chromatin[55]. The N-terminus of the SRC proteins contain three structured domains: bHLH, PAS-A, and PAS-B. In general, PAS domains have well-characterized roles in protein-protein interactions, ligand binding, and signaling, among others[62]. However, for the SRC PAS domains the only interaction that is structurally characterized is that of SRC1 PAS-B with STAT6[63].
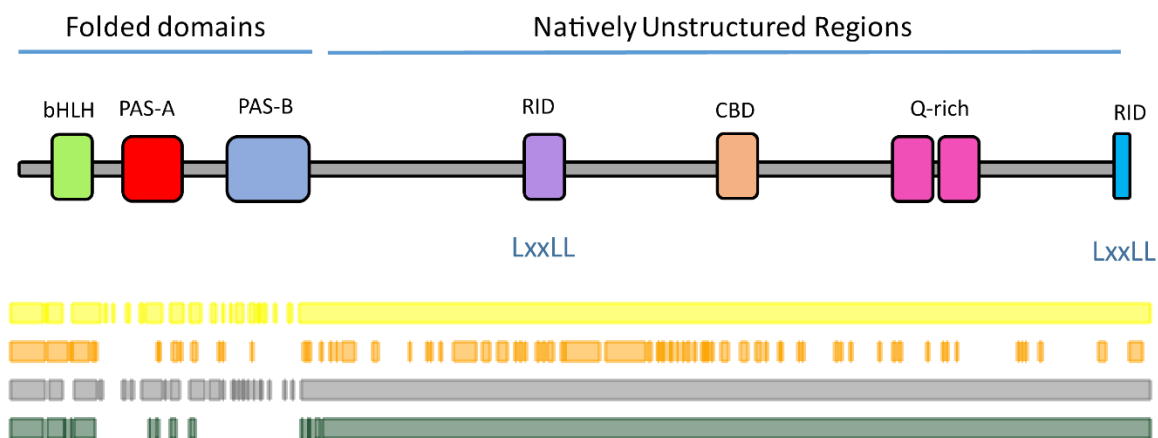
Figure 1.5 Domains and disorder in SRC proteins.

Shown top is a cartoon schematic of the domains and motifs found in the SRC proteins. Regions containing folded or natively unstructured domains are delineated by a blue bar. The RID regions contain the LxxLL motifs involved in interactions with NR LBDs. Shown below are protein disorder predictions for SRC1 from various software with predictions colored based on the software used for generating the prediction. Note that most predicted ordered segments map to the three N-terminal domains.

SRC PAS domains

Part of the following thesis will look specifically at the interactions of NRs with the PAS-B domain of SRC1. PAS-B is part of a family of small, multipurpose domains, known as Per-ARNT-Sim (PAS) domains, that are found throughout all domains of life. They are primarily involved in protein-protein interactions but can have functions as varied as light-sensing and small molecule detection[62]. Despite having widely divergent sequences (less than 20% average sequence identity among individual domains) they have a conserved fold, consisting of an antiparallel β-sheet of five strands with α-helices on either side of the sheet[64].

The SRC PAS domains are structurally similar to those found other bHLH-PAS type TFs, such as aryl hydrocarbon receptor nuclear translocator (ARNT) and circadian locomotor output cycles kaput (CLOCK), which primarily engage in self-interactions via their exposed β-sheet surface to facilitate homodimerization[65]. However, in the case of the SRC PAS domains, the limited evidence available suggests that their primary interactions are with other transcription factors. Structural work showed that PAS-B can directly bind to a peptide fragment of the STAT-6 transcription factor via a conserved groove formed by the domain's α-helices[60]. Additionally, PAS-B domains in other bHLH-PAS transcription factors have been shown to bind short helical motifs in other coactivators such as TRIP230[67]. This suggests that PAS domains play an important role in transcription factor and coactivator interactions by directly binding short peptides that form α-helices.

Summary

Recruitment of coactivators by NRs is a critical step in gene regulation in humans. However, understanding how they interact is complicated by the fact that these are large proteins with multiple functionally and structurally distinct domains. For known NR-coactivator pairs it is often not feasible to solve a structure of the entire protein complex bound to DNA. In these cases, the critical interacting regions of the proteins need to be determined by other means. In my work I employ a variety of techniques, such as biochemical pulldown assays, mutagenesis, and NMR to define these regions. Furthermore, these interactions are governed by the molecular features of the individual domains, but these features remain uncharacterized for many NRs and coactivators. In particular, I am interested in how structural dynamics, intrinsic disorder, and ligand binding play a part in these protein-protein interactions and I used different techniques, such as NMR, mass spectrometry (MS), and isothermal titration calorimetry (ITC), to characterize these different phenomena. Thus, a multi-disciplinary approach allowed me to gain key insights into which domains make direct contact and what molecular features allow for these interactions.

The overarching goal of my thesis was to gain these types of insights with regards to the domains of several important NRs and their coactivators. For the ligand-binding domain of the nuclear receptor Ftz-F1 and the PAS-B domain of the coactivator SRC1, I characterized their biophysical features, such as structural dynamics and ligand binding, that could play significant roles in their interactions with a number of their respective partners. In the case of the NR Nurr1, I discovered that its AF1 domain can interact with the SRC1 PAS-B domain and characterize some of the structural features that allow for binding of an intrinsically disordered AF1 domain. Collectively, my work uncovered a number of novel features that explains how these NRs and coactivators interact with each other and thus may be relevant to other, similar systems.

CHAPTER 2: Dynamics in the ligand-binding domain of the orphan receptor Ftz-F1

Introduction

Ftz-F1 is a member of the NR5A sub-family of NRs that is found in *Drosophila* and plays an important role in regulating segmentation during embryo development[68] (Fig. 1.2). Ftz-F1 is expressed in all somatic cells but works in tandem with the fushi tarazu homeodomain protein (Ftz) to produce expression patterns of alternating stripes that facilitate segmentation. Initially, Ftz-F1 was found colocalized with Ftz in the genome and the two proteins were found to work synergistically to induce gene expression[69]. Further work showed this was the result of cooperative association on DNA and direct binding of a LxxLL motif in Ftz with the LBD of Ftz-F1[68].

Structural studies of the Ftz-F1 LBD in complex with the Ftz LxxLL peptide showed the Ftz-F1 AF2 helix (α12) in a position typical of activated NRs and the LxxLL peptide bound to the canonical hydrophobic groove and charge clamp[70]. The Ftz-F1 LBD had high overall structural similarity to the other characterized NR5A members, SF-1 and LRH-1, with its 12 helices forming the canonical three-layer α-helical sandwich[71]. However, in the Ftz-F1 crystal structure, the α-helix 6 (α6) was seen inserted into a hydrophobic cavity that normally represents the LBP in classical NRs. In contrast, the LRH-1 and SF-1 crystal structures have a large, empty LBP (in mouse LRH-1) or a LBP filled with phospholipids (in human SF-1/LRH-1). Initially, the SF-1 and LRH-1 LBDs were found bound to bacterial phospholipids leading to controversy over the NRs status as ligand-binding. However, more recent studies have shown that they bind phospholipids that are physiologically relevant in humans, such as PIP_3, supporting the significance of their ligand-binding abilities[72–74]. For Ftz-F1, the appearance of α6 in the LBP, along with the fact that

no endogenous ligand had been found for Ftz-F1, led to claims that it acted in a ligand-independent manner[70].

A previous post-doc in the lab, Zhonglei Chen, purified ^{15}N labelled Ftz-F1 LBD and collected NMR spectra of the LBD in the presence and absence of the Ftz peptide (Fig. 2.1A). Triple resonance experiments allowed for assignment of both the apo and holo proteins and subsequent chemical shift perturbation (CSP) analysis revealed that α12 in the apo Ftz-F1 LBD remains in an active conformation, poised to bind the Ftz peptide. Additionally, this analysis showed that the peptide binds to the LBD in the canonical hydrophobic groove, confirming that the binding interface seen in the previously solved crystal structure is also seen in solution (Fig. 2.1B). Unexpectedly, inspection of the HSQC of the bound LBD revealed multiple resonances that were broadened, suggesting those residues were undergoing motions on the microsecond to millisecond timescale. These resonances mapped to residues found in α6, in direct contrast to the Ftz-F1 LBD crystal structure in which α6 is well packed in the LBP (Fig. 2.2).
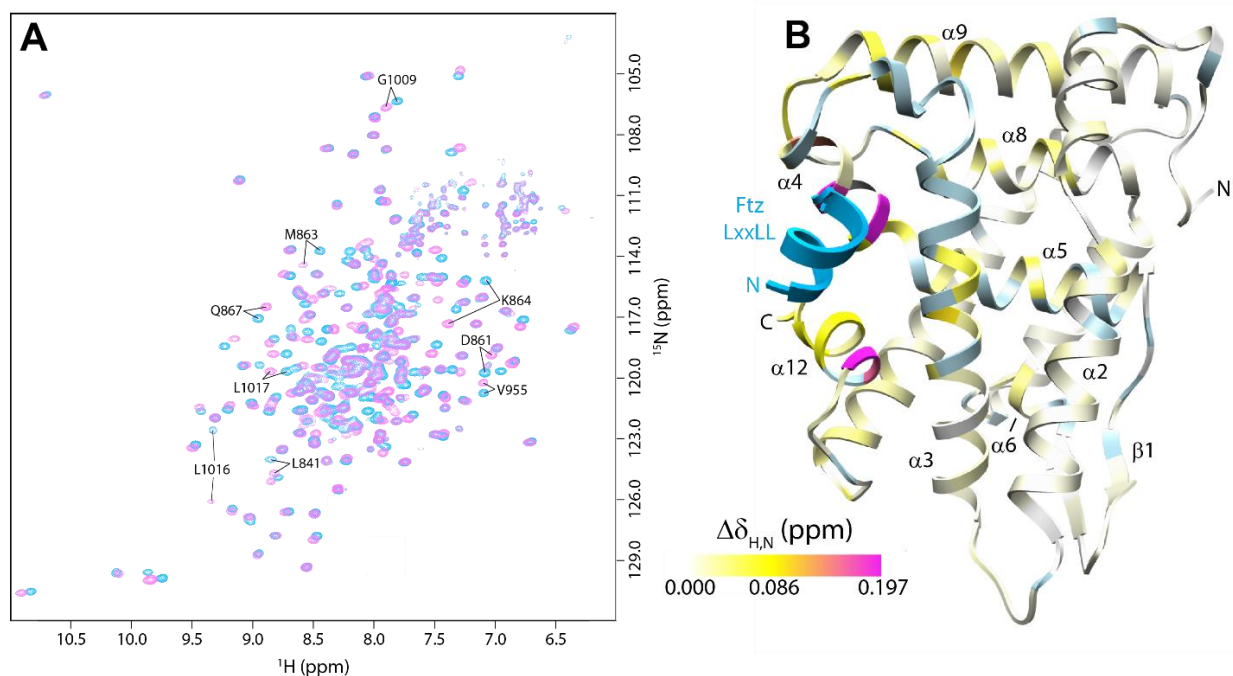
Figure 2.1 Assessment of LxxLL binding to the Ftz-F1 LBD by NMR.
(A) ^{1}H$-^{15}$N correlated spectra of apo (blue) and Ftz LxxLL-peptide bound Ftz-F1 LBD
(magenta). Select backbone correlations that exhibit significant chemical shift changes upon
peptide binding are annotated. (B) Backbone chemical shift perturbations caused by Ftz
LxxLL peptide binding mapped onto the polypeptide backbone of the LBD crystal structure.
The intermediate and high values in the color key are $\langle \Delta\delta_{H,N} \rangle + 1\sigma$ (yellow) and $\langle \Delta\delta_{H,N} \rangle + 3\sigma$
(magenta), respectively, where $\langle \Delta\delta_{H,N} \rangle$ is the average value and σ is the standard deviation.
Colors are linearly interpolated to reflect the corresponding values. Prolines and residues with
unassigned resonances are colored light blue. Data collected by Zhonglei Chen. Data were
processed and backbone resonance assignments were made by Ishwar Radhakrishnan.
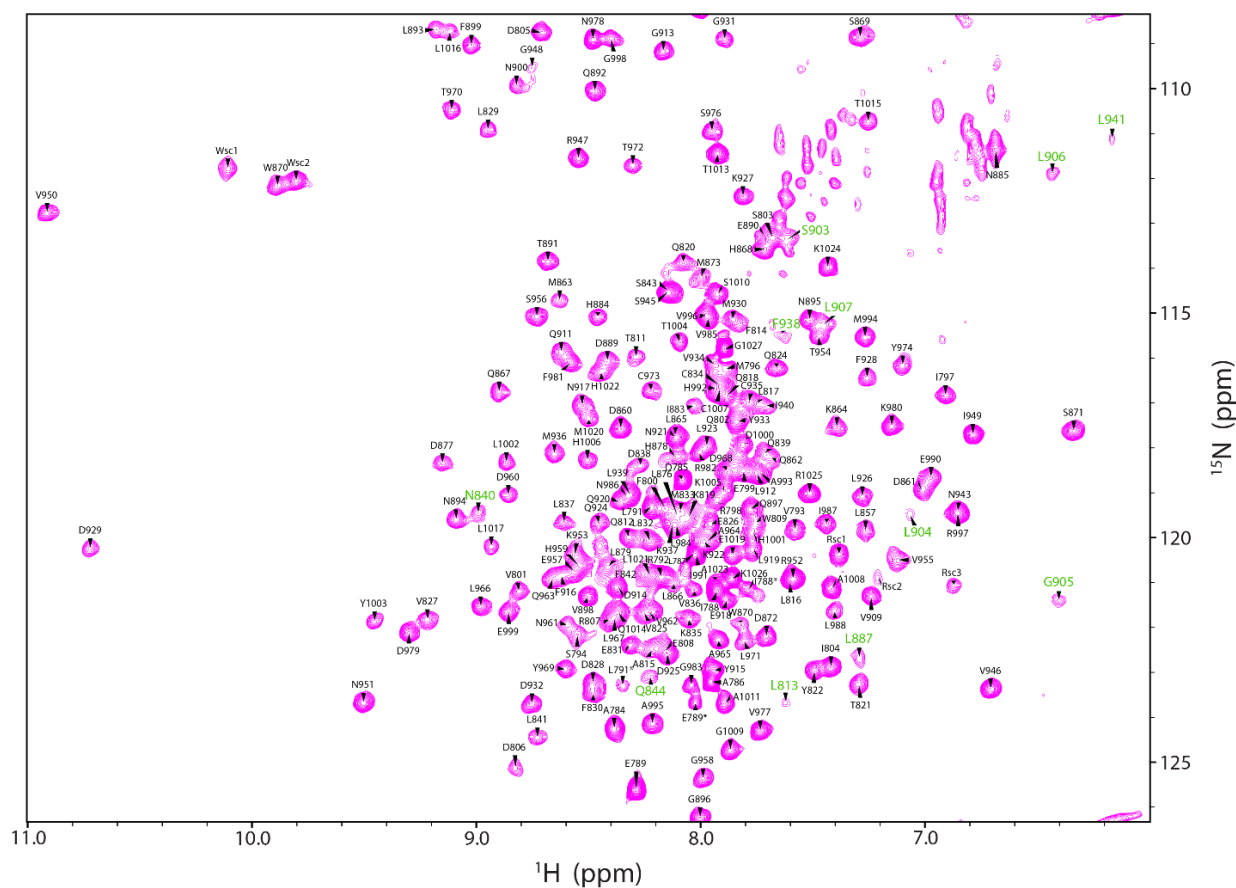
Figure 2.2 Broadened resonances in the HSQC spectrum of the LBD.
An expanded plot of the $^1H-^{15}N$ correlated spectrum of Ftz-F1 LBD recorded at 29 °C. Backbone resonance assignments are annotated. Annotations with asterisks belong to a minor conformer resulting from *cis-trans* proline isomerization at the N-terminus. Selected residues that exhibit substantial resonance broadening are colored green. Data collected by Zhonglei Chen and processed by Ishwar Radhakrishnan.

Given the importance of α6 in determining if the Ftz-F1 ligand-binding pocket is accessible, the following work attempts to characterize the conformation and dynamics of α6 in solution. Results will be described from several different NMR experiments known as CPMG (Carr-Purcell-Meiboom-Gill) relaxation dispersion, T_1, T_2, and NOE (nuclear Overhauser effect) experiments acquired for ^{15}N spins which characterize motions on a wide range of timescales from ps-ns to μs-ms at the leve of protein backbone for each residue. Additionally, the implications of mutagenesis experiments and homology modelling will be discussed with regards to the conformation and occupancy of α6 in the LBP.

Experimental Methods

Production of the wild-type and mutant Ftz-F1 LBDs, Ftz LxxLL peptide, and mouse SF1 LBD

The coding sequence for *Drosophila melanogaster* Ftz-F1 LBD (residues 785-1027) was amplified by PCR, cloned into the pMCSG7 expression vector[75], sequenced to verify identity, and expressed in *E. coli* BL21(DE3) cells at 16 °C. After 20 h, the cells were harvested, resuspended in lysis buffer (50 mM Tris, pH 8.5, 200 mM NaCl, 1 mM BME, 0.5 mM TCEP, 5 mM Imidazole, 0.1% Triton X-100) and lysed by sonication for 30 min. The lysate was clarified by centrifugation at 12,000 rpm for 25 min and the soluble supernatant was incubated with a Ni^{2+}-NTA resin (Qiagen), washed with buffer (50 mM Tris, pH 8.5, 200 mM NaCl, 1 mM BME, 0.5 mM TCEP, 5 mM imidazole) and eluted with the same buffer containing 250 mM imidazole. Eluted fractions of the protein were dialyzed overnight at 4 °C against the lysis buffer lacking detergent and imidazole and incubated in the presence of 1:20 (w/w) TEV protease:protein to simultaneously remove the His$_6$-tag. The dialysate was either concentrated and subjected to size exclusion chromatography

or was incubated with Ni^{2+} resin and the flow-through collected. The sample was then dialyzed into the appropriate buffers and concentrated for further studies.

The G913Y Ftz-F1 LBD mutant was produced by mutating the expression vector encoding the wild-type protein using the QuikChange site-directed mutagenesis protocol (Agilent Genomics). Following verification of the DNA sequence, the mutant protein was expressed and purified using a similar protocol as the wild-type protein.

The *Drosophila melanogaster* Ftz LxxLL peptide corresponding to residues 103-119 (EERPSTLRALLTNPVKKY) was synthesized using automated methods (University of Utah Peptide Facility) and purified by reversed phase HPLC. The sequence contains a non-native tyrosine residue at the C-terminus to facilitate concentration measurements. The identity of the peptide was verified by mass spectrometry.

The mouse SF1 LBD (residues 221-462) for phospholipid binding assays was produced using the same approach as described above for the *Drosophila* Ftz-F1 LBD.

Fluorescence Thermal Shift (FTS) assays

FTS assays with the wild-type and the G913Y LBD mutant were performed on a Bio-Rad IQ5 Real-Time PCR machine at a protein concentration of 1.8 μM in the absence or presence of 50 M Ftz LxxLL peptide in 20 mM sodium phosphate buffer (pH 7.2) containing 50 mM NaCl, 1 mM TCEP, and 5x SYPRO Orange dye. Fluorescence data were acquired by varying the temperature from 25 °C to 95 °C at the rate of 0.5 °C/s. Data were processed using scripts written in R developed by myself.

NMR sample preparation and NMR spectroscopy

Uniformly ^{15}N and/or ^{13}C-labeled samples of wild-type and mutant Ftz-F1 LBD were prepared as described above except cells were grown in media with U-^{15}N-ammonium sulfate and/or U-^{13}C-glucose (Cambridge Isotopes) as the sole nitrogen and carbon sources; isotope incorporation was verified by mass spectrometry. The sample concentration for NMR studies were in the 0.4-0.7 mM range and the samples were prepared in 20 mM sodium phosphate buffer (pH 6.8) containing 50 mM NaCl, 1 mM TCEP, and 0.5 mM EDTA. NMR data were acquired principally on an Agilent DirectDrive 600 MHz spectrometer equipped with a cold probe at 29 °C. NMR data were processed using Felix 98.0 and analyzed using NMRFAM-Sparky. Backbone assignments were accomplished manually using TROSY versions of 2D ^{1}H-^{15}N HSQC, 3D HNCA, HN(CO)CA, HNCACB, CBCA(CO)NH, HNCO, and ^{15}N-edited NOESY-HSQC (all part of the Agilent BioPack library).

CPMG relaxation dispersion measurements[76,77] were performed for wild-type Ftz-F1 LBD at field strengths of 14.1 and 16.45 T at 29 °C; a separate measurement was conducted at 24 °C at 14.1 T. In addition to a reference spectrum, spectra with a constant time CPMG delay (T_{CP}) of 40 ms were recorded with the following pulsing rates: 50, 100 (x2), 150, 200, 300, 400 (x2), 600, 800, 1000 Hz (at 14.1 T) and 50 (x2), 150, 250, 350, 450, 550, 650, 750, 850 Hz (16.45 T). Spectra were acquired with a recycle delay of 2.5 s to mitigate the effects of heating from radiofrequency pulses; compensatory pulses were applied prior to the recycle delay to maintain similar levels of heating in all the spectra. Experiments with different pulsing rates were acquired in an interleaved mode in randomized order to minimize variation in experimental conditions. Peak heights were measured in NMRFAM-SPARKY[78] and reformatted for analysis using the relax software suite[79].

The $R_{2,\text{eff}}$ dispersion curves were evaluated for consistency across all three data sets. The curves were fitted using the relax software suite assuming no exchange, 2-site fast exchange, and 2-site slow exchange analytic models; the best model was chosen based on the Akaike Information Criterion. The curves were fit after clustering spatially proximal residues (local clustering) followed by global clustering of all exchanging residues. The goodness-of-fits between the various clustering schemes were evaluated by comparing χ^2 values. Monte Carlo simulations were used to model uncertainties in the fitted values.

{^{1}H}-^{15}N heteronuclear Overhauser effect (NOE) experiments were conducted at 14.1 T following the approach of Ferrage et al[80]. Experiments were performed in triplicate with a recycle delay of 6.5 s (either greater than or ca. 5*(^{15}N T$_1$)). To minimize variation in experimental conditions, the reference and steady-state experiments were performed in an interleaved mode. ^{1}H pulses in the steady-state experiment were centered on the ^{1}H^N region (8.2 ppm) and applied at a lower power over 4 s of the recycle delay; water magnetization was carefully preserved through flip-back pulses in the reference experiment. Peak heights were measured in NMRFAM-SPARKY and uncertainties in the measurements were propagated to the calculated NOE values.

Results

Characterization of slow timescale motions in the LBD

Because the broadened resonances in the Ftz-F1 HSQC spectrum are indicative of slow timescale motions (microsecond-millisecond), I initially used ^{15}N CPMG relaxation dispersion experiments[77,81] to look at dynamics. These report on slow timescale motions by showing systematic variation in effective R_2 rates when conformational exchange is occurring near the rate of CPMG pulses. Data collection was performed at two different magnetic field strengths (14.1

and 16.45 T) to account for potential variation in the observed exchange rate as a function of field strength. While most residues did not show any signs of exchange in these experiments, there were a subset of residues that appeared to exhibit exchange induced dispersion. These residues could be put into subsets based on location, which mapped to different regions of the domain (Fig 2.3).
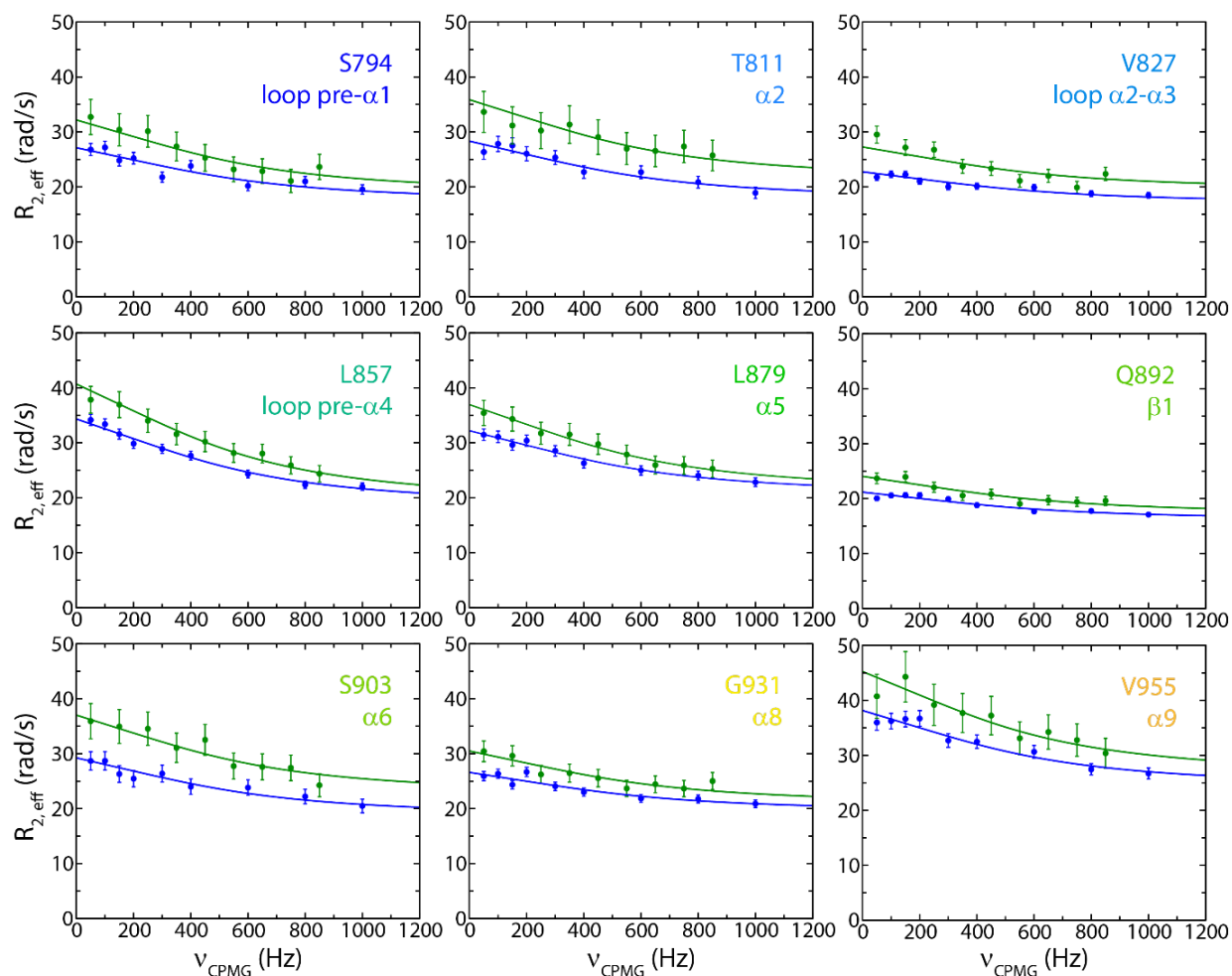
Figure 2.3 Microsecond to millisecond time scale motions in the Ftz-F1 LBD in complex with the Ftz LxxLL peptide.

Representative $R_{2,eff}$ dispersion curves as a function of CPMG pulse repetition rate from different parts of the domain (residue annotations are color ramped from blue to red to reflect their position within the domain). The measurements were performed at 14.1 T (blue symbols and lines) and 16.45 T (green symbols and lines). The fitted values are drawn as solid lines and fitting uncertainties are shown as error bars. Calculated exchange rates are the result of a fitting process that assumed a shared exchange rate for all the residues deemed to undergo exchange. Data analysis performed by Ishwar Radhakrishnan.

Visual inspection of these groupings revealed a high concentration of dynamic residues in α6 and in helices lining the LBP. These included α2, α3, α5, and the loop connecting α5 to the β-hairpin. Another grouping of dynamic residues is found in a region that includes the β-hairpin the C-terminus of α2, and the loop that follows α2. Similarly, a cluster of residues exists in the segment spanning the C-terminus of α3 and the succeeding loop segment including the short α3' helix and up to the beginning of α4, as well as in the segment at the C- and N-termini of α8 and α9. Two additional, smaller clusters of residues are present, one at the N-terminus of the LBD and C-terminus of α9 and the other at the C-terminus of α1 and N-terminus of α8 (Fig 2.4).
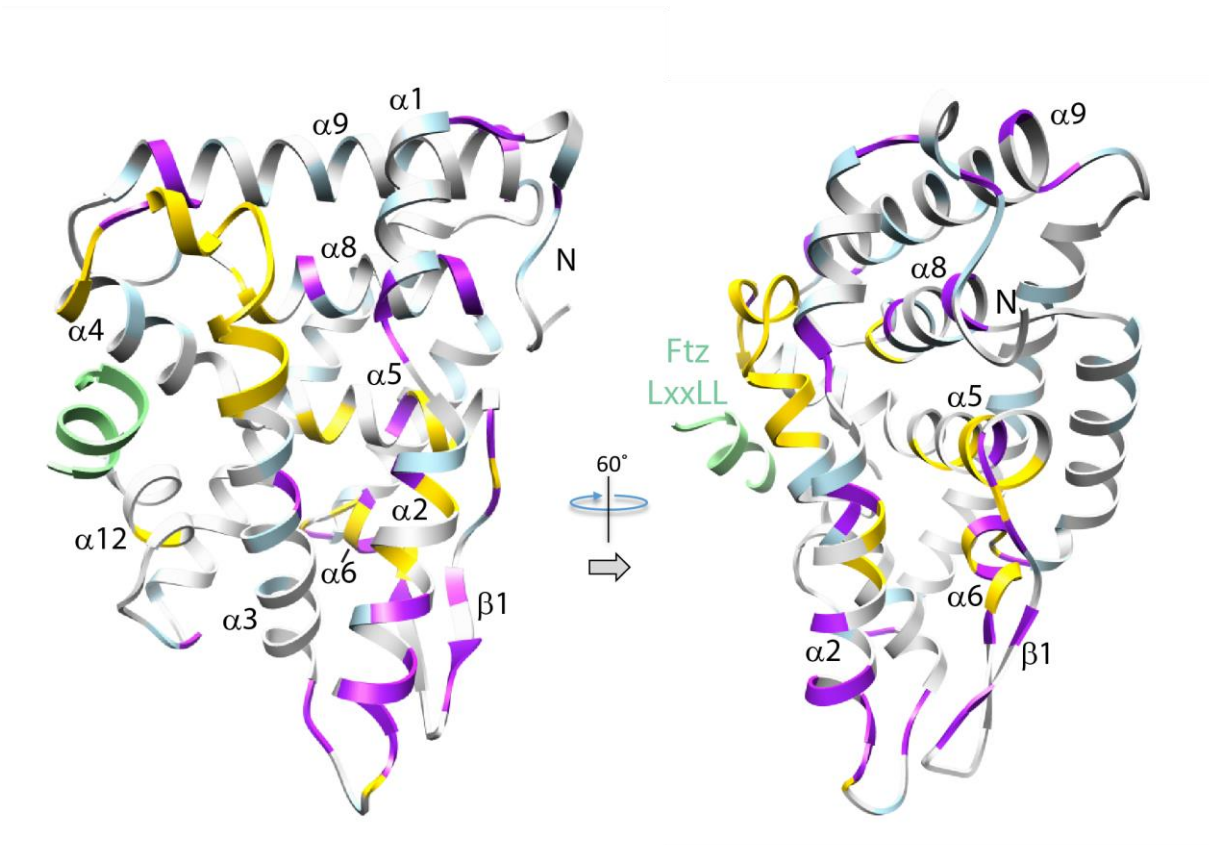
Figure 2.4 Sites undergoing conformational exchange in the Ftz peptide-bound Ftz-F1 LBD.

Residues are colored based on whether exchange could be quantified via CPMG relaxation dispersion experiments (purple) or could not be quantified but are suspected to undergo exchange as implied by severe or substantial resonance broadening (gold). Prolines and residues with overlapped resonances are colored light blue.

Next, variations of R_{eff} as a function of CPMG pulse rate were fit to different exchange models to quantitatively assess the rate of exchange. Data were fit to both a two-site slow exchange and two-site fast exchange model with the best fits coming from the fast exchange model. However large uncertainties existed for individually fit residues, so residues were clustered into five sets based on the previously mentioned structural groups to try and improve fits. Fitting of these groups gave better fits with exchange rates that ranged from 2080 ± 310 s^{-1} to 3750 ± 730 s^{-1}. I also attempted fitting with all data merged into a single group which ended up producing the smallest deviations and resulted in a global exchange rate of 3530 ± 90 s^{-1} (Fig 2.3). To increase my confidence in these results I also collected CPMG data at a lower temperature (24 °C instead of 29 °C). Fitting all the resultant data globally as before yielded a lower exchange rate of 1720 ± 60 s^{-1}, as would be expected at a lower temperature (Fig. 2.5). Together these data suggest that conformational dynamics are occurring in the Ftz-F1 LBD on the millisecond timescale, primarily in and around α6.
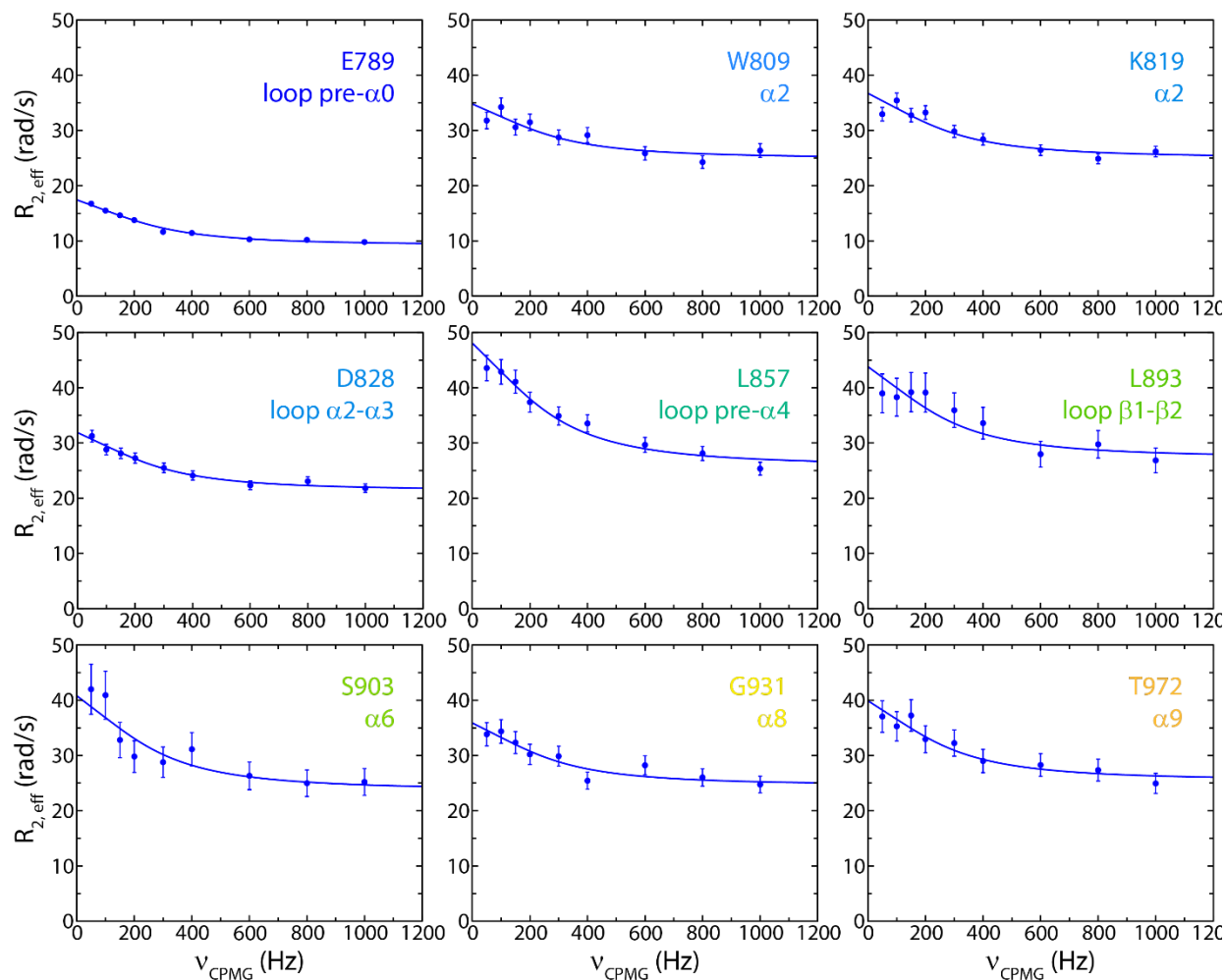
Figure 2.5 CPMG measurements performed at a lowered temperature.
A representative set of R_{2eff} dispersion curves recorded at 24 °C as a function of CPMG pulse repetition rate from different parts of the domain. Depicted as in Figure 2.3; residue annotations are color ramped from blue to red to reflect their position within the domain. The fitted values are drawn as solid lines and fitting uncertainties are shown as error bars. Calculated exchange rates are the result of a fitting process that assumed a shared exchange rate for all the residues deemed to undergo exchange. Data analysis performed by Ishwar Radhakrishnan.

Characterization of fast timescale dynamics

Since CPMG experiments only report on motions occurring on a slow timescale, ^{1}H-^{15}N heteronuclear NOE experiments were performed next to assess motions, such as sidechain rearrangements and fluctuations in loop structures, that occur on a fast timescale (picoseconds to nanoseconds)[80,82]. In these experiments low NOE values are indicators of the presence of fast timescale motions while high NOE values indicate that these motions are absent. NOE values for residues at the N-termini (-0.02 ± 0.55; reported values are the average and standard deviation for each segment) and C-termini (0.26 ± 0.41) as well as at the loop regions between α2-α3 (0.65 ± 0.07) and α11- α12 (0.43 ± 0.15) NOE values were low, which is expected due to their propensity for rapid fluctuation. On the other hand, residues that were in helices that would not undergo such motions had high NOE values, close to the limiting value (0.79 ± 0.10) (Fig 2.6A). Based on the apparently tight packing of α6 and its B-factors (46.5 ± 3.6 Å^2) that are in line with other static helices (51.0 ± 8.8 Å^2), the crystal structure of the LBD would seem to indicate that α6 should have similarly high NOE values (Fig 2.6B)[70]. However, residues in α6 had lower than expected NOE values for a helix (0.63 ± 0.11), further suggesting that this helix is dynamic beyond what would be expected based on the crystal structure (Fig 2.6A).
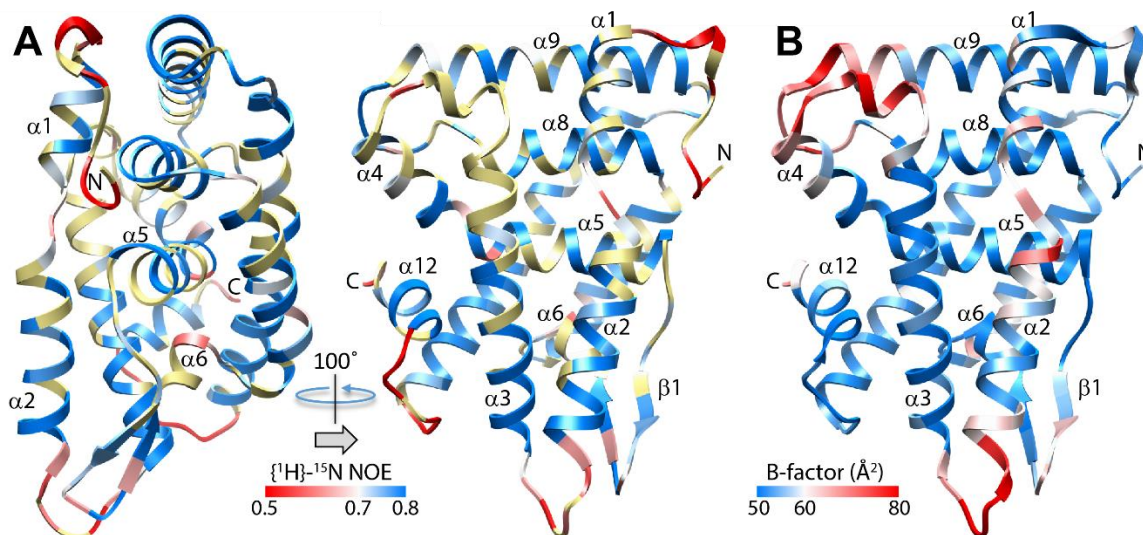
Figure 2.6 Picosecond to nanosecond time scale motions in the Ftz-F1 LBD in complex with the Ftz LxxLL peptide.
Views of the LBD with (A) $\{^1H\}-^{15}N$ heteronuclear NOEs and (B) crystallographic B-factors mapped onto the polypeptide backbone. Unassigned residues, prolines, and residues for which NOE data could not be reliably quantified are colored yellow. Colors and the corresponding values are as depicted by the color keys; colors are linearly interpolated between the values displayed in the keys.

It should be mentioned that other regions in the domain have a discrepancy between the average NOE values and B-factors of their residues. Specifically, the polypeptide segments at/near the N-terminus as well as the loop linking α11 and α12 also exhibit low NOE values (-0.02 ± 0.55 and 0.43 ± 0.15, respectively) but moderate values for the B-factors (51.6 ± 4.0 Å^2 and 50.3 ± 2.7 Å^2, respectively. However, both deviations from the expected values can be understood in light of crystal contacts that likely stabilize these regions in the crystal structure and be absent in solution. There are no such crystal contacts in the case of α6 to explain the discrepancy between NOEs and B-factors. A more reasonable explanation, that aligns with both the CPMG and NOE data, is that α6 samples multiple conformations that likely exist both inside and outside of the LBP. This would result in slow timescale exchange as the helix enters and exits the pocket and fast timescale exchange as the helix partially unravels when free in solution. The crystal structure could then be interpreted as a view of one of the conformations (the highly populated conformation) that was selected for during the crystallization process.

Assessing the location of α6 in the major conformer

A key question that arises from this model is which conformation, that of α6 located in the LBP or exposed to solvent, represents the major conformer? Unfortunately, broadening of resonances in α6 prevented collection of internuclear NOEs to determine the exact location of the helix. Instead, I addressed this question by designing a mutant that would perturb resonances for residues in α6 differently based on α6's majority location. Specifically, a glycine to tyrosine (G913Y) mutant was made for a solvent exposed residue in α7 that was predicted not to affect the stability of the domain based on FoldX calculations[83]. The mutant had a melting temperature slightly lower than wild-type (45.2 ± 0.4 vs. 46.5 ± 0.3 °C) and bound to the Ftz LxxLL peptide efficiently and

had a melting temperature equivalent to wild-type in the bound state (49.9 ± 0.9 vs. 49.6 ± 0.3 °C)

(Fig 2.7A).
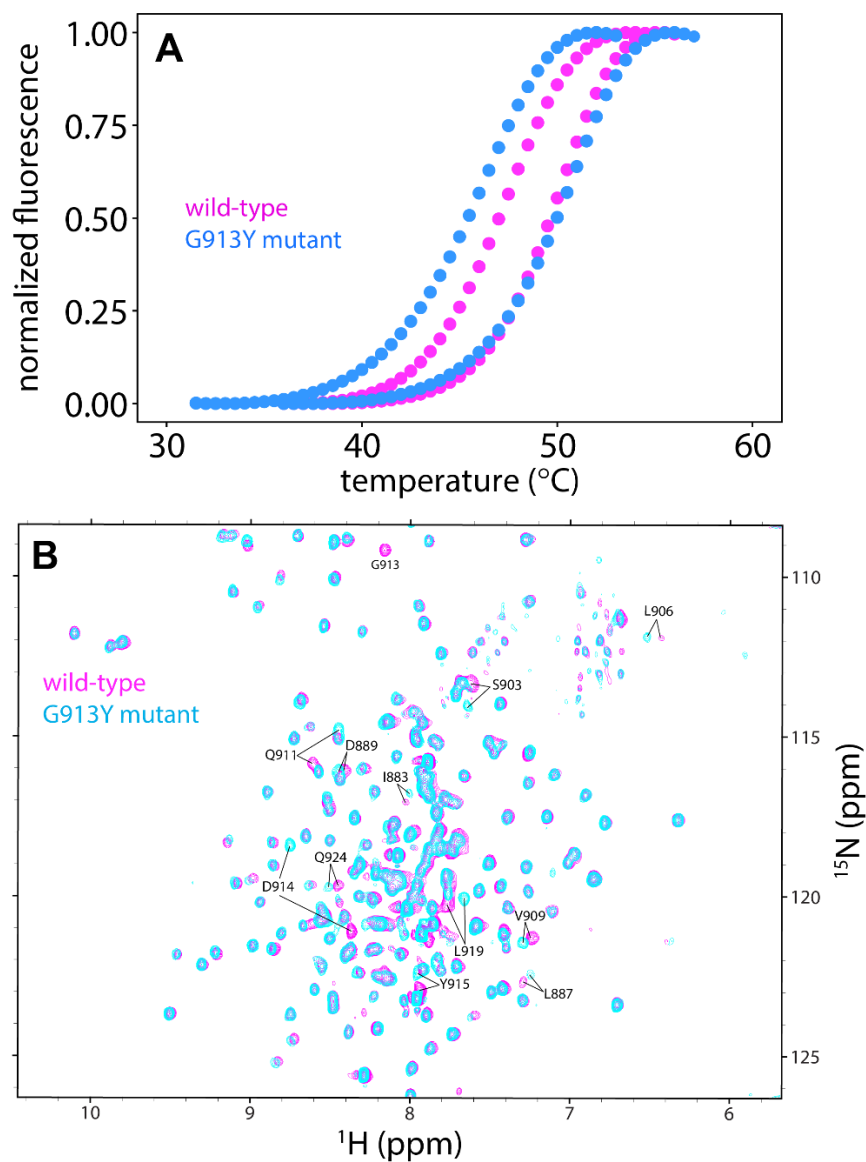
Figure 2.7 Characterization and comparison of the Ftz-F1 LBD G913Y mutant with wild-type.

(A) Increases in the melting temperature of the G913Y mutant and the wild-type protein in the absence and presence of the Ftz LxxLL peptide. (B) The ^{1}H–^{15}N correlated spectrum of the mutant in complex with the Ftz peptide overlaid on top of the corresponding wild-type spectrum. Selected resonances that are perturbed by the mutation are annotated.

HSQC spectra collected for the G913Y mutant were compared to spectra of the wild-type LBD collected under similar conditions (Fig 2.7B). To help in interpreting the CSP data, a homology model was made of the Ftz-F1 LBD based on the mouse LRH1 LBD structure which has α6 located outside of the LBP and would likely be representative of a solvent exposed α6 conformation in the Ftz-F1 LBD. As expected, mapping of CSPs from the G913Y mutant onto the LBD structure showed changes to residues on either side of the mutation. However, there were significant CSPs associated with residues in α6, the C-terminus of α5, and the α5-β1 linker region, which would be consistent with a major conformer similar that seen in the Ftz-F1 LBD crystal structure and not the LRH1 based homology model (Fig 2.8). When assigning backbone resonances for the mutant to confirm the identity of shifted peaks I also noted that ^{13}C$^{\alpha}$ shifts for residues around α6 were very similar to wild type which further suggested that the α6 primary state was located in the LBP.
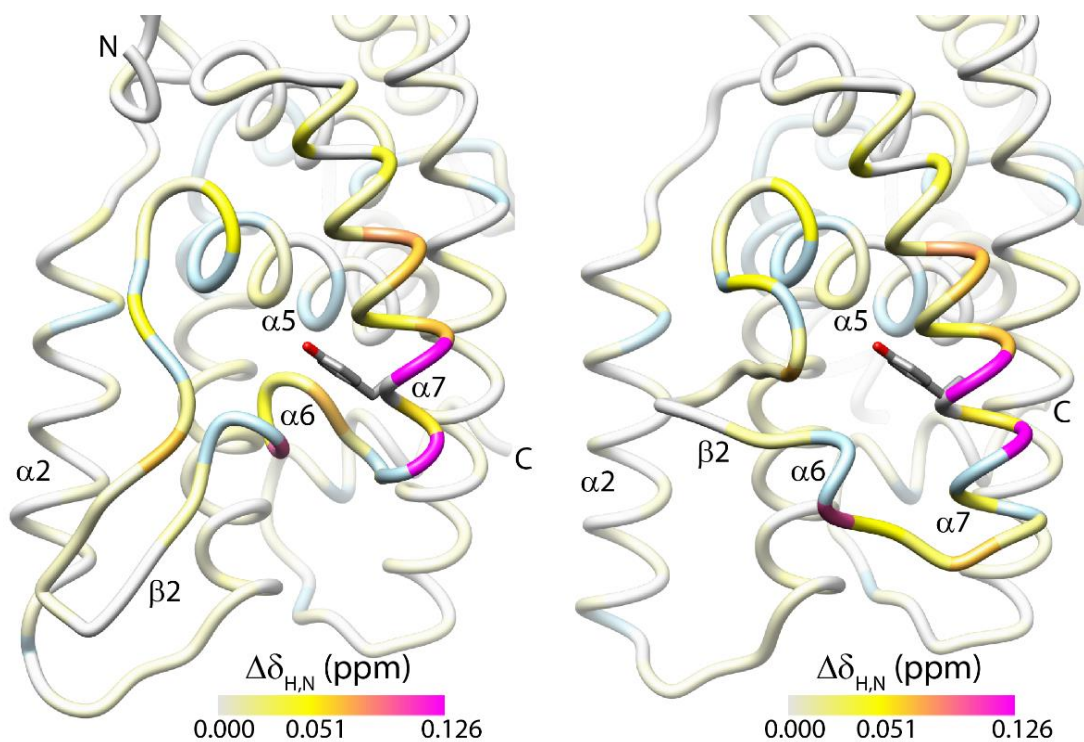
Figure 2.8 Major conformer in solution resembles the conformation captured in the crystal for the Ftz-F1 LBD.
Chemical shift deviations ($\Delta\delta_{H,N}$) between the Ftz peptide-bound forms of the wild-type Ftz-F1 LBD and its G913Y mutant. The deviations are mapped onto the polypeptide backbone of the crystal structure (PDB entry 2XHS; left) and a homology model for the same sequence based on the crystal structure of the mouse LRH1 ortholog (PDB entry 1YOK; right). The computationally predicted location of the tyrosine side chain in the mutant is depicted for reference. The intermediate and high values in the color key represent $\langle\Delta\delta_{H,N}\rangle + 1\sigma$ (yellow) and $\langle\Delta_{\delta H,N}\rangle + 3\sigma$ (magenta), respectively. Colors are linearly interpolated to reflect the corresponding values. Prolines and residues with unassigned resonances are colored light blue.

Conclusions

A recurring theme from solution NMR studies of NR LBDs is the localized or widespread resonance broadening in the apo protein, likely caused by slow timescale motions[38,84,85]. Ligand binding leads to complete quenching of dynamics in some cases, as evidenced by the narrowing of linewidths or recovery of 'missing' resonances, but only partial quenching in others. Often, residues that comprise the LBP are affected by resonance broadening in the apo state, and it stands to reason that the slow timescale dynamics reflect the sampling of multiple conformational states to facilitate ligand binding.

My studies of what is widely regarded as a *bona fide* orphan receptor in which the LBP is occupied by a helical segment (α6) within the protein reveals a surprisingly dynamic α6 helix with motions occurring on both the ps-ns and the μs-ms timescales. The presence of slow motions was not expected because the receptor is not known to be regulated by small molecule ligands, which would require eviction of the helix and remodeling of the LBP. Therefore, what might be the purpose of such motions? One possibility is that these intrinsic motions could promote dissociation of the bound Ftz LxxLL motif, not unlike the coupling noted previously between motional processes and the catalytic rate in enzymes[86]. Indeed, in my studies I find that the motions on the slow timescale are not just confined to α6 and segments in direct contact with the helix (*i.e.* residues that comprise the LBP); rather, they pervade through the domain, extending up to the LxxLL binding surface, hinting at the presence of a communication network involving residue-level contacts that connects these sites, akin to those described previously for the orthologous LRH1 receptor[87]. Besides identifying regions affected by these motions, my studies yielded quantitative insights into the motional timescales with parsimonious models used for data fitting indicating exchange in the 10^{-3} s range. This yields an association rate of ca. 10^9 M^{-1} s^{-1},

since the equilibrium dissociation constant of the Ftz-F1 LBD-Ftz LxxLL peptide complex was previously measured to be ca. 1 μM[70]. Because the inferred association rate is at the higher end of the spectrum, I suggest that the μs-ms timescale motions within the LBD may have a role in the dissociation of Ftz LxxLL peptide, although further studies are required to conclusively address this issue.

The detection of fast motions in the Ftz-F1 α6 helix is also unexpected. Although, the helix is short (5 residues long) and bears the hallmarks of both an α-helix and a 3_{10}-helix in terms of hydrogen bonding interactions, how a secondary structural element deeply embedded in the hydrophobic core could undergo motions on the ps-ns timescale is confounding. However, since the LBD samples at least two distinct conformations, I surmise that the fast timescale dynamics and the resulting low values for the ^{1}H-^{15}N heteronuclear NOEs for α6 most likely emanates from a second (minor) conformer (we note that the measured values for the NOEs are heavily skewed towards the major conformer, which I showed was consistent with the conformation observed in the crystal). The α6 helix in several NR5A orthologs is largely solvent-accessible and adopts a strikingly different conformation than the one described for *Drosophila* Ftz-F1[71,73,88]. In crystal structures featuring NR5A receptors with an empty LBP, such as the apo human LRH1 LBD and the mouse LRH1 LBD, the α6 helix is characterized respectively by a lack of detectable density or by unusually high temperature factors, implying a dynamic helix[72,87]. I propose that the minor conformer for Ftz-F1 LBD in solution adopts an analogous conformation (Fig. 2.9). In support of this model, the α6 helix and the segment leading to α7 has the sequence LMSLGLLGVP that features several residues that are poor helix formers, likely destabilizing the helix upon removal from the pocket, thereby contributing to the enhanced flexibility noted for this region in my studies.
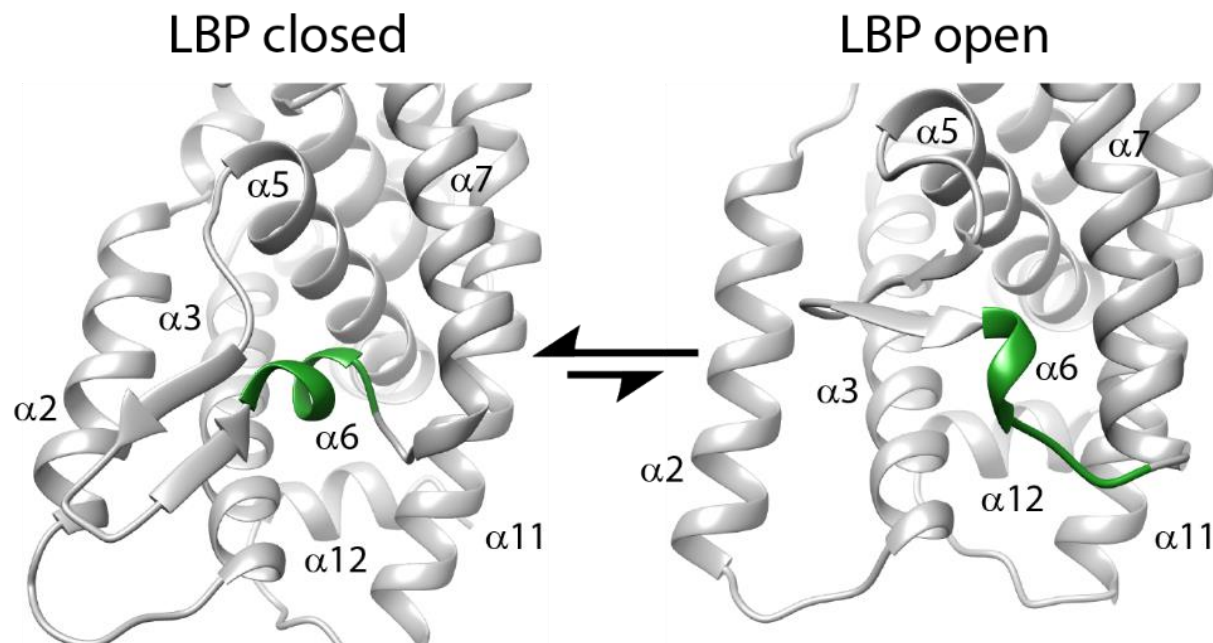
Figure 2.9 Structural model for exchange between two states for the Ftz-F1 LBD in complex with the Ftz LxxLL motif.
Because no high-resolution structure for the open complex of the Ftz-F1 LBD has been described, a model for the open complex was generated via homology modeling using the structure of the mouse LRH1 ortholog (PDB entry 1YOK) as the template in the SWISS-MODEL Web server. Model generated by Ishwar Radhakrishnan.

To decisively test the aforementioned model, I conducted a well-established assay to assess whether Ftz-F1 LBD could bind to phospholipids. If the α6 helix indeed spent time outside the LBP, this should allow ligands (such as phospholipids attributed to its orthologs in mouse and human) to enter and occupy the pocket. Consistent with my model, Ftz-F1 LBD bound to the same phospholipids previously shown to bind and function as *bona fide* ligands for its orthologs, such as mouse SF1 LBD (Fig 2.10)[89,90]. Given that the LBP has all the molecular determinants for binding phospholipids, I conclude that they most likely bind to the pocket upon the extrusion of α6 helix.
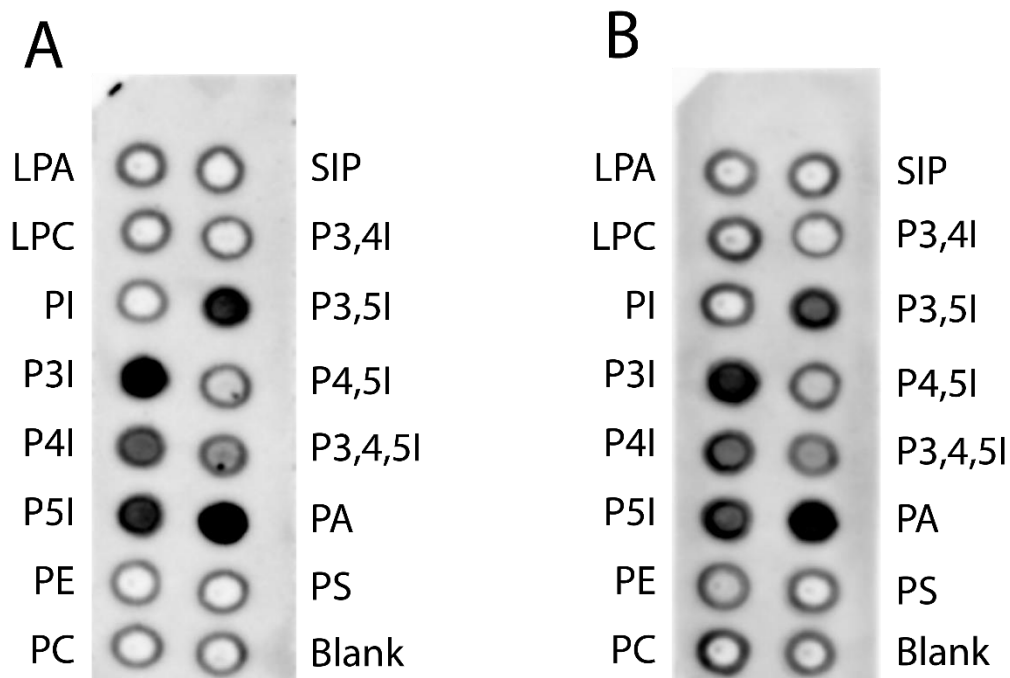
Figure 2.10 Analysis of phospholipid binding by NR5A members.
Purified His6-tagged proteins including (**A**) *Drosophila* Ftz-F1 LBD and (**B**) mouse SF1 LBD were incubated with a panel of phospholipids that were immobilized on to a surface. Protein binding was assessed using an antibody directed against the His6-tag.

Assuming the α6 helix in Ftz-F1 LBD samples two conformations as described above, a key question relates to how it might switch between the two states. Although additional studies are required to answer this question, I was intrigued by the high concentration of residues near/at the C-terminal region of α2, the α5- α1 loop and the β-hairpin that show evidence of exchange. I hypothesize that these regions undergo correlated motions facilitating the switch from the major conformation to the minor conformation accompanied by the eviction of the helix from the LBP. Indeed, molecular dynamics simulations of liganded nuclear receptors have suggested that analogous regions of these receptors undergo localized changes in conformation for ligand entry and exit[91,92]. Additionally, HDX data for the human LRH1 receptor also suggested conformational fluctuations on the seconds (or faster) timescale in these regions[93].

Nuclear receptors have evoked intense interest since their discovery, as these proteins bind to an extraordinarily diverse range of small molecules to control many aspects of development and cellular physiology and have inspired many successful drug discovery programs. A key question in NR biology relates to their molecular evolution and functional diversification. In the prevailing view, many extant nuclear receptors are thought to have descended from an ancestral liganded receptor having undergone subtle changes to the LBP to accommodate diverse ligands while others evolved ligand-independence by acquiring mutations that stabilized the active conformation involving the C-terminal helix[19]. Members of the NR5A subfamily exhibit the full range of functional diversification described for the NR superfamily with most higher vertebrate NR5A receptors showing ligand dependence whereas arthropods are thought to be ligand-independent[72–74]. Interestingly, some of the greatest variability within the subfamily at the sequence level lies in the vicinity of the α6 helix. The unexpected presence and extent of dynamics described for the

Drosophila receptor by my studies thus alludes to potential variation in the canonical mechanism of ligand binding and activation.

CHAPTER 3: A Novel Ligand-Binding Domain-Independent Mechanism of Coactivator

Recruitment by the Nurr1 Nuclear Receptor

Introduction

Nuclear receptors (NRs) constitute one of the largest families of transcription factors that are unique to metazoans. NRs regulate a wide variety of physiological processes including growth, development, and cellular homeostasis[16,21,94–96]. To facilitate these different regulatory roles, NRs evolved to interact with a variety of coactivator proteins, which then facilitate the assembly of chromatin-modifying complexes and transcriptional machinery to specific genes[97–99]. However, in some cases, the specific domains or protein segments, in both NRs and coactivators, that facilitate coactivator recruitment are not know. Discovering these regions and charactering their interactions is an important part of understanding how NRs regulate transcription[21,53].

NRs share a conserved architecture that consists of three domains connected by flexible linkers. The central DNA-binding domain (DBD) targets the NR to a specific region of the genome while the N-terminal activation function 1 (AF1) domain and the C-terminal ligand-binding domain (LBD), which also functions as a transactivation domain, are involved in the recruitment of coactivator [100] (Fig. 1.1). Among the two activation domains, the structure and function LBD is the better understood. It is highly conserved and numerous structural studies have shown that restructuring of the C-terminal helix (designated activation function 2 or AF2), which often occurs in response to ligand binding, allows for recruitment of coactivators via a direct interaction between the LBD and a LxxLL motif[15,31,100,101]. The AF1 domain and its interactions with coactivators, on the other hand, is less well characterized. Cell-based assays have shown that the AF1 domains of NRs are important for full target gene activation; in some cases, the activation is predominantly or exclusively mediated by these domains[102–104]. Progress towards a molecular-

level understanding of the mechanism by which AF1 domains engage with coactivators has been limited likely owing to the domain being unstructured according to NMR studies[47,105] and also as evidenced by the lack of any detectable electron density in crystal structures of full-length NRs bound to DNA[50,51]. CryoEM characterizations of full-length receptors in complex with coactivators and DNA also feature blobs of density for this region, precluding higher-resolution insights into AF1-coactivator interactions[52,106].

Among the various coactivators recruited by NRs, members of the steroid receptor coactivator (SRC) family were the first discovered and the most extensively studied[61]. As implied by the name, they were initially shown to mediate transcriptional activation by the steroid receptor family of NRs and have since been found to be recruited by other NR families[57]. Structurally, these proteins are broadly divided into ordered and disordered regions. At their N-termini they contain a basic Helix-Loop-Helix (bHLH) domain and two Per-ARNT-Sims (PAS) domains, known as PAS-A and PAS-B. The rest of the protein following this region is largely disordered and contains a number of short motifs involved in protein-protein interactions[64]. Like other coactivators, SRC proteins often bind directly to NRs and then recruit chromatin-modifying complexes as well as the transcription machinery[3,56]. Normally, NR-SRC recruitment occurs via an interaction between several LxxLL motifs found in the SRC C-terminal region and a NR LBD[34]. However, there is evidence suggesting that NR AF1 domains also play important roles in recruiting the SRC coactivators[107–109].

The NR4A subfamily of NRs represents an excellent system in which to look for potential interactions between NR AF1 domains and the SRC coactivators. This subfamily consists of 3 paralogs (Nur77, Nurr1, and NOR1) that are expressed in a tissue-specific manner and play a variety of regulatory roles in human physiology[110,111]. In particular, Nurr1 is significant as a *bona*

fide target for Parkinson's disease that is characterized by impaired motor function due to progressive degeneration of dopaminergic neurons and reduced dopamine levels[112,113]. Importantly, their AF1 domains are thought to play the primary role in recruiting the SRC family of coactivators[102,103,114]. Crystal structures of Nurr1 and Nur77 LBDs revealed that the surface where LxxLL motifs normally bind is interrupted by hydrophilic residues and a reversal of a canonical lysine/glutamate charge clamp, precluding the canonical LxxLL-LBD interaction[41,42]. Cell-based assays have also shown that the AF1 domain is the primary determinant of the NR4A members ability to activate transcription and that the SRC coactivators play a role in mediating this activity[103,114,115].

Given these data, I wanted to test whether the NR4A AF1 domains could directly recruit SRC proteins and, if so, structurally and functionally characterize the interaction. I hypothesized that they would bind one of the structured domains at the SRC N-termini, since AF1 domains are often disordered and disordered-ordered interactions are a common theme among transcription factors and coactivators[116,117]. Furthermore, several studies have shown a short peptide in the transcription factor STAT6 can directly bind to the PAS-B domain of SRC1, showing there is precedent for transcription factors to recruit SRC proteins by an interaction with one of their N-terminal domains[60,63,118]. To test this hypothesis, I looked to see if the AF1 domain of the NR4A receptor Nurr1 could bind to one of the N-terminal domains of the coactivator SRC1. Here I show that a short segment of the Nurr1 AF1 domain can directly interact with a conserved surface on the SRC1 PAS-B domain, suggesting a novel mechanism for recruitment of SRC coactivators by NR AF1 domains.

Experimental Methods

Construct design and creation

For study of the Nurr1 AF1, the DNA coding sequence for residues 1-165 of human Nurr1 were cloned into a pMCSG7 vector using ligation-independent cloning[75]. Additionally, two shortened peptide segments of the human Nurr1 AF1 were designed that consisted of residues 28-67 and 30-51. For study of the SRC1 PAS-B and NTD domains, the DNA coding sequence for residues 260-369 (PAS-B) and 1-369 (NTD) were cloned into pMCSG7 and pMCSG24 vectors.

Two fusion constructs were designed of the SRC1 PAS-B domain and shortened segments of the Nurr1 AF1 domain. Initially, a DNA coding sequence for the SRC1 residues 260-385, followed directly by the Nurr1 residues 28-67 was synthesized and was inserted into a pMCSG7 vector. Unfortunately, solubility issues and peak crowding made this construct difficult to work with. Subsequently, the Nurr1 residue 52 in the fusion construct was mutated to a stop codon to create a DNA sequence that coded for SRC1 residues 260-385, followed directly by the Nurr1 residues 28-51. In these fusion constructs SRC1 residues 370-385 were chosen to act as the linker between the PAS-B domain and the Nurr1 AF1 segments because previous X-ray crystallography and NMR studies as well as disorder prediction software suggested this region is disordered in solution[60].

Single- and double-site mutants of the Nurr1 AF1 domain (1-165) were produced by mutation of the vectors encoding the wild-type protein using the QuikChange site-directed mutagenesis protocol (Agilent Genomics). A Nurr1 AF1 mutant, in which residues 28-67 were deleted was also created.

Protein and peptide production and purification

Prior to use, all DNA constructs were sequenced to verify identity. Constructs were transformed into *E. coli* BL21 (DE3) cells which were grown at 37 °C until they reached an $A_{600 \, nm}$ of 0.6, at which time they were moved to 16 °C, induced with 1 mM IPTG, and proteins were expressed overnight. The next day cells were pelleted and were stored at -80 °C until use.

Nurr1 AF1 wild-type and mutant, SRC1 PAS-B wild-type, and PAS-B-Nurr1 fusion proteins, expressed using pMCSG7 vectors, were all purified in the following manner. Note that Nurr1 AF1 was purified under denaturing conditions. Cells pellets from 0.5 l or 1 l cultures were resuspended in 40ml of lysis buffer (300 mM NaCl, 50 mM Tris, 1 mM TCEP, pH 7.2) and sonicated for 25 minutes. The resulting cell lysate was then clarified by centrifugation at 12,000 rpm for 20 min and the soluble supernatant was put on to a column containing 1-2 ml of Ni^{2+}-NTA resin (Qiagen) and incubated at 4 °C for one hour. The supernatant was then allowed to flow through the column and two 10ml washes were performed with both a high-salt wash buffer and a wash buffer. The protein was then eluted in four 4 ml fractions with an elution buffer and the purity of the fractions was analyzed by running SDS-PAGE gels. Clean fractions were combined and were put through two rounds of dialysis at room temperature overnight to remove imidazole and urea in the case of Nurr1 AF1 proteins. The His_6-tag for PAS-B and PAS-B-Nurr1 fusion proteins was cleaved during dialysis and the His_6-tag for Nurr1 AF1 proteins was cleaved after dialysis, in both cases using 1:25 (w/w) TEV protease.

Proteins were then subjected to different forms of chromatography to enhance purity. Nurr1 AF1 proteins and PAS-B wild-type proteins were immediately put through HPLC following cleavage and protein-containing fractions were lyophilized and stored for future use. When ready for use, lyophilized Nurr1 AF1 wild-type and mutant proteins and SRC1 PAS-B wild-type proteins

were resuspended in the desired buffer containing also 6M guanidine hydrochloride and then put through two rounds of dialysis overnight at 4 °C to remove the guanidine hydrochloride. In the case of PAS-B-Nurr1 fusion proteins the Ni^{2+}-NTA resin used in the purification was cleaned using 6M Gdn-HCl, 500mM imidazole, and 0.1% Triton-X and regenerated with Nickel Sulfate and the proteins were put back through the column to remove any uncleaved protein. The fully cleaved protein was then further purified by way of size-exclusion chromatography using a Superdex75 column and SEC running buffer (150mM NaCl, 25mM Tris, 1mM TCEP, pH7.5).

SRC1 PAS-B and SRC1 NTD wild-type proteins and GST, expressed using pMCSG24 vectors, were all purified in the following manner for use in pulldown assays. Buffers used for the different purifications are listed in the table below. Cell pellets from 15ml cultures were resuspended in 15 ml of lysis buffer and lysed by sonication for 25 minutes. The resulting cell lysate was then clarified by centrifugation at 12,000 rpm for 20 min and 10ml of the soluble supernatant was incubated with 350 µl of glutathione-sepharose resin at 4 °C for 1 hour. The supernatant was then allowed to flow through the column and one 5ml wash was performed with a high-salt wash buffer and two 5ml washes were performed with wash buffer. The resin was then resuspended in 1ml of wash buffer containing protease inhibitors (Leupeptin 10 µg/ml, Pepstatin A 1 µg/ml, PMSF 100 µM) and stored at 4 °C. The purity of the protein on the resin was analyzed by running SDS-PAGE gels.

Synthesized peptides of the Nurr1 AF1 28-67 and 30-51 protein sequences were purchased in crude form and were purified by HPLC using a C18 column and an aqueous buffer containing 0.1% trifluoroacetic acid and a gradient of 5% to 95% acetonitrile. The peptide-containing fractions were lyophilized and stored for later use.

GST pulldown assay

All steps in this assay were performed at 4 °C in the cold room. In preparation, Nurr1 AF1 proteins were dialyzed into GST pulldown buffer (100-150 mM NaCl, 50 mM Tris, 1 mM TCEP, pH 7.2), put through a 0.2um filter, and their concentration adjusted to 10-30 μM. Glutathione-sepharose resin containing purified proteins was washed one time with pulldown buffer for equilibration. All washes were performed by resuspending resin in buffer, spinning down resin for one minute at 14,000 rpm, and carefully pipetting off supernatant. For each Nurr1 AF1 construct, 1 ml of AF1 protein at 10-30 μM was incubated with 40 μl (bed volume) of glutathione-sepharose resin containing either SRC1 PAS-B, SRC1 NTD, or GST for 1 hour. The resin was then washed once with pulldown buffer containing 0.2% NP-40 and two times with pulldown buffer. The resin was then resuspended in Tris-Tricine loading buffer, boiled for 10 minutes, and run on an SDS-PAGE gel, along with the input proteins. Band intensity for gels were quantified using the AzureSpot software (Azure Biosystems) and normalized based on the relative intensities of input proteins.

NMR sample preparation

All ^{15}N- and/or ^{13}C-labeled proteins were produced similarly to unlabeled protein except that they were grown in media containing ^{15}N ammonium sulfate and/or ^{13}C glucose as their sole nitrogen/carbon sources. Labeled proteins were dialyzed into NMR buffer (50 mM NaCl, 20 mM NaPO4, 1 mM DTT, pH 6.8) and concentrated to between 0.1mM-1mM. Final samples were then prepared by adding in EDTA (0.5 mM) and D$_2$O (5%) and adjusting the pH to 6.8. For titrations, peptides were resuspended in NMR buffer and the sample pH was adjusted after every titration point.

NMR data collection and analysis

NMR data were acquired a Bruker Neo 600 MHz spectrometer equipped with a QCI-F cryoprobe at temperatures ranging from 25-30 °C. NMR data were processed using Topspin and NMRpipe and analyzed using NMRFAM-Sparky[78]. 2D ^{1}H-^{15}N HSQC spectra were collected for peptide titrations. Chemical shift perturbations were calculated in R using the equation

$$CSP = \left(1/2\left[\Delta\delta_H^2 + \left(0.04(\Delta\delta_N^2)\right)\right]\right)^{1/2}$$

in which δ_H and δ_N are the proton and nitrogen chemical shifts, respectively. The CSPs for each residue were then mapped onto structures for analysis using Chimera. For backbone assignments, 2D ^{1}H-^{15}N HSQC, 3D HNCA, HNCACB, CBCA(CO)NH, and HNCO experiments were collected for both SRC1 PAS-B and the PAS-B-Nurr1 28-51 fusion proteins. Backbone resonances were assigned using a mixture of automated assignments from the I-PINE web server and assignments made by hand[119]. Of note, although the NMR structure of PAS-B had already been solved bound to STAT6, there were no backbone assignments for the apo protein, necessitating *de novo* sequence-specific assignments.

Secondary structure predictions for the SRC1 PAS-B-Nurr1 28-51 fusion protein were generated by running the NMRFAM GetSBY applet in SPARKY[78]. Assignments from 2D ^{1}H-^{15}N HSQC, 3D HNCA, HNCACB, CBCA(CO)NH, and HNCO experiments were used as inputs. To measure the affinity of PAS-B and the Nurr1 30-51 peptide, CSPs were first measured for the most perturbed peaks. These were then individually plotted as a function of concentration and fit to the following equation

$$\Delta\delta_{obs} = \frac{\Delta\delta_{max}\left\{([P]_t + [L]_t + K_d) - [([P]_t + [L]_t + K_d)^2 - 4[P]_t[L]_t]^{\frac{1}{2}}\right\}}{2[P]_t}$$

in which $[P]_t$ is the total concentration of protein, $[L]_t$ is the total concentration of ligand, $\Delta\delta_{obs}$ is difference in the chemical shift from the free to the observed state, and $\Delta\delta_{max}$ is the maximum observed shift upon saturation[120]. Data was fit, analyzed, and plotted using a python script written by myself.

Size exclusion chromatography

To assess relative affinities, Nurr1 AF1 wild-type and Nurr1 AF1 Δ28-67 in complex with SRC1 PAS-B wild-type as well as apo SRC1 PAS-B were subjected to size-exclusion chromatography. For both complexes and for apo SRC1 PAS-B 200 μl of proteins at a concentration of 100 μM were injected onto a Superdex75 column. Samples were initially in a buffer containing 100 mM NaCl, 50 mM NaPO4, 1 mM BME, at pH 7.5. The running buffer contained 150 mM NaCl, 25 mM MOPS, and was at pH 7.5. Chromatograms were visualized using an R script written by myself.

Isothermal Titration Calorimetry

In preparation for ITC, proteins and peptides were either resuspended or dialyzed into ITC buffer (100 mM NaCl, 50 mM NaPO4, 1 mM BME, pH 7.5). Concentrations of proteins was determined by measurements of absorbance at 280 nm at several different dilution factors. Experiments were performed on a MicroCal (GE Healthcare) ITC200 isothermal titration calorimeter and data was analyzed using MicroCal ITC Origin 7.0. SRC1 PAS-B was added to the well at concentrations between 40-50 μM and Nurr1 AF1 wild-type and 28-67 peptides were added to the syringe at concentrations ranging from 0.8-2 mM. Baseline heats of dilution was subtracted from the resulting thermograms, which were then integrated, and blanks and obvious outliers were

subtracted. The binding isotherms were fit using a one site model. Due to the micromolar affinity of the complex and solubility limitations of the individual proteins C values were relatively low and thus reliable N values were not able to be obtained[121]. Experiments were done in triplicate to obtain mean K_d values and standard errors.

Sequence alignments

A list of species was assembled that included diverse representatives from the vertebrate subphylum. Full length human Nurr1 and human SRC1 sequences were then used as queries in BLAST searches that were restricted to these species. The resulting orthologous sequences were then aligned using the Clustal-Ω web server and analyzed using the Jalview software[122,123].

Structure predictions

Sequences for PAS-B (260-369) and the Nurr1 AF1 minimal binding region (30-51) were entered into the AlphaFold-Multimer (v2.1.0) software, which was run using the Google Colab python notebook created by the DeepMind team[124].

Mass Spectrometry

To confirm their identity and purity prior to use, purified peptides were resuspended in a 50:50 mix of acetonitrile and water, mixed 1:1 with α-Cyano-4-hydroxycinnamic acid (CHCA) at saturating concentration, and analyzed by MALDI mass spectrometry on a Bruker AutoFlex III Smartbeam.

To confirm the identity of wild-type and mutant proteins after cloning and purification, proteins were subjected to LC-ESI-MS using an Agilent 6545 QTOF instrument. Runs were

performed by first injecting 7 μl of protein samples in native buffers onto a guard column using a running buffer of consisting of H_2O and 0.1% formic acid. Proteins were then eluted with a running buffer of 100% acetonitrile for detection. The resulting spectra were deconvoluted and analyzed using Agilent's MassHunter Bioconfirm software.

Results

The Nurr1 AF1 domain directly interacts with the SRC1 PAS-B domain

To investigate if the AF1 domain of an NR4A receptor could directly bind to one of the structured SRC domains, pulldowns were performed with the individual domains[60,63]. Constructs were designed of the Nurr1 AF1 domain (AF1), the SRC1 PAS-B domain (PAS-B), and the entire SRC1 N-terminal region (NTR) that contains the bHLH, PAS-A, and PAS-B domains and the resulting proteins were purified from *E. coli*. Pulldowns were performed by binding GST-tagged PAS-B and SRC1 NTR to GST resin, incubating the resin with AF1, followed by washes and assessment by SDS-PAGE to observe protein still bound to the resin. These pulldowns revealed that AF1 binds the NTR as well as PAS-B alone (Fig. 3.1). Interestingly AF1 appeared to bind PAS-B and NTR with comparable efficiency. This, along with the fact that the STAT6 transcription factor also bound PAS-B suggests that an interaction between the Nurr1 AF1 domain and SRC1 N-terminal region is mediated solely by the PAS-B domain[63].
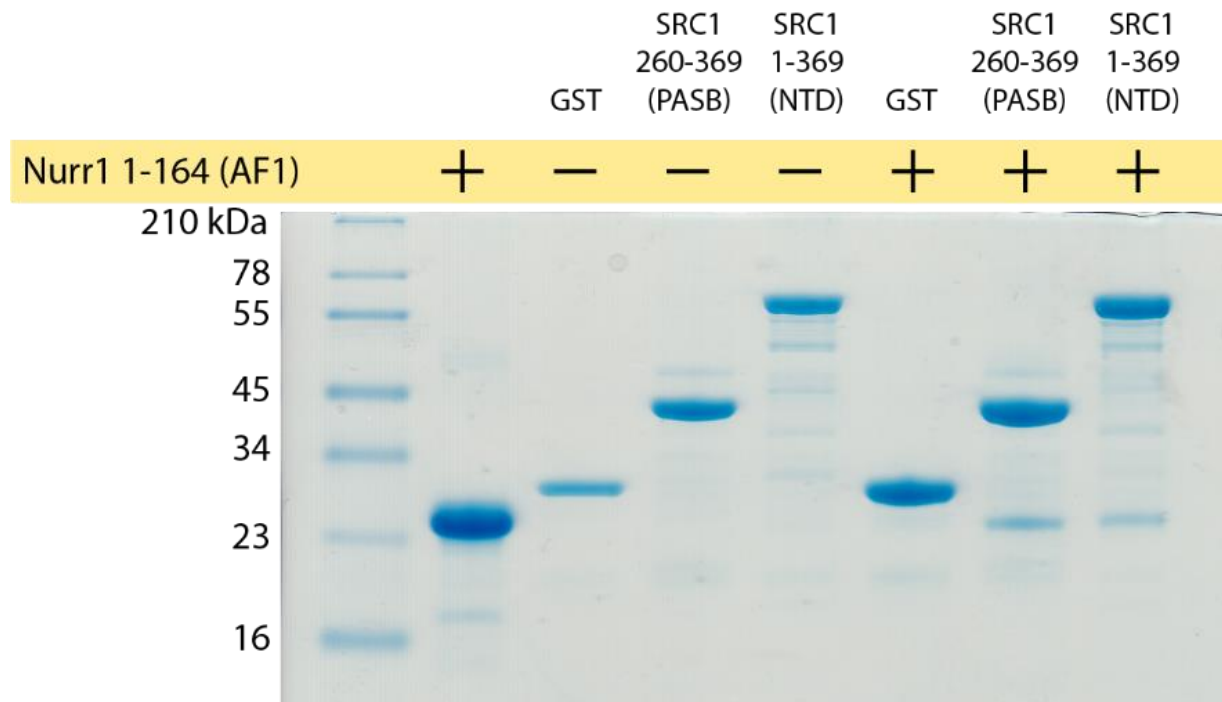
Figure 3.1 Pulldown of Nurr1 AF1 by SRC1 PAS-B confirms a direct interaction.
GST-pulldown assay using purified Nurr1 AF1 and GST-tagged SRC1 PAS-B or SRC1 NTD.
Retention of Nurr1 AF1 after washes shows the AF1 domain directly interacts with the PAS-B
domain and NTD. GST was used as a negative control.

NMR titrations were performed next with ^{15}N-labeled AF1 and unlabeled PAS-B to further assess the interaction. To facilitate analysis of the titration spectra, the backbone resonances of the apo-AF1 were first assigned using standard double- and triple-resonance experiments on an ^{15}N, ^{13}C-labeled sample. An additional benefit of the assignment procedure is the ability to predict protein secondary structure based on chemical shifts. These predictions (Fig. 3.2), along with the narrow range of proton chemical shifts in the apo HSQC spectrum (Fig. 3.3A), show that AF1 is disordered in solution. Following assignment, unlabeled PAS-B was titrated in with ^{15}N AF1 and multiple peaks were perturbed, further confirming that AF1 directly interacts with PAS-B (Fig. 3.3A).
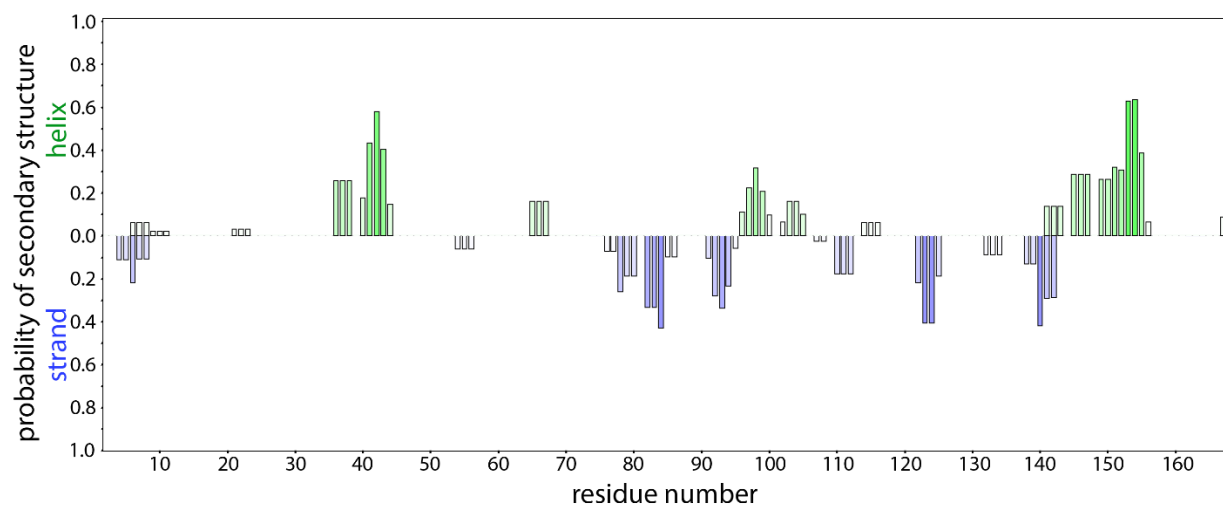
Figure 3.2 Nurr1 AF1 is unstructured.

A bar plot showing backbone chemical shift-based predictions of secondary structure, specifically, α-helices and β-strands by the GetSBY module in the NMRFAM-Sparky suite for apo-Nurr1 AF1[78].
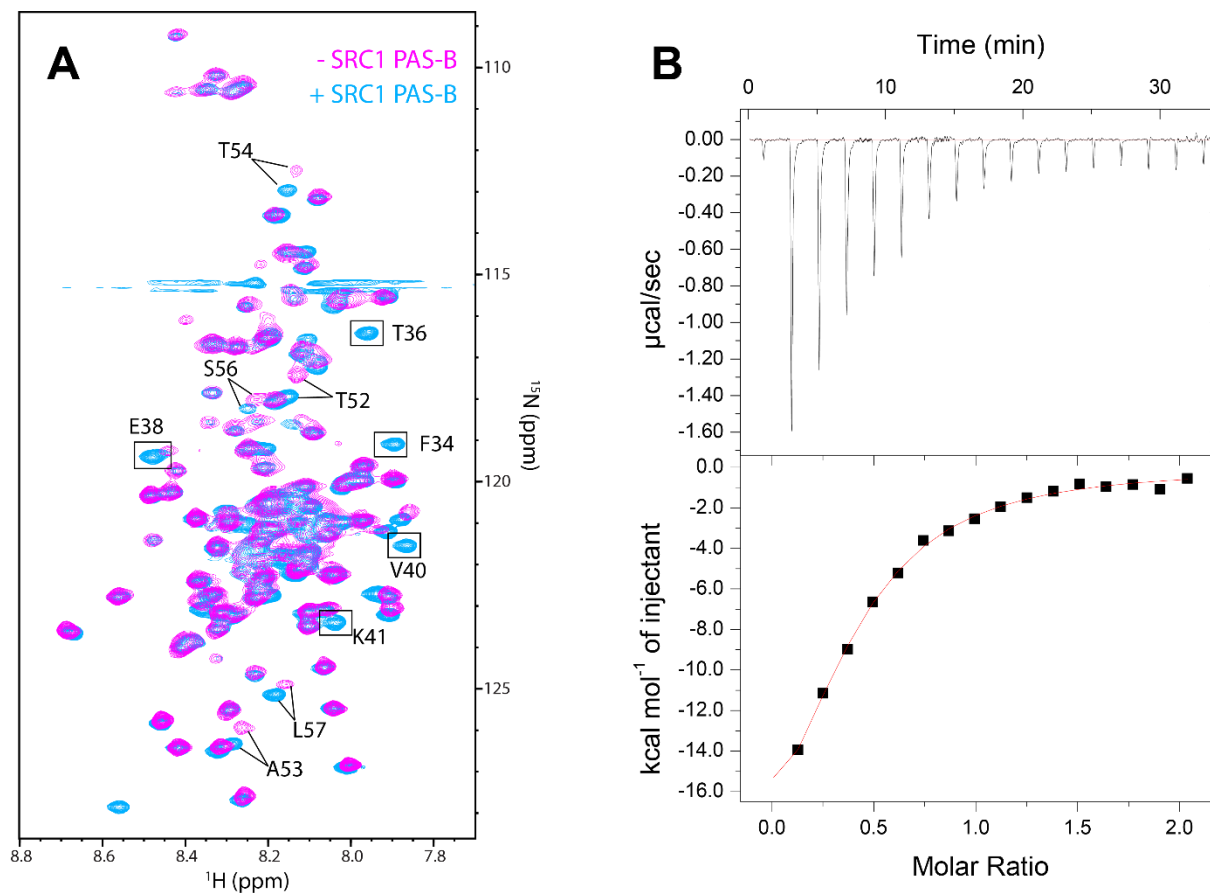
Figure 3.3 NMR and ITC titrations confirm binding and reveal affinity of interaction of AF1 with PAS-B.
(A) ^{1}H-^{15}N HSQC spectrum of ^{15}N-labeled full-length Nurr1 AF1 in the absence (cyan) and presence (magenta) of equimolar, unlabeled SRC1 PAS-B. Select residues from the 28-67 AF1 segment that shift (arrows) or broaden (boxes) upon PAS-B addition are highlighted. (B) Representative ITC results from a titration of full-length AF1 into PAS-B. Runs were performed in triplicate to calculate K_d values. Top panel shows a raw thermogram after baseline correction. Bottom panel shows fitting of the resulting integrated isotherm.

The affinity of the interaction was then assessed by performing isothermal titration calorimetry (ITC). For these experiments AF1 was titrated into PAS-B and the resulting isotherm was fit to determine the binding affinity. Multiple replicate ITC experiments produced an average K_d of 8.1 (±1.2) μM, revealing the interaction to be of moderate affinity, which confirms the NMR titration data that suggested an interaction in fast exchange (Fig. 3.3B). It should be noted, an affinity of this magnitude is in line with interactions seen for a number of other transcription factor-coactivator pairs which are characterized by multivalent interactions that are of moderate or weak affinity[125–128].

Nurr1 residues 30-51 are the primary determinant for binding of PAS-B

Next, I sought to determine the specific AF1 residues that were involved in the interaction with PAS-B. To do this, the movement or broadening of peaks in the ^{15}N AF1 HSQC spectra upon addition of PAS-B was assessed and plotted by residue. This revealed a region from approximately residue 28 to 67 that contained most peak perturbations, suggesting that this region was critical for PAS-B binding (Fig. 3.4A).
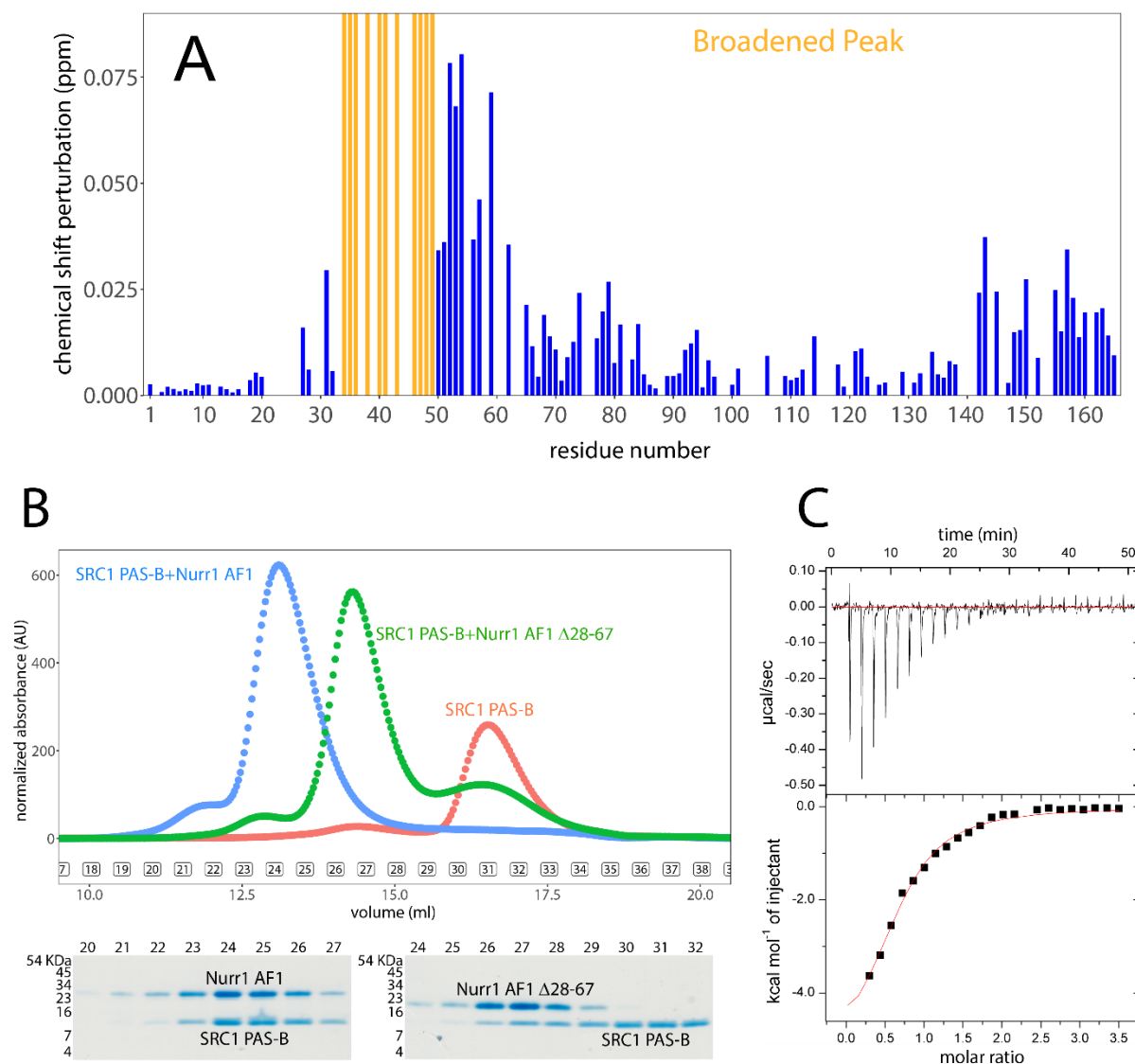
Figure 3.4 Identification of AF1 residues 28-67 as critical for binding PAS-B.
(A) Plot of CSPs/broadening per residue of AF1 based on spectra from the titration of PAS-B into ^{15}N-labeled AF1 (see figure 3.2A). Bars in blue represent the calculated CSP value and bars in orange indicate peak broadening. (B) Overlaid chromatograms from SEC runs of apo PAS-B, PAS-B + wild-type AF1, and PAS-B + AF1 deletion mutant (Δ28-67). Results of SDS-PAGE gels of eluted fractions are shown in bottom panels. (C) Representative ITC results from a titration of the 28-67 AF1 peptide into PAS-B. Runs were performed in triplicate to calculate K_d values. Top panel shows a raw thermogram after baseline correction. Bottom panel shows fitting of the resulting integrated isotherm.

To test this, a mutant of AF1 was made that lacked residues 28-67 (AF1 Δ28-67) and the mutant and wild-type AF1 domains were each put through size-exclusion chromatography in complex with PAS-B. Chromatograms and SDS-PAGE gels of fractions from the runs showed wild-type AF1 co-eluted with PAS-B, as expected, but that deletion of residues 28-67 largely disrupted co-elution, supporting the idea that these residues are largely responsible for the interaction (Fig. 3.4B). To more quantitatively assess the binding affinity, ITC experiments were then performed with PAS-B and Nurr1 28-67. The experiments were performed in a manner similar to those with full length AF1 and produced an average K_d of 10.9 ($\pm$1.8) μM (Fig. 3.4C). This affinity is largely in agreement with that of the full-length AF1, confirming that this region contains the minimal segment needed for PAS-B binding.

To facilitate structural studies of bound AF1, I sought to determine if a shorter segment existed that contained all residues that are critical for PAS-B binding. In the AF1 CSP map residues 30-51 contained most of the peaks that broadened, suggesting that this region undergoes conformational exchange and is the primary region making direct contact with PAS-B (Fig 3.4A). Thus, a peptide of AF1 residues 30-51 was synthesized and NMR titrations were performed with ^{15}N-PAS-B to assess if shortening the peptide altered its ability to bind. A multipoint NMR titration of the 30-51 peptide with ^{15}N-PAS-B was performed to estimate the affinity and determine if it was similar to the longer AF1 constructs. This revealed an apparent average K_d of 10.5 ($\pm$3.6) μM (Fig. 3.5). While this affinity is at the lower limit of what can be accurately measured by NMR, the variance in fitted values for each residue is low and the average value closely matches the K_d measured by ITC for the full length AF1, supporting the idea that shortening the peptide has not dramatically altered its affinity and that the 30-51 segment is the primary determinant of PAS-B binding.
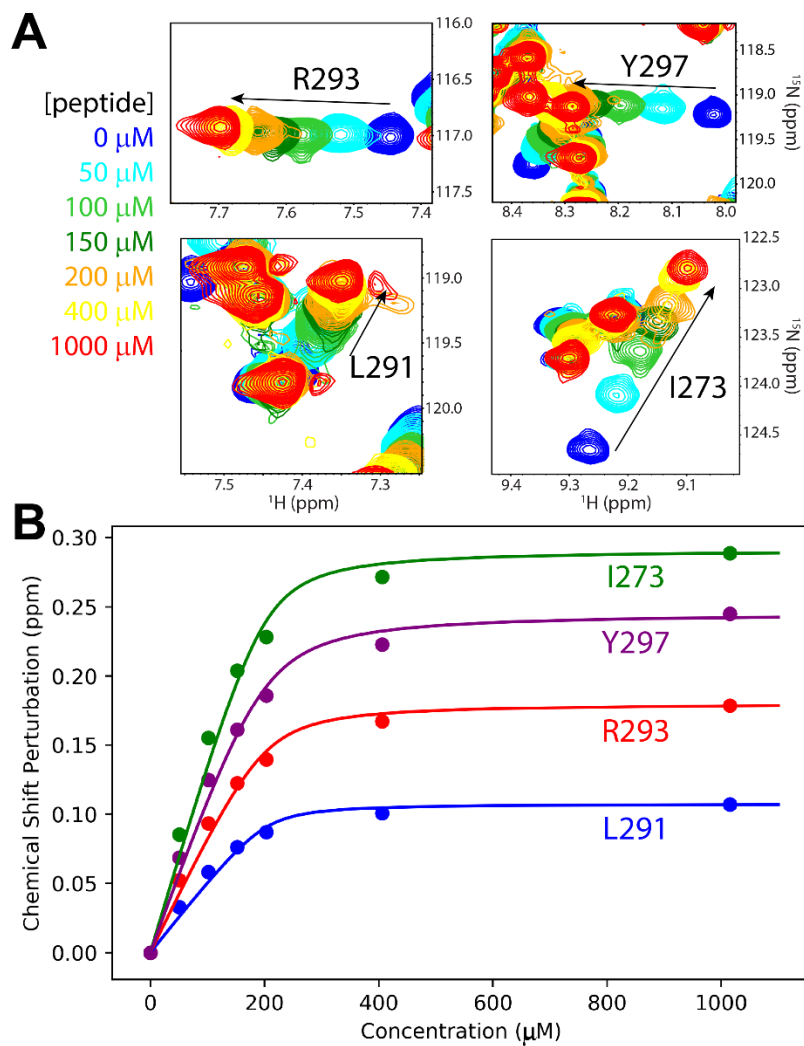
Figure 3.5 Assessing the affinity of the 30-51 AF1 peptide for PAS-B.
(A) Overlaid spectra from the titration that show peaks for the indicated residue shifting as the concentration of the 30-51 AF1 peptide increases (increasing from blue to red). (B) CSPs from peaks of select residues (colored) plotted as a function of increasing concentration of the 30-51 AF1 peptide during its titration into ^{15}N-labeled PAS-B. Solid lines represent fits that were used to determine K_d values.

Nurr1 AF1 binds to a conserved hydrophobic cleft on PAS-B

To determine where Nurr1 AF1 binds on PAS-B, NMR titrations were performed with ^{15}N-labeled PAS-B and both the 28-67 and 30-51 AF1 peptides. As expected, titrations for both peptides exhibited peak shifts, confirming binding in fast exchange (Fig 3.6A,C). Since the structure of SRC1 PAS-B bound to a peptide from the transcription factor STAT6 had already been solved, the CSPs of peaks corresponding to each residue could be mapped onto the structure of PAS-B[60,63]. Titrations for both peptides resulted in similar CSP maps, supporting the idea that the 30-51 AF1 segment contains the residues making primary contact with PAS-B (Fig 3.6 B,D). In the structure with PAS-B, the STAT6 peptide is seen wedged in a hydrophobic cleft created by residues from the second beta sheet (β2) and second alpha helix (α2) of PAS-B. Interestingly, mapping the CSPs from the AF1 titration onto the structure revealed that AF1 appears to bind PAS-B at the same location as STAT6, with peaks having the largest CSPs corresponding to residues in β2 and α2 (Fig 3.6 B,D). The fact that short peptides from two different transcription factors bind to the same surface of PAS-B points to the possibility that binding to this surface represents a more general mechanism for recruitment of SRC1 by transcription factors. To assess the biological significance of β2 and α2, sequences from SRC1 paralogs were collected and aligned (Fig 3.7). This showed that β2 and α2 are two of the most highly conserved segments in PAS-B, further highlighting the functional significance of this surface.
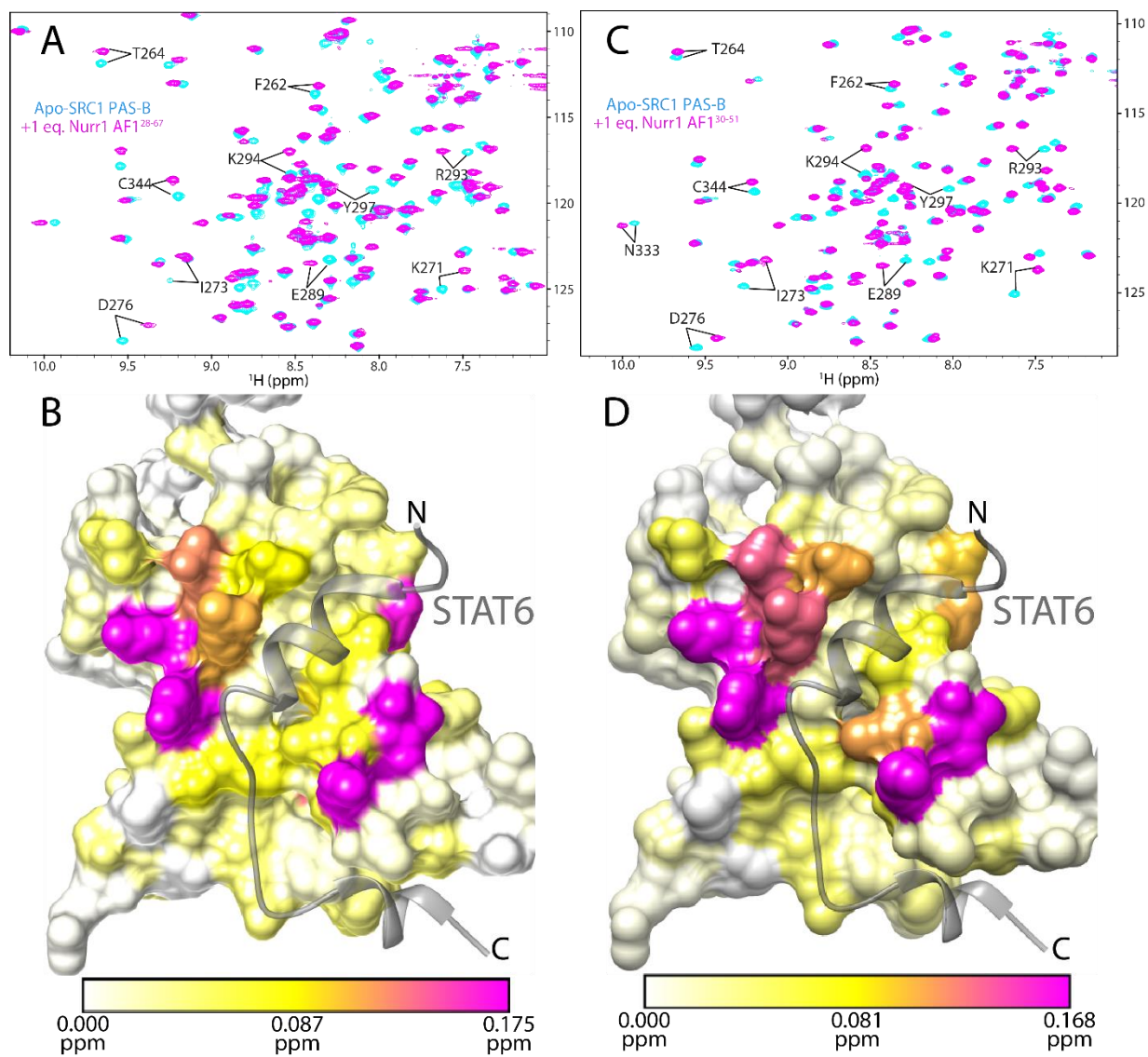
Figure 3.6 The 28-67 and 30-51 AF1 peptides bind similarly to a hydrophobic cleft in PAS-B.

Above are overlaid ^{1}H-^{15}N HSQC spectra of ^{15}N-labeled PAS-B in the absence (cyan) and presence (magenta) of unlabeled (A) 28-67 AF1 peptide and (C) 30-51 AF1 peptide. Shifting of highly perturbed peaks is highlighted with arrows. Below are CSPs from PAS-B HSQC spectra from titrations with the (B) 28-67 AF1 peptide and (D) 30-51 AF1 peptide mapped onto the structure (PDB entry 5NWM)[63] of PAS-B bound to STAT6 (shown in grey). The intermediate and high values in the color key represent $\langle \Delta\delta_{H,N} \rangle + 1\sigma$ (yellow) and $\langle \Delta_{\delta H,N} \rangle + 3\sigma$ (magenta), respectively. Colors are linearly interpolated to reflect the corresponding values.
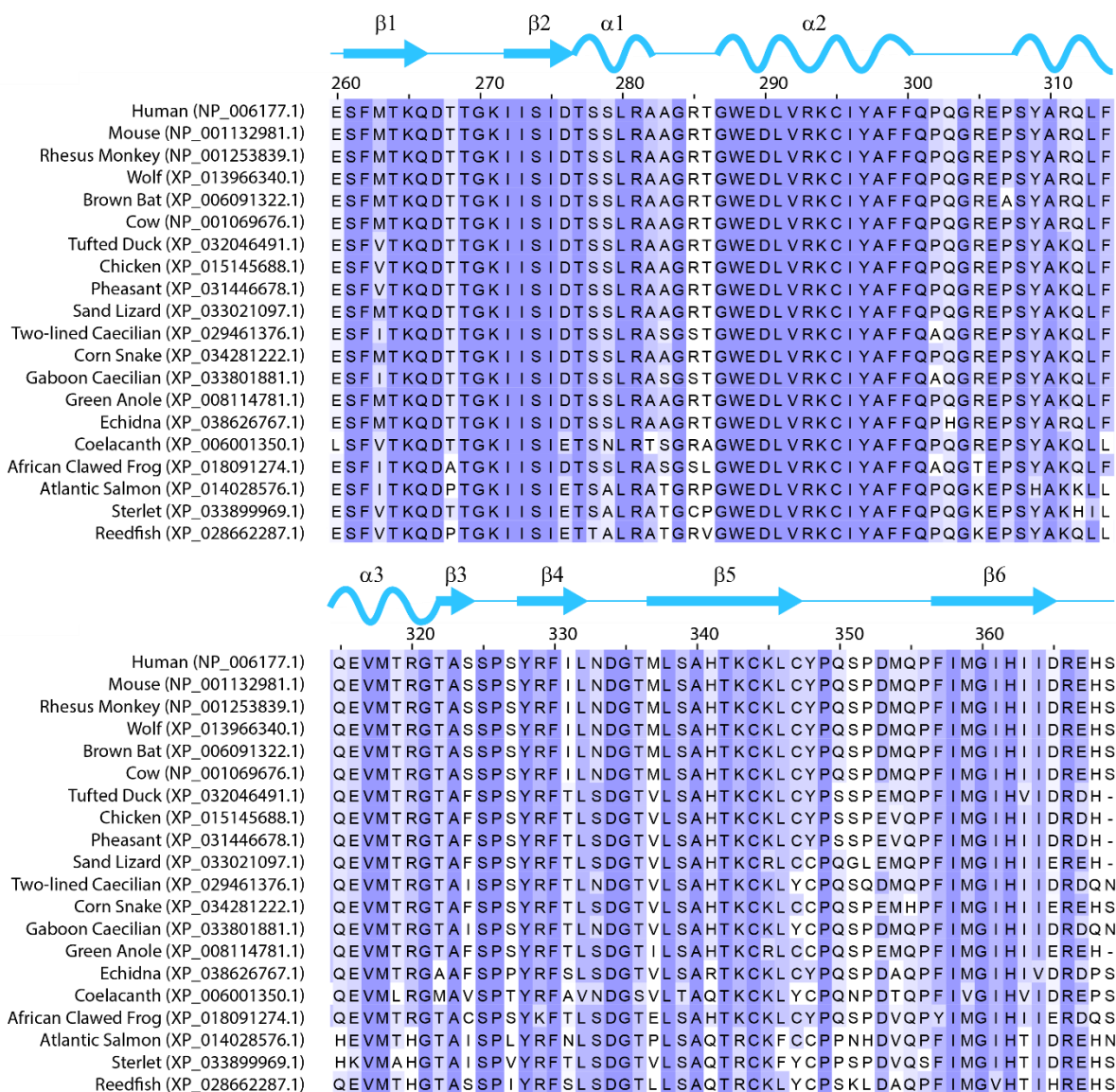
Figure 3.7 The Nurr1 AF1 binding site is highly conserved in SRC1 orthologs.

A Clustal Ω-guided multiple sequence alignment of the PAS-B domain (residues 260-369) of SRC1 orthologs. The locations of various secondary structural elements within PAS-B are shown in the form of a cartoon on top of the alignment. The Nurr1 AF1 binding site on PAS-B comprises primarily of residues immediately preceding the β2 strand, β2, and the α2 helix. Residues are colored by conservation (darker blue for strongly conserved residues) based on the BLOSUM62 scoring matrix in Jalview[123].

Experimental results support a predicted structure of the AF1 and PAS-B complex

To gain deeper insights into AF1 conformation, I performed structural studies. Initial attempts to crystallize PAS-B with both the 30-51 and 28-67 peptides proved unsuccessful. Therefore, a fusion construct of PAS-B and Nurr1 AF1 residues 28-51 was generated to facilitate structural analysis by NMR. Backbone resonances for residues in the linker and AF1 peptide regions were assigned and mapping of CSPs for PAS-B associated peaks onto the PAS-B structure revealed that the linker does not appear to affect the peptide binding footprint (Fig 3.8B). Overlaying strongly perturbed peaks from the apo spectrum with those from the 30-51 titration and the fusion protein showed that the peaks moved along the same vector, confirming that the fusion peptide has the same PAS-B binding mode as the free peptide and represents a more occupied state (Fig. 3.8A). Unfortunately, limited solubility of the fusion protein and broadening of the peptide resonances prevented collection of the intramolecular NOEs needed to solve a high-resolution NMR structure. However, assignment of the backbone resonances did allow for secondary structure predictions to be made for the fusion protein (Fig. 3.9A). This revealed that the PAS-B region contains the expected sequence of α-helices and β-sheets, that the linker region is unstructured as expected, and that, interestingly, a segment within the AF1 peptide likely forms a helix upon binding.
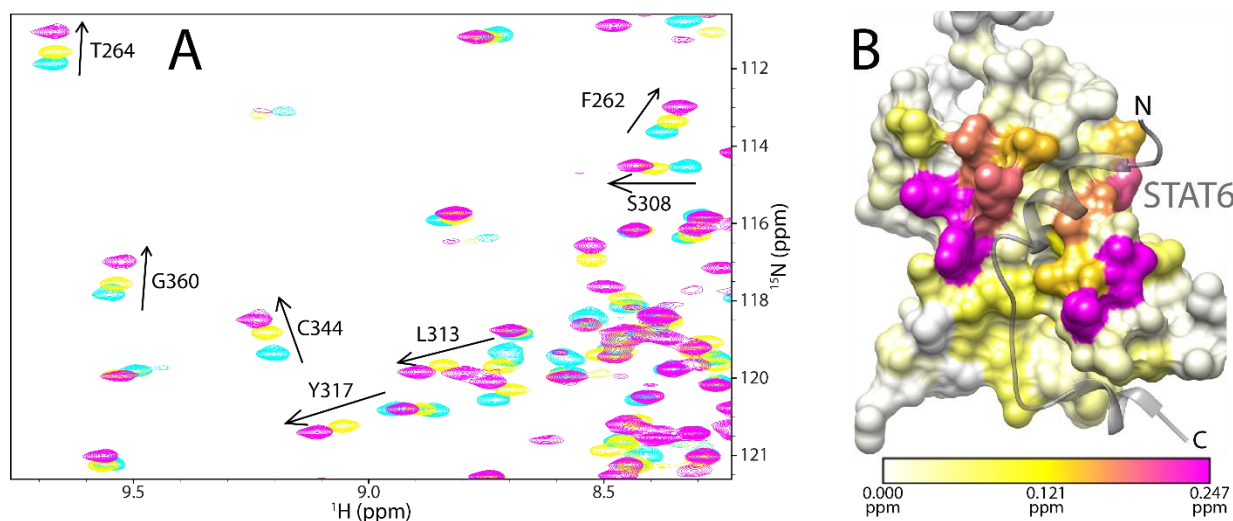
Figure 3.8 Fusing the 30-51 AF1 peptide to PAS-B does not affect binding.
(A) Overlaid HSQC spectra of apo ^{15}N-PAS-B (cyan), ^{15}- PAS-B + equimolar 30-51 AF1 peptide (yellow), and the ^{15}N-PAS-B-Nurr1 30-51 fusion. Arrows show peaks moving towards their location in the fusion spectra which represents the highest occupancy state. (B) CSPs from PAS-B Nurr1 30-51 fusion spectra mapped onto the structure (PDB entry 5NWM)[63] of PAS-B bound to STAT6 (shown in grey). The intermediate and high values in the color key represent $\langle\Delta\delta_{H,N}\rangle + 1\sigma$ (yellow) and $\langle\Delta_{\delta H,N}\rangle + 3\sigma$ (magenta), respectively. Colors are linearly interpolated to reflect the corresponding values.
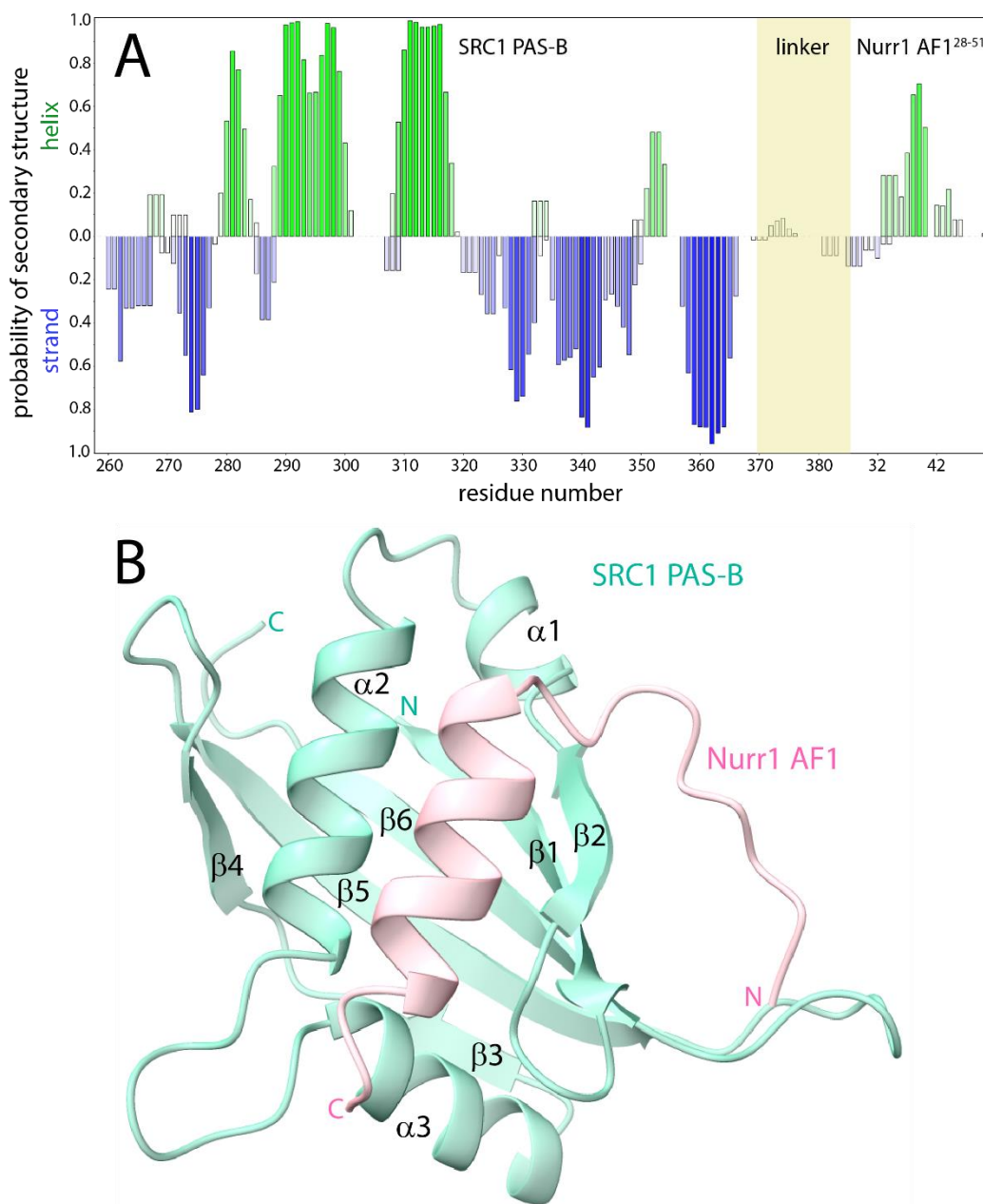
Figure 3.9 Nurr1 AF1 engages the SRC1 PAS-B hydrophobic cleft via a helical motif.
(A) A bar plot showing backbone chemical shift-based predictions of secondary structure for the SRC1 PAS-B-Nurr1 AF1[28-51] fusion by the GetSBY module in the NMRFAM-Sparky suite. (B) A model of the interacting regions of the Nurr1 AF1[28-51]-SRC1 PAS-B complex predicted by AlphaFold-Multimer (v.2.1.0)[124,129]; calculations were performed with the AlphaFold Colab notebook using default settings and the sequences as separate chains (https://colab.research.google.com/github/deepmind/alphafold/blob/main/notebooks/AlphaFold.ipynb). Figure prepared using ChimeraX[130].

Since traditional structural techniques proved intractable for this system, AlphaFold-Multimer was used to gain further insight into the structure of the complex[124,129,131]. Running the PAS-B domain and the AF1 28-51 peptide through the AlphaFold-Multimer workflow produced a structure that looked consistent with my experimental data. The Nurr1 peptide was folded into an α-helix as predicted and was bound to the hydrophobic cleft formed by β2 and α2, as was shown by CSP mappings (Fig. 3.9B). Mutants of Nurr1 AF1 were then designed to test AlphaFold's prediction that residues 30-51 formed a helix upon binding and the helix's predicted orientation. The predicted helix has a hydrophobic surface that fits in between β2 and α2, which includes residues F39, F42, and L46 (Fig. 3.10AB). F39 was mutated to a proline to test the effect of helix disruption and F42 and L46 were mutated to arginines to confirm that these residues were oriented facing into the hydrophobic cleft and were central to the interaction. Additionally, F34 and L35, which are in an extended conformation N-terminal to the helix, were simultaneously mutated to arginines to test their contributions to binding. (Fig. 3.10C). These mutants were all made in the background of full-length AF1 and the mutated proteins were purified and used in pulldown assays with PAS-B. Comparing the retention of mutants with wild-type AF1 revealed that all mutants disrupted binding to varying extents (Fig. 3.10D). The F39P mutant almost eliminated binding, which supports the prediction of helix formation, as disrupting the helix would eliminate multiple important contacts. The F42R mutant disrupted most of the binding and the L46R mutant partially disrupted binding, confirming that these residues play important roles in the interaction and supporting the prediction that places these residues in the cleft of β2 and α2. The F34R, L35R mutant had a weak effect on binding, suggesting these residues play a less significant role in stabilizing the interaction. Although arginine mutants were used in these binding assays to produce more dramatic effects, alanine mutants of F39 and F42 also show diminished binding (Fig. 3.11)
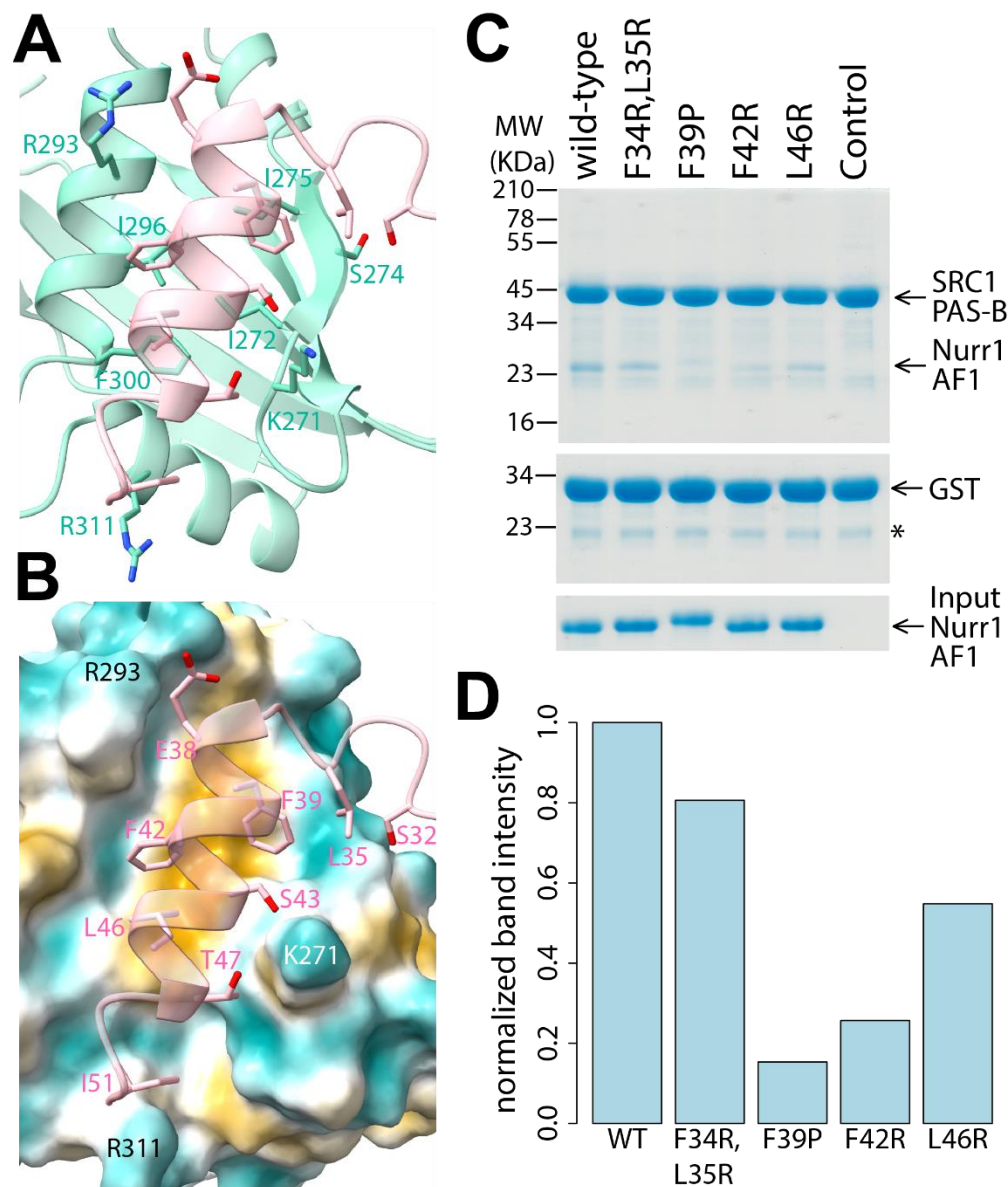
Figure 3.10 Functional analysis of Nurr1 AF1-SRC1 PAS-B interactions.
Views of the intermolecular interface showing (A) the interacting side chains of residues in PAS-B and AF1 and (B) the interacting side chains of the AF1 residues with PAS-B rendered as a molecular surface color coded according to the molecular lipophilicity potential (dark gold: most lipophilic; dark teal: most hydrophilic) in ChimeraX[130]. Intermolecular interactions were analyzed using MONSTER (https://numonster.northwestern.edu)[132]. (C) GST-pulldown assays conducted with purified Nurr1 AF1 wild-type or mutant proteins and GST-PAS-B. Input and bound proteins were resolved by SDS-PAGE and stained using Coomassie Brilliant Blue. The results of the assay are shown in the top gel along with the GST control in the middle gel while the input Nurr1 AF1 proteins are shown in the bottom gel. (D) Bar graphs quantifying the amount of bound protein in the GST-pulldown assays after normalizing for the input proteins.
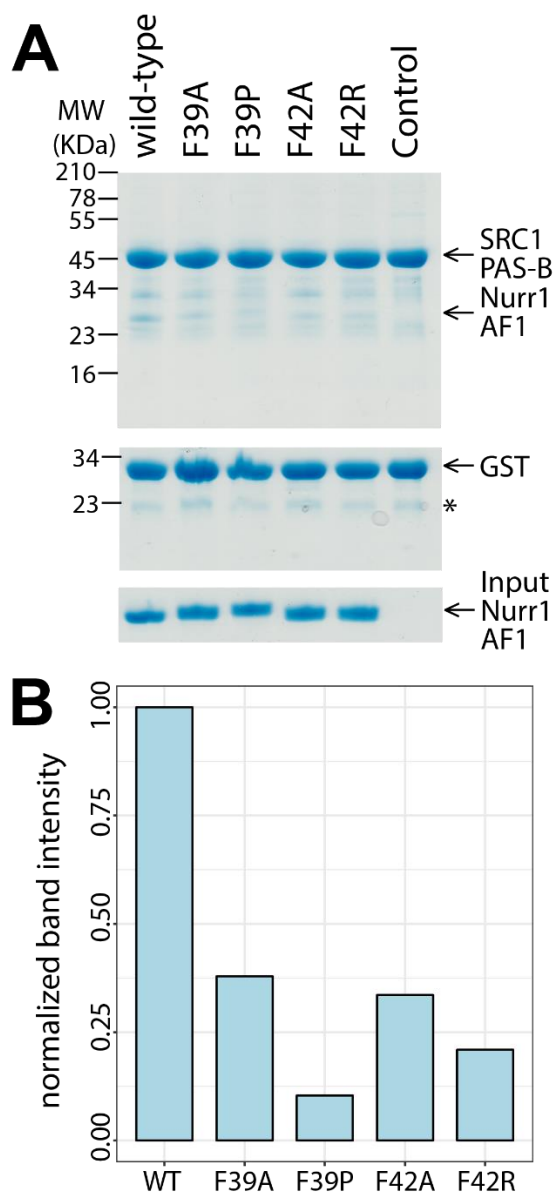
Figure 3.11 Functional analysis of Nurr1 AF1 alanine mutants and comparison with the proline and arginine mutants.
(A) GST-pulldown assays conducted with purified Nurr1 AF1 wild-type or mutant proteins and GST-PAS-B. Input and bound proteins were resolved by SDS-PAGE and stained using Coomassie Brilliant Blue. The results of the assay are shown in the top gel along with the GST control in the middle gel while the input Nurr1 AF1 proteins are shown in the bottom gel. (B) Bar graphs quantifying the amount of bound protein in the GST-pulldown assays after normalizing for the input proteins**.**

Conclusions

Learning how transcription factors recruit coactivators is key to understanding gene regulation. For NRs, understanding the specific role their AF1 domains play in this process is a long-standing challenge[46]. Here I sought to gain insight into why the AF1 domain of Nurr1 is important for recruiting SRC1. My work shows that the NR4A receptor Nurr1 can directly recruit SRC1 via an interaction between their respective AF1 and PAS-B domains, potentially representing a novel mechanism for SRC recruitment by NRs.

Initial characterization of the interaction showed it to be of moderate affinity, with a measured K_d of ~8 μM. Binding of these affinities often play a role in the assembly of large transcriptional complexes with multiple binding sites[125–128,133]. Multivalent systems can produce binding profiles that appear cooperative in nature and allow for more precise control of complex assembly[134]. Similarly, low to moderate affinity binding can mediate phase separation in transcriptional condensates, which has been shown to be important for the activity of multiple NRs[54,135]. It is also possible that a post-translational modification or binding of a ligand could increase the affinity of the interaction *in vivo*. The NR4A family members including Nurr1 have been shown to be phosphorylated by a variety of intracellular kinases[136,137]. Similarly, PAS domains are known to bind small molecules and transduce signals[138]. All of these remain possibilities until the association of Nurr1 and SRC1 are understood more fully in the context of an assembled transcriptional complex in the cell.

I then looked to biochemically and biophysically characterize how AF1 binds PAS-B. Initially, I used ITC, mutagenesis, and NMR experiments to show that residues 30-51 bind PAS-B similarly to full-length AF1 and are the primary determinant of binding. Structure determination by crystallography or NMR was unsuccessful, so I turned to a mixture of *in silico* and experimental

techniques to gain structural information for the complex. Putting the PAS-B and AF1 30-51 sequences through AlphaFold-Multimer produced a predicted structure in which the 28-51 peptide was helical and bound to the expected surface of PAS-B[124]. In support of this prediction, assignment of a fusion construct of PAS-B and the 28-51 peptide revealed that a portion of the 30-51 segment likely forms a helix upon binding and mutating a residue in the center of the helix to a proline (F39P) almost completely disrupted the interaction. This is not surprising, as disorder to helix transitions are a common theme in interactions between unstructured and structured domains[139,140]. Furthermore, mutating residues that were predicted to be at the interface partially disrupted binding, suggesting that AlphaFold correctly aligned the helix with PAS-B.

From the perspective of PAS-B, my NMR data shows that both the full-length AF1 and the 30-51 AF1 peptide bind a hydrophobic cleft created by β2 and α2, which are two of the most conserved segments in the PAS-B domain. This was striking, as the only other transcription factor known to interact with SRC1 PAS-B, STAT6, binds this same hydrophobic cleft, also via a short helical peptide[63]. This suggests binding to the conserved surface of PAS-B represents a more general mechanism for recruitment of SRC1 by transcription factors. This argument is further supported by the fact that other transcriptional regulators, such as ARNT, have been shown to recruit proteins via interactions between short helices and their PAS domains[141]. It is also interesting to note that in the SRC1 paralogs, SRC2 and SRC3, the PAS-B domain is not highly conserved, having only 38-52% sequence identity. This could indicate that the PAS-B domain may play a role in determining the specificity of SRC proteins for different transcription factor targets. Future studies should investigate the ability of other NR4A receptor AF1 domains to interact with the PAS domains of the SRC proteins.

A direct interaction between the Nurr1 AF1 domain and the SRC1 PAS-B domain represents a novel way for NRs to recruit an SRC coactivator and has unique implications for future studies of both NR AF1 domains and the SRC coactivators. One implication could be the establishment of new paradigms, like that seen with binding of LxxLL motifs to NR LBDs. My work showing that a helical motif binds a conserved surface of the SRC1 PAS-B domain suggests a mechanism for recruitment of SRC proteins that could be found in other NRs and lends credence to the developing paradigm of interactions of helical motifs and PAS domains as a part of transcriptional regulation[64]. Conversely, it is important to remember that biology does not always follow a simple set of rules. In contrast with the conserved LBDs, NR AF1 domains are more varied in both length and sequence and the mechanisms by which they recruit coactivators are largely unknown[46,49]. The ability of the Nurr1 AF1 domain to recruit an SRC coactivator by interacting with one of its PAS domains represents a previously unseen mode of AF1 functioning. Additionally, structure-function analysis has provided key insights into how an unstructured AF1 domain interacts with a structured coactivator domain. This highlights that AF1 domains likely have diverse ways of functioning and to uncover them they must each be approached differently based on the unique biological context in which they function.

CHAPTER 4: Characterization of a novel ligand-binding pocket in the SRC1 PAS-B domain

Introduction

The SRC family of proteins are an important class of transcriptional coactivators that facilitate gene activation. These multi-domain proteins function by binding to transcription factors located at gene enhancer regions and, subsequently, helping to assemble molecular complexes that alter chromatin states and initiate transcription[3,56,61,142]. SRCs were first discovered for their role in nuclear receptor mediated gene activation but since then have been shown to interact with numerous other transcription factors[56,143,144]. There are three SRC members in humans, SRC1, SRC2, and SRC3, which have a conserved domain architecture. The majority of the protein is unstructured and contains the well-studied LxxLL motifs that are used to recruit NRs and two additional segments used in interactions with the CBP/p300 and the CARM1/PRMT1 coactivators. However, at the N-terminus there reside three structured domains known as bHLH, PAS-A, and PAS-B.

PAS domains have a broad spectrum of molecular roles in cells, including facilitating protein-protein interactions, sensing changes in concentrations of light, oxygen, metals, and various metabolites[62,138,145]. In humans they are especially well known for playing roles in regulation and signaling in processes such as transcription, kinase signaling pathways, and metabolism[62,65]. Often these roles, frequently found in PAS domains that are quite distantly related, involve binding of a ligand. In the case of the SRCs, studies with a closely related PAS-B domain, that of the transcription factor HIF2, showed that it could bind artificial small molecules, which affected its ability to form heterodimers[146].

To date, the only high-resolution structure of one of the SRC PAS domains is that of SRC1 PAS-B bound to the transcription factor STAT6[60,63]. PAS domains usually consist of two to three alpha helices layered next to a five-strand beta sheet with domains that bind ligands usually having a pocket located in between the alpha helix and beta sheet layers. In this structure, the SRC1 PAS-B domain harbors a small pocket located between the alpha helices and the beta sheet. Given the propensity of PAS domains for ligand binding, I investigated if any small molecules could bind this pocket. Using a combination of *in vitro* techniques, I show that various hydrophobic small molecules can in fact bind to PAS-B. Furthermore, I use NMR to characterize the site-specific binding (one covalently) of two prostaglandin molecules to its internal pocket.

Methods

Protein design, expression, and purification

For study of the SRC1 PAS-B domain, the DNA coding sequence for residues 260-369 (PAS-B) were cloned into the pMCSG7 vector[75]. A fusion construct was designed of the SRC1 PAS-B domain and a shortened segment of the Nurr1 AF1 domain, consisting of residues 28-51. In this fusion construct SRC1 residues 370-385 were chosen to act as the linker between the PAS-B domain and the Nurr1 AF1 segments because previous X-ray crystallography and NMR studies as well as disorder prediction software suggested this region is disordered in solution[60].

Prior to use all DNA constructs were sequenced to verify identity. Constructs were transformed into *E. coli* BL21 (DE3) cells which were grown at 37 °C until they reached an O.D. of 0.6, at which time they were moved to 16 °C, induced with 1 mM IPTG, and proteins were expressed overnight. The next day cells were pelleted and were stored at -80 °C until use.

SRC1 PAS-B wild-type, and PAS-B-Nurr1 fusion proteins were purified in the following manner. Cells pellets from 0.5 l or 1 l cultures were resuspended in 40 ml of lysis buffer (300 mM NaCl, 50 mM HEPES, 1 mM TCEP, pH 7.5) and sonicated for 25 minutes. The resulting cell lysate was then clarified by centrifugation at 12,000 rpm for 20 min and the soluble supernatant was put on to a column containing 1-2 ml of Ni^{2+}-NTA resin (Qiagen) and incubated at 4 °C for one hour. The supernatant was then allowed to flow through the column and two 10 ml washes were performed with both high-salt wash buffer (lysis buffer + 300 mM NaCl, 10 mM imidazole) and wash buffer (lysis buffer + 10 mM imidazole). The protein was then eluted in four 4 ml fractions with an elution buffer (lysis buffer + 275 mM imidazole) and the purity of the fractions was analyzed by running SDS-PAGE gels. Clean fractions were combined and were put through two rounds of dialysis at room temperature overnight to remove imidazole. The His_6-tag was cleaved during dialysis using 1:25 (w/w) TEV protease.

After dialysis, cleaved PAS-B protein was immediately put through HPLC following protein-containing fractions were lyophilized and stored for future use. When ready for SRC1 PAS-B was resuspended in the desired buffer containing also 6 M guanidine hydrochloride and then put through two rounds of dialysis overnight at 4 °C to remove the guanidine hydrochloride.

After dialysis and cleavage of PAS-B-Nurr1 fusion proteins, the Ni^{2+}-NTA resin used in the purification was cleaned using 6 M Gdn-HCl, 500 mM imidazole, and 0.1 % Triton-X and regenerated with Nickel Sulfate and the proteins were put back through the column to remove any uncleaved protein. The fully cleaved protein was then further purified by way of size-exclusion chromatography using a Superdex75 column and SEC running buffer (150 mM NaCl, 25 mM Tris, 1 mM TCEP, pH 7.5).

NMR sample preparation and data collection

All [15]N- and/or ^{13}C-labeled proteins were produced similarly to unlabeled protein except that they were grown in media containing ^{15}N ammonium sulfate and/or ^{13}C glucose as their sole nitrogen/carbon sources. Labeled proteins were dialyzed into NMR buffer (50 mM NaCl, 20 mM NaPO4, 1 mM DTT, pH 6.8) and concentrated to between 0.1 mM-0.25 mM. Final samples were then prepared by adding in EDTA (0.5 mM) and D_2O (5 %) and adjusting the pH to 6.8. NMR data were acquired a Bruker Neo 600 MHz spectrometer equipped with a QCI-F cryoprobe at temperatures ranging from 25-30 °C. NMR data were processed using Topspin and NMRpipe and analyzed using NMRFAM-Sparky[78]. 2D ^{1}H-^{15}N HSQC spectra were collected for ligand titrations. Chemical shift perturbations were calculated in R using the equation

$$CSP = \left(1/2\left[\Delta\delta_H^2 + \left(0.04(\Delta\delta_N^2)\right)\right]\right)^{1/2}$$

in which δ_H and δ_N are the proton and nitrogen chemical shifts, respectively. The CSPs for each residue were then mapped onto structures for analysis using Chimera.

Ligand titrations

Ligands were received as dry powder or in ethanol solutions that were blown dry and then resuspended in DMSO or water to concentrations of several hundred micromolar to 50 mM. Subsequently, they were aliquoted and stored at -20 °C until used. For NMR titrations, non-covalently binding ligands were titrated into NMR samples, the sample's pH was adjusted, and data was collected. In the case of multiple titration points, additional ligand was added, and the sample pH was readjusted after collection of each data point. DMSO caused small CSPs in the PAS-B spectra above concentrations of 3%-5% and in these cases a reference spectrum was used to prevent misattribution of ligand CSPs to DMSO. Ligands that had the potential for covalent

modification were incubated with PAS-B overnight at 4 °C in buffer containing little or no reducing agent prior to NMR or MS data collection. In the case of PEA and dopamine, ligands were resuspended in DMSO and added to PAS-B NMR samples in an anaerobic glove box to prevent oligomerization and degradation. In cases were ligands had light sensitivities, such as with PGA1, samples were prepared in low-light conditions.

Mass Spectrometry

To assess covalent modification of PAS-B and fusion PAS-B by PGJ2, proteins were subjected to LC-ESI-MS using an Agilent 6545 QTOF instrument. Runs were performed by first injecting 7 μl of protein samples in native buffers onto a guard column using a running buffer of consisting of H_2O and 0.1% formic acid. Proteins were then eluted with a running buffer of 100 % acetonitrile for detection. The resulting spectra were deconvoluted and analyzed using Agilent's MassHunter Bioconfirm software.

Results

In silico and in vitro analyses show that SRC1 PAS-B contains a pocket that can bind ligands
Inspection of the PAS-B crystal structure bound to both STAT6 and stapled peptides revealed a cavity in the center of the domain (Fig 4.1). It is largely lined by bulky hydrophobic residues such as F262, I296, F300, F314, and I358, with the notable exception of C344 that resides deep in the pocket. Characterization by the CASTp pocket-finding algorithm showed the cavity to have a volume of ~280-380 Å^3 [147]. Interestingly, the pocket is directly adjacent to the binding site for peptides from both the STAT6 and Nurr1 transcription factors, with residues forming the mouth of the pocket also making direct contacts with the bound peptides[60].
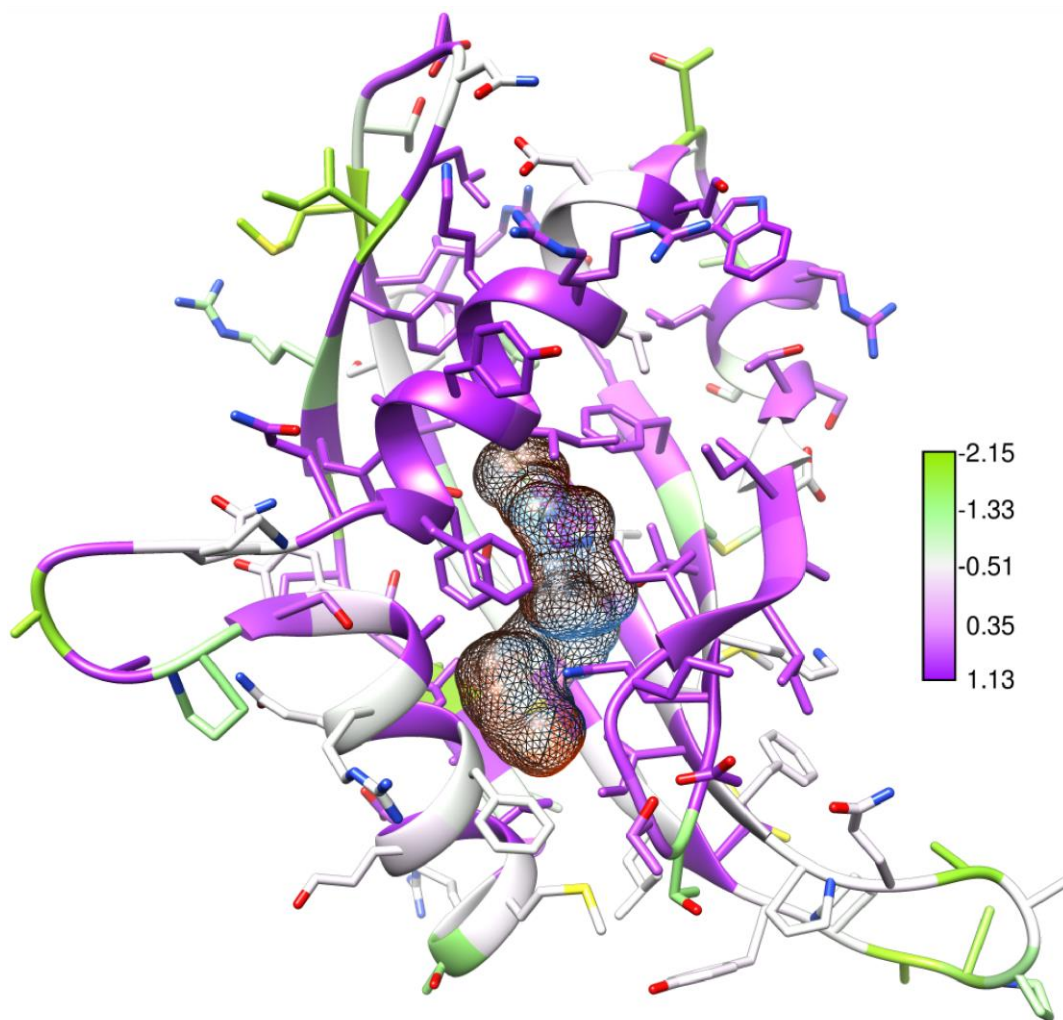
Figure 4.1 A pocket in the interior of SRC1 PAS-B.
The internal cavity in PAS-B depicted via mesh surface based on volume calculations using the CASTp software[147]. Residues are colored according to conservation, with residues in purple being invariant. CASTp and sequence analysis performed by Ishwar Radhakrishnan.

To test if PAS-B could bind small molecules, I first compiled a short list of candidate small molecule ligands that are thought to play a role in regulating NR4A receptors, which are known to recruit SRC1[115,148,149]. I chose to look at 6-mercaptopurine (6-MP) and dihydroxyindole (DHI), both potential regulators of the NR4A family of NRs, as well as β-phenylethylamine (PEA), a precursor of dopamine which is regulated by NR4A receptors, and indole, which is a structural analog of dopamine. NMR titrations were then performed with these candidate ligands to test for binding *in vitro*. For these experiments ^{15}N isotopically labeled PAS-B was purified, ligands were titrated into the resultant PAS-B NMR sample, and ^{1}H-^{15}N HSQC spectra were collected for each sample. These spectra were compared to the apo PAS-B spectrum to quantitate any chemical shift perturbations (CSPs), which were then mapped onto the PAS-B structure. Titrations with 6-MP and DHI did not produce any significant CSPs. However, titrations with PEA and indole both produced CSPs that primarily mapped to residues in the binding pocket, with the strongest shifts coming from indole (Fig. 4.2A). This confirms that an internal pocket is present in PAS-B in solution and can be bound by small molecule ligands (Fig. 4.2B).

Small Molecule	Ratio (protein to ligand)	Protein	Binding (Yes/No)
6-mercaptopurine	1 to 10	SRC1 PAS-B	N
dihydroxyindole	1 to 1	SRC1 PAS-B	N
dimethylsulfoxide	2.5%, 5%	SRC1 PAS-B	Y
indole	1 to 10, 1 to 40	SRC1 PAS-B	Y
L-tyrosine	1 to 1, 1 to 5	SRC1 PAS-B	N
dopamine	1 to 10	SRC1 PAS-B	N
phenylethylamine	1 to 10, 1 to 30	SRC1 PAS-B	Y
tyramine	1 to 10	SRC1 PAS-B	N
tryptamine	1 to 10	SRC1 PAS-B	N
phenylethylamine HCl	1 to 10, 1 to 40	SRC1 PAS-B	N
epinephrine	1 to 10	SRC1 PAS-B	N
pyrodoxine	1 to 10	SRC1 PAS-B	N
lidocaine	1 to 10	SRC1 PAS-B	N
nornicotine	1 to 10	SRC1 PAS-B	N
carbamazepine	1 to 10	SRC1 PAS-B	N
diacetin	1 to 10	SRC1 PAS-B	N
all-trans retinoic acid	1 to 10	PAS-B Nurr1 28-51 Fusion	N
arachidonic acid	1 to 10	PAS-B Nurr1 28-51 Fusion	N

Table 4.1 Small molecule titrations with SRC1 PAS-B and PAS-B Nurr1 Fusion
List of small molecules that were used in NMR titrations with SRC1 PAS-B and the PAS-B Nurr1 28-51 fusion. Included is the name of the molecule, the ratio of protein to small molecule, the protein used in the titration, and whether or not binding was observed via chemical shift perturbations.
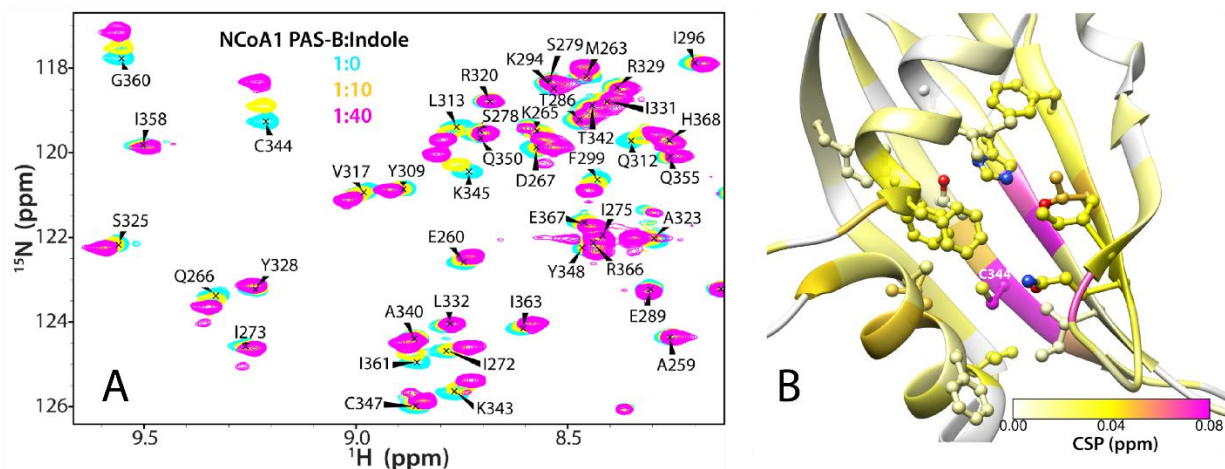
Figure 4.2 NMR titrations confirm indole binds to the PAS-B pocket.
(A) Overlaid ^{1}H-^{15}N HSQC spectra from titrations of indole into ^{15}N-labeled PAS-B. (B) CSPs from the final point (1:40) of indole titrations with PAS-B mapped onto the structure of PAS-B (PDB entry 5NWM)[63]. Residues forming the pocket are shown via ball and stick representation. The intermediate and high values in the color key represent $\langle\Delta\delta_{H,N}\rangle + 1\sigma$ (yellow) and $\langle\Delta_{\delta H,N}\rangle + 3\sigma$ (magenta), respectively. Colors are linearly interpolated to reflect the corresponding values.

Ligand docking and NMR titrations show that prostaglandins can bind PAS-B

To identify candidate endogenous ligands for the SRC1 PAS-B domain, an unbiased *in silico* screen of compounds in the small molecule pathway database (https://smpdb.ca) by pursuing a flexible docking approach using RosettaLigand. Given the relatively small size of the pocket, we only considered ligands with molecular weight <375 Da. Over 2000 compounds were screened in this manner and a handful of the top scoring candidates were evaluated for binding by solution NMR. Although none of these compounds produced any appreciable changes to the HSQC spectrum of ^{15}N-PAS-B, visual inspection of the docked poses of ligands with top scores revealed a recurring theme of carboxylate and aliphatic moieties of fatty acids engaging respectively with the side chains of conserved arginine and phenylalanine residues (R311 and F314) in the α3 helix. This led us to investigate if prostaglandins, which are fatty acid-derived signaling molecules, could bind PAS-B. NMR titrations were thus performed with ^{15}N-PAS-B and seven different prostaglandin derivatives. Titrations with some of the prostaglandins were also performed with a fusion construct of PAS-B and the Nurr1 28-51 that binds to the transcription factor binding site on PAS-B to test if prostaglandin binding might affect transcription factor recruitment. Since two of the molecules, Prostaglandin A1 (PGA1) and 15-deoxy-$\Delta^{12,14}$-prostaglandin J2 (15d-PGJ2), are known to covalently bind to cysteine residues, likely as a mechanism for regulating protein function, these molecules were incubated with PAS-B in buffer containing little to no reducing agent prior to data collection for some of their titration series[26,150,151]. All seven molecules induced significant CSPs (Table 4.1), with Prostaglandin B1 (PGB1) and PGJ2 inducing the strongest perturbations (Fig. 4.3A,B), confirming that prostaglandins can bind PAS-B.

	Ligand	Mean CSP ± Standard Deviation for All Peaks (ppm)	Mean CSP ± Standard Deviation for Top 5% of Most Perturbed Peaks (ppm)
PAS-B	PGB1	0.052 ± 0.055	0.229 ± 0.023
	PGE2	0.031 ± 0.031	0.129 ± 0.016
	PGA1	0.036 ± 0.046	0.198 ± 0.029
	PGE1	0.032 ± 0.033	0.134 ± 0.021
PAS-B-Nurr1 Fusion	PGB1	0.026 ± 0.026	0.114 ± 0.036
	PGE2	0.009 ± 0.007	0.031 ± 0.007
	PGA1	0.014 ± 0.019	0.083 ± 0.027
	8-iso PGE1	0.013 ± 0.010	0.041 ± 0.019
	PGD1	0.008 ± 0.004	0.019 ± 0.002
	PGJ2	0.038 ± 0.064	0.270 ± 0.076

Table 4.2 CSPs from titrations of prostaglandins with PAS-B
Mean CSPs based on HSQC of titrations of various prostaglandins with either wild-type PAS-B or the PAS-B Nurr1 30-51 fusion. Mean values were calculated by averaging CSPs from the top five percent most perturbed residues with calculated standard deviations shown.
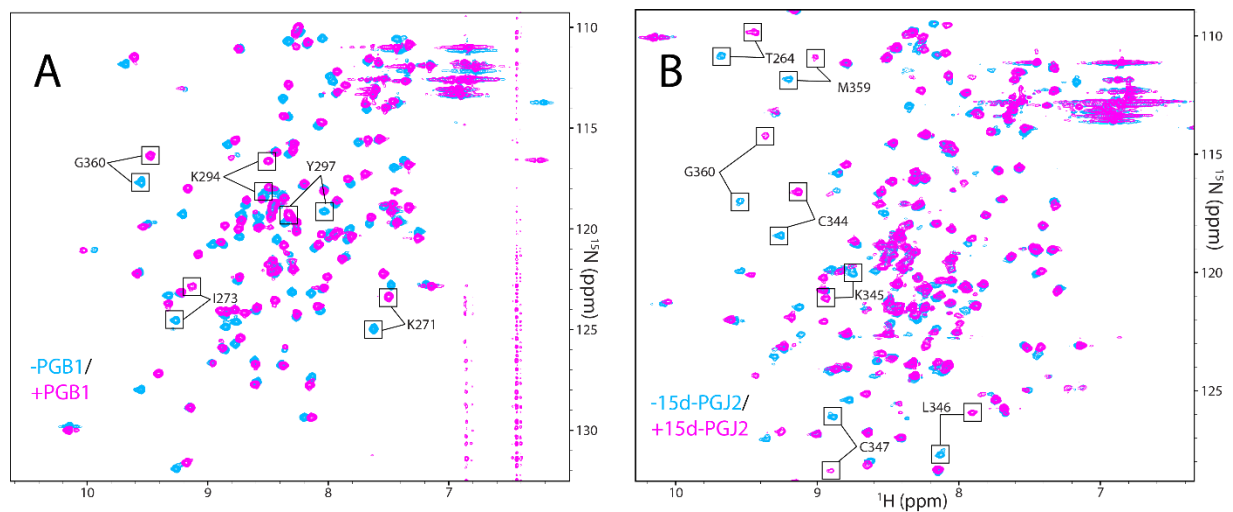
Figure 4.3 PGB1 and PGJ2 cause significant changes to the PAS-B and PAS-B-Nurr1 fusion HSQC spectra.

Overlaid ^{1}H-^{15}N HSQC spectra of (A) ^{15}N-labeled PAS-B in the absence (cyan) and presence (magenta) of PGB1 and (B) ^{15}N-labeled PAS-B-Nurr1 fusion in the absence (cyan) and presence (magenta) of PGJ2. Shifts of highly perturbed peaks are highlighted with arrows and boxes.

PGB1 binds the PAS-B pocket with moderate affinity and can interfere with transcription factor binding

Mapping of CSPs onto the PAS-B structure revealed that residues inside the pocket, such as C344, G360, and L313 were strongly perturbed by the titration of PGB1 into ^{15}N-PAS-B, suggesting that PGB1 does insert into the pocket as predicted (Fig. 4.4A). Interestingly, residues at or around the mouth of the pocket such as K271, I273, D276, D290, R293, K294, and Y297 were also strongly perturbed. This suggests the possibility that PGB1 is forming an encounter complex on the surface of PAS-B before inserting into the pocket or that binding of PGB1 inside the pocket causes significant conformational changes in residues around the mouth of the pocket or that there are multiple binding modes. Assessment of CSPs based on titration concentration suggested the range of affinity of PGB1 for PAS-B to be in the hundreds of micromolar.
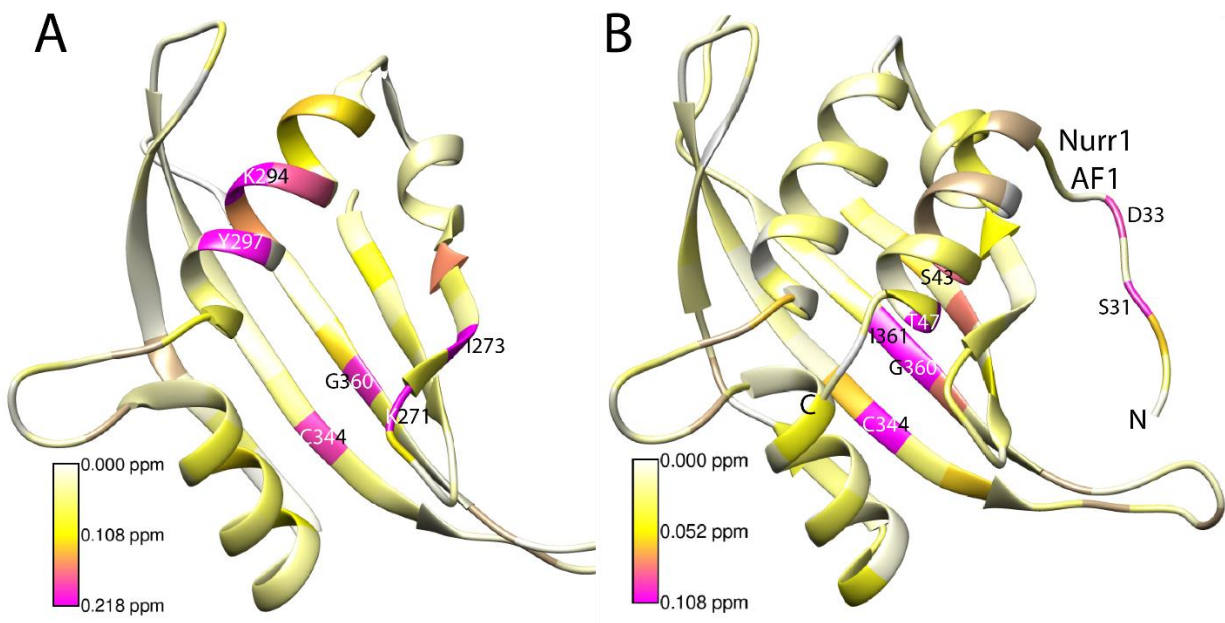
Figure 4.4 PGB1 binds wild-type and fusion PAS-B.
CSPs from titrations of PGB1 with (A) wild-type PAS-B and (B) the PAS-B Nurr1 30-51 fusion mapped onto the structure of PAS-B (PDB entry 5NWM)[63]. The intermediate and high values in the color key represent $\langle\Delta\delta_{H,N}\rangle + 1\sigma$ (yellow) and $\langle\Delta_{\delta H,N}\rangle + 3\sigma$ (magenta), respectively. Colors are linearly interpolated to reflect the corresponding values.

PAS-B Residue	Calculated Affinity (µM)
I273	1186.36
K271	1055.61
K294	282.19
Y297	270.42
G360	1905.00

Table 4.3 Estimated Binding Affinity of PGB1 for PAS-B
Shown are calculated affinities for individual PAS-B residues based on titrations performed with PGB1. PAS-B was at a concentration of 0.2mM and titration points for PGB1 were 0:1, 1:1, 1:2, 1:5, and 1:10.

Previously, I had shown that residues 28-51 of the transcription factor Nurr1 bound PAS-B and facilitated the recruitment SRC1 by Nurr1. To test if PGB1 binding would affect PAS-B's ability to be recruited by Nurr1 AF1, I titrated PGB1 with ^{15}N labeled fusion protein of PAS-B and Nurr1 AF1 28-51 (fusion PAS-B). Mapping of the resulting CSPs onto a structure of both PAS-B and Nurr1 AF1 28-51 showed that PGB1 still binds the pocket of fusion PAS-B and has a significant effect on residues in the 28-51 peptide region (Fig. 4.4B) Since I had previously assigned these residues in the apo spectrum of Nurr1 1-164, I compared the directionality of the CSPs for the 28-51 peptide with the location of the corresponding peaks in the apo 1-164 spectrum. For many residues in the peptide, vectors could be drawn from their location in the fusion PAS-B spectrum, to the fusion PAS-B + PGB1 spectrum, to the apo Nurr1 1-164 spectrum, showing that PGB1 is lowering the occupancy of bound 28-51 for the fusion PAS-B. This provides good evidence that PGB1 likely disrupts the ability of Nurr1 to bind PAS-B.

PGJ2 covalently modifies a cysteine in the PAS-B pocket

We next sought to determine the mode of binding for PGJ2, since it is known to covalently modify cysteines. Initially, NMR titrations of PGJ2 were performed with PAS-B in the presence of excess reducing agent (5 mM TCEP). CSPs from this titration revealed that PGJ2 binds most prominently to residues such as I271, I273, R293, K294, and Y297, at and around the mouth of the pocket of PAS-B, with much subtler perturbations present for some residues in the pocket such as L313, V317, and C344. This suggests that under reducing conditions PGJ2 is bound primarily to the surface of PAS-B and, to a lesser extent, the interior of the pocket (Fig. 4.5).

Next, PGJ2 was titrated in with PAS-B-Nurr1 fusion in the absence of reducing agent. The resulting sample was then subjected to ESI mass spectrometry under non-native conditions,

revealing that PGJ2 covalently modified PAS-B at a single site with ~95% occupancy (Fig. 4.6A). An NMR spectrum of this sample was then collected and the resulting CSPs revealed that many of the strongest perturbations were associated with residues located around C344 in the pocket, suggesting that, upon entering the pocket, PGJ2 covalently modifies C344 (Fig. 4.6B). There is another candidate cysteine nearby for PGJ2 modification, C347, that faces away from the pocket and was strongly perturbed during the titration. However, C347 is not surrounded by strong CSPs like C344 and is located significantly farther from the binding surface of PGJ2 under reducing conditions, making it an unlikely site of modification. Further analysis of the CSPs show that, unlike for PGB1, the group of residues affected by covalent PGJ2 binding extends from the pocket down to the bottom of α-helix 3 and further along β-strand 5, indicating that covalently bound PGJ2 likely exists in a unique conformation that is more deeply buried in the domain. This conclusion is further evidenced by the fact that the residues in the Nurr1 28-51 peptide portion of the fusion PAS-B are not significantly affected by PGJ2 modification. To confirm the modification site, further experiments involving mutation of C344 could be performed, with the expected outcome being that incubation with PGJ2 would no longer modify PAS-B.
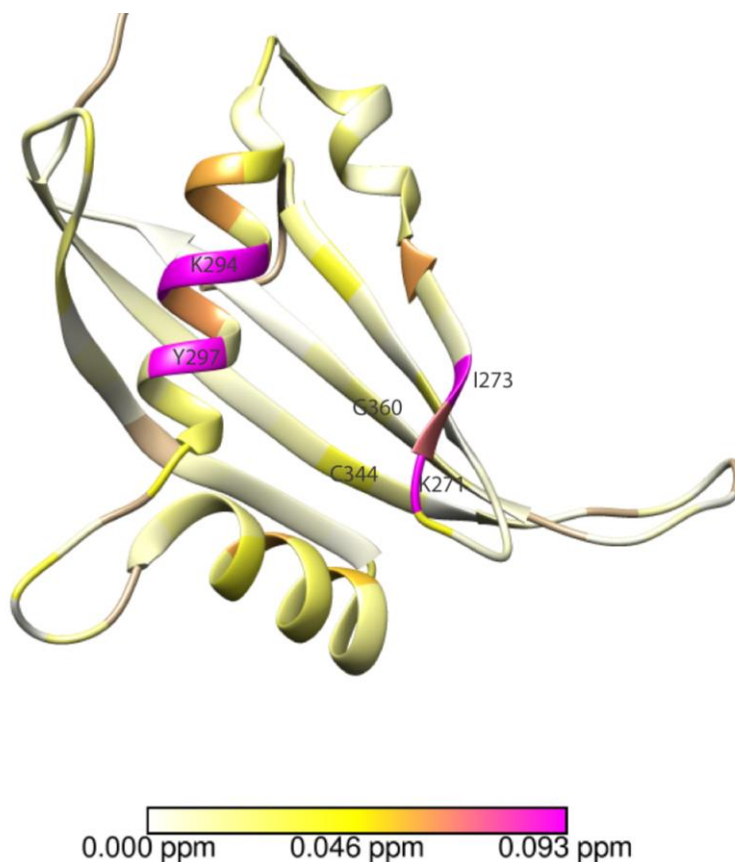
Figure 4.5 PGJ2 binds wild-type PAS-B under reducing conditions.
CSPs from titration of PGJ2 with wild-type PAS-B under reducing conditions (5mM TCEP) mapped onto the structure of PAS-B (PDB entry 5NWM)[63]. The intermediate and high values in the color key represent $\langle\Delta\delta_{H,N}\rangle + 1\sigma$ (yellow) and $\langle\Delta_{\delta H,N}\rangle + 3\sigma$ (magenta), respectively. Colors are linearly interpolated to reflect the corresponding values.
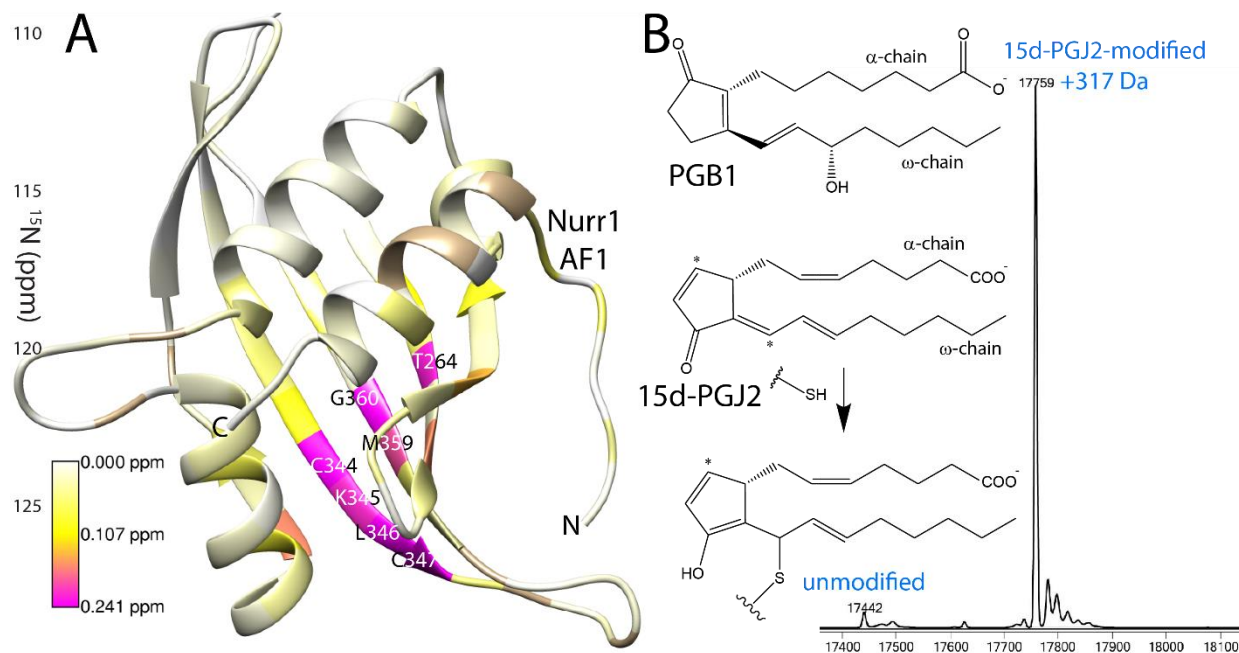
Figure 4.6 PGJ2 covalently modifies a single cysteine in PAS-B.
(A) CSPs from titrations of PGJ2 the PAS-B Nurr1 30-51 fusion mapped onto the AlphaFold Multimer predicted structure of PAS-B and the Nurr1 30-51 peptide[124]. The intermediate and high values in the color key represent $\langle\Delta\delta_{H,N}\rangle + 1\sigma$ (yellow) and $\langle\Delta_{\delta H,N}\rangle + 3\sigma$ (magenta), respectively. Colors are linearly interpolated to reflect the corresponding values. (B) Mass spectrum of PAS-B Nurr1 30-51 fusion after modification with PGJ2. Additionally, chemical structure of 15d-PGJ2 before and after covalent modification of a cysteine (asterisks: thiol-reactive centers).

Conclusions

PAS domains are ubiquitous in all domains of life and have a wide variety of functions. When investigating the structure of the transcriptional coactivator SRC1, I noted that its PAS-B domain had an internal cavity with no reported function in the literature. Given the history of ligand-binding in PAS domains, and in particular, in the closely related HIF-2 PAS-B domain, I investigated whether SRC1 PAS-B could bind to small molecules[62,146].

Inspection of the SRC1 PAS-B structure revealed a cavity in this location, suggesting that PAS-B might have a similar capacity to bind ligands. The pocket is lined with sidechains that are highly conserved among SRC1 orthologs and hydrophobic, hinting at the potential biological significance of the pocket. To test ligand binding, I performed NMR titrations with various small molecules and found that indole bound the pocket with millimolar affinity. While it is unlikely that this represents a biologically relevant interaction, it confirms that small, hydrophobic molecules can enter the PAS-B pocket, as the structure would suggest.

Docking simulations with a variety of small molecule ligands revealed that fatty acids might be able to bind PAS-B by inserting one of their hydrophobic tails into the pocket, while the carboxylic acid of the other chain could engage with a conserved arginine on the surface of the domain. An important class of fatty acids known as prostaglandins (PGs) have a number of connections to SRC biology and were a compelling target that could be used to test this prediction. Prostaglandins are the products of arachidonic acid metabolism and play an important role in inducing and resolving the inflammatory response by binding to G protein-coupled receptors (GPCRs)[152,153]. Certain PGs serve as endogenous ligands by binding covalently to nuclear receptors, including PPARγ and Nurr1, which often work in conjunction with SRC coactivators to alter transcription[150,151,154,155]. Thus, I tested the ability of seven different PGs to interact with PAS-

B and found that all seven bound to or in the vicinity of the pocket, with PGB1 and PGJ2 inducing the most significant CSPs.

NMR titrations showed that PGB1 binds PAS-B specifically with a K_d that is likely in the range of hundreds of micromolar and mapping of the CSPs onto the PAS-B structure shows that PGB1 binding has widespread effects in the pocket and on the surface of the domain around the mouth of the pocket. The extent of these CSPs on the surface PAS-B indicated that PGB1 may form an encounter complex before entering the pocket or that binding may cause conformational changes that propagate beyond the binding site.

The mouth of the pocket is located near the binding site of the STAT-6 and Nurr1 transcription factors, suggesting that ligand binding could affect these interactions. This was confirmed experimentally when an NMR titration of PGB1 with the fusion PAS-B showed the fused Nurr1 28-51 peptide being displaced. This suggests a potential role for ligand-binding in the regulation of SRC1's interactions with transcription factors which would further implicate the SRC1 PAS-B domains as a valuable target for drug design[64,118,146].

Currently, only Nurr1 and STAT6 are known to be directly recruited via the PAS-B domain. However, SRC1 is a target for numerous other transcription factors and there is a distinct possibility that the coactivator might be recruited through this surface, leaving the door open for future studies to investigate how ligand binding might alter their recruitment[156]. A recent study showed that another PG, PGA1, can bind the ligand-binding domain of Nurr1 and enhance its ability to activate transcription[151]. How this might interplay with PG binding by PAS-B is unclear and would require further study.

The other ligand with significant binding, PGJ2, was studied under both reducing and non-reducing conditions, due to its ability to covalently modify cysteines. Under reducing conditions,

it bound largely to the exterior of PAS-B, around the mouth of the pocket, and under non-reducing

conditions NMR and MS suggest it singly modified C344. It is interesting to note that C344 is

conserved throughout SRC1 orthologs which hints at its potential biological relevance (Fig 3.7).

NMR CSPs suggest that, when covalently bound, PGJ2 is inserted deeper into the pocket, such

that one of hydrophobic tails resides near α-helix 3 and β-sheet 4. Structures of PGJ2 bound to

PAS-B from docking and molecular dynamics simulations suggest that this is a structurally

plausible conformation and also show that among the two reactive centers, C13 rather than C9 of

PGJ2, is most likely to covalently modify C344 (Fig. 4.6).

Assessing the non-reducing and reducing data and structure of docked PGJ2 together leads

to a model in which PGJ2 forms an encounter complex at the mouth of the pocket and, upon

entering the pocket, C13 covalently modifies C344. While this model does require that PGJ2 is

inserted unexpectedly deep into the PAS-B pocket and move through some energetically

unfavorable conformations to perform covalent modification, some of my data suggests that

intrinsic dynamics present in PAS-B could facilitate this. It should be noted in passing that a

number of peaks in the PAS-B spectrum, corresponding to residues located near the tail of PGJ2,

have secondary peaks that suggest structural motions occur in PAS-B on the microsecond-

millisecond timescale[6]. Further investigation of these dynamics could lead to a better

understanding of how the domain binds ligands.

While the possibility of a novel role for ligands in SRC function is intriguing, it is important

to emphasize that more work needs to be done to determine if ligand binding via the PAS-B domain

has biological significance. A search for potential ligands would also benefit from a better

understanding of the transcription factors that bind directly to PAS-B, such that specific cellular

pathways could be correlated with this interaction and small molecules involved in those pathways

could tested for their ability to bind PAS-B[98,157]. Finally, this work suggests that it will be important to further structurally characterize other PAS domains involved in gene regulation. Identifying PAS domains with pockets like that found in SRC1 PAS-B could lead to novel gene regulatory mechanisms and additional targets for drug screening studies.

CHAPTER 5: Conclusions and Prospectus

Conclusions

Recruitment of transcriptional machinery to genes by TFs is a key step in gene regulation. For NRs this process often starts with the recruitment of coactivators, which help assemble large multi-protein complexes including the transcriptional machinery[2,4,158]. The molecular details of coactivator recruitment can be difficult to unravel because NRs and coactivators have multiple domains, each with different binding partners, and whose interactions involve complex phenomena such as allostery, ligand binding, and intrinsic disorder[14,38,46,49,158–160]. Understanding how a NR recruits a coactivator requires a stepwise process of determining which of their domains interact and then characterizing the biophysical factors that govern these interactions using a mix of biochemical and structural techniques. This work represents several such steps for a mix of NRs and coactivators.

Finding the cognate ligand for a NR LBD, or establishing it as ligand-independent, is fundamental to understanding the NR's biological role[21,32,161]. Often crystal structures are used to assess if and how a ligand might bind an LBD[72,73,162,163]. This was the case for Ftz-F1, which was deemed as likely an orphan receptor based on a crystal structure of its LBD, which showed one of its own helices (α6) packed tightly into the LBD's canonical ligand-binding pocket[70]. However, my work in Chapter 2 shows that motions on multiple timescales pervade the domain in solution. In particular, a unique combination of slow and fast time scale motions found in α6 and the residues surrounding it, suggest that the ligand-binding pocket is not as tightly packed in solution and that the segment spanning α6 likely spends some of its time exposed to solvent. This opens the possibility that either binding of a novel ligand or intrinsic motions in the pocket could regulate

coactivator recruitment by Ftz-F1 and is an important reminder that nuclear receptors cannot be deemed orphan receptors based solely on crystal structures.

In Chapter 3, I looked at a different NR-coactivator pair by studying how the NR Nurr1 recruits the coactivator SRC1. Previous work had shown that both Nurr1's AF1 domain and SRC1 are required to activate transcription but there was no concrete biochemical or structural data showing how SRC1 was recruited[102,103,107]. Additionally, there are currently no interaction paradigms, like that of LxxLL motifs for LBDs, that would suggest an obvious binding spot on SRC1 for AF1[46,100]. I hypothesized that the unstructured AF1 domain would interact with one of the structured domains at the SRC1 N-terminus and then used biochemical and NMR experiments to show that AF1 interacts directly with the PAS-B domain. This represents a novel, ligand-independent mechanism for recruitment of SRC proteins by nuclear receptors and is one of the few instances in which a direct interaction of an AF1 domain with a coactivator has been established[46,48,164]. I attempted to structurally characterize the AF1 and PAS-B complex using both crystallography and NMR, but was unable to get a high-resolution structure due to the complex's poor solubility and dynamic nature. To get around this I used a combination of biochemistry, NMR, and state-ofthe-art structure prediction software to develop a structural model of the complex[124]. This revealed that Nurr1 likely forms a helix upon binding and interacts with a conserved surface on PAS-B. Intriguingly, this was the same surface bound by a helical peptide from the STAT6 transcription factor, suggesting that binding to the PAS-B domain may represent a conserved way for transcription factors to recruit SRC1[60,63].

While studying the interactions of AF1 with PAS-B and looking closely at the PAS-B structure it became apparent that a small hydrophobic pocket exists in the center of the domain. Ligand binding is a common theme among PAS domains, with the ligand binding pocket typically

located in the same location as was seen in PAS-B, which prompted my investigation in Chapter 4 into the ligand binding capacity of PAS-B[138,146]. A mixture of *in vitro* and *in silico* screens revealed that small hydrophobic molecules could enter the pocket, with prostaglandins turning out to be the best binders. Interestingly, the mouth of the pocket is located right next to the transcription factor binding site on PAS-B, meaning that ligand binding could easily alter the affinity of interactions occurring at that surface[60,63]. These findings hint at potential novel regulatory mechanisms connected to SRC1 and provide a compelling new target for drug development[118].

Prospectus

This work contributes toward the broader goal of understanding the molecular factors that work together during NR-based gene regulation. Each chapter adds to our knowledge about certain NRs and coactivators, which in turn reveals new questions and avenues of study. Here are some of the outstanding questions and experiments that could be explored based on the work in this thesis.

A key remaining question from Chapter 2 is if the dynamics in α6 allow for binding of an unknown ligand or inherently alter the binding of the LxxLL coactivator peptide. *In vitro* assays, such as the thermal-shift assays I used with Ftz-F1 mutants, or *in silico* methods, like those employed in Chapter 4, could be used to screen small molecule libraries to try and find a native Ftz-F1 ligand. If a ligand were to be discovered, it is likely it would be some form of lipid, given that I provide some preliminary support for Ftz-F1 lipid binding with dot blot assays and other NR5A family members are known to bind phospholipids[71,73,90]. In this case, the groundwork for biochemical and structural studies is already laid, since I have established a protocol in the lab for successfully loading phospholipids into LBDs and studying them by native MS. If the dynamics in α6 inherently alter the binding of the LxxLL peptide, then extended MD simulations could be

used to look for allosteric pathways that connect α6 with the LxxLL binding site[91,92,160]. If a pathway is found, mutants could be designed disrupt the pathway and their effect on LxxLL binding could tested using thermal-shift assays. A broader continuation of this project could be to investigate Ftz-F1 orthologs to find the sequence determinants for ligand binding in NR5A receptors, which are known to have different ligand binding abilities in members from humans and mice[72,74,88]. Orthologs of Ftz-F1 have sequences in the ligand-binding pocket and in α6 that are similar to NR5A members that have both empty and ligand-bound pockets. Crystallization of these members could allow correlations to be drawn between sequence, pocket shape, and occupancy for NR5A members.

Because the interaction between Nurr1 AF1 and SRC1 PAS-B revealed multiple novel binding mechanisms for two unique classes of proteins, there are many directions the work in Chapter 3 could be taken. First, the interaction between AF1 and PAS-B could be investigated *in vivo*. Using a Nurr1 Δ30-51 deletion mutant, co-immunoprecipitation could be used to confirm the interaction occurs in cells and a Nurr1-based reporter assay could be used to confirm the significance of the interaction for Nurr1 based transcriptional activation. Given the moderate affinity of the interaction, it would also be interesting to investigate if post-translational modifications could alter its affinity. Phosphorylation is a known modification for NR AF1 domains and would be easy to begin working on, as I have already established a protocol for use of the $MKK6_a$ kinase in the lab and have shown that it successfully modifies the glucocorticoid receptor AF1[46,165]. More broadly, two important questions naturally arise from this work: are there other transcription factors that bind the conserved surface on PAS-B and are there other nuclear receptor AF1 domains that bind structured coactivator domains. In the future, cell-based studies

using co-immunoprecipitation experiments could help elucidate if other NR AF1 domains interact with the PAS domains of different SRCs.

In the short term, the work in Chapter 4 could be continued by characterizing how ligand binding occurs in the pocket. The presence of secondary peaks for many of residues in the PAS-B spectra suggests that the domain exists in multiple conformations in solution, which might play a role in ligand binding. MD simulations could be run to determine how structural motions affect the binding and orientation of ligands in the pocket[91,92]. CPMG and NOE experiments, like those performed in Chapter 2, could then be used to characterize dynamics in PAS-B, which could support or refute the results of the simulations[160,166,167]. However, the most interesting next step in this project would be to look for a native ligand for PAS-B. One way this could be done is to look for other transcription factors that bind PAS-B and investigate molecules that are involved in their regulatory pathways. An obvious place to start would be to investigate the systems that are regulated by the STAT6 transcription factor[60,63]. However, there could be many undiscovered interactions between transcription factors and PAS-B and uncovering them will take a substantial amount of work. A less speculative approach could involve *in vitro* screens using small molecule libraries in conjunction with high-throughput experiments, such as fluorescence thermal-shift assays, to detect ligand binding[168].

Significance

NRs and coactivators are common regulators of genes and their interactions are central parts of assembling transcriptional and regulatory complexes. Gaining insight into this assembly process is important for understanding how its disruption leads to disease, for offering new targets for therapeutics that alter gene regulation, and ultimately for building a more complete picture of

cellular function. However, many NR-coactivator interactions are difficult to visualize structurally, even with powerful new techniques such as cryoEM, and are often governed by complex biophysical phenomena and regulatory mechanisms. Subsequently, there are many NR-coactivator pairs that are generally known to associate but which lack a detailed explanation of how their interaction occurs and are regulated. This thesis work addresses this gap in knowledge for several important NR-coactivator pairs by discovering a new class of molecular binding partners, prostaglandins, for the important coactivator SRC1, as well as answering a longstanding question about how SRC1 and Nurr1 interact by discovering a novel interaction between their PAS-B and AF1 domains. Additionally, this work characterizes some of the key biophysical and structural features, such as conformational dynamics, natively disordered regions, and conserved binding surfaces that are found in key functional domains of the NRs Ftz-F1 and Nurr1 and the coactivator SRC1 (Fig. 5.1). These characterizations give a more complete picture of some of the molecular mechanisms that facilitate the association of these NR-coactivator pairs and in other cases will provide new testable hypotheses for how these interactions occur.
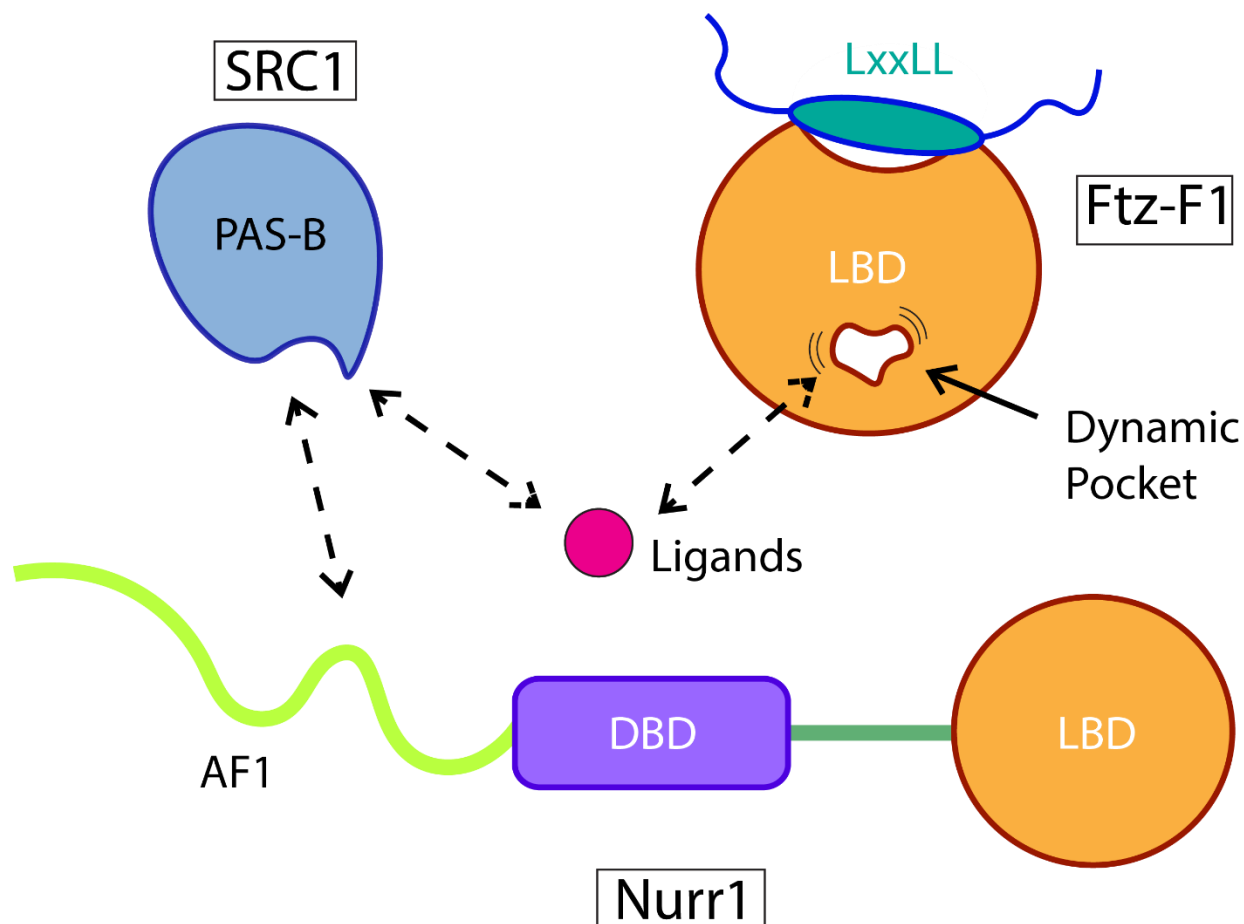
Figure 5.1 Overview of insights into nuclear receptor and coactivator interactions.
A cartoon schematic highlighting several key insights into nuclear receptor and coactivator interactions. Specifically, domains from the proteins SRC1, Nurr1, and Ftz-F1 that were studied in this thesis are represented in relation to these insights. For the Ftz-F1 LBD, dynamics in its ligand-binding pocket likely allow for a key helix to exit the pocket and the domain to bind ligands, which could affect its interactions with a coactivator LxxLL motif. Similarly, discovery of a pocket in the SRC1 PAS-B domain revealed a novel ligand-binding ability that could influence transcription factor binding. Finally, PAS-B was found to bind the AF1 domain of the nuclear receptor Nurr1, explaining the ability of Nurr1 to recruit SRC1 independent of its LBD.

REFERENCES

1.   Clapier, C. R., Iwasa, J., Cairns, B. R. & Peterson, C. L. Mechanisms of action and regulation of ATP-dependent chromatin-remodelling complexes. *Nature Reviews Molecular Cell Biology* vol. 18 407–422 (2017).
2.   Abdella, R. *et al.* Structure of the human Mediator-bound transcription preinitiation complex. *Science (80-.).* **372**, 52–56 (2021).
3.   Leo, C. & Chen, J. D. The SRC family of nuclear receptor coactivators. *Gene* vol. 245 1–11 (2000).
4.   Lambert, S. A. *et al.* The Human Transcription Factors. *Cell* vol. 172 650–665 (2018).
5.   Henzler-Wildman, K. & Kern, D. Dynamic personalities of proteins. *Nature* **450**, 964–972 (2007).
6.   Palmer, A. G. A dynamic look backward and forward. *J. Magn. Reson.* **266**, 73–80 (2016).
7.   Wodak, S. J. *et al.* Allostery in Its Many Disguises: From Theory to Applications. *Structure* vol. 27 566–578 (2019).
8.   Olsson, M. H. M., Parson, W. W. & Warshel, A. Dynamical contributions to enzyme catalysis: Critical tests of a popular hypothesis. *Chemical Reviews* vol. 106 1737–1756 (2006).
9.   Brath, U., Akke, M., Yang, D., Kay, L. E. & Mulder, F. a a. Functional Dynamics of Human FKBP12 Revealed by Methyl C Rotating Frame Relaxation Dispersion NMR Spectroscopy Functional Dynamics of Human FKBP12 Revealed by Methyl Rotating Frame Relaxation Dispersion NMR Spectroscopy. *J. Am. Chem. Soc.* 5718–5727 (2006) doi:10.1021/ja0570279.
10.   Jensen, M. R., Zweckstetter, M., Huang, J. R. & Blackledge, M. Exploring free-energy landscapes of intrinsically disordered proteins at atomic resolution using NMR spectroscopy. *Chemical Reviews* vol. 114 6632–6660 (2014).
11.   Arai, M., Sugase, K., Dyson, H. J. & Wright, P. E. Conformational propensities of intrinsically disordered proteins influence the mechanism of binding and folding. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 9614–9619 (2015).
12.   Keul, N. D. *et al.* The entropic force generated by intrinsically disordered segments tunes protein function. *Nature* vol. 563 584–588 (2018).
13.   Oldfield, C. J. & Dunker, A. K. Intrinsically Disordered Proteins and Intrinsically Disordered Protein Regions. *Annu. Rev. Biochem.* **83**, 553–584 (2014).
14.   Weis, W. I. & Kobilka, B. K. The Molecular Basis of G Protein-Coupled Receptor Activation. *Annual Review of Biochemistry* vol. 87 897–919 (2018).
15.   Zhi, X., Zhou, X. E., Melcher, K. & Xu, H. E. Structures and regulation of non-X orphan nuclear receptors: A retinoid hypothesis. *J. Steroid Biochem. Mol. Biol.* **157**, 27–40 (2016).
16.   Evans, R. M. & Mangelsdorf, D. J. Nuclear receptors, RXR, and the big bang. *Cell* (2014) doi:10.1016/j.cell.2014.03.012.
17.   Markov, G. V. & Laudet, V. Origin and evolution of the ligand-binding ability of nuclear receptors. *Molecular and Cellular Endocrinology* vol. 334 21–30 (2011).
18.   Laudet, V., Auwerx, J., Gustafsson, J.-A. & Walter Wahli. Letter to the Editor A Unified Nomenclature System for the Nuclear Receptor Superfamily. *Cell* **97**, 161–163 (1999).
19.   Bridgham, J. T. *et al.* Protein evolution by molecular tinkering: Diversification of the

nuclear receptor superfamily from a ligand-dependent ancestor. *PLoS Biol.* **8**, (2010).

20. Nuclear Receptors Nomenclature Committee. A Unified Nomenclature System for the Nuclear Receptor Superfamily. *Cell* **97**, 161–163 (1999).

21. Mullican, S. E., DiSpirito, J. R. & Lazar, M. A. The orphan nuclear receptors at their 25-year reunion. *J. Mol. Endocrinol.* **51**, (2013).

22. Lavery, D. N. & McEwan, I. J. Structure and function of steroid receptor AF1 transactivation domains: Induction of active conformations. *Biochemical Journal* vol. 391 449–464 (2005).

23. Sever, R. & Glass, C. K. Signaling by nuclear receptors. *Cold Spring Harb. Perspect. Biol.* **5**, (2013).

24. Weikum, E. R., Knuesel, M. T., Ortlund, E. A. & Yamamoto, K. R. Glucocorticoid receptor control of transcription: Precision and plasticity via allostery. *Nature Reviews Molecular Cell Biology* vol. 18 159–174 (2017).

25. McEwen, B. S. *et al.* Mechanisms of stress in the brain. *Nat. Neurosci.* **18**, 1353–1363 (2015).

26. Itoh, T. *et al.* Structural basis for the activation of PPARγ by oxidized fatty acids. *Nat. Struct. Mol. Biol.* **15**, 924–931 (2008).

27. Rhinn, M. & Dollé, P. Retinoic acid signalling during development. *Development* **139**, 843–858 (2012).

28. Ruggiero, C., Doghman, M. & Lalli, E. How genomic studies have improved our understanding of the mechanisms of transcriptional regulation by NR5A nuclear receptors. *Mol. Cell. Endocrinol.* **408**, 138–144 (2015).

29. Berger, J. P., Akiyama, T. E. & Meinke, P. T. PPARs: Therapeutic targets for metabolic disease. *Trends in Pharmacological Sciences* vol. 26 244–251 (2005).

30. Santos, R. *et al.* A comprehensive map of molecular drug targets. *Nat. Rev. Drug Discov.* **16**, 19–34 (2016).

31. Helsen, C. & Claessens, F. Looking at nuclear receptors from a new angle. *Mol. Cell. Endocrinol.* **382**, 97–106 (2014).

32. Shang, J. & Kojetin, D. J. Structural mechanism underlying ligand binding and activation of PPARγ. *Structure* **29**, 940-950.e4 (2021).

33. Heidari, Z. *et al.* Definition of functionally and structurally distinct repressive states in the nuclear receptor PPARγ. *Nat. Commun.* **10**, (2019).

34. Brzozowski, A. M. *et al.* Molecular basis of agonism and antagonism in the oestrogen receptor. *Nature* **389**, 753–758 (1997).

35. Bourguet, W. *et al.* Crystal structure of a heterodimeric complex of RAR and RXR ligand-binding domains. *Mol. Cell* **5**, 289–298 (2000).

36. Egea, P. F. Crystal structure of the human RXRalpha ligand-binding domain bound to its natural ligand: 9-cis retinoic acid. *EMBO J.* **19**, 2592–2601 (2000).

37. Gronemeyer, H., Gustafsson, J.-A. & Laudet, V. Principles for modulation of the nuclear receptor superfamily. *Nat. Rev. Drug Discov.* **3**, 950–964 (2004).

38. Kojetin, D. J. *et al.* Structural mechanism for signal transduction in RXR nuclear receptor heterodimers. *Nat. Commun.* **6**, 8013 (2015).

39. Zhang, J., Simisky, J., Tsai, F. T. F. & Geller, D. S. A critical role of helix 3-helix 5 interaction in steroid hormone receptor function. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 2707–2712 (2005).

40. Scheepstra, M. *et al.* Ligand Dependent Switch from RXR Homo- to RXR-NURR1

Heterodimerization. *ACS Chem. Neurosci.* **8**, 2065–2077 (2017).

41. Puigserver, P. *et al.* Insulin-regulated hepatic gluconeogenesis through FOXO1-PGC-1α interaction. *Nature* **423**, 550–555 (2003).

42. Flaig, R., Greschik, H., Peluso-Iltis, C. & Moras, D. Structural basis for the cell-specific activities of the NGFI-B and the Nurr1 ligand-binding domain. *J. Biol. Chem.* **280**, 19250–19258 (2005).

43. de Vera, I. M. S. *et al.* Defining a Canonical Ligand-Binding Pocket in the Orphan Nuclear Receptor Nurr1. *Structure* **27**, 66-77.e5 (2019).

44. Belorusova, A. Y. *et al.* Solution Behavior of the Intrinsically Disordered N-Terminal Domain of Retinoid X Receptor ?? in the Context of the Full-Length Protein. *Biochemistry* **55**, 1741–1748 (2016).

45. McEwan, I. J., Lavery, D., Fischer, K. & Watt, K. Natural disordered sequences in the amino terminal domain of nuclear receptors: lessons from the androgen and glucocorticoid receptors. *Nucl. Recept. Signal.* **5**, (2007).

46. Simons, S. S., Edwards, D. P. & Kumar, R. Minireview: Dynamic Structures of Nuclear Hormone Receptors: New Promises and Challenges. *Mol. Endocrinol.* **28**, 173–182 (2014).

47. Belorusova, A. Y. *et al.* Solution Behavior of the Intrinsically Disordered N-Terminal Domain of Retinoid X Receptor α in the Context of the Full-Length Protein. *Biochemistry* **55**, 1741–1748 (2016).

48. De Mol, E. *et al.* Regulation of Androgen Receptor Activity by Transient Interactions of Its Transactivation Domain with General Transcription Regulators. *Structure* (2018) doi:10.1016/j.str.2017.11.007.

49. Simons, S. S. & Kumar, R. Variable steroid receptor responses: Intrinsically disordered AF1 is the key. *Molecular and Cellular Endocrinology* vol. 376 81–84 (2013).

50. Chandra, V. *et al.* Multidomain integration in the structure of the HNF-4α nuclear receptor complex. *Nature* **495**, 394–398 (2013).

51. Chandra, V. *et al.* Structure of the intact PPAR-γ-RXR-α nuclear receptor complex on DNA. *Nature* **456**, 350–356 (2008).

52. Yu, X. *et al.* Structural Insights of Transcriptionally Active, Full-Length Androgen Receptor Coactivator Complexes. *Mol. Cell* **79**, 812-823.e4 (2020).

53. Tsafou, K., Tiwari, P. B., Forman-Kay, J. D., Metallo, S. J. & Toretsky, J. A. Targeting Intrinsically Disordered Transcription Factors: Changing the Paradigm. *Journal of Molecular Biology* vol. 430 2321–2341 (2018).

54. Shrinivas, K. *et al.* Enhancer Features that Drive Formation of Transcriptional Condensates. *Mol. Cell* **75**, 549-561.e7 (2019).

55. McKenna, N. J. & O'Malley, B. W. Combinatorial control of gene expression by nuclear receptors and coregulators. *Cell* vol. 108 465–474 (2002).

56. Yao, T. P., Ku, G., Zhou, N., Scully, R. & Livingston, D. M. The nuclear hormone receptor coactivator SRC-1 is a specific target of p300. *Proc. Natl. Acad. Sci. U. S. A.* **93**, 10626–10631 (1996).

57. Zhang, H. *et al.* Differential gene regulation by the SRC family of coactivators. *Genes Dev.* **18**, 1753–1765 (2004).

58. Gao, Z. *et al.* Coactivators and Corepressors of NF-κB in IκBα Gene Promoter. *J. Biol. Chem.* **280**, 21091–21098 (2005).

59. Lee, S. K., Kim, H. J., Kim, J. W. & Lee, J. W. Steroid receptor coactivator-1 and its

family members differentially regulate transactivation by the tumor suppressor protein p53. *Mol. Endocrinol.* **13**, 1924–1933 (1999).

60. Razeto, A. *et al.* Structure of the NCoA-1/SRC-1 PAS-B Domain Bound to the LXXLL Motif of the STAT6 Transactivation Domain. *J. Mol. Biol.* **336**, 319–329 (2004).

61. Xu, J., Wu, R. C. & O'Malley, B. W. Normal and cancer-related functions of the p160 steroid receptor co-activator (SRC) family. *Nature Reviews Cancer* vol. 9 615–630 (2009).

62. Gu, Y. Z., Hogenesch, J. B. & Bradfield, C. A. The PAS superfamily: Sensors of environmental and developmental signals. *Annual Review of Pharmacology and Toxicology* vol. 40 519–561 (2000).

63. Russo, L., Giller, K., Pfitzner, E., Griesinger, C. & Becker, S. Insight into the molecular recognition mechanism of the coactivator NCoA1 by STAT6. *Sci. Rep.* (2017) doi:10.1038/s41598-017-17088-5.

64. Partch, C. L. & Gardner, K. H. Coactivator recruitment: A new role for PAS domains in transcriptional regulation by the bHLH-PAS family. *Journal of Cellular Physiology* vol. 223 553–557 (2010).

65. Kewley, R. J., Whitelaw, M. L. & Chapman-Smith, A. The mammalian basic helix-loop-helix/PAS family of transcriptional regulators. *International Journal of Biochemistry and Cell Biology* vol. 36 189–204 (2004).

66. Lodrini, M. *et al.* p160/SRC/NCoA coactivators form complexes via specific interaction of their PAS-B domain with the CID/AD1 domain. *Nucleic Acids Res.* (2008) doi:10.1093/nar/gkn029.

67. Beischlag, T. V. *et al.* Recruitment of Thyroid Hormone Receptor/Retinoblastoma-interacting Protein 230 by the Aryl Hydrocarbon Receptor Nuclear Translocator Is Required for the Transcriptional Response to Both Dioxin and Hypoxia. *J. Biol. Chem.* **279**, 54620–54628 (2004).

68. Heffer, A., Grubbs, N., Mahaffey, J. & Pick, L. The evolving role of the orphan nuclear receptor ftz-f1, a pair-rule segmentation gene. *Evol. Dev.* **15**, 406–417 (2013).

69. Yu, Y. *et al.* The nuclear hormone receptor Ftz-F1 is a cofactor for the Drosophila homeodomain protein Ftz. *Nature* **385**, 552–555 (1997).

70. Yoo, J. *et al.* Crystal structure of Fushi tarazu factor 1 ligand binding domain/Fushi tarazu peptide complex identifies new class of nuclear receptors. *J. Biol. Chem.* **286**, 31225–31231 (2011).

71. Wang, W. *et al.* The crystal structures of human steroidogenic factor-1 and liver receptor homologue-1. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 7505–7510 (2005).

72. Sablin, E. P., Krylova, I. N., Fletterick, R. J. & Ingraham, H. A. Structural basis for ligand-independent activation of the orphan nuclear receptor LRH-1. *Mol. Cell* **11**, 1575–1585 (2003).

73. Krylova, I. N. *et al.* Structural analyses reveal phosphatidyl inositols as ligands for the NR5 orphan receptors SF-1 and LRH-1. *Cell* **120**, 343–355 (2005).

74. Ortlund, E. a *et al.* Modulation of human nuclear receptor LRH-1 activity by phospholipids and SHP. *Nat. Struct. Mol. Biol.* **12**, 357–363 (2005).

75. Eschenfeldt, W. H., Lucy, S., Millard, C. S., Joachimiak, A. & Mark, I. D. A family of LIC vectors for high-throughput cloning and purification of proteins. *Methods Mol. Biol.* **498**, 105–115 (2009).

76. Palmer, A. G., Kroenke, C. D. & Loria, J. P. Nuclear magnetic resonance methods for

quantifying microsecond-to-millisecond motions in biological macromolecules. *Methods Enzymol.* (2001).

77. Loria, J. P., Rance, M. & Palmer, A. G. A Relaxation-Compensated Carr - Purcell - Meiboom - Gill Sequence for Characterizing Chemical Exchange by NMR Spectroscopy. *J. Am. Chem. Soc.* **121**, 2331–2332 (1999).

78. Lee, W., Tonelli, M. & Markley, J. L. NMRFAM-SPARKY: Enhanced software for biomolecular NMR spectroscopy. *Bioinformatics* **31**, 1325–1327 (2015).

79. Bieri, M. & Gooley, P. R. Automated NMR relaxation dispersion data analysis using NESSY. *BMC Bioinformatics* **12**, 421 (2011).

80. Ferrage, F., Piserchio, A., Cowburn, D. & Ghose, R. On the measurement of 15N–{1H} nuclear Overhauser effects. *J. Magn. Reson.* **192**, 302–313 (2008).

81. Palmer, A. G. Probing molecular motion by NMR. *Curr. Opin. Struct. Biol.* **7**, 732–737 (1997).

82. Stone, M. J. *et al.* Backbone dynamics of the Bacillus subtilis glucose permease IIA domain determined from 15N NMR relaxation measurements. *Biochemistry* **31**, 4394–4406 (1992).

83. Schymkowitz, J. *et al.* The FoldX web server: An online force field. *Nucleic Acids Res.* **33**, 382–388 (2005).

84. Michiels, P. *et al.* Assignment of the orphan nuclear receptor Nurr1 by NMR. *Biomol. NMR Assign.* **4**, 101–105 (2010).

85. Hughes, T. S. *et al.* Ligand and receptor dynamics contribute to the mechanism of graded PPARγ agonism. *Structure* **20**, 139–150 (2012).

86. Boehr, D. D., Mcelheny, D., Dyson, H. J. & Wright, P. E. *The Dynamic Energy Landscape of Dihydrofolate Reductase Catalysis*. https://www.science.org.

87. Musille, P. M., Kossmann, B. R., Kohn, J. A., Ivanov, I. & Ortlund, E. A. Unexpected allosteric network contributes to LRH-1 co-regulator selectivity. *J. Biol. Chem.* **291**, 1411–1426 (2016).

88. Sablin, E. P. *et al.* Structure of Liver Receptor Homolog-1 (NR5A2) with PIP3 hormone bound in the ligand binding pocket. *J. Struct. Biol.* **192**, 342–348 (2015).

89. Blind, R. D. *et al.* The signaling phospholipid PIP$_3$ creates a new interaction surface on the nuclear receptor SF-1. *Proc. Natl. Acad. Sci.* **111**, 15054–15059 (2014).

90. Sablin, E. P. *et al.* Structure of SF-1 bound by different phospholipids: evidence for regulatory ligands. *Mol. Endocrinol.* **23**, 25–34 (2009).

91. Martínez, L., Webb, P., Polikarpov, I. & Skaf, M. S. Molecular dynamics simulations of ligand dissociation from thyroid hormone receptors: Evidence of the likeliest escape pathway and its implications for the design of novel ligands. *J. Med. Chem.* **49**, 23–26 (2006).

92. Genest, D. *et al.* Ligand-escape pathways from the ligand-binding domain of PPARγ receptor as probed by molecular dynamics simulations. *Eur. Biophys. J.* **37**, 369–379 (2008).

93. Musille, P. M. *et al.* Antidiabetic phospholipid–nuclear receptor complex reveals the mechanism for phospholipid-driven gene regulation. *Nat. Struct. Mol. Biol.* **19**, 532–537 (2012).

94. Mangelsdorf, D. J. *et al.* The nuclear receptor superfamily: The second decade. *Cell* **83**, 835–839 (1995).

95. Chambon, P. The nuclear receptor superfamily: A personal retrospect on the first two

decades. *Molecular Endocrinology* vol. 19 (2005).

96. McKenna, N. J. *et al.* Minireview: Evolution of NURSA, the Nuclear Receptor Signaling Atlas. *Molecular Endocrinology* vol. 23 740–746 (2009).

97. Xu, W. Nuclear receptor coactivators: The key to unlock chromatin. in *Biochemistry and Cell Biology* vol. 83 418–428 (2005).

98. Lonard, D. M. & O'Malley, B. W. The Expanding Cosmos of Nuclear Receptor Coactivators. *Cell* vol. 125 411–414 (2006).

99. York, B. & O'Malley, B. W. Steroid Receptor Coactivator (SRC) family: Masters of systems biology. *Journal of Biological Chemistry* vol. 285 (2010).

100. Huang, P., Chandra, V. & Rastinejad, F. Structural overview of the nuclear receptor superfamily: insights into physiology and therapeutics. *Annu. Rev. Physiol.* **72**, 247–72 (2010).

101. Khorasanizadeh, S. & Rastinejad, F. Visualizing the architectures and interactions of nuclear receptors. *Endocrinology* vol. 157 4212–4221 (2016).

102. Wansa, K. D. S. A., Harris, J. M., Yan, G., Ordentlich, P. & Muscat, G. E. O. The AF-1 Domain of the Orphan Nuclear Receptor NOR-1 Mediates Trans-activation, Coactivator Recruitment, and Activation by the Purine Anti-metabolite 6-Mercaptopurine. *J. Biol. Chem.* **278**, 24776–24790 (2003).

103. Senali Abayratna Wansa, K. D., Harris, J. M. & Muscat, G. E. O. The activation function-1 domain of Nur77/NR4A1 mediates trans-activation, cell specificity, and coactivator recruitment. *J. Biol. Chem.* **277**, 33001–33011 (2002).

104. Chamberlain, N. L., Whitacre, D. C. & Miesfeld, R. L. Delineation of two distinct type 1 activation functions in the androgen receptor amino-terminal domain. *J. Biol. Chem.* **271**, 26772–26778 (1996).

105. De Mol, E. *et al.* EPI-001, A Compound Active against Castration-Resistant Prostate Cancer, Targets Transactivation Unit 5 of the Androgen Receptor. *ACS Chem. Biol.* **11**, 2499–2505 (2016).

106. Yi, P. *et al.* Structure of a Biologically Active Estrogen Receptor-Coactivator Complex on DNA. *Mol. Cell* **57**, 1047–1058 (2015).

107. Castro, D. S., Arvidsson, M., Bolin, M. B. & Perlmann, T. Activity of the Nurr1 carboxyl-terminal domain depends on cell type and integrity of the activation function 2. *J. Biol. Chem.* **274**, 37483–37490 (1999).

108. Khan, S. H. *et al.* Binding of the N-terminal region of coactivator TIF2 to the intrinsically disordered AF1 domain of the glucocorticoid receptor is accompanied by conformational reorganizations. *J. Biol. Chem.* **287**, 44546–44560 (2012).

109. Alen, P., Claessens, † Frank, Verhoeven, G., Rombauts, W. & Peeters, B. *The Androgen Receptor Amino-Terminal Domain Plays a Key Role in p160 Coactivator-Stimulated Gene Transcription. MOLECULAR AND CELLULAR BIOLOGY* vol. 19 (1999).

110. Ohkura, N., Hijikuro, M., Yamamoto, A. & Miki, K. Molecular cloning of a novel thyroid/steroid receptor superfamily gene from cultured rat neuronal cells. *Biochemical and Biophysical Research Communications* vol. 205 1959–1965 (1994).

111. Law, S. W., Conneely, O. M., DeMayo, F. J. & O'Malley, B. W. Identification of a new brain-specific transcription factor, nurr1. *Mol. Endocrinol.* **6**, 2129–2135 (1992).

112. Jakaria, M. *et al.* Molecular Insights into NR4A2(Nurr1): an Emerging Target for Neuroprotective Therapy Against Neuroinflammation and Neuronal Cell Death. *Molecular Neurobiology* vol. 56 5799–5814 (2019).

113. Hegarty, S. V., Sullivan, A. M. & O'Keeffe, G. W. Midbrain dopaminergic neurons: A review of the molecular circuitry that regulates their development. *Developmental Biology* vol. 379 123–138 (2013).

114. Kurakula, K., Koenis, D. S., van Tiel, C. M. & de Vries, C. J. M. NR4A nuclear receptors are orphans but not lonesome. *Biochimica et Biophysica Acta - Molecular Cell Research* (2014) doi:10.1016/j.bbamcr.2014.06.010.

115. Maira, M., Martens, C., Batsché, É., Gauthier, Y. & Drouin, J. Dimer-Specific Potentiation of NGFI-B (Nur77) Transcriptional Activity by the Protein Kinase A Pathway and AF-1-Dependent Coactivator Recruitment. *Mol. Cell. Biol.* **23**, 763–776 (2003).

116. Zor, T., Mayr, B. M., Jane Dyson, H., Montminy, M. R. & Wright, P. E. Roles of phosphorylation and helix propensity in the binding of the KIX domain of CREB-binding protein by constitutive (c-Myb) and inducible (CREB) activators. *J. Biol. Chem.* **277**, 42241–42248 (2002).

117. Wright, P. E. & Dyson, H. J. Intrinsically disordered proteins in cellular signalling and regulation. *Nature Reviews Molecular Cell Biology* vol. 16 18–29 (2015).

118. Lee, Y. *et al.* Targeted Inhibition of the NCOA1/STAT6 Protein-Protein Interaction. *J. Am. Chem. Soc.* **139**, 16056–16059 (2017).

119. Lee, W. & Markley, J. L. PINE-SPARKY.2 for automated NMR-based protein structure research. *Bioinformatics* **34**, 1586–1588 (2018).

120. Williamson, M. P. Using chemical shift perturbation to characterise ligand binding. *Progress in Nuclear Magnetic Resonance Spectroscopy* (2013) doi:10.1016/j.pnmrs.2013.02.001.

121. Turnbull, W. B. & Daranas, A. H. On the Value of c: Can Low Affinity Systems Be Studied by Isothermal Titration Calorimetry? *J. Am. Chem. Soc.* **125**, 14859–14866 (2003).

122. Sievers, F. *et al.* Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* **7**, 539 (2011).

123. Waterhouse, A. M., Procter, J. B., Martin, D. M. A., Clamp, M. & Barton, G. J. Jalview Version 2-A multiple sequence alignment editor and analysis workbench. *Bioinformatics* **25**, 1189–1191 (2009).

124. Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).

125. Xie, T., Zmyslowski, A. M., Zhang, Y. & Radhakrishnan, I. Structural Basis for Multi-specificity of MRG Domains. *Structure* **23**, 1049–1057 (2015).

126. Wang, F. *et al.* Structures of KIX domain of CBP in complex with two FOXO3a transactivation domains reveal promiscuity and plasticity in coactivator recruitment. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 6078–6083 (2012).

127. Currie, S. L. *et al.* ETV4 and AP1 Transcription Factors Form Multivalent Interactions with three Sites on the MED25 Activator-Interacting Domain. *J. Mol. Biol.* **429**, 2975–2995 (2017).

128. Pacheco, D. *et al.* Transcription Activation Domains of the Yeast Factors Met4 and Ino2: Tandem Activation Domains with Properties Similar to the Yeast Gcn4 Activator. *Mol. Cell. Biol.* **38**, (2018).

129. Evans, R. *et al. Protein complex prediction with AlphaFold-Multimer.* https://doi.org/10.1101/2021.10.04.463034 (2021) doi:10.1101/2021.10.04.463034.

130. Pettersen, E. F. *et al.* UCSF ChimeraX: Structure visualization for researchers, educators, and developers. *Protein Sci.* **30**, 70–82 (2021).

131. Varadi, M. *et al.* AlphaFold Protein Structure Database: Massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Res.* **50**, (2022).

132. Salerno, W. J., Seaver, S. M., Armstrong, B. R. & Radhakrishnan, I. MONSTER: Inferring non-covalent interactions in macromolecular structures from atomic coordinate data. *Nucleic Acids Res.* **32**, (2004).

133. Tuttle, L. M. *et al.* Gcn4-Mediator Specificity Is Mediated by a Large and Dynamic Fuzzy Protein-Protein Complex. *Cell Rep.* **22**, (2018).

134. Erlendsson, S. & Teilum, K. Binding Revisited—Avidity in Cellular Function and Signaling. *Frontiers in Molecular Biosciences* vol. 7 (2021).

135. Trojanowski, J., Frank, L., Rademacher, A., Grigaitis, P. & Rippe, K. Transcription activation is enhanced by multivalent interactions independent of liquid-liquid phase separation. *bioRxiv* 2021.01.27.428421 (2021) doi:10.1101/2021.01.27.428421.

136. Zhang, T. *et al.* Nurr1 is phosphorylated by ERK2 in vitro and its phosphorylation upregulates tyrosine hydroxylase expression in SH-SY5Y cells. *Neurosci. Lett.* **423**, (2007).

137. Jacobsen, K. X. *et al.* A Nurr1 point mutant, implicated in Parkinson's disease, uncouples ERK1/2-dependent regulation of tyrosine hydroxylase transcription. *Neurobiol. Dis.* **29**, 117–122 (2008).

138. Henry, J. T. & Crosson, S. Ligand-binding PAS domains in a genomic, cellular, and structural context. *Annual Review of Microbiology* vol. 65 261–286 (2011).

139. Staby, L. *et al.* Eukaryotic transcription factors: Paradigms of protein intrinsic disorder. *Biochemical Journal* vol. 474 2509–2532 (2017).

140. Dyson, H. J. & Wright, P. E. NMR illuminates intrinsic disorder. *Curr. Opin. Struct. Biol.* **70**, 44–52 (2021).

141. Partch, C. L., Card, P. B., Amezcua, C. A. & Gardner, K. H. Molecular basis of coiled coil coactivator recruitment by the aryl hydrocarbon receptor nuclear translocator (ARNT). *J. Biol. Chem.* **284**, 15184–15192 (2009).

142. Kim, J. H., Li, H. & Stallcup, M. R. *CoCoA, a Nuclear Receptor Coactivator which Acts through an N-Terminal Activation Domain of p160 Coactivators knockout experiments (Torchia et al as well as in vitro tran-scription systems (Liu et al., 2001) have demonstrated the critical roles of these coactivators in mediating NR function. p160 coactivators also bind to and enhance the. Molecular Cell* vol. 12 (2003).

143. Voegel, J. J., Heine, M. J. S., Zechel, C., Chambon, P. & Gronemeyer, H. TIF2, a 160 kDa transcriptional mediator for the ligand-dependent activation function AF-2 of nuclear receptors. *EMBO J.* **15**, 3667–3675 (1996).

144. Oñate, S. A., Tsai, S. Y., Tsai, M. J. & O'Malley, B. W. Sequence and characterization of a coactivator for the steroid hormone receptor superfamily. *Science (80-.).* **270**, 1354–1357 (1995).

145. Stuffle, E. C., Johnson, M. S. & Watts, K. J. PAS domains in bacterial signal transduction. *Current Opinion in Microbiology* vol. 61 8–15 (2021).

146. Scheuermann, T. H. *et al. Artificial ligand binding within the HIF2 PAS-B domain of the HIF2 transcription factor.* www.pnas.org/cgi/content/full/ (2009).

147. Tian, W., Chen, C., Lei, X., Zhao, J. & Liang, J. CASTp 3.0: computed atlas of surface

topography of proteins. *Nucleic Acids Res.* **46**, W363–W367 (2018).

148. Maira, M., Martens, C., Philips, A. & Drouin, J. *Heterodimerization between Members of the Nur Subfamily of Orphan Nuclear Receptors as a Novel Mechanism for Gene Activation*. *MOLECULAR AND CELLULAR BIOLOGY* vol. 19 (1999).

149. Jiang, L. *et al.* Structural basis of binding of homodimers of the nuclear receptor NR4A2 to selective Nur-responsive DNA elements. *J. Biol. Chem.* **294**, 19795–19803 (2019).

150. Ju, Z. *et al.* Design of PPAR-γ agonist based on algal metabolites and the endogenous ligand 15-deoxy-Δ 12, 14 -prostaglandin J 2. *Eur. J. Med. Chem.* **157**, 1192–1201 (2018).

151. Rajan, S. *et al.* PGE1 and PGA1 bind to Nurr1 and activate its transcriptional function. *Nat. Chem. Biol.* **16**, 876–886 (2020).

152. Ricciotti, E. & FitzGerald, G. A. Prostaglandins and Inflammation. *Arterioscler. Thromb. Vasc. Biol.* **31**, 986–1000 (2011).

153. Figueiredo-Pereira, M. E., Rockwell, P., Schmidt-Glenewinkel, T. & Serrano, P. Neuroinflammation and J2 prostaglandins: linking impairment of the ubiquitin-proteasome pathway and mitochondria to neurodegeneration. *Front. Mol. Neurosci.* **7**, (2015).

154. Jang, Y., Kim, W., Leblanc, P., Kim, C. H. & Kim, K. S. Potent synthetic and endogenous ligands for the adopted orphan nuclear receptor Nurr1. *Experimental and Molecular Medicine* vol. 53 19–29 (2021).

155. Shiraki, T. *et al.* α,β-Unsaturated ketone is a core moiety of natural ligands for covalent binding to peroxisome proliferator-activated receptor γ. *J. Biol. Chem.* **280**, 14145–14153 (2005).

156. Dasgupta, S. & O'malley, B. W. Transcriptional coregulators: Emerging roles of SRC family of coactivators in disease pathology. *Journal of Molecular Endocrinology* vol. 53 R47–R59 (2014).

157. Lonard, D. M. & O'Malley, B. W. Nuclear receptor coregulators: Modulators of pathology and therapeutic targets. *Nature Reviews Endocrinology* vol. 8 598–604 (2012).

158. Reiter, F., Wienerroither, S. & Stark, A. Combinatorial function of transcription factors and cofactors. *Current Opinion in Genetics and Development* vol. 43 73–81 (2017).

159. Rosenbaum, M. I., Clemmensen, L. S., Bredt, D. S., Bettler, B. & Strømgaard, K. Targeting receptor complexes: a new dimension in drug discovery. *Nature Reviews Drug Discovery* (2020) doi:10.1038/s41573-020-0086-4.

160. Pastor, N. & Amero, C. Information flow and protein dynamics: the interplay between nuclear magnetic resonance spectroscopy and molecular dynamics simulations. *Front. Plant Sci.* **6**, 306 (2015).

161. Zhi, X. *et al.* Structural basis for corepressor assembly by the orphan nuclear receptor TLX. *Genes Dev.* **29**, 440–450 (2015).

162. Iyer, A. K. & McCabe, E. R. B. Molecular mechanisms of DAX1 action. *Molecular Genetics and Metabolism* vol. 83 60–73 (2004).

163. Li, Y., Lambert, M. H. & Xu, H. E. Activation of nuclear receptors: A perspective from structural genomics. *Structure* vol. 11 741–746 (2003).

164. Kumar, R. *et al.* TATA box binding protein induces structure in the recombinant glucocorticoid receptor AF1 domain. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 16425–16430 (2004).

165. Khan, S. H., McLaughlin, W. A. & Kumar, R. Site-specific phosphorylation regulates the structure and function of an intrinsically disordered domain of the glucocorticoid receptor.

Sci. Rep. **7**, (2017).

166. Granata, D., Camilloni, C., Vendruscolo, M. & Laio, A. Characterization of the free-energy landscapes of proteins by NMR-guided metadynamics. *Proc. Natl. Acad. Sci.* **110**, 6817–6822 (2013).

167. Troussicot, L., Guilli??re, F., Limongelli, V., Walker, O. & Lancelin, J. M. Funnel-metadynamics and solution NMR to estimate protein-ligand affinities. *J. Am. Chem. Soc.* **137**, 1273–1281 (2015).

168. Niesen, F. H., Berglund, H. & Vedadi, M. The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability. *Nat. Protoc.* **2**, 2212–2221 (2007).

169. Walters, K. A., Simanainen, U. & Handelsman, D. J. Molecular insights into androgen actions in male and female reproductive function from androgen receptor knockout models. *Human Reproduction Update* vol. 16 543–558 (2010).

170. Watson, P. A., Arora, V. K. & Sawyers, C. L. Emerging mechanisms of resistance to androgen receptor inhibitors in prostate cancer. *Nature Reviews Cancer* vol. 15 701–711 (2015).

171. Panel, C. & Genome, H. Androgen-independent prostate cancer Androgen-independent prostate cancer. **1**, 34–45 (2001).

172. Powell, S. M. *et al.* Mechanisms of androgen receptor signalling via steroid receptor coactivator-1 in prostate. *Endocrine-Related Cancer* vol. 11 117–130 (2004).

173. Lavery, D. N. & McEwan, I. J. Structural characterization of the native NH2-terminal transactivation domain of the human androgen receptor: A collapsed disordered conformation underlies structural plasticity and protein-induced folding. *Biochemistry* **47**, 3360–3369 (2008).

174. Jenster, G., Van der Korput, H. A. G. M., Trapman, J. & Brinkmann, A. O. Identification of two transcription activation units in the N-terminal domain of the human androgen receptor. *J. Biol. Chem.* **270**, 7341–7346 (1995).

175. He, B., Kemppainen, J. A. & Wilson, E. M. FXXLF and WXXLF sequences mediate the NH2-terminal interaction with the ligand binding domain of the androgen receptor. *J. Biol. Chem.* **275**, 22986–22994 (2000).

176. Callewaert, L., Van Tilborgh, N. & Claessens, F. Interplay between two hormone-independent activation domains in the androgen receptor. *Cancer Res.* **66**, 543–553 (2006).

177. Claessens, F. *et al.* Diverse roles of androgen receptor (AR) domains in AR-mediated signaling. *Nuclear receptor signaling* vol. 6 (2008).

APPENDIX 1: Interactions of the androgen receptor Tau-1 region with the SRC1 PAS-B domain

Introduction

The role of the androgen receptor in human physiology and disease

The primary function of AR is to regulate the development and maintenance of male specific traits. This includes differentiation and development of reproductive organs in the fetus, pubertal sexual maturation, and the maintenance of bone and muscle mass[169]. However, AR is probably most known for its role in prostate cancer. Like androgen insensitivity disorder, prostate cancer is caused by several different mutations to the AR gene and results in tens of thousands of male deaths every year. Initially, prostate cancer can be treated by androgen deprivation therapy via surgical or chemical castration but if the cancer persists it will progress to a castration-resistant form of the disease that is difficult to treat. One of the factors that is thought to play a role in resistance is the production of AR splice variants that contain only the NTD and DBD which has led to a unique interest in the AR NTD[170–172].

The role of the androgen receptor NTD in gene regulation

The domain architecture of AR is typical of a nuclear receptor, with a LBD and DBD that are structurally homologous to other NRs and a long, largely unstructured NTD[173]. However, early work by Guido et al. showed AR could activate transcription independent of its LBD (a fact that, in retrospect, makes sense in light of the AR NTD-DBD splice variants found in prostate cancer patients), highlighting the unique significance of the NTD in AR[174]. Further work revealed that most coactivator LxxLL motifs could not bind the AR LBD with the same high affinity as other NR LBDs[109]. Instead, the NTD appeared to play a crucial role in recruiting the SRC coactivators.

Additionally, the AR LBD had unique structural features that allowed for binding of an FxxLF motif and in 2000 Wilson et al. showed that an FxxLF motif within the first 30 residues of the AR own NTD bound to the AR LBD and potentiated the LBD's activity in a ligand dependent manner[175].

All these findings suggested that a complex set of interactions, centered around its NTD, facilitate AR's activity. Subsequently, a significant amount of work was done to try and determine the which regions of the NTD and SRC proteins are critical for gene activation and, potentially, interact directly. In 1995 Jenster et al. established that were two regions in the NTD, Tau-1 and Tau-5, that were especially important for transactivation[174]. Further studies have since revolved around this paradigm. Additional work over the next two decades attempted to narrow down and confirm the exact regions involved in AR mediated coactivator recruitment[176]. However, the specific molecular mechanisms that underlie the recruitment of the SRC coactivators by AR remains to be determined.

Given my work showing an interaction between the Nurr1 AF1 and the PAS-B domain of SRC1 as well as the evidence that the AR NTD is involved in direct recruitment of SRC coactivators via their N-terminal region (where PAS-B is located), it seemed reasonable to posit that AR might be interacting with SRC1 via the same mechanism as Nurr1[109]. Furthermore, the core region of Tau-1 appeared to have sequence similarities with a peptide from the transcription factor STAT6 which had been shown to directly bind the PAS-B domain, leading us to specifically hypothesize that the AR Tau-1 also directly with the SRC1 PAS-B domain[60]. The following work sought to test this hypothesis and characterize any interactions that exist between the AR Tau-1 and SRC1 PAS-B domain.

Methods

Protein design, expression, and purification

For study of the SRC1 PAS-B domain, the DNA coding sequence for residues 260-369 of SRC1 was cloned into the pMCSG7 vector[75]. For study of the AR Tau-1 region, the DNA coding sequence residues 161-205 of AR, with an additional tyrosine added to the end of the sequence to allow for measurements of protein concentration, was cloned into the pMCSG23 vector. Prior to use all DNA constructs were sequenced to verify identity. Constructs were transformed into *E. coli* BL21 (DE3) cells which were grown at 37 °C until they reached an O.D. of 0.6, at which time they were moved to 16 °C, induced with 1mM IPTG, and proteins were expressed overnight. The next day cells were pelleted and were stored at -80 °C until use.

SRC1 PAS-B was purified in the following manner. Cells pellets from 0.5 L or 1 L cultures were resuspended in 40 ml of lysis buffer (300 mM NaCl, 50 mM HEPES, 1 mM TCEP, pH 7.5) and sonicated for 25 minutes. The resulting cell lysate was then clarified by centrifugation at 12,000 rpm for 20 min and the soluble supernatant was put on to a column containing 1-2 ml of Ni2+-NTA resin (Qiagen) and incubated at 4 °C for one hour. The supernatant was then allowed to flow through the column and two 10ml washes were performed with both high-salt wash buffer (lysis buffer + 300 mM NaCl, 10 mM imidazole) and wash buffer (lysis buffer + 10 mM imidazole). The protein was then eluted in four 4ml fractions with an elution buffer (lysis buffer + 275 mM imidazole) and the purity of the fractions was analyzed by running SDS-PAGE gels. Clean fractions were combined and were put through two rounds of dialysis at room temperature overnight to remove imidazole. The His6-tag was cleaved during dialysis using 1:25 (w/w) TEV protease. After dialysis, cleaved PAS-B protein was immediately put through HPLC following protein-containing fractions were lyophilized and stored for future use. When ready for SRC1

PAS-B was resuspended in the desired buffer containing also 6 M guanidine hydrochloride and then put through two rounds of dialysis overnight at 4 °C to remove the guanidine hydrochloride. AR Tau-1 was purified in the following manner. Cells pellets from 1 L cultures were resuspended in 40 ml of lysis buffer (300 mM NaCl, 50 mM Tris, 1 mM TCEP, pH 8) and sonicated for 25 minutes. The resulting cell lysate was then clarified by centrifugation at 12,000 rpm for 20 min and the soluble supernatant was put on to a column containing 1-2 ml of Ni^{2+}-NTA resin (Qiagen) and incubated at 4 °C for one hour. The supernatant was then allowed to flow through the column and two 10 ml washes were performed with both high-salt wash buffer (lysis buffer + 300 mM NaCl, 10 mM imidazole) and wash buffer (lysis buffer + 10 mM imidazole). The protein was then eluted in four 4 ml fractions with an elution buffer (lysis buffer + 275 mM imidazole) and the purity of the fractions was analyzed by running SDS-PAGE gels. Clean fractions were combined and were put through two rounds of dialysis at room temperature overnight to remove imidazole. The His6-tag was cleaved after dialysis using 1:25 (w/w) TEV protease in conical so as not to lose low molecular weight peptide. After cleavage the Ni^{2+}-NTA resin used in the purification was cleaned using 6 M Gdn-HCl, 500 mM imidazole, and 0.1 % Triton-X and regenerated with Nickel Sulfate. The sample was put back through the cleaned resin to remove MBP and any uncleaved protein. For this, urea and NaCl were added to the cleaved Tau-1 sample to final concentrations of 500 mM and 300 mM, respectively, and was incubated with cleaned resin for ~25 minutes. The sample was collected as flow through and the resin was washed with 8 ml of dialysis buffer with urea and NaCl added to final concentrations of 500 mM and 300 mM. After subtractive purification, Tau-1 was put through HPLC following protein-containing fractions were lyophilized and stored for future use. When ready for use, Tau-1 was resuspended in the desired buffer.

NMR data collection and analysis

NMR data were acquired a Bruker Neo 600 MHz spectrometer equipped with a QCI-F cryoprobe

at temperatures ranging from 25-30 °C. NMR data were processed using Topspin and NMRpipe

and analyzed using NMRFAM-SPARKY[78]. 2D 1H-15N HSQC spectra were collected for peptide

titrations. Chemical shift perturbations were calculated in R using the equation

$$CSP = \left(1/2\left[\delta_H^2 + \left(0.04(\delta_N^2)\right)\right]\right)^{1/2}$$

in which δ_H and δ_N are the proton and nitrogen chemical shifts, respectively. The CSPs for each

residue were then mapped onto structures for analysis using Chimera. For backbone assignments,

2D 1H-15N HSQC, 3D HNCA, HNCACB, CBCA(CO)NH, and HNCO experiments were

collected for both AR Tau-1. Backbone resonances were assigned using a mixture of automated

assignments from the I-PINE web server and assignments made by hand. Secondary structure

predictions for the AR Tau-1 were generated by running the NMRFAM TALOS applet in

SPARKY. Assignments from 2D 1H-15N HSQC, 3D HNCA, HNCACB, CBCA(CO)NH, and

HNCO experiments were used as inputs. To measure the affinity of PAS-B and the Tau-1 peptide,

CSPs were first measured for the most perturbed peaks. These were then individually plotted as a

function of concentration and fit to the following equation

$$\Delta\delta_{obs} = \frac{\Delta\delta_{max}\left\{([P]_t + [L]_t + K_d) - [([P]_t + [L]_t + K_d)^2 - 4[P]_t[L]_t]^{\frac{1}{2}}\right\}}{2[P]_t}$$

in which $[P]_t$ is the total concentration of protein, $[L]_t$ is the total concentration of ligand, $\Delta\delta_{obs}$ is

difference in the chemical shift from the free to the observed state, and $\Delta\delta_{max}$ is the maximum

observed shift upon saturation[120]. Data was fit, analyzed, and plotted using a python script written

by myself.

Results

AR Tau-1 forms a moderate affinity complex with SRC1 PAS-B

To test the hypothesis that Tau-1 could bind PAS-B, NMR titrations were performed with the two

proteins. Initially, constructs of Tau-1 (161-205) and PAS-B (260-369) were designed and purified

from *E. coli*, with PAS-B was purified as described previously and Tau-1 was purified as an MBP-

tagged fusion to prevent degradation. For these titrations unlabeled Tau-1 was titrated into ^{15}N

labeled PAS-B and the resulting HSQC spectrum was collected and compared with apo ^{15}N-PAS-

B. Significant chemical shift perturbations were seen upon addition of Tau-1, confirming that Tau-

1 can indeed bind PAS-B. Multi-point titrations showed that peaks were shifting towards the bound

state, indicating that binding was in fast-exchange and likely had binding constant in micromolar

range (Ap. Fig. 1.1A).
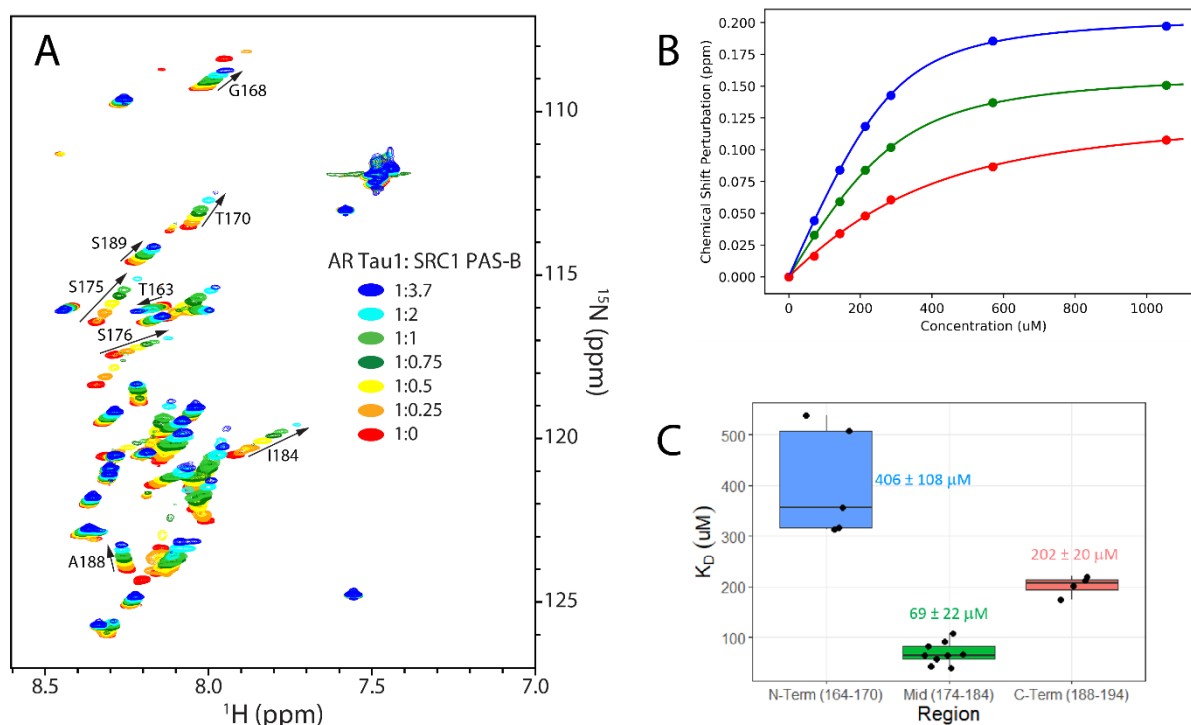
Appendix Figure 1.1 Assessment of the Tau-1 and PAS-B interaction by NMR.
(A) Overlaid spectra from the titration of SRC1 PAS-B into ^{15}N ^{13}C labeled AR Tau-1 with peaks shifting from red to blue as the concentration of the Tau-1 peptide increases. The direction of peak movements is shown for select residues with arrows (B) CSPs from peaks of select residues plotted as a function of increasing concentration of PAS-B during the titration. Solid lines represent fits that were used to determine K_d values. (C) Scatterplot of fitted K_d values for individual residues with quartile values for each group shown via a boxplot. Residues are grouped as being N-terminal (blue), in the middle (green), or C-terminal (red) based on their location in the Tau-1 sequence.

Since micromolar binding can be to low affinity for accurate measurements by ITC, the affinity of the complex was assessed using an extensive multi-point NMR titration. In this case, 15N labeled Tau-1 was titrated into unlabeled PAS-B, with HSQC spectra collected for each titration point. CSPs were calculated for the most perturbed peaks and fit as a function of increasing concentration to obtain Kd values for individual residues (Ap. Fig. 1.1B). Interestingly, these produced a range of Kd values that grouped into three distinct clusters based on the residues' location in the Tau-1 peptide (Ap. Fig. 1.1C). The N-terminal residues (164-170) had an average K_d of ~ 350 μM, the middle of the peptide (174-184) had an average K_d of ~60μM, and C-terminal region had an average K_d of ~200 μM. This could indicate that the peptide has multiple stages of binding, but more in-depth biophysical analysis would need to be done to determine the cause of this variation. Averaging of the individual K_d values for the most perturbed resonances located in the center of the peptide produced a collective K_d of 68.6 (±22.4) μM, indicating that this is a moderate affinity interaction.

AR Tau-1 binds a conserved surface of PAS-B, likely as a helix

To determine the binding surface on PAS-B for Tau-1 CSPs from titrations with unlabeled Tau-1 and ^{15}N labelled PAS-B were mapped onto the PAS-B structure (Ap. Fig. 1.2). The most significant CSPs were found on the second beta strand and second alpha helix of the domain which is the binding site previously characterized for the STAT6 peptide and the same site where I found the Nurr1 peptide bound[60,63]. Furthermore, directly comparing the CSPs mapped onto PAS-B for the Tau-1 and Nurr1 peptides revealed an almost identical pattern, suggesting that they have a similar binding mode.
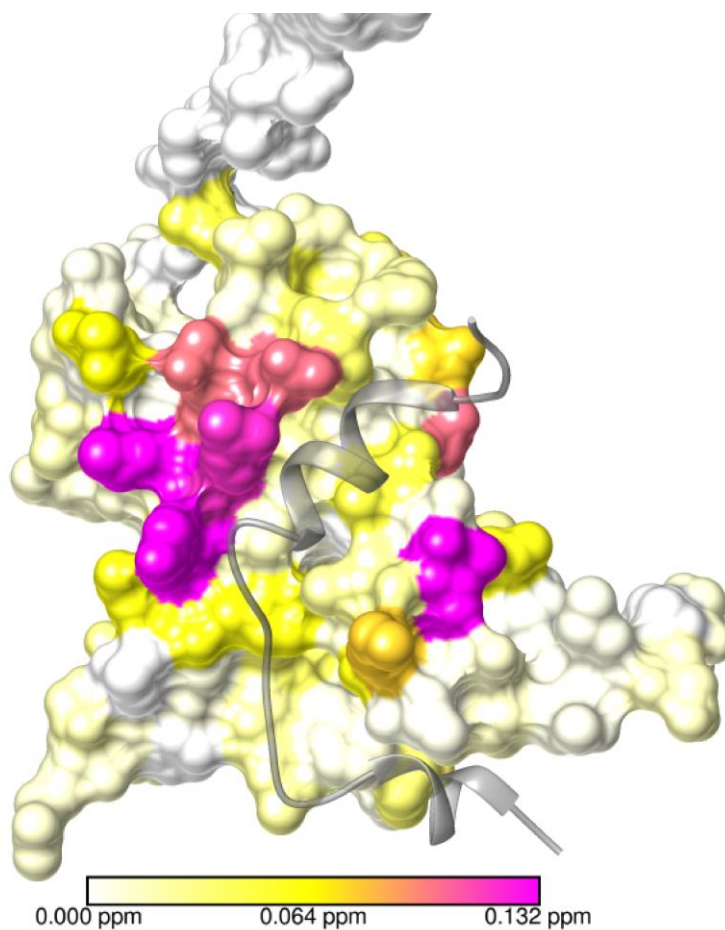
Appendix Figure 1.2 Tau-1 binds a hydrophobic cleft on PAS-B.

CSPs from the PAS-B HSQC spectra from titrations with the AR Tau-1 peptide mapped onto the structure (PDB entry 5NWM)[63] of PAS-B bound to STAT6 (shown in grey). The intermediate and high values in the color key represent $\langle \Delta\delta_{H,N} \rangle + 1\sigma$ (yellow) and $\langle \Delta_{\delta H,N} \rangle + 3\sigma$ (magenta), respectively. Colors are linearly interpolated to reflect the corresponding values.

To further investigate its binding mode, I attempted to use NMR to predict the backbone conformation and region of binding for the Tau-1 peptide. For these experiments, ^{15}N ^{13}C labelled Tau-1 peptide was first purified and the backbone resonances were assigned using standard triple resonance techniques. The resulting chemical shifts were put into the TALOS prediction software to generate secondary structure predictions[78]. This revealed that the N- and C-termini are likely unstructured but that, interestingly, the middle of the peptide has a significant likelihood of forming a helix (Ap. Fig. 1.3). This lends weight to the idea that the Tau-1 peptide binds in a similar manner to the Nurr1 peptide, which is similarly predicted to form a helix. Assignment of the Tau-1 spectrum also allowed the CSPs for labelled Tau-1 to map per residue, revealing that, although almost all the residues in the peptide are perturbed upon binding, the resonances of residues in the center of peptide have the largest shifts. This corresponds with both the region predicted to form a helix and that had the highest calculated affinities, suggesting that the interaction between PAS-B and Tau-1 is driven by a helical region in the center of the peptide which has the strongest affinity for the surface of PAS-B.
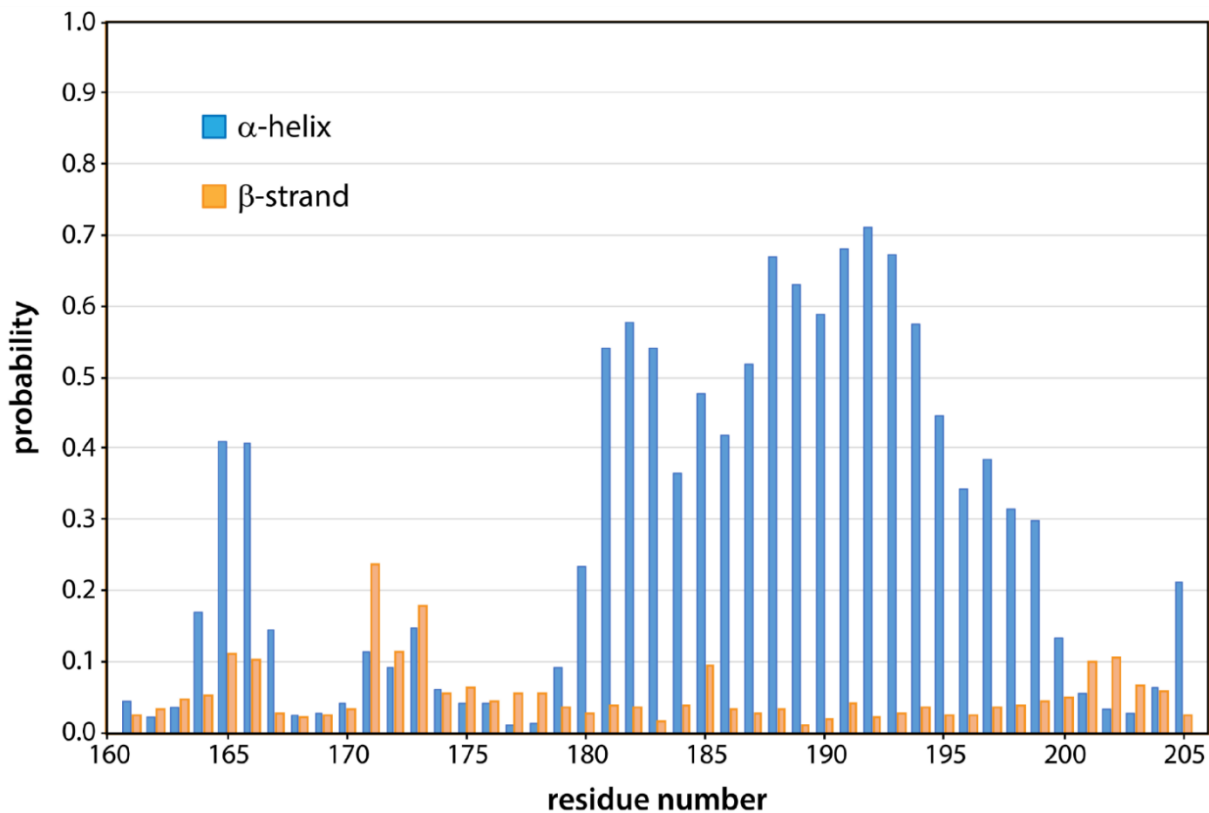
Appendix Figure 1.3 Portions of Tau-1 are predicted to form a helix in solution.
Secondary structure prediction for the apo Tau-1 peptide based on NMR chemical shift data using the TALOS software[78]. Bars indicate the likelihood of forming a helix (blue) or strand (orange).

Conclusions

Despite the physiological and medical significance of the AR, progress has been slow coming in understanding the fundamental, yet surprisingly complicated, molecular mechanisms underlying its function. The Tau-1 region has long been known to play an important part in AR activation but how it actualizes this is still unclear[173,177]. Recruitment of SRC coactivators is a critical step for gene activation by NRs so it seems reasonable to hypothesize that Tau-1 allows for SRC recruitment[172]. This work supports a model in which AR recruits SRC1 via a direct interaction between the SRC1 PAS-B domain and the AR Tau-1 region.

NMR titrations confirmed that Tau-1 can directly bind PAS-B, with CSPs for the assigned peptide showing that the majority of the 161-205 segment is affected by the interaction. Additionally, fitting of CSP data revealed that the center of the peptide is likely the most critical for binding, with the residues in this segment having an average K_d of ~68 μM. This could make sense in light of secondary structure predictions for Tau-1 that suggest this segment may exist, at least partially, as a helix in solution. As was mentioned previously, helix formation of an unstructured region upon binding a structured domain, ostensibly via conformational selection in this case, is a common structural motif. The most pertinent example of this, of course, is the STAT6 peptide which binds to the same location on PAS-B. Interestingly, an alignment of the Tau-1 and STAT6 peptides reveals many sequence identities and similarities, making it reasonable to suppose the two peptides could bind PAS-B in a similar conformation. Notable among these is the presence of an LKDIL segment in Tau-1, which is similar to the LxxLL motif that frequently plays a role in helical peptide binding and further supports the idea of Tau-1 binding in such a conformation[60,175]. Conversely, NMR data of PAS-B during Tau-1 titrations show that Tau-1 binds the highly conserved hydrophobic surface formed by β2 and α2. Along with STAT6, I previously showed

that Nurr1 AF1 domain bound this surface and argued this site represents a general mechanism for SRC1 to recruited by transcription factors. It should be noted that in both the cases further *in vivo* studies need to be done to establish the significance of these interactions. However, if biologically relevant, the binding of Tau-1 further suggests the PAS-B domain is a core part of SRC1's regulatory function.

CURRICULUM VITAE

Nicolas Daffern
Evanston, IL 60201 • (918) 231-5121 • nicolasdaffern2015@u.northwestern.edu

Education
Northwestern University, Evanston, IL
Graduation: April 2022
Doctoral Candidate in the Interdisciplinary Biological Sciences program (IBiS)
GPA: 3.82
Oklahoma State University, Stillwater, OK
Graduation: May 2015
Bachelor of Science in Biochemistry
Minors: Microbiology, Chemistry, Spanish
GPA: 3.55

Research
Radhakrishnan Laboratory, Northwestern University   June 2016 – Present
- Discovered and characterized novel protein-protein and protein-ligand interactions involving nuclear receptor activation domains and the SRC family of transcriptional coregulators
- Studied structural dynamics in Ftz-F1 nuclear receptor ligand-binding domain
- Applied a variety of biochemical and biophysical techniques including NMR spectroscopy, X-ray crystallography, ITC, Thermal Shift Assays, Pulldowns, MALDI/ESI Mass Spectrometry, SEC MALS, Fluorescence Anisotropy
- Cloned and purified a wide variety of wild-type and mutant proteins in *E. coli*
- Performed computational studies such as phylogenetic analysis and molecular dynamics simulation

IBiS Doctoral Program Rotations, Northwestern University   September 2015 – June 2016
- Studied the physical properties of methylated DNA using magnetic tweezers (Marko Laboratory)
- Purified the Type IIIA CRISPR Csm complex for studies by EM (Mondragón Laboratory)
- Crossed *C. elegans* mutant strains to look at phenotype variability (Andersen Laboratory)

Hoff Laboratory, Oklahoma State University   February 2014 – March 2015
- Examined additivity of mutational effects on kinetics in Photoactive Yellow Protein
- Purified mutant proteins and collect data on chromophore pK_a and photocycle kinetics

Additional Training

Trainee, NIGMS Molecular Biophysics Training Program   December 2016 – December 2018
- Participated in advanced courses and hands-on workshops to gain experience with a wide variety of biophysical techniques
- Presented research and interacted with experts in the field at program meetings and symposiums

Argon LS-CAT Data Collection Workshop   March 3rd, 2018
- Hands on training in protein x-ray crystallography data collection and structure determination

Telluride School of Biomolecular Structure and Dynamics   July 23rd – 30th, 2017
- Attended lectures on computational and experimental techniques for studying protein dynamics
- Developed and presented plans in groups for addressing fundamental questions in the field

Teaching and Mentorship

Student Mentoring in Radhakrishnan Lab   October 2016 – Present
- Oversaw projects and taught students experimental techniques
- 4 Northwestern undergraduate students (3 months – 2 years)
    - One student awarded a Beckman Scholarship based on lab work
- 1 High school student (6 months)
    - Selected as finalist for Intel International Science and Engineering Fair based on lab work
- 1 IBiS program doctoral rotation student (3 months)
- 2 Visiting undergraduate students (1 summer each)

Completed Course on Teaching in STEM (Northwestern CIRTL)   June 2019 – August 2019
- Watched virtual lectures and attended weekly group sessions to learn teaching strategies

Teaching Assistant   January 2018 – March 2019
- Bioinformatics class (Biol Sci 323) for Northwestern graduate and undergraduate students
- Gave weekly presentations reviewing lecture materials and assisted students with coding

Leadership and Organizations

Northwestern Biophysics Journal Club   September 2018 – September 2019
- Chair, Organizing Committee

Northwestern Annual Biophysics Symposium   September 2017 – June 2017
- Co-Chair, Symposium Organizing Committee

OSU Biochemistry Club   March 2013 – June 2015
- Member

Wilderness Guide for OSU's *Outdoor Adventure Program*   August 2010 – June 2015
- Led groups of up to 30 participants on multi-day wilderness expeditions

Honors and Awards

Northwestern University Nominee for the HHMI's Gilliam Fellowship (2018)
Robert H. Lurie Cancer Center Travel Award to present at 59th ENC Conference (2018)
Award for Poster (2nd Place) at Northwestern Annual Biophysics Symposium (2017)
Paul F. Kruse Biochemistry and Molecular Biology Award (2015)
Academic Excellence Scholarship (2010-2015)
Dean's Honor Roll (2012-2014)
Humphrey's International Study Scholarship (2013)

Posters and Presentations

Poster – Gateway NMR meeting (virtual). "*Ligand-Independent Regulatory Mechanisms in Nuclear Receptors*" Fall 2020

Podium Presentation – Northwestern University, Annual Biophysics Symposium. *"Molecular Mechanism of Ligand-Independent Activation by Nuclear Receptors"*. Fall 2020

Podium Presentation – Northwestern University, IBiS Annual Retreat. *"Molecular Mechanism of Ligand-Independent Activation by Nuclear Receptors"*. Summer 2019

Podium Presentation – Data Blitz Short Talk, 16th Annual Chicago Biomedical Consortium Symposium. "*Conformation and Dynamics of the Orphan Nuclear Receptor Ftz-F1*". Fall 2018

Poster – 29th Annual Lurie Cancer Center Scientific Posters Session. "*Conformation and Dynamics of the Orphan Nuclear Receptor Ftz-F1*". Summer 2018

Poster – 59th Experimental Nuclear Magnetic Resonance Conference (ENC). "*Conformation and Dynamics of the Orphan Nuclear Receptor Ftz-F1*". Spring 2018

Poster – Northwestern University, 8th Annual Biophysics Symposium. "*Conformation, Dynamics, and Evolution of the NR5A Sub-family of Nuclear Receptors*". Summer 2017

Poster – Oklahoma State University, 12th Annual Research Symposium in Biological Sciences. "*Studies of Additivity in the Function of Photoactive Yellow Protein*". Spring 2015

Publications

Daffern, N., Chen, Z., Zhang, Y., Pick, L., and Radhakrishnan, I. "Solution Nuclear Magnetic Resonance Studies of the Ligand-Binding Domain of an Orphan Nuclear Receptor Reveal a Dynamic Helix in the Ligand-Binding Pocket." *Biochemistry*. 54(13), 1977-1986 (2018).

Daffern, N.*, Nordyke, C.*, Zhang, M.*, Palmer, AG., Straub, J. "Chemical Exchange in a Biphasic Potential: A Tutorial." *The Biophysicist*. In Press (2022) (* denotes equal contributions)

Marcum, RD., Hsieh, J., Giljen, M., Justice, E., **Daffern, N.**, Zhang, Y., Radhakrishnan, I., "A Capped Tudor Domain within a Core Subunit of the Sin3L/Rpd3L Histone Deacetylase Complex Binds to Nucleic Acid G-Quadruplexes." *Journal of Biological Chemistry.* 282(2), (2022)

Manuscripts in Preparation

Daffern, N. and Radhakrishnan, I. "The Nurr1 Activation Function 1 (AF1) Domain Recruits Nuclear Receptor Coactivators via a Novel Molecular Mechanism."

Daffern, N., Kelly, K., Villegas, J., and Radhakrishnan, I. "Characterization of a Novel Ligand-binding Pocket in the SRC1 PAS-B Domain."

Referees

Ishwar Radhakrishnan, i-radhakrishnan@northwestern.edu, Thesis Advisor

Kelly Mayo, k-mayo@northwestern.edu, Chair, Thesis Committee

John Marko, john-marko@northwestern.edu, Member, Thesis Committee