Distribution Agreement

In presenting this thesis or dissertation as a partial fulfillment of the requirements for an advanced degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis or dissertation in whole or in part in all forms of media, now or hereafter known, including display on the world wide web. I understand that I may select some access restrictions as part of the online submission of this thesis or dissertation. I retain all ownership rights to the copyright of the thesis or dissertation. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation.

Signature:

_____ _____

Emily Rye Weikum Date

ALLOSTERIC MODULATION OF NUCLEAR RECEPTOR FUNCTION
By

Emily Rye Weikum
Doctor of Philosophy

Graduate Division of Biological and Biomedical Science
Biochemistry, Cell, and Developmental Biology

_____
Eric A. Ortlund, Ph.D.
Advisor

_____
Xiaodong Cheng, Ph.D.
Committee Member

_____
Graeme L. Conn, Ph.D.
Committee Member

_____
Anita H. Corbett, Ph.D.
Committee Member

_____
Nicholas T. Seyfried, Ph.D.
Committee Member

Accepted:

_____
Lisa A. Tedesco, Ph.D.
Dean of the James T. Laney School of Graduate Studies

_____
Date

ALLOSTERIC MODULATION OF NUCLEAR RECEPTOR FUNCTION

By

Emily Rye Weikum
B. S., Kennesaw State University, 2012

Advisor: Eric A. Ortlund, Ph.D.

An abstract of
A dissertation submitted to the Faculty of the
James T. Laney School of Graduate Studies of Emory University
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy
in Graduate Division of Biological and Biomedical Science
Biochemistry, Cell, and Developmental Biology
2017

ABSTRACT

Allosteric Modulation of Nuclear Receptor Function
By Emily Rye Weikum

Nuclear receptors are a family of ligand-regulated transcription factors that control specific gene programs across numerous biological processes. The assembly of distinct transcriptional complexes drives regulatory specificity, each complex attuned to a particular gene-, cell- and physiologic-context. These distinct complexes are influenced by allosteric effectors, such as DNA sequence and ligands, which modulate nuclear receptor function. These collective works utilize structural biology and biochemistry to examine these allosteric effectors of nuclear receptor function. We explore the idea that different DNA sequences alter nuclear receptor structure. We show that the glucocorticoid receptor can interact directly with a sequence within inflammatory gene promoters. This finding represents a paradigm shift in our understanding of how the glucocorticoid receptor could repress transcription at these sites. We also show the first reported crystal structure of germ cell nuclear factor bound to its DNA response element. This nuclear receptor is critical in development and understanding the DNA binding properties of this protein can gleam insight into its function. In addition to DNA sequences, we also structurally characterize the glucocorticoid receptor ligand-binding domain in complex with a widely used and potent clinical ligand. As there are only a few GR ligand binding domain structures reported, this work provided structural mechanisms driving this highly stabilizing ligand. Furthermore, this work also reports the first glucocorticoid receptor structure in complex with a peptide from the atypical coregulator, small heterodimer partner. Collectively, this work reviews the idea that these allosteric modifications drive different NR surfaces that are read by coregulator proteins, resulting in alternative transcriptional programs.

ALLOSTERIC MODULATION OF NUCLEAR RECEPTOR FUNCTION

By

Emily Rye Weikum
B. S., Kennesaw State University, 2012

Advisor: Eric A. Ortlund, Ph.D.

A dissertation submitted to the Faculty of the
James T. Laney School of Graduate Studies at Emory University
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy
in Biochemistry
2017

Table of Contents

Chapter 4: Structural analysis of the glucocorticoid receptor ligand-binding domain in complex with triamcinolone acetonide and a fragment of the atypical coregulator, SHP

Chapter 6: Discussion

ABBREVIATIONS

AF-1/Tau1/τ1: activator function surface-1

AF-2: activator function surface-1

AncGR2: ancestral glucocorticoid receptor 2

AR: androgen receptor

AP-1: activator protein-1

BP: base pairs

ChIP-seq: chromatin immunoprecipitation followed by deep sequencing

COUP-TF: chicken ovalbumin upstream promoter transcription factors

CTE: C-terminal extension

DAX: dosage-sensitive sex reversal-adrenal hypoplasia congenital critical region on the X chromosome, gene 1

DBD: DNA binding domain

D box: distal box

Dex: dexamethasone

DHS: DNase I hypersensitive sites

DR0: direct repeat binding sequence with no spacer

DSF: differential scanning fluorimetry

ER: estrogen receptors

FAIRE: formaldehyde-assisted isolation of regulatory elements and deep sequencing

Ftz-F1: *fushi tarazu* factor 1 domain

FXR: farnesoid X receptors

GBS: glucocorticoid receptor binding sequence

GC: glucocorticoid

GCNF: germ cell nuclear factor

GR: glucocorticoid receptor

GRE: glucocorticoid response elements

GOR: glucocorticoid receptor occupied region

HATs: histone acetyltransferases

HDACs: histone deacetyltrasnfereases

HDX-MS: hydrogen deuterium exchange mass spectrometry

HNF4: hepatocyte nuclear factor 4

HSP: heat shock protein

IPTG: isopropyl β-D-1-thiogalactopyranoside

IR-GBS: inverted repeat glucocorticoid receptor binding sequence

LBD: ligand binding domain

LBP: ligand binding pocket

LRH-1: liver receptor homolog-1

LXR: liver X receptors

MD: molecular dynamics

MF: mometasone furoate

MR: mineralocorticoid receptor

NCoR: nuclear receptor corepressor-1

NF-κB: nuclear factor kappa-beta

NMR: nuclear magnetic resonance

NR: nuclear receptor

NRS: negative regulatory DNA sequence motifs

NTD: N-terminal domain

OD: optical density

P box: proximal box

PDB: protein data bank

PPAR: peroxisome proliferator activated receptors

PR: progesterone receptor

PTMs: post-translational modifications

RAR: retinoic acid receptors

RE: response elements

REV-ERB: reverse-Erb receptors

SHP: small heterodimer protein

SMRT: nuclear receptor corepressor-2

RMSD: root mean squared deviation

ROR: retinoic acid related receptors

RXR: retinoid X receptors

SF-1: steroidogenic factor-1

SGRMs: selective glucocorticoid receptor modulators

SRC: steroid receptor coregulator

TA: triamcinolone acetonide

TALENs: transcription activator-like effector nucleases

Tau2/τ2: trans-activating domain 2

TF: transcription factor

TR: thyroid hormone receptor

TRF: transcriptional regulatory factor

VDR: vitamin D receptors

CHAPTER 1: INTRODUCTION

Nuclear Receptor Superfamily

The nuclear receptor (NR) superfamily is comprised of a family of transcription factors (TFs) that play an important role in a number of biological processes including metabolism, reproduction, and inflammation[1,2]. The first member of this family was cloned in 1985 but today the family has expanded to include 48 members in humans[3,4]. Most NRs are regulated endogenously by small lipophilic ligands such as steroids, retinoids, and phospholipids, but this protein family also contains "orphan" members for which no ligand has yet been identified[5]. Ligand binding induces conformational changes within the receptor, which in turn binds specific DNA sequences throughout the genome[6,7]. Once DNA bound, coregulator proteins, chromatin remodelers, and the general transcriptional machinery are recruited to the DNA in order to activate or repress target gene expression[8-10]. Since NRs are responsible for regulating thousands of genes, their activity is tightly controlled[11,12]. If left unchecked, aberrant NR activity can underlie numerous diseases such as cancer, diabetes, and chronic inflammation. Therefore, NR biology is a critical field of study[13,14].

Our knowledge of the NR family has drastically expanded within the last decade due to advancements in genome-wide methodologies, structural studies of receptor domains and full-length complexes, and identification of new coregulator proteins that modulate receptor activity[3,15]. This work has laid the foundation for pharmaceutical companies and academic researchers to develop synthetic ligands that target these receptors[16,17]. Yet, due to the large array of genes regulated by these proteins, drugs that target NRs tend to have unwanted side effects[16,18]. For this reason, more research is required to understand all the mechanisms that guide NR regulation. This expanding view of NR regulation could pave the way for future

therapeutics. Here, we introduce this protein family and focus on the structural mechanisms of nuclear receptor action.

Classifications

NRs are divided into 7 subfamilies[19,20]. A list of receptors, subfamily, and their ligands are shown in **Table 1**.

Subgroup 0: This group includes the atypical NRs, dosage-sensitive sex reversal-adrenal hypoplasia congenital critical region on the X chromosome, gene 1 (DAX) and small heterodimer partner (SHP)[21,22]. These two proteins are unique in their structures and contain only a ligand-binding domain (LBD) that folds in a manner consistent with the rest of the family[23-25]. Their LBDs also contain motifs that are commonly seen in NR coactivators[26]. These motifs interact with other NR LBDs to regulate transcription[27-31].

Subgroup 1: This large family is formed by thyroid hormone receptors (TR)[32], retinoic acid receptors (RAR)[33], peroxisome proliferator activated receptors (PPAR)[34], reverse-Erb receptors (REV-ERB)[35], retinoic acid related receptors (ROR)[35], farnesoid X receptors (FXR)[36], liver X receptors (LXR)[37], and vitamin D receptors (VDR)[38]. These receptors are regulated by a variety of signaling molecules including thyroid hormone, fatty acids, bile acids, and sterols.

Subgroup 2: This subfamily contains notable orphan receptors such as retinoid X receptors (RXR)[39], chicken ovalbumin upstream promoter transcription factors (COUP-TF)[40], and hepatocyte nuclear factor 4 (HNF4)[41]. RXR is of particular importance as it forms heterodimeric complexes with many NRs and is the only receptor in the group with a known activating ligand[42].

Subgroup 3: This group comprises the steroid receptors (SRs), which are key regulators of a host of metabolic, reproductive, and developmental processes[43]. The SR family includes the androgen receptor (AR)[44], progesterone receptor (PR)[45], glucocorticoid receptor (GR)[46], mineralocorticoid receptor (MR)[47], and two closely-related estrogen receptors (ER)[48]. Cholesterol-derived hormones, like cortisol and estrogen, regulate SRs through direct binding. A detailed review on the glucocorticoid receptor can be found in Chapter 2.

Subgroup 4: This group contains the orphan nuclear receptors nerve growth factor 1B (NGF1-B), nurr-related factor-1 (NURR1), and neuron-derived orphan receptor-1 (NOR-1). These proteins are required for neuron development and maintenance[49].

Subgroup 5: This group contains steroidogenic factor 1 (SF-1)[50] and liver receptor homolog-1 (LRH-1)[51]. Though generally still classified as orphan receptors, evidence suggests these proteins are regulated by phospholipids[29,52]. LRH-1 and SF-1 are vital in development and metabolism[51,53]. A more detailed look into LRH-1's role in development can be found in the introduction of Chapter 6.

Subgroup 6: This group contains only one receptor, germ cell nuclear factor (GCNF)[54], an orphan receptor that has a critical role in development[55]. This protein remains in its own category due to a critical difference in its LBD; it does not contain an activator function helix (AF-H) and is known to drive gene silencing[56]. A more detailed look into GCNF biology can be found in the introduction of Chapter 6.

Structural Insight into Nuclear Receptor Action

X-ray crystal structures of nuclear receptors, both full-length and discrete domains, have provided critical information on how ligands are recognized, bind their DNA response element(s), dimerize, and interact with coregulators.

Overall Architecture

Despite diversity in the size, shape, and charges of activating ligands, almost all members of the nuclear receptor superfamily share a common modular domain structure[15,57]. Except for the atypical receptors SHP and DAX, the overall architecture is comprised of five domains: A-E (**Figure 1.1a**). Each of these subdomains plays a specific role in receptor biology[58]. The size of NRs can vary but steroid receptors are generally around 100 kD and the remainder of the family around 66 kD (**Figure 1.1b**).

A/B: N-terminal Domain (NTD): The NTD is a highly disordered domain that has little sequence conservation between NRs, which explains why the NTD is not amenable to structural analysis[58]. Additionally, there is a large disparity in the size of this domain (**Figure 1.1b**).

The NTD contains the activator function-1 region (AF-1), which interacts with a variety of coregulator proteins in a cell and promoter-specific manner[59]. For all NRs, the majority of the domain is disordered. However, for GR this region could adopt a more alpha-helical content when coregulators are bound[60]. This region also gives rise to multiple isoforms through alternative splicing, as seen in TR and GR[32,46]. Finally, the NTD is the target for numerous post-translational modifications including phosphorylation,

SUMOylation, and acetylation[61]. These modifications have varying effects, both driving and repressing transcription.

C: DNA Binding Domain (DBD): This region is the most conserved among all nuclear receptor domains[62]. The DBD has two subdomains that each contain four cysteine residues that coordinate a zinc ion to create the canonical DNA-binding zinc finger motif (**Figure 1.2**)[63]. Each zinc finger is then followed by an amphipathic helix and a peptide loop[64,65]. The first subdomain contains the DNA reading helix, which sits within the major groove to make base-specific interactions with the DNA[66]. The second subdomain helix makes nonspecific contacts with the DNA backbone. The peptide loop in this subdomain contains the distal box, or "D box," that contains residues for receptor dimerization[67-69]. Some NRs, like LRH-1 and GCNF, contain a DBD C-terminal extension (CTE) that makes additional base-specific contacts within the DNA minor groove[70,71].

D: Hinge Region: The hinge region is a short, flexible linker between the DBD and the LBD[58]. This region has the least sequence and size conservation between nuclear receptors. Like the NTD, this region is also a site for regulatory PTMs. The hinge can also contain a nuclear localization signal.

E: Ligand Binding Domain (LBD): The LBD is a complex allosteric signaling domain that not only binds to ligands but also interacts directly with coregulator proteins[72,73]. This structurally conserved domain commonly contains 11 α-helices and 4 β-strands that fold into three parallel layers to form an alpha helical sandwich (**Figure 1.3**)[74]. This folding creates a hydrophobic ligand-binding pocket (LBP) at the base of the receptor[72,75,76]. Superposition of NR LBD structures reveals that the top part of the receptor is most similar where as the base,

which contains the LBP, is more variable[15,74]. This variability across NRs at the ligand-binding region allows NRs to recognize a diverse cadre of ligands.

The LBD contains another activation function surface (AF-2), which is comprised of helices 3, 4, and 12. Helix 12, or the activation function helix (AF-H) has been shown to be conformationally dynamic upon ligand binding, altering the AF-2 to facilitate binding to different coregulator proteins[72,74,77].

NR-Ligand Interactions

Nuclear receptors bind directly to a variety of small, lipophilic ligands, such as steroids, thyroid hormone, retinoids, and lipids that can diffuse easily across the cell membrane[5]. Of the 48 human NRs, 24 have known ligands and the remaining 24 are classified as "orphans" or "adopted orphans". In the absence of ligand, NRs tend to be unstable, explaining the dearth of apo-NR LBD structures[77-79]. Ligand binding greatly increases the stability of the LBD, evidenced by changes in NMR spectra between liganded and unliganded PPARs and less proteolytic cleavage seen in the ER ligand-bound versus apo state[76-81]. This stabilization, among other factors, facilitates coregulator binding[82].

Ligands bind the receptor within the LBP at the base of the LBD. This pocket is comprised of ~75% hydrophobic residues, but also contains critical polar residues that make key hydrogen bonding interactions to the ligand[73,74]. These hydrogen bonds help position the ligand in the correct orientation. For example, SRs use a conserved glutamine on H3 and arginine on H5 to lock the ligand's A ring in place (**Figure 1.4a,b**)[83,84]. A striking example of the importance of these hydrogen bond networks in the LBP is seen in FXR and LXR ligands; though similar, these ligands are bound in completely opposite orientations due to

the hydrogen bonding network within the LBP (**Figure 1.4c,d**)[85,86]. These differences ensure the natural ligands are bound by the correct receptor. Ligand selection is further achieved by a dramatic difference in the size of ligand binding pockets across NRs. For example, SR LBPs tend to be 400-600 Å, while 700-850 Å for FXR and LXR, and almost 1300 Å for PPARs (**Figure 1.4e**)[83,85,87]. In these cases, a significant component of ligand selection stems from steric selection.

NR-DNA Interactions

Nuclear receptor DBDs bind to a variety of DNA response elements (REs) whose nucleotide sequences can take the form of a palindrome, direct repeat, or extended monomeric sites (**Figure 1.5**)[63,67]. The SRs bind palindromic repeats (**Figure 1.5a**). These palindromes contain two AGGACA repeats that can be separated by a spacer region that varies in length. The length of this spacer has been shown to allosterically modulate SRs, resulting in varied transcriptional outputs[88-90]. However, the most common spacer length is 3 bp[68,91,92]. Receptors that bind direct repeats include the RXR-RAR heterodimer, GCNF, and VDR (**Figure 1.5b**)[93-95]. These sequences are composed of two AGGTCA sites separated by a spacer sequence from 0-5bp long. Finally, LRH-1 and SF-1 are examples of receptors that bind extended half-site sequences (**Figure 1.5c**)[71,96]. These REs contain one AGGTCA site as well an A/T rich sequence directly upstream.

NRs can form monomers, dimers, or heterodimers

NRs are generally found as monomers in solution but upon DNA binding can form higher order complexes. NRs can be monomeric on DNA but are more often found as

homodimers or in heterodimeric complexes with RXR[3]. This increases overall size and complexity of NRs, allowing new surfaces to be accessed for PTMs or coregulator binding[46].

LRH-1, NGF1B, and SF-1 are among the few NRs that bind DNA as monomers[71,96]. These receptors utilize the CTE within their DBDs to facilitate additional DNA contacts within the minor groove, expanding their DNA footprint. Members of the SR subfamily commonly form homodimers. The ER LBD structure shows H8, H9, H10, and loops 8-9 from each monomer interacts to form a homodimer (**Figure 1.6a**)[84]. This is in contrast with the GR dimer, which showed a unique dimer interface not seen in other NR structures (**Figure 1.6c**)[97]. Finally, the rest of the NR superfamily commonly forms heterodimers with RXR[3,98]. Similar to the ER structure, the dimer interface is formed between H7, H9, H10, H11, and loops 8-9 (**Figure 1.6b**)[98].

NR-Coregulator Interactions

After DNA binding, NRs recruit a variety of proteins collectively known as coregulators[8,99]. To date, there are approximately 200 different coregulator proteins, which fall into two main categories: coactivators and corepressors[8,9]. These interact directly with NRs at the AF-1 and AF-2 surfaces[59]. Since the AF-1 lies within the unstructured NTD we have not been able to obtain structural information about these interactions[58,60]. However, almost all NR LBD structures are co-crystallized with fragments of NR-interaction domains of coregulator proteins[59].

Coactivator proteins interact with NRs via an α-helix containing a short LXXLL motif (L- leucine, X- any amino acid)[26,82]. This motif interacts with the NR AF-2 surface[55]. The coregulator's leucine residues lie within the hydrophobic groove of the AF-2 surface and

the ends of the peptide are generally held in place by a charge clamp formed by a lysine on the NR's H3 and a glutamate on H12 that cap the helix dipole (**Figure 1.7a**)[82].

Corepressors contain conserved LXXX(I/L)XXX(I/L) motif (referred to as CoRNR box) (L- leucine, I- isoleucine, X- any amino acid)[100,101]. These extended motifs interact at the same hydrophobic AF surface but their length inhibits the canonical charge clamp formation (**Figure 1.7b**)[102,103].

The discrimination between either coactivator or corepressor binding has been linked to the conformational flexibility of H12[6,74]. Originally, the "mouse-trap" model was proposed[74]. This model was based on the structures of apo RXR and ligand-bound RAR (**Figure 1.7c**)[95,105]. It was posited that upon agonist binding, there was a large structural rearrangement of H12, causing it to snap shut[84]. However, this phenomenon was only observed for a few proteins[84]. Other NR LBD structures, like LRH-1 in both the apo and ligand bound state, did not demonstrate large movements in H12[106]. This suggested another model was possible. The current favored model is the "dynamic stabilization model," which suggests that H12 is not in one fixed position, but rather is dynamic[79,107]. Ligand binding stabilizes the helix into a more fixed conformation. Methods that measure dynamics of H12, such as NMR and HDX, have been pivotal in providing evidence to support this model[76,81,108].

Nuclear Receptor Signaling

Nuclear Receptor Mechanism of Action

NRs have been classified as into four mechanistic subtypes types, I-IV (**Figure 1.8**):

Type I Nuclear Receptors: Receptors of this group are SRs and are activated by cholesterol-derived steroidal hormones, such as estrogens, androgen, progestagens, and corticoids[43]. These receptors are sequestered to the cytoplasm bound to chaperone proteins but upon ligand activation, they exchange their chaperone proteins and undergo nuclear translocation. In the nucleus, SRs generally bind as homodimers to DNA REs that consist of two inverted repeats (**Figure 1.8a**)[109,110].

Type II Nuclear Receptors: Receptors of this type, such as RAR and LXR, are often retained in the nucleus, regardless of the presence of activating ligand[10]. Upon ligand binding, the receptor is released from a corepressor complex and swapped for coactivators and the transcriptional machinery. These receptors commonly form heterodimers with RXR on direct repeat DNA REs (**Figure 1.8b**)[3].

Type III Nuclear Receptors: This type of NR, such as VDR, has a similar mechanism of action to Type II NRs but instead form homodimers on their REs, which are direct repeat sequences (**Figure 1.8c**)[63].

Type IV Nuclear Receptors: This type of NR has a similar mechanism of action to Type II NRs but instead bind to DNA as a monomer and recognize extended half-sites within REs (**Figure 1.8d**)[71,96]. Examples of Type IV include LRH-1 and SF-1.

Transactivation and Transrepression

NRs modulate transcription through many distinct mechanisms that ultimately result in either activation or repression of specific gene programs[1]. As stated above, transcriptional activation is achieved by ligand binding converting the receptor from an inactive to active state[7]. In this state, NRs recruit coactivator proteins, which are typically scaffolds that initiate

the formation of large protein complexes[46]. Within these complexes are histone modifying enzymes such as histone acetyltransferases (HATs) and histone methyltransferases (HMs)[112,113]. These proteins facilitate the opening of chromatin, making it accessible to additional regulatory proteins. Finally, the general transcriptional machinery and RNA Polymerase II are recruited to drive transcription (**Figure 1.9a**)[111].

Conversely, NRs can repress transcription by two different mechanisms[114]. First, NRs can bind to corepressors in their apo state as shown in mechanistic type II-IV receptors[114]. These corepressor proteins recruit histone modifying enzymes such as histone deacetylases (HDACs)[8], which act in opposition of HATs to restrict chromatin and block the transcriptional machinery from accessing the DNA (**Figure 1.9b**)[113,115]. Second, NRs can interact with "negative DNA response elements" (nGRE)[116,117]. Binding to these elements results in NRs adopting different conformations than when bound to "positive" DNA response elements and eventually recruit corepressor proteins to block transcription[118]. This is best exemplified for GR (see Chapter 2).

Nuclear Receptors as critical Pharmaceutical Targets

Aberrant nuclear receptor signaling pathways contributes to numerous disease states such as cancer, diabetes, obesity, and others[14,17]. For this reason, NRs are major pharmaceuticals targets[17]. Initial ligand design has been quite simple as NR LBPs are enclosed and are amenable to binding a variety of ligands[74]. However, due to the breadth and complexity of NR biology, designing ligands with limited cross-reactivity has proven quite difficult[1]. Despite these issues, NR-targeting ligands make up 10-20% of current FDA-approved drugs have a worldwide market of 30 billion dollars per year[119].

Historically, there have been two main approaches for identifying NR ligands[18]. First, NR ligands were isolated from human tissue extracts[120]. For example, the study of the adrenal gland led to the discovery of a compound effective at blocking inflammation. This compound was later discovered to be cortisol, the endogenous ligand for GR[120]. Later, synthesis of cortisol sparked the development of the synthetic compounds dexamethasone and prednisolone[121]. Second, compounds were identified by connecting ligand effects were connected with protein biology[17]. For example, thiazolidinediones showed promise in treating diabetes[122]. These effects were later linked to PPARγ signaling[122]. The newest generation of NR ligands are termed "selective nuclear receptor modulators", which are designed against a single NR to partially or selectively activate a subset of signaling pathways. These idea is to separate the beneficial outcomes of treatment from the less desirable side effects[123]. Collectively, these ligands are called are a popular idea for targeting ER and GR[124,125]. Due to the complexity of NR signaling, these compounds have been largely unsuccessful thus far but efforts remain to develop these ligands.

Questions and Hypotheses Addressed In This Work

Because of their critical roles in a number of biological processes, NRs have to be tightly regulated to ensure the proper genes expressed or repressed? What guides these proteins to either turn a gene off or on is the major question addressed in this work. We explore how this is achieved through a review of the glucocorticoid receptor in **Chapter 2**. Here, we propose that DNA sequences, ligands, PTMs, and other TFs act as allosteric modulators that ultimately guide GR action. We explore these modulators in subsequent chapters. In **Chapter 3** we show that DNA sequences can have a profound effect on how GR regulates genes. Moreover, we identify a novel mechanism for how the GR turns genes off.

Chapter 4 explores how a popular clinical GR ligand, triamcinolone acetonide, interacts with GR. Here we use structural biology to explore how this ligand stabilizes the LBD so greatly. As there are only a few structures GR-ligand complexes available, this work provides valuable information that can be used to target GR in the future. In **Chapter 5** we report the first crystal structure of GCNF and LRH-1 in complex with their DNA response element within the *Oct4* promoter. These NRs are critical to reciprocally regulate *Oct4* during development from the same DNA sequence. This work further highlights the importance of how DNA sequences have such a profound effect on transcriptional outputs. Finally, in **Chapter 6** we discuss how these collective works support our overarching hypotheses and present ideas for the future of this work and the NR field.

Tables and Figures

Family	Common Name	Abbreviation	Gene Name	Ligand
0B	Dosage-sensitive sex reversal-adrenal hypoplasia congenital critical region on the X chromosome, gene 1	DAX1	*NR0B1*	Orphan
	Short heterodimeric partner	SHP	*NR0B2*	Orphan
1A	Thyroid hormone receptor-α	TRα	*THRA*	Thyroid hormones
	Thyroid hormone receptor-β	TRβ	*THRB*	Thyroid hormones
1B	Retinoic acid receptor-α	RARα	*RARA*	Retinoic acids
	Retinoic acid receptor-β	RARβ	*RARB*	Retinoic acids
	Retinoic acid receptor-γ	RARγ	*RARG*	Retinoic acids
1C	Peroxisome proliferator-activated receptor-α	PPARα	*PPARA*	Fatty acids
	Peroxisome proliferator-activated receptor-β	PPARβ	*PPARD*	Fatty acids
	Peroxisome proliferator-activated receptor-γ	PPARγ	*PPARG*	Fatty acids
1D	Reverse-Erb-α	REV-ERBα	*NR1D1*	Heme
	Reverse-Erb-β	REV-ERBβ	*NR1D2*	Heme
1F	Retinoic acid related orphan-α	RORα	*RORA*	Sterols
	Retinoic acid related orphan-β	RORβ	*RORB*	Sterols
	Retinoic acid related orphan-γ	RORγ	*RORC*	Sterols
1H	Farnesoid X receptor	FXRα	*NR1H4*	Bile Acids
	Farnesoid X receptor-β	FXRβ	*NR1H5P*	Orphan
	Liver X receptor-α	LXRα	*NR1H3*	Oxysterols

	Liver X receptor-β	LXRβ	*NR1H2*	Oxysterols
1I	Vitamin D receptor	VDR	*VDR*	1α,25-dihydroxyvitamin D3
	Pregnane X receptor	PXR	*NR1I2*	Endobiotics and xenobiotics
	Constitutive androstane receptor		*NR1I3*	Xenobiotics
2A	Hepatocyte nuclear factor-4-α	HNF4α	*HNF4A*	Fatty acids
	Hepatocyte nuclear factor-4-γ	HNF4γ	*HNF4G*	Fatty acids
2B	Retinoid X receptor-α	RXRα	*RXRA*	9-cis retinoic acid
	Retinoid X receptor-β	RXRβ	*RXRB*	9-cis retinoic acid
	Retinoid X receptor-γ	RXRγ	*RXRG*	9-cis retinoic acid
2C	Testicular receptor 2	TR2	*NR2C1*	Orphan
	Testicular receptor 4	TR4	*NR2C2*	Orphan
2E	Tailless homolog orphan receptor	TLX	*NR2E1*	Orphan
	Photoreceptor-cell-specific nuclear receptor	PNR	*NR2E3*	Orphan
2F	Chicken ovalbumin upstream promoter-transcription factor α	COUP-TFα	*NR2F1*	Orphan
	Chicken ovalbumin upstream promoter-transcription factor β	COUP-TFβ	*NR2F2*	Orphan
	Chicken ovalbumin upstream promoter-transcription factor γ	COUP-TFγ	*NR2F6*	Orphan
3A	Estrogen receptor-α	ERα	*ESR1*	Estrogens
	Estrogen receptor-β	ERβ	*ESR2*	Estrogens
3B	Estrogen-related receptor-α	ERRα	*ESRRA*	Orphan
	Estrogen-related receptor-β	ERRβ	*ESRRB*	Orphan

	Estrogen-related receptor-γ	ERRγ	*ESRRG*	Orphan
3C	Androgen receptor	AR	*AR*	Androgens
	Glucocorticoid receptor	GR	*NR3C1*	Glucocorticoids
	Mineralocorticoid receptor	MR	*NR3C2*	Mineralocorticoids and glucocorticoids
	Progesterone receptor	PR	*PGR*	Progesterone
4A	Nerve growth factor 1B	NGF1-B	*NR4A1*	Orphan
	Nurr-related factor 1	NURR1	*NR4A2*	Orphan
	Neuron-derived orphan receptor 1	NOR-1	*NR4A3*	Orphan
5A	Steroidogenic factor 1	SF-1	*NR5A1*	Phospholipids
	Liver receptor homolog-1	LRH-1	*NR5A2*	Phospholipids
6A	Germ cell nuclear factor	GCNF	*NR6A1*	Orphan

Table 1.1: Nuclear Receptor Superfamily

Table of human nuclear receptors, gene name, and their activating ligands.

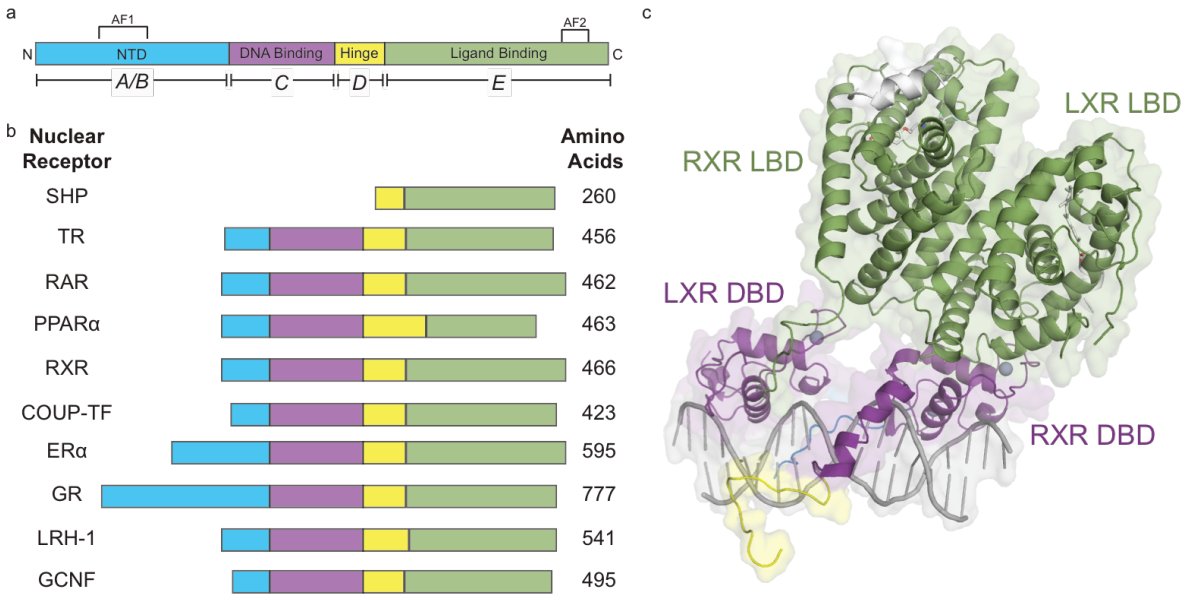


Figure 1.1: Modular Domain Structure of NRs.

(a) Basic modular domain structure of NRs is comprised of an unstructured NTD that contains the AF-1 surface, a zinc finger DBD, a flexible hinge region, and a LBD that binds to ligands and interacts with coregulator proteins through the AF-2 surface. (b) General domain size and amino acid length of a variety of NRs. The DBD and LBDs are the most conserved regions where as the other domains are more variable in length and sequence composition. (c) Example of full-length NR structure shows LXR-RXR heterodimer (PDB: 4NQA) (DBD colored purple, hinge region in yellow, and LBD in green) .

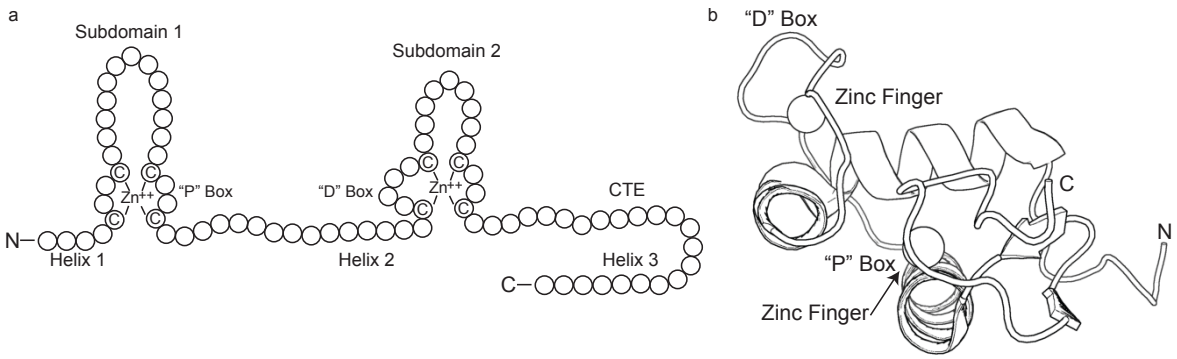


Figure 1.2: NR DNA Binding Domains.

(a) Cartoon representation of NR DBDs indicating important motifs. This domain contains two subdomains, each containing one zinc finger. The first subdomain residues interact with the DNA major groove to make base specific interactions on genomic response elements. The second subdomain participates in DBD dimerization and makes non-specific contacts with the DNA backbone. Some NRs, like LRH-1 and GCNF, also contain C-terminal extensions (CTEs) that make base specific contacts with the minor groove. (b) Cartoon representation of folded NR DBD highlighting the important regions. (PDB: 3FYL).

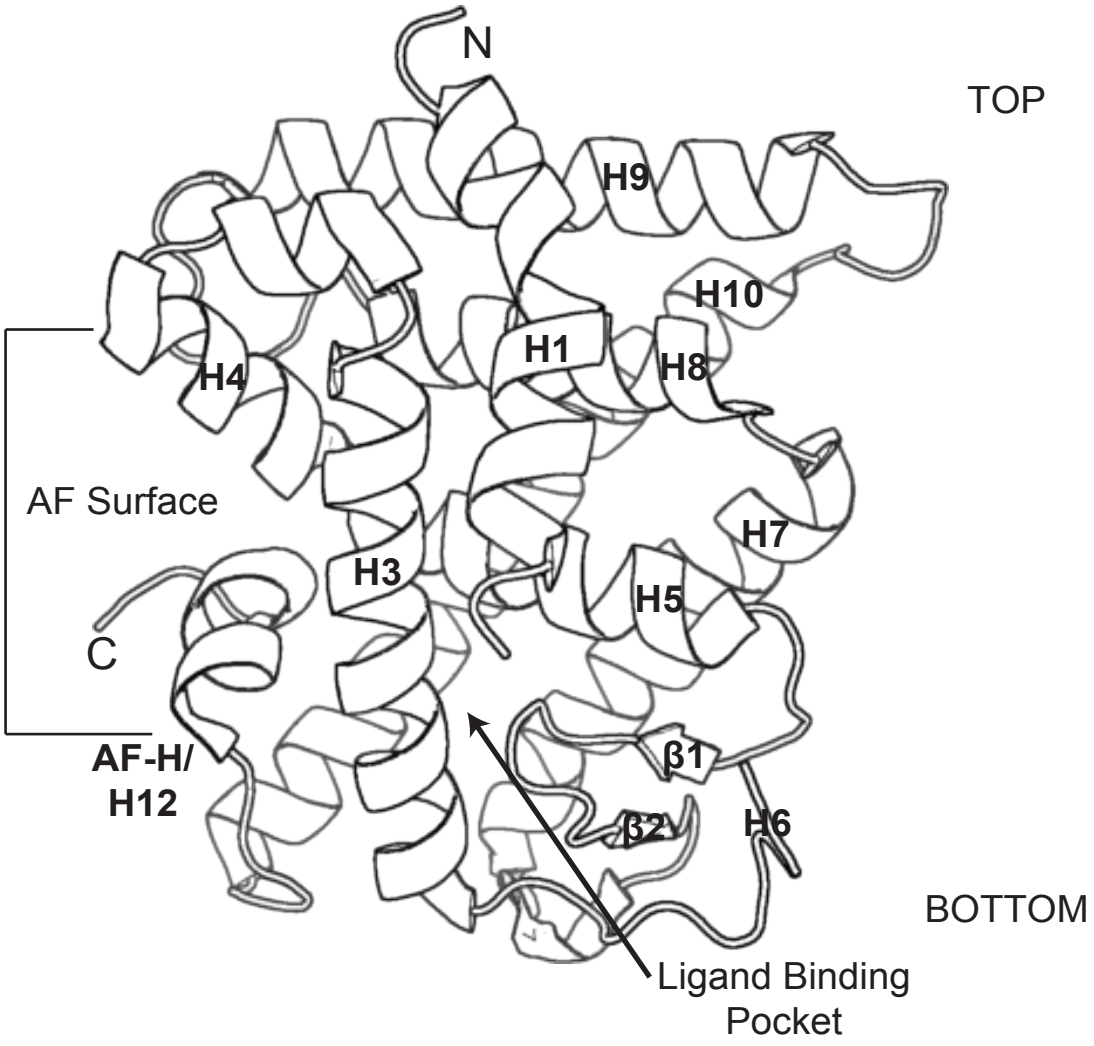


Figure 1.3: NR Ligand Binding Domains.

Cartoon representation of the structurally conserved NR LBD. This domain is comprised of 11 α-helices and 4 β-strands that fold into three layers of a helical sandwich bundle. This fold creates a hydrophobic ligand binding pocket at the bottom of the receptor. This domain also contains the AF surface, comprised of H3, H4, and the AF-H, which interacts with coregulator proteins (PDB: 1PZL).

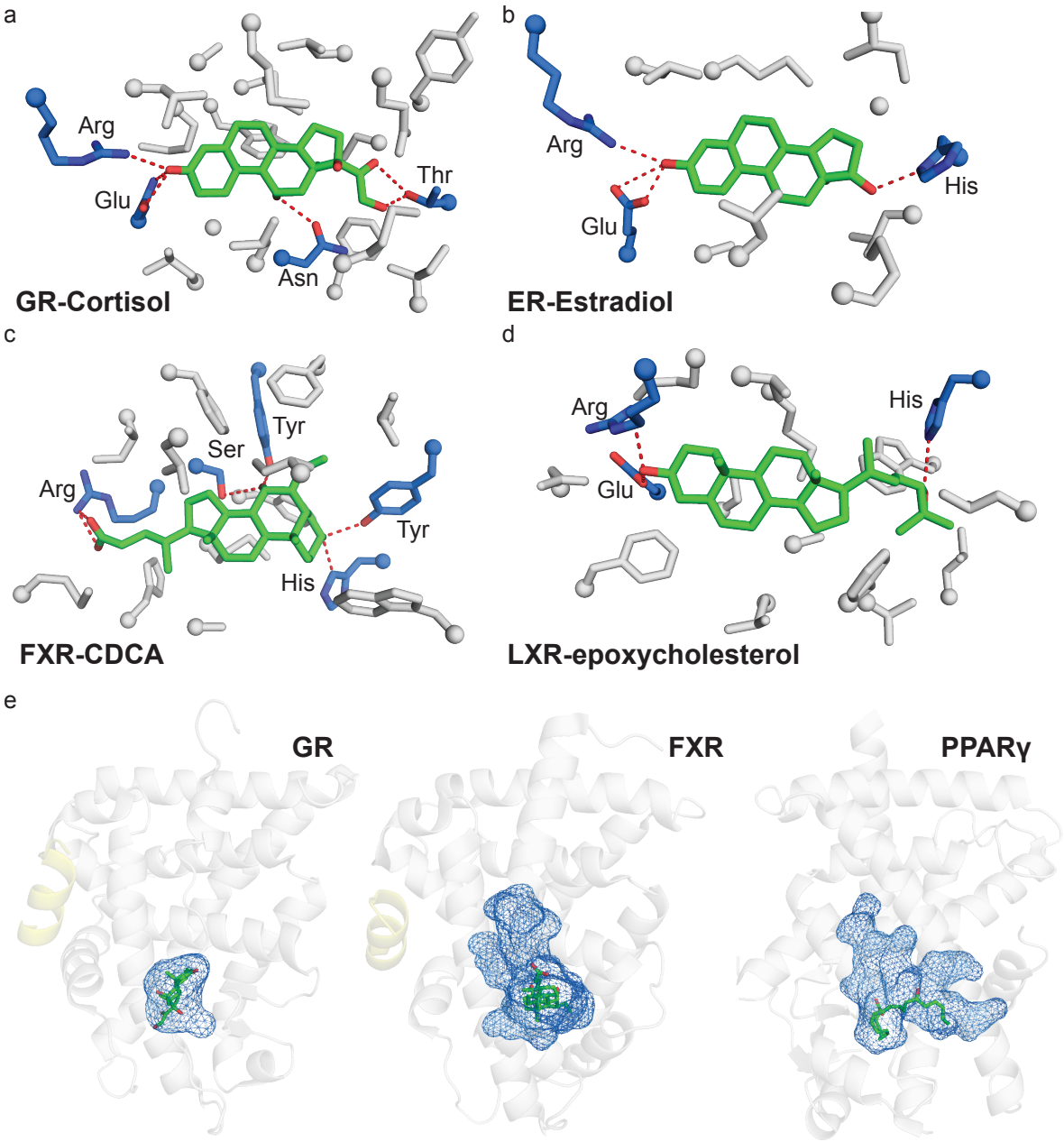


Figure 1.4: NR Ligand-Interactions.

Close up view of SR LBPs showing that (a) GR LBD-cortisol (PDB: 4P6X) and (b) ER LBD-estradiol (PDB: 1ERE) use conserved Glu and Arg residues (blue sticks) to make hydrogen bonding interactions (red) with steroid ligands. These interactions help orient the ligand within the pocket. (c) Close up views of FXR LBD-CDCA (PDB: 1OT7) and (d) LXR

LBD-epoxycholesterol (PDB: 1P8D) show, despite similar ligands, the receptors orient them in opposite directions. This allows natural ligands to discriminate between NRs whose LBDs are highly conserved (e) Comparisons of ligand cavity sizes between GR (PDB: 4P6X), FXR (PDB: 1OT7), and PPAR (PDB: 5AZV).

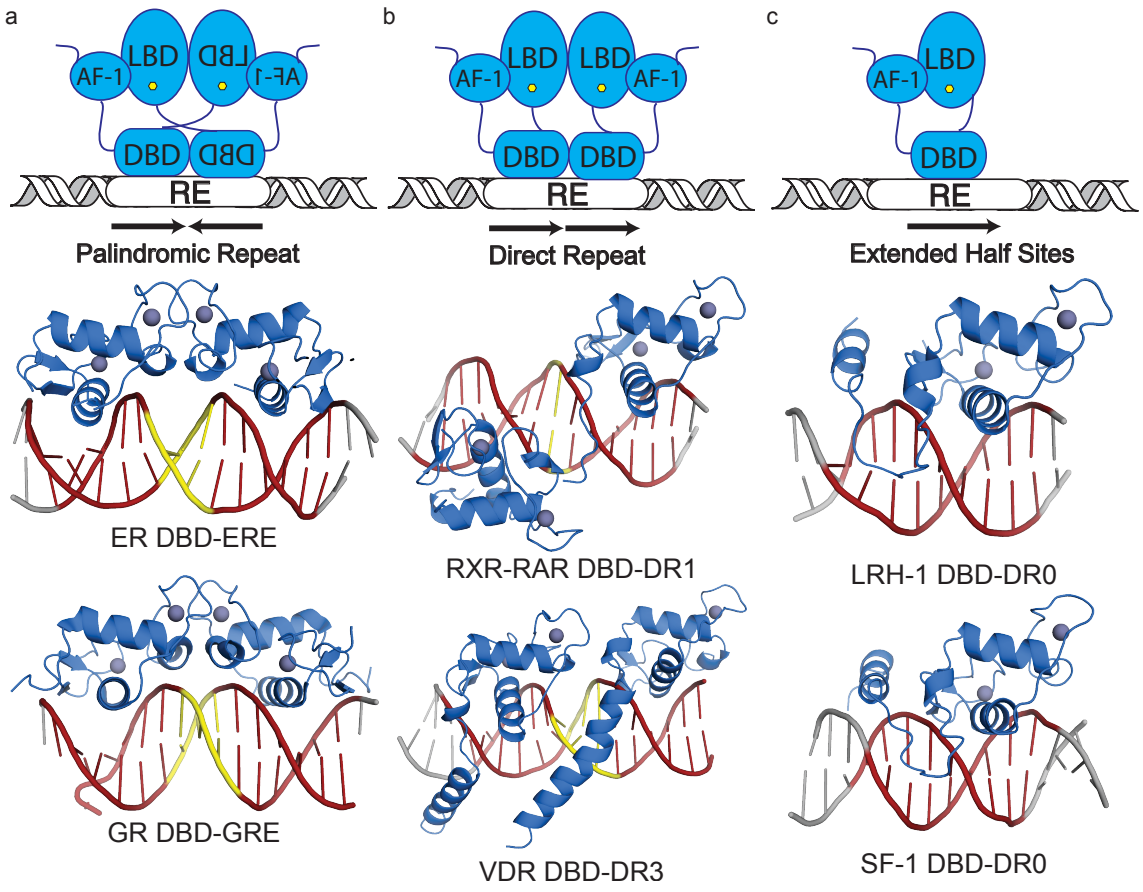


Figure 1.5: Genomic Response Elements.

Nuclear receptors bind to genomic response elements that come in a variety of forms. (a) Members of the SR subfamily bind to palindromic repeats (shown as red DNA cartoon). These repeats are separated by different spacer lengths (shown as yellow DNA cartoon). As examples, the ER DBD-ERE and GR DBD-GRE crystal structures are shown below. (b) Most other NRs bind to direct repeats, which can also be separated by spacers from 0-5 bp. The structures of the RXR-RAR DBD heterodimer is shown in complex with a DR with 1 bp spacer (DR1) and the VDR homodimer DBD is shown in complex with a DR with 3 bp spacer (DR3). (c) Though rare, some NRs bind to DNA as a monomer to extended half site

sequences. Examples include LRH-1 DBD and SF-1 DBD. (PDBs, from left to right: top row – 4AA6, 1DSZ, 5L0M; bottom row – 3FYL, 1KB4, 2FF0)

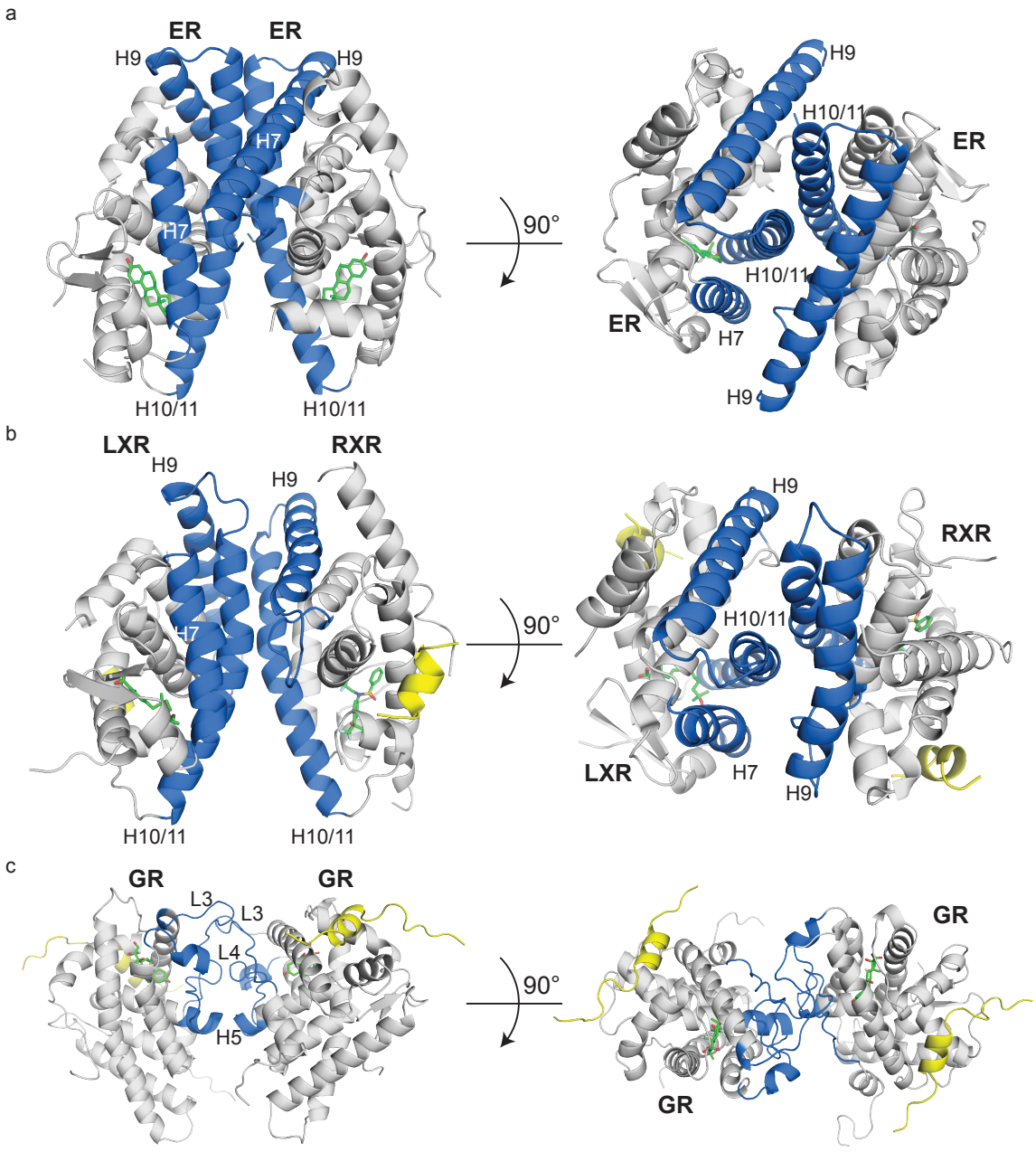


Figure 1.6: NR Dimerization Interfaces.

Many NRs utilize the H10/H11 surface to form homodimers or heterodimers. (a) ER LBD homodimer shows dimerization occurs between H7, H9, H10/11 (PDB: 1ERE). (b) The LXR-RXR LBD heterodimer shows a similar dimerization interface (PDB: 1UHL). (c)

Unlike the other two, the GR LBD homodimer structure revealed a novel dimerization interface (PDB: 1M2Z). The dimerization interface is colored blue, ligands are shown as sticks (green) and coregulator peptides are colored yellow.

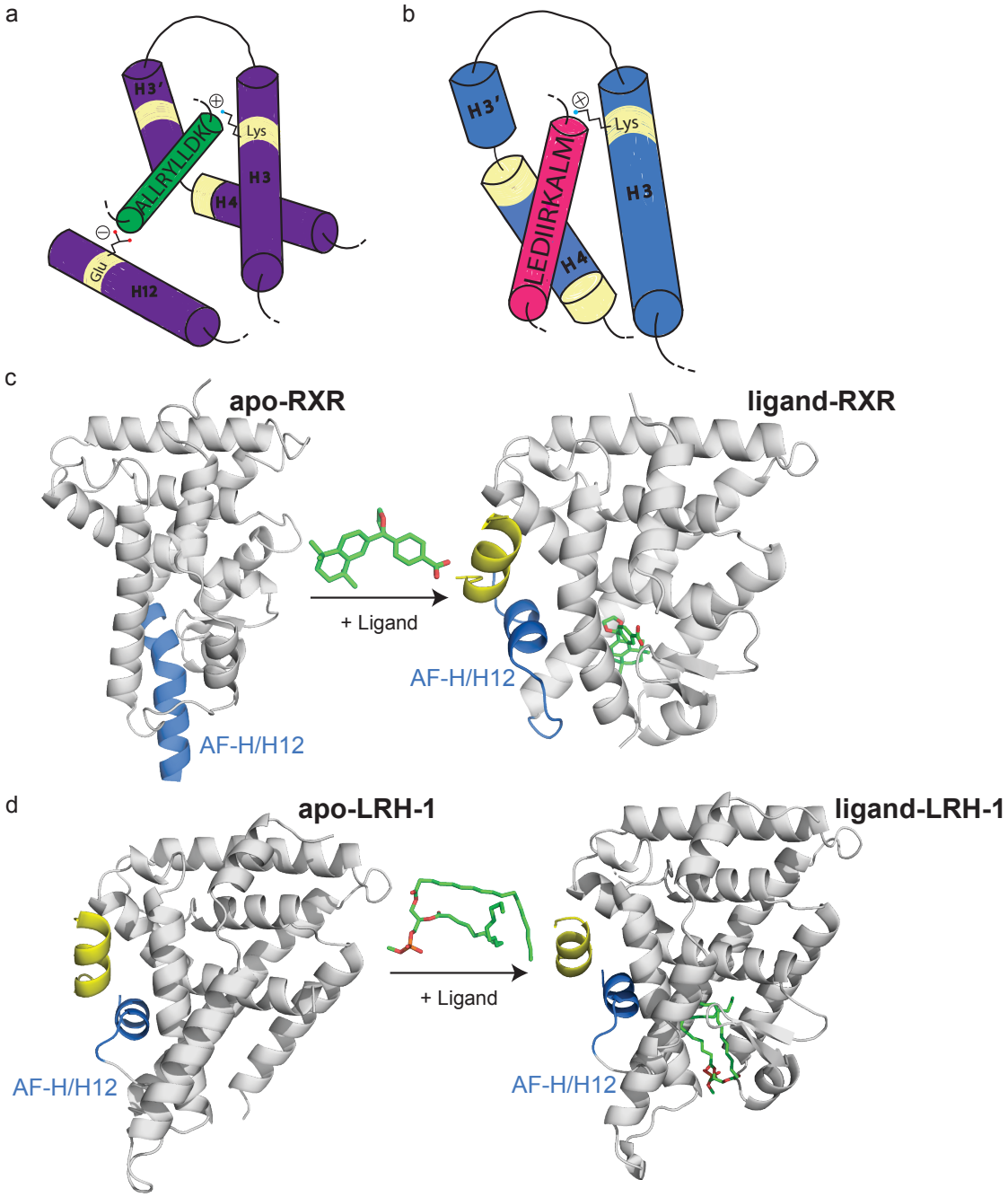


Figure 1.7: NR Coregulator Interactions.

(a) Cartoon representation of the coregulator LXXLL peptide (green) interacting with the AF surface (purple). The peptide is held in place by a conserved charge clamp interaction formed

by a glutamate on H12 and a lysine on H4. (b) Cartoon representation of corepressor peptides (pink) interacting with the AF surface (blue). Corepressors contain extended (L/I)XX(I/V)I or LXXX(I/L)XXX(I/L) motifs that do not allow for the charge clamp formation. (c,d) The basis of the "mouse-trap" model was made by comparing the apo (c) and ligand bound (d) structures of RXR. Upon ligand binding a large rearrangement of H12 is seen. (PDBs: 1LBD, 1MVC). (e,f) The more favored "dynamic stabilization" model of NR activation suggests H12 does not undergo such a large conformational change, but instead H12 flexible and ligand binding simply stabilizes the helix. This model was proposed after other apo NR structures, did not show H12 displaced and upon ligand binding there was little change in the location of this helix. (PDBs: 4DOR, 4PLE). Coregulator peptides are colored blue and ligands are shown as sticks (green).

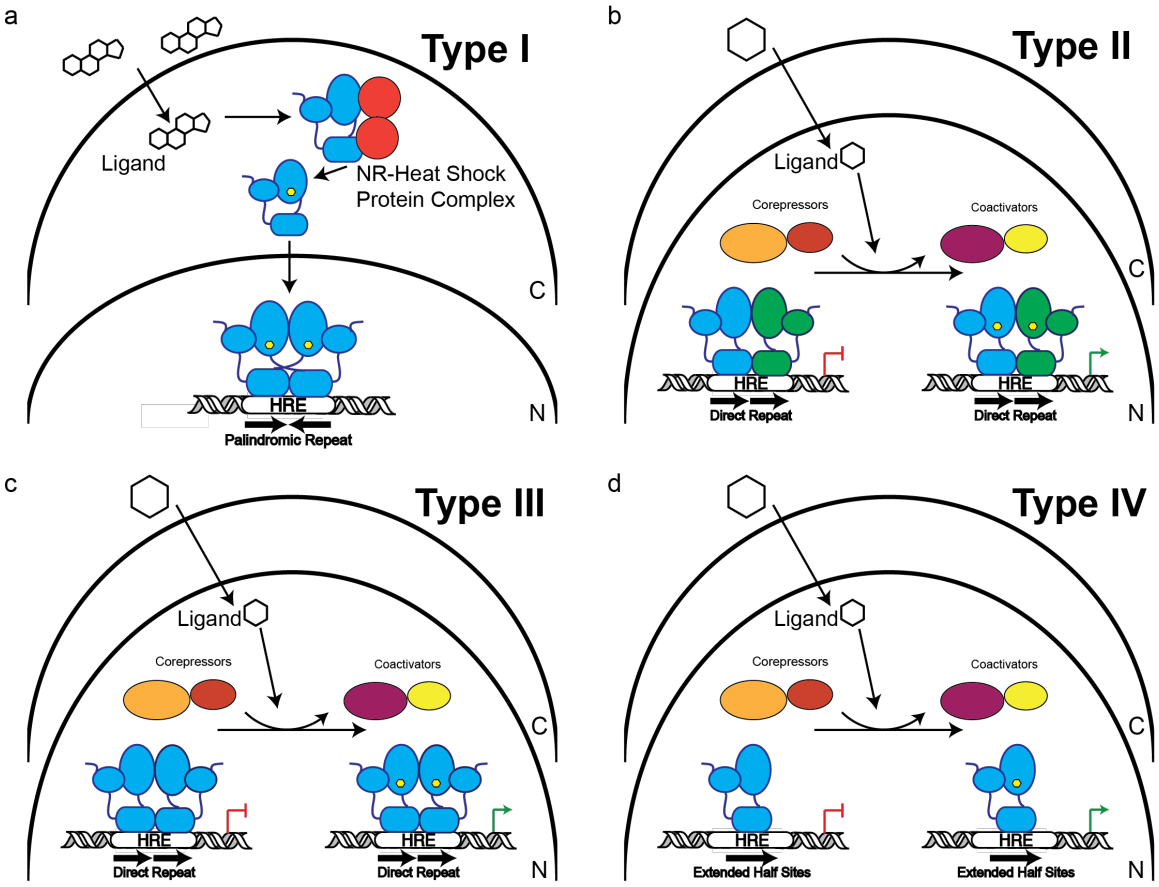


Figure 1.8: Schematic of NR signaling mechanisms.

(a) Type I receptors reside in the cytoplasm (C) in complex with chaperone proteins. Upon ligand binding (hexagon) the receptor is released from this complex and is trafficked into the nucleus (N) where they typically bind to palindromic repeat response elements (HREs) as a homodimer to regulate transcription. These NRs are commonly SRs that are activated by steroid hormones. (b) Type II receptors are localized in the nucleus. In their unliganded state they interact with corepressor proteins but upon ligand binding are exchanged for coactivators. NRs in this group generally form heterodimeric complexes with RXR. (c) Similar to Type II receptors, Type III receptors reside in the nucleus and undergo interchange

from being bound by corepressors and coactivators. These receptors bind to direct repeat HREs as homodimers. (d) Type IV receptors are almost identical to Type III except they bind HREs that are extended half sites as monomers.

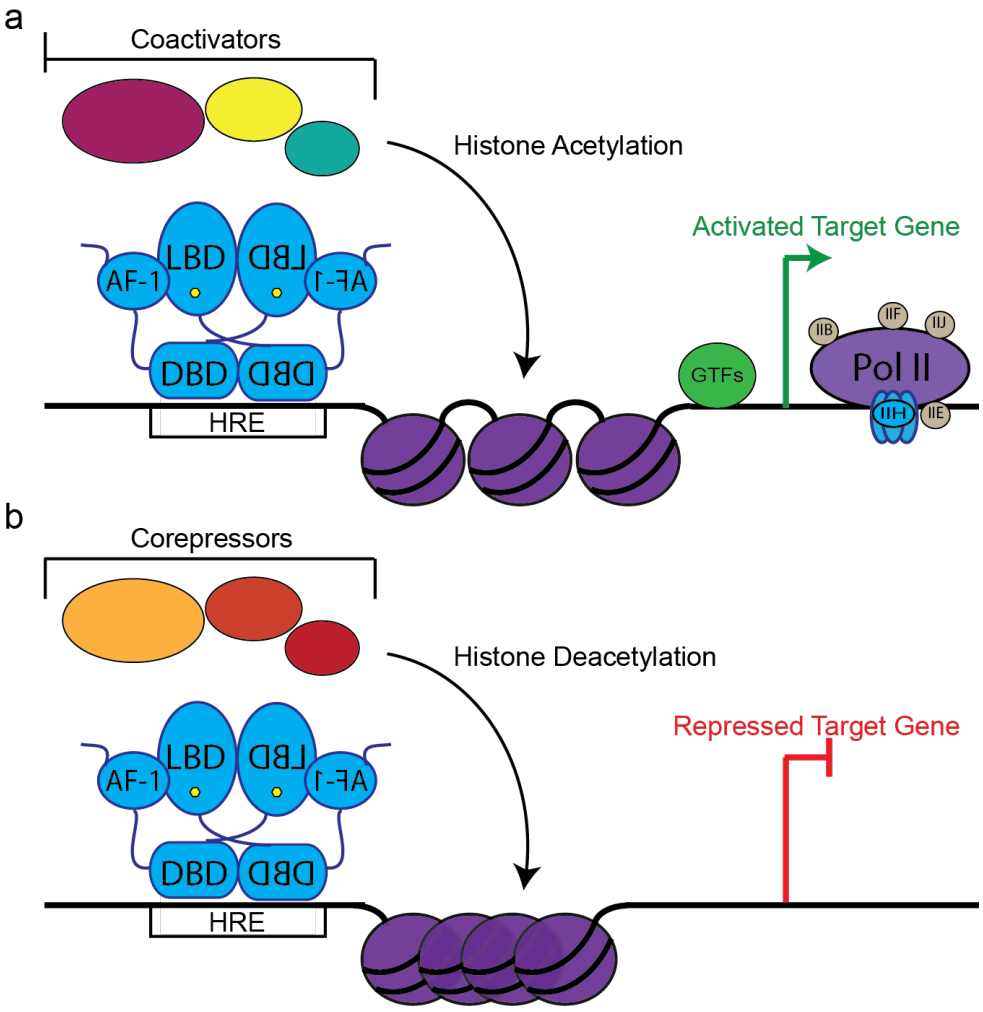


Figure 1.9: NRs Both Activate and Repress Transcription.

(a) To activate gene expression, NRs interact with their RE. DNA bound NRs recruit coactivator proteins, which in turn recruit histone-modifying enzymes. These histone-modifying enzymes are commonly histone acetylases, which acetylate histone tails. This modification is a mark of active chromatin. Ultimately, the general transcriptional machinery and RNA Polymerase are recruited to drive gene expression. (b) To repress transcription, NRs recruit corepressor proteins. These proteins recruit other histone-modifying enzymes that aim to reverse histone acetylation and instead restrict the chromatin. This condensation

prevents the transcriptional machinery from accessing the DNA, thus repressing gene expression.

References

1. Kininis, M. and W. L. Kraus (2008). "A global view of transcriptional regulation by nuclear receptors: gene expression, factor localization, and DNA sequence analysis." Nucl Recept Signal **6**: e005.
2. Gustafsson, J. A. (2016). "Historical overview of nuclear receptors." J Steroid Biochem Mol Biol **157**: 3-6.
3. Mangelsdorf, D. J., C. Thummel, M. Beato, P. Herrlich, G. Schutz, K. Umesono, B. Blumberg, P. Kastner, M. Mark, P. Chambon and R. M. Evans (1995). "The nuclear receptor superfamily: the second decade." Cell **83**(6): 835-839.
4. Govindan, M. V., M. Devic, S. Green, H. Gronemeyer and P. Chambon (1985). "Cloning of the human glucocorticoid receptor cDNA." Nucleic Acids Res **13**(23): 8293-8304.
5. Xu, E. H. and M. H. Lambert (2003). "Structural insights into regulation of nuclear receptors by ligands." Nucl Recept Signal **1**: e004.
6. Nagy, L. and J. W. Schwabe (2004). "Mechanism of the nuclear receptor molecular switch." Trends Biochem Sci **29**(6): 317-324.
7. Li, Y., M. H. Lambert and H. E. Xu (2003). "Activation of nuclear receptors: a perspective from structural genomics." Structure **11**(7): 741-746.
8. Lonard, D. M. and W. O'Malley B (2007). "Nuclear receptor coregulators: judges, juries, and executioners of cellular regulation." Mol Cell **27**(5): 691-700.
9. McKenna, N. J., R. B. Lanz and B. W. O'Malley (1999). "Nuclear receptor coregulators: cellular and molecular biology." Endocr Rev **20**(3): 321-344.

10. Tata, J. R. (2002). "Signalling through nuclear receptors." Nat Rev Mol Cell Biol **3**(9): 702-710.

11. Aranda, A. and A. Pascual (2001). "Nuclear hormone receptors and gene expression." Physiol Rev **81**(3): 1269-1304.

12. Meier, C. A. (1997). "Regulation of gene expression by nuclear hormone receptors." J Recept Signal Transduct Res **17**(1-3): 319-335.

13. Tenbaum, S. and A. Baniahmad (1997). "Nuclear receptors: structure, function and involvement in disease." Int J Biochem Cell Biol **29**(12): 1325-1341.

14. Stigsdotter, A. and L. Backman (1989). "Multifactorial memory training with older adults: how to foster maintenance of improved performance." Gerontology **35**(5-6): 260-267.

15. Huang, P., V. Chandra and F. Rastinejad (2010). "Structural overview of the nuclear receptor superfamily: insights into physiology and therapeutics." Annu Rev Physiol **72**: 247-272.

16. Xu, H. E. (2015). "Family reunion of nuclear hormone receptors: structures, diseases, and drug discovery." Acta Pharmacol Sin **36**(1): 1-2.

17. Gronemeyer, H., J. A. Gustafsson and V. Laudet (2004). "Principles for modulation of the nuclear receptor superfamily." Nat Rev Drug Discov **3**(11): 950-964.

18. Moore, J. T., J. L. Collins and K. H. Pearce (2006). "The nuclear receptor superfamily and drug discovery." ChemMedChem **1**(5): 504-523.

19. Germain, P., B. Staels, C. Dacquet, M. Spedding and V. Laudet (2006). "Overview of nomenclature of nuclear receptors." Pharmacol Rev **58**(4): 685-704.

20. Owen, G. I. and A. Zelent (2000). "Origins and evolutionary diversification of the nuclear receptor superfamily." Cell Mol Life Sci **57**(5): 809-827.

21. Suntharalingham, J. P., F. Buonocore, A. J. Duncan and J. C. Achermann (2015). "DAX-1 (NR0B1) and steroidogenic factor-1 (SF-1, NR5A1) in human disease." Best Pract Res Clin Endocrinol Metab **29**(4): 607-619.

22. Crawford, P. A., C. Dorn, Y. Sadovsky and J. Milbrandt (1998). "Nuclear receptor DAX-1 recruits nuclear receptor corepressor N-CoR to steroidogenic factor 1." Mol Cell Biol **18**(5): 2949-2956.

23. Lee, Y. S., D. Chanda, J. Sim, Y. Y. Park and H. S. Choi (2007). "Structure and function of the atypical orphan nuclear receptor small heterodimer partner." Int Rev Cytol **261**: 117-158.

24. Seol, W., H. S. Choi and D. D. Moore (1996). "An orphan nuclear hormone receptor that lacks a DNA binding domain and heterodimerizes with other receptors." Science **272**(5266): 1336-1339.

25. Zhi, X., X. E. Zhou, Y. He, C. Zechner, K. M. Suino-Powell, S. A. Kliewer, K. Melcher, D. J. Mangelsdorf and H. E. Xu (2014). "Structural insights into gene repression by the orphan nuclear receptor SHP." Proc Natl Acad Sci U S A **111**(2): 839-844.

26. Heery, D. M., E. Kalkhoven, S. Hoare and M. G. Parker (1997). "A signature motif in transcriptional co-activators mediates binding to nuclear receptors." Nature **387**(6634): 733-736.

27. Borgius, L. J., K. R. Steffensen, J. A. Gustafsson and E. Treuter (2002). "Glucocorticoid signaling is perturbed by the atypical orphan receptor and corepressor SHP." J Biol Chem **277**(51): 49761-49766.

28. Johansson, L., A. Bavner, J. S. Thomsen, M. Farnegardh, J. A. Gustafsson and E. Treuter (2000). "The orphan nuclear receptor SHP utilizes conserved LXXLL-related motifs for interactions with ligand-activated estrogen receptors." Mol Cell Biol **20**(4): 1124-1133.

29. Ortlund, E. A., Y. Lee, I. H. Solomon, J. M. Hager, R. Safi, Y. Choi, Z. Guan, A. Tripathy, C. R. Raetz, D. P. McDonnell, D. D. Moore and M. R. Redinbo (2005). "Modulation of human nuclear receptor LRH-1 activity by phospholipids and SHP." Nat Struct Mol Biol **12**(4): 357-363.

30. Zhang, H., A. Slewa, E. Janssen, I. Skaland, Y. Yu, E. Gudlaugsson, W. Feng, K. Kjellevold, H. Soiland and J. P. Baak (2011). "The prognostic value of the orphan nuclear receptor DAX-1 (NROB1) in node-negative breast cancer." Anticancer Res **31**(2): 443-449.

31. Nedumaran, B., G. S. Kim, S. Hong, Y. S. Yoon, Y. H. Kim, C. H. Lee, Y. C. Lee, S. H. Koo and H. S. Choi (2010). "Orphan nuclear receptor DAX-1 acts as a novel corepressor of liver X receptor alpha and inhibits hepatic lipogenesis." J Biol Chem **285**(12): 9221-9232.

32. Brent, G. A. (2012). "Mechanisms of thyroid hormone action." J Clin Invest **122**(9): 3035-3043.

33. Tang, X. H. and L. J. Gudas (2011). "Retinoids, retinoic acid receptors, and cancer." Annu Rev Pathol **6**: 345-364.

34. Tyagi, S., P. Gupta, A. S. Saini, C. Kaushal and S. Sharma (2011). "The peroxisome proliferator-activated receptor: A family of nuclear receptors role in various diseases." J Adv Pharm Technol Res **2**(4): 236-240.

35. Kojetin, D. J. and T. P. Burris (2014). "REV-ERB and ROR nuclear receptors as drug targets." Nat Rev Drug Discov **13**(3): 197-216.

36. Li, G. and L. G. G (2015). "Farnesoid X receptor, the bile acid sensing nuclear receptor, in liver regeneration." Acta Pharm Sin B **5**(2): 93-98.

37. Hong, C. and P. Tontonoz (2014). "Liver X receptors in lipid metabolism: opportunities for drug discovery." Nat Rev Drug Discov **13**(6): 433-444.

38. Haussler, M. R. (1986). "Vitamin D receptors: nature and function." Annu Rev Nutr **6**: 527-562.

39. Evans, R. M. and D. J. Mangelsdorf (2014). "Nuclear Receptors, RXR, and the Big Bang." Cell **157**(1): 255-266.

40. Park, J. I., S. Y. Tsai and M. J. Tsai (2003). "Molecular mechanism of chicken ovalbumin upstream promoter-transcription factor (COUP-TF) actions." Keio J Med **52**(3): 174-181.

41. Watt, A. J., W. D. Garrison and S. A. Duncan (2003). "HNF4: a central regulator of hepatocyte differentiation and function." Hepatology **37**(6): 1249-1253.

42. Mangelsdorf, D. J. and R. M. Evans (1995). "The RXR heterodimers and orphan receptors." Cell **83**(6): 841-850.

43. Fuller, P. J. (1991). "The steroid receptor superfamily: mechanisms of diversity." FASEB J **5**(15): 3092-3099.

44. Gao, W., C. E. Bohl and J. T. Dalton (2005). "Chemistry and structural biology of androgen receptor." Chem Rev **105**(9): 3352-3370.

45. Jacobsen, B. M. and K. B. Horwitz (2012). "Progesterone receptors, their isoforms and progesterone regulated transcription." Mol Cell Endocrinol **357**(1-2): 18-29.

46. Weikum, E. R., M. T. Knuesel, E. A. Ortlund and K. R. Yamamoto (2017). "Glucocorticoid receptor control of transcription: precision and plasticity via allostery." Nat Rev Mol Cell Biol **18**(3): 159-174.

47. Gomez-Sanchez, E. and C. E. Gomez-Sanchez (2014). "The multifaceted mineralocorticoid receptor." Compr Physiol **4**(3): 965-994.

48. Heldring, N., A. Pike, S. Andersson, J. Matthews, G. Cheng, J. Hartman, M. Tujague, A. Strom, E. Treuter, M. Warner and J. A. Gustafsson (2007). "Estrogen receptors: how do they signal and what are their targets." Physiol Rev **87**(3): 905-931.

49. Maruyama, K., T. Tsukada, N. Ohkura, S. Bandoh, T. Hosono and K. Yamaguchi (1998). "The NGFI-B subfamily of the nuclear receptor superfamily (review)." Int J Oncol **12**(6): 1237-1243.

50. Hoivik, E. A., A. E. Lewis, L. Aumo and M. Bakke (2010). "Molecular aspects of steroidogenic factor 1 (SF-1)." Mol Cell Endocrinol **315**(1-2): 27-39.

51. Fayard, E., J. Auwerx and K. Schoonjans (2004). "LRH-1: an orphan nuclear receptor involved in development, metabolism and steroidogenesis." Trends Cell Biol **14**(5): 250-260.

52. Sablin, E. P., R. D. Blind, I. N. Krylova, J. G. Ingraham, F. Cai, J. D. Williams, R. J. Fletterick and H. A. Ingraham (2009). "Structure of SF-1 bound by different phospholipids: evidence for regulatory ligands." Mol Endocrinol **23**(1): 25-34.

53. Val, P., A. M. Lefrancois-Martinez, G. Veyssiere and A. Martinez (2003). "SF-1 a key player in the development and differentiation of steroidogenic tissues." Nucl Recept **1**(1): 8.

54. Zechel, C. (2005). "The germ cell nuclear factor (GCNF)." Mol Reprod Dev **72**(4): 550-556.

55. Wang, Q. and A. J. Cooney (2013). "Revisiting the role of GCNF in embryonic development." Semin Cell Dev Biol **24**(10-12): 679-686.

56. Okumura, L. M., B. J. Lesch and D. C. Page (2013). "The ligand binding domain of GCNF is not required for repression of pluripotency genes in mouse fetal ovarian germ cells." PLoS One **8**(6): e66062.

57. Rastinejad, F., P. Huang, V. Chandra and S. Khorasanizadeh (2013). "Understanding nuclear receptor form and function using structural biology." J Mol Endocrinol **51**(3): T1-T21.

58. Pawlak, M., P. Lefebvre and B. Staels (2012). "General molecular biology and architecture of nuclear receptors." Curr Top Med Chem **12**(6): 486-504.

59. Kumar, R. and E. B. Thompson (2003). "Transactivation functions of the N-terminal domains of nuclear hormone receptors: protein folding and coactivator interactions." Mol Endocrinol **17**(1): 1-10.

60. Lavery, D. N. and I. J. McEwan (2005). "Structure and function of steroid receptor AF1 transactivation domains: induction of active conformations." Biochem J **391**(Pt 3): 449-464.

61. Anbalagan, M., B. Huderson, L. Murphy and B. G. Rowan (2012). "Post-translational modifications of nuclear receptors and human disease." Nucl Recept Signal **10**: e001.

62. Danielsen, M. (2001). "Bioinformatics of nuclear receptors." Methods Mol Biol **176**: 3-22.

63. Gronemeyer, H. and D. Moras (1995). "Nuclear receptors. How to finger DNA." Nature **375**(6528): 190-191.

64. Helsen, C., S. Kerkhofs, L. Clinckemalie, L. Spans, M. Laurent, S. Boonen, D. Vanderschueren and F. Claessens (2012). "Structural basis for nuclear hormone receptor DNA binding." Mol Cell Endocrinol **348**(2): 411-417.

65. Hard, T., E. Kellenbach, R. Boelens, B. A. Maler, K. Dahlman, L. P. Freedman, J. Carlstedt-Duke, K. R. Yamamoto, J. A. Gustafsson and R. Kaptein (1990). "Solution structure of the glucocorticoid receptor DNA-binding domain." Science **249**(4965): 157-160.

66. Schwabe, J. W., L. Chapman, J. T. Finch and D. Rhodes (1993). "The crystal structure of the estrogen receptor DNA-binding domain bound to DNA: how receptors discriminate between their response elements." Cell **75**(3): 567-578.

67. Khorasanizadeh, S. and F. Rastinejad (2001). "Nuclear-receptor interactions on DNA-response elements." Trends Biochem Sci **26**(6): 384-390.

68. Luisi, B. F., W. X. Xu, Z. Otwinowski, L. P. Freedman, K. R. Yamamoto and P. B. Sigler (1991). "Crystallographic analysis of the interaction of the glucocorticoid receptor with DNA." Nature **352**(6335): 497-505.

69. Watson, L. C., K. M. Kuchenbecker, B. J. Schiller, J. D. Gross, M. A. Pufall and K. R. Yamamoto (2013). "The glucocorticoid receptor dimer interface allosterically transmits sequence-specific DNA signals." Nat Struct Mol Biol **20**(7): 876-883.

70. Solomon, I. H., J. M. Hager, R. Safi, D. P. McDonnell, M. R. Redinbo and E. A. Ortlund (2005). "Crystal structure of the human LRH-1 DBD-DNA complex reveals Ftz-F1 domain positioning is required for receptor activity." J Mol Biol **354**(5): 1091-1102.

71. Weikum, E. R., M. L. Tuntland, M. N. Murphy and E. A. Ortlund (2016). "A Structural Investigation into Oct4 Regulation by Orphan Nuclear Receptors, Germ Cell Nuclear Factor (GCNF), and Liver Receptor Homolog-1 (LRH-1)." J Mol Biol **428**(24 Pt B): 4981-4992.

72. Moras, D. and H. Gronemeyer (1998). "The nuclear receptor ligand-binding domain: structure and function." Curr Opin Cell Biol **10**(3): 384-391.

73. Weatherman, R. V., R. J. Fletterick and T. S. Scanlan (1999). "Nuclear-receptor ligands and ligand-binding domains." Annu Rev Biochem **68**: 559-581.

74. Wurtz, J. M., W. Bourguet, J. P. Renaud, V. Vivat, P. Chambon, D. Moras and H. Gronemeyer (1996). "A canonical structure for the ligand-binding domain of nuclear receptors." Nat Struct Biol **3**(1): 87-94.

75. Nolte, R. T., G. B. Wisely, S. Westin, J. E. Cobb, M. H. Lambert, R. Kurokawa, M. G. Rosenfeld, T. M. Willson, C. K. Glass and M. V. Milburn (1998). "Ligand binding and co-activator assembly of the peroxisome proliferator-activated receptor-gamma." Nature **395**(6698): 137-143.

76. Johnson, B. A., E. M. Wilson, Y. Li, D. E. Moller, R. G. Smith and G. Zhou (2000). "Ligand-induced stabilization of PPARgamma monitored by NMR spectroscopy: implications for nuclear receptor activation." J Mol Biol **298**(2): 187-194.

77. Nagy, L. and J. W. Schwabe (2004). "Mechanism of the nuclear receptor molecular switch." Trends Biochem Sci **29**(6): 317-324.

78. Keidel, S., P. LeMotte and C. Apfel (1994). "Different agonist- and antagonist-induced conformational changes in retinoic acid receptors analyzed by protease mapping." Mol Cell Biol **14**(1): 287-298.

79. Pissios, P., I. Tzameli, P. Kushner and D. D. Moore (2000). "Dynamic stabilization of nuclear receptor ligand binding domains by hormone or corepressor binding." Mol Cell **6**(2): 245-253.

80. Gee, A. C. and J. A. Katzenellenbogen (2001). "Probing conformational changes in the estrogen receptor: evidence for a partially unfolded intermediate facilitating ligand binding and release." Mol Endocrinol **15**(3): 421-428.

81. Kallenberger, B. C., J. D. Love, V. K. Chatterjee and J. W. Schwabe (2003). "A dynamic mechanism of nuclear receptor activation and its perturbation in a human disease." Nat Struct Biol **10**(2): 136-140.

82. Darimont, B. D., R. L. Wagner, J. W. Apriletti, M. R. Stallcup, P. J. Kushner, J. D. Baxter, R. J. Fletterick and K. R. Yamamoto (1998). "Structure and specificity of nuclear receptor-coactivator interactions." Genes Dev **12**(21): 3343-3356.

83. He, Y., W. Yi, K. Suino-Powell, X. E. Zhou, W. D. Tolbert, X. Tang, J. Yang, H. Yang, J. Shi, L. Hou, H. Jiang, K. Melcher and H. E. Xu (2014). "Structures and mechanism for the design of highly potent glucocorticoids." Cell Res **24**(6): 713-726.

84. Brzozowski, A. M., A. C. Pike, Z. Dauter, R. E. Hubbard, T. Bonn, O. Engstrom, L. Ohman, G. L. Greene, J. A. Gustafsson and M. Carlquist (1997). "Molecular basis of agonism and antagonism in the oestrogen receptor." Nature **389**(6652): 753-758.

85. Mi, L. Z., S. Devarakonda, J. M. Harp, Q. Han, R. Pellicciari, T. M. Willson, S. Khorasanizadeh and F. Rastinejad (2003). "Structural basis for bile acid binding and activation of the nuclear receptor FXR." Mol Cell **11**(4): 1093-1100.

86. Williams, S., R. K. Bledsoe, J. L. Collins, S. Boggs, M. H. Lambert, A. B. Miller, J. Moore, D. D. McKee, L. Moore, J. Nichols, D. Parks, M. Watson, B. Wisely and T. M. Willson (2003). "X-ray crystal structure of the liver X receptor beta ligand binding domain: regulation by a histidine-tryptophan switch." J Biol Chem **278**(29): 27138-27143.

87. Egawa, D., T. Itoh, Y. Akiyama, T. Saito and K. Yamamoto (2016). "17-OxoDHA Is a PPARalpha/gamma Dual Covalent Modifier and Agonist." ACS Chem Biol **11**(9): 2447-2455.

88. Schwabe, J. W. and D. Rhodes (1991). "Beyond zinc fingers: steroid hormone receptors have a novel structural motif for DNA recognition." Trends Biochem Sci **16**(8): 291-296.

89. Lefstin, J. A., J. R. Thomas and K. R. Yamamoto (1994). "Influence of a steroid receptor DNA-binding domain on transcriptional regulatory functions." Genes Dev **8**(23): 2842-2856.

90. Schwabe, J. W., L. Chapman and D. Rhodes (1995). "The oestrogen receptor recognizes an imperfectly palindromic response element through an alternative side-chain conformation." Structure **3**(2): 201-213.

91. Lefstin, J. A. and K. R. Yamamoto (1998). "Allosteric effects of DNA on transcriptional regulators." Nature **392**(6679): 885-888.

92. Meijsing, S. H., M. A. Pufall, A. Y. So, D. L. Bates, L. Chen and K. R. Yamamoto (2009). "DNA binding site sequence directs glucocorticoid receptor structure and activity." Science **324**(5925): 407-410.

93. Rastinejad, F., T. Wagner, Q. Zhao and S. Khorasanizadeh (2000). "Structure of the RXR-RAR DNA-binding complex on the retinoic acid response element DR1." EMBO J **19**(5): 1045-1054.

94. Rastinejad, F., T. Perlmann, R. M. Evans and P. B. Sigler (1995). "Structural determinants of nuclear receptor assembly on DNA direct repeats." Nature **375**(6528): 203-211.

95. Shaffer, P. L. and D. T. Gewirth (2002). "Structural basis of VDR-DNA interactions on direct repeat response elements." EMBO J **21**(9): 2242-2252.

96. Little, T. H., Y. Zhang, C. K. Matulis, J. Weck, Z. Zhang, A. Ramachandran, K. E. Mayo and I. Radhakrishnan (2006). "Sequence-specific deoxyribonucleic acid (DNA) recognition by steroidogenic factor 1: a helix at the carboxy terminus of the DNA binding domain is necessary for complex stability." Mol Endocrinol **20**(4): 831-843.

97. Bledsoe, R. K., V. G. Montana, T. B. Stanley, C. J. Delves, C. J. Apolito, D. D. McKee, T. G. Consler, D. J. Parks, E. L. Stewart, T. M. Willson, M. H. Lambert, J. T. Moore, K. H. Pearce and H. E. Xu (2002). "Crystal structure of the glucocorticoid receptor ligand binding domain reveals a novel mode of receptor dimerization and coactivator recognition." Cell **110**(1): 93-105.

98. Svensson, S., T. Ostberg, M. Jacobsson, C. Norstrom, K. Stefansson, D. Hallen, I. C. Johansson, K. Zachrisson, D. Ogg and L. Jendeberg (2003). "Crystal structure of the

heterodimeric complex of LXRalpha and RXRbeta ligand-binding domains in a fully agonistic conformation." EMBO J **22**(18): 4625-4633.

99. Millard, C. J., P. J. Watson, L. Fairall and J. W. Schwabe (2013). "An evolving understanding of nuclear receptor coregulator proteins." J Mol Endocrinol **51**(3): T23-36.
100. Hu, X. and M. A. Lazar (1999). "The CoRNR motif controls the recruitment of corepressors by nuclear hormone receptors." Nature **402**(6757): 93-96.
101. Webb, P., C. M. Anderson, C. Valentine, P. Nguyen, A. Marimuthu, B. L. West, J. D. Baxter and P. J. Kushner (2000). "The nuclear receptor corepressor (N-CoR) contains three isoleucine motifs (I/LXXII) that serve as receptor interaction domains (IDs)." Mol Endocrinol **14**(12): 1976-1985.
102. Nagy, L., H. Y. Kao, J. D. Love, C. Li, E. Banayo, J. T. Gooch, V. Krishna, K. Chatterjee, R. M. Evans and J. W. Schwabe (1999). "Mechanism of corepressor binding and release from nuclear hormone receptors." Genes Dev **13**(24): 3209-3216.
103. Perissi, V., L. M. Staszewski, E. M. McInerney, R. Kurokawa, A. Krones, D. W. Rose, M. H. Lambert, M. V. Milburn, C. K. Glass and M. G. Rosenfeld (1999). "Molecular determinants of nuclear receptor-corepressor interaction." Genes Dev **13**(24): 3198-3208.
104. Egea, P. F., A. Mitschler and D. Moras (2002). "Molecular recognition of agonist ligands by RXRs." Mol Endocrinol **16**(5): 987-997.
105. Bourguet, W., M. Ruff, P. Chambon, H. Gronemeyer and D. Moras (1995). "Crystal structure of the ligand-binding domain of the human nuclear receptor RXR-alpha." Nature **375**(6530): 377-382.
106. Musille, P. M., B. R. Kossmann, J. A. Kohn, I. Ivanov and E. A. Ortlund (2016). "Unexpected Allosteric Network Contributes to LRH-1 Co-regulator Selectivity." J

Biol Chem **291**(3): 1411-1426.

107. Bruning, J. B., M. J. Chalmers, S. Prasad, S. A. Busby, T. M. Kamenecka, Y. He, K. W. Nettles and P. R. Griffin (2007). "Partial agonists activate PPARgamma using a helix12 independent mechanism." Structure **15**(10): 1258-1271.
108. Simons, S. S., Jr., D. P. Edwards and R. Kumar (2014). "Minireview: dynamic structures of nuclear hormone receptors: new promises and challenges." Mol Endocrinol **28**(2): 173-182.
109. O'Malley, B. W. and M. J. Tsai (1992). "Molecular pathways of steroid receptor action." Biol Reprod **46**(2): 163-167.
110. Pratt, W. B., M. D. Galigniana, Y. Morishima and P. J. Murphy (2004). "Role of molecular chaperones in steroid receptor action." Essays Biochem **40**: 41-58.
111. Acevedo, M. L. and W. L. Kraus (2004). "Transcriptional activation by nuclear receptors." Essays Biochem **40**: 73-88.
112. Bulynko, Y. A. and B. W. O'Malley (2011). "Nuclear receptor coactivators: structural and functional biochemistry." Biochemistry **50**(3): 313-328.
113. Gadaleta, R. M. and L. Magnani (2014). "Nuclear receptors and chromatin: an inducible couple." J Mol Endocrinol **52**(2): R137-149.
114. Glass, C. K. and M. G. Rosenfeld (2000). "The coregulator exchange in transcriptional functions of nuclear receptors." Genes Dev **14**(2): 121-141.
115. Auger, A. P. and H. M. Jessen (2009). "Corepressors, nuclear receptors, and epigenetic factors on DNA: a tail of repression." Psychoneuroendocrinology **34 Suppl 1**: S39-47.
116. Hudson, W. H., C. Youn and E. A. Ortlund (2013). "The structural basis of direct glucocorticoid-mediated transrepression." Nat Struct Mol Biol **20**(1): 53-58.

117. Surjit, M., K. P. Ganti, A. Mukherji, T. Ye, G. Hua, D. Metzger, M. Li and P. Chambon (2011). "Widespread negative response elements mediate direct repression by agonist- liganded glucocorticoid receptor." Cell **145**(2): 224-241.

118. Hua, G., L. Paulen and P. Chambon (2016). "GR SUMOylation and formation of an SUMO-SMRT/NCoR1-HDAC3 repressing complex is mandatory for GC-induced IR nGRE-mediated transrepression." Proc Natl Acad Sci U S A **113**(5): E626-634.

119. Sladek, F. M. (2003). "Nuclear receptors as drug targets: new developments in coregulators, orphan receptors and major therapeutic areas." Expert Opin Ther Targets **7**(5): 679-684.

120. Kendall, E. C. (1951). "The development of cortisone as a therapeutic agent." Indian Med J **45**(10): 239-241.

121. Becker, D. E. (2013). "Basic and clinical pharmacology of glucocorticosteroids." Anesth Prog **60**(1): 25-31; quiz 32.

122. Bermudez, V., F. Finol, N. Parra, M. Parra, A. Perez, L. Penaranda, D. Vilchez, J. Rojas, N. Arraiz and M. Velasco (2010). "PPAR-gamma agonists and their role in type 2 diabetes mellitus management." Am J Ther **17**(3): 274-283.

123. Burris, T. P., L. A. Solt, Y. Wang, C. Crumbley, S. Banerjee, K. Griffett, T. Lundasen, T. Hughes and D. J. Kojetin (2013). "Nuclear receptors and their selective pharmacologic modulators." Pharmacol Rev **65**(2): 710-778.

124. Clark, A. R. and M. G. Belvisi (2012). "Maps and legends: the quest for dissociated ligands of the glucocorticoid receptor." Pharmacol Ther **134**(1): 54-67.

125. Diez-Perez, A. (2006). "Selective estrogen receptor modulators (SERMS)." Arq Bras Endocrinol Metabol **50**(4): 720-734.

CHAPTER 2: GLUCOCORTICOID RECEPTOR CONTROL OF TRANSCRIPTION: PRECISION AND PLASTICITY VIA ALLOSTERY

Emily R. Weikum[1], Matthew T. Knuessel[2], Eric A. Ortlund[1], Keith R. Yamamoto[21]

1) Department of Biochemistry, Emory University School of Medicine, Atlanta GA 30322 USA

2) Department of Cellular and Molecular Pharmacology, University of California San Francisco School of Medicine, San Francisco, CA 94143, USA

Weikum, ER, Knuessel MT, Ortlund EA, Yamamoto, KR (2017) Glucocorticoid receptor control of transcription: Precision and plasticity via allostery. NRMCB 18(3).

The glucocorticoid receptor is a ligand-regulated transcription factor that controls gene expression in a variety of biological processes. We sought to expand the current simplistic view of GR signaling by reviewing the complexity of this transcription factor. GR regulates specific gene expression profiles in a context dependent manner, which is governed by numerous allosteric effectors including chromatin state, DNA sequence, and post-translational modifications. These allosteric modifications result in different GR surfaces that are read by coregulator proteins, driving alternative transcriptional outcomes. This review was accepted for publication in Nature Reviews Molecular and Cellular Biology in December 2016.

---

[a] Wrote the paper: ERW MTK EAO KRY.

Abstract

The glucocorticoid receptor (GR) is a constitutively expressed transcriptional regulatory factor (TRF) that controls many distinct gene networks, each uniquely determined by particular cellular and physiologic contexts. The precision of GR-mediated responses appears to depend on combinatorial, context-specific assembly of GR-nucleated transcriptional regulatory complexes at genomic response elements. In turn, evidence suggests that context-driven plasticity is conferred by integration of multiple signals, each serving as an allosteric effector of GR conformation, a key determinant of regulatory complex composition and activity. This structural and mechanistic perspective on GR regulatory specificity likely extends to other eukaryotic transcriptional regulatory factors.

Introduction

Control of gene transcription is critical for development, physiology, and homeostasis. Thus, aberrant transcriptional regulation commonly drives disease processes. Transcriptional regulatory factors (TRFs) play a critical role in this process by recognizing specific DNA sequences to activate or repress expression of specific genes. One of the most well-characterized metazoan TRFs is the glucocorticoid receptor (GR), the founding member of the nuclear receptor superfamily, members of which evolved to bind specific small lipophilic signaling molecules[1].

Expressed in virtually all vertebrate cells, GR directly up- and down-regulates thousands of genes distinct to the cell type, governing various aspects of development, metabolism, stress response, inflammation, and other key tissue and organismal processes. GR is encoded by the *NR3C1* gene located on chromosome 5 (5q31) and is closely related to its paralogs *NR3C2* the mineralocorticoid receptor (MR), *NR3C3* the progesterone receptor (PR), and *NR3C4* the androgen receptor (AR). These four nuclear receptors share a common domain structure consisting of an N-terminal domain (NTD), a zinc finger DNA binding domain (DBD), a hinge region, and a C-terminal ligand binding domain (LBD). Embedded within these domains are regions that confer regulatory activity; for GR, these are denoted as AF1/Tau1/τ1, Tau2/τ2, and AF2 (**Figure 2.1a**). While the DBD and LBD are highly conserved between GR, MR, PR, and AR, the N-terminal AF1 domains and the surrounding NTD regions within these four genes are more divergent in sequence and size. In addition, alternative splicing and translational start sites produce multiple isoforms of the NR3C family members. The predominant and best-studied isoform of GR, GRα, is a 777 amino acid polypeptide in humans. Other less abundant and less well characterized, but likely functional,

isoforms of GR have been described and are reviewed in[2,3]. For example, a GRβ isoform carries LBD alterations that change its ligand binding properties and may compete with GRα in certain experimental settings[4].

GR activity is gated by the endogenous steroid hormone cortisol in humans, or by exogenous glucocorticoid drugs, such as dexamethasone. In the absence of ligand, apo-GR is monomeric in the cytoplasm, where it associates with molecular chaperone complexes containing heat shock protein 90 (Hsp90), heat shock protein 70 (Hsp70), and other factors[5]; this interaction promotes high-affinity hormone binding while inactivating other receptor activities, such as nuclear localization and DNA binding. Glucocorticoid binding provokes GR conformational changes that render multiple functional domains active, including nuclear localization sequences within the hinge and LBD regions. After translocation into the nucleus, GR associates with specific genomic glucocorticoid response elements (GREs)[6] and nucleates assembly of transcriptional regulatory complexes, comprised of GR, other TRFs, and coregulatory factors, which together activate or repress transcription of glucocorticoid-responsive genes[7,8] (**Figure 2.1b**). Although GR forms stable complexes with DNA *in vitro*, it appears to exchange within seconds *in vivo*[9,10], implying that distinct GR molecules establish initial genomic contact and regulate transcription, and that GR is actively disassembled from chromatin *in vivo*[11,12]. The role of this striking *in vivo* dynamics in the mechanisms of regulation is not known.

GR operates context-specifically - it regulates gene networks that are precisely determined in a given context, yet displays remarkable plasticity as a function of cell type and physiological state (recently reviewed in[13]), leading to diverse outcomes. For example, GR-mediated gene expression governs apoptosis in the context of hematopoietic T cells[14] but

increases adipogenesis, lipolysis, and differentiation in adipose cells[15]. How can both precision and plasticity of GR-regulated transcription be achieved? Addressing this apparent paradox of precision and plasticity is in fact the overarching challenge for all of eukaryotic transcriptional regulation. GR provides a striking framework in which to address the challenge because it is expressed ubiquitously in vertebrate cells, and the GR-regulated gene networks are strongly cell type-specific.

Here, we first discuss how GR interacts with the genome *in vivo* and *in vitro*, highlighting the context-specificity of the *in vivo* interactions. Second, we document the importance of context in specifying GR activity, and discuss how different "surfaces" appear to define the regulatory logic of GR function. Third, we consider how four classes of signals, including DNA binding, ligand binding, post-translational modifications (PTMs), and interactions with other, non-GR transcription factors are integrated to impact GR structure and function. Fourth, we examine multiple classes of coregulatory factors that associate with GR interaction surfaces to assemble transcriptional regulatory complexes and impose enzymatic actions that modulate transcription. Finally, we present a model that accounts for the precision and plasticity of eukaryotic transcriptional regulation. In this model, TRFs act as scaffolds whose conformations are altered allosterically by signaling inputs, while coregulators serve both as readers that associate with surfaces induced on TRF scaffolds by those allosteric signals, and as enzymes that modulate target gene transcription.

GR-genome interactions

TRFs, including GR, obtain a portion of their regulatory specificity by interacting with specific genomic loci. These interactions can either be direct GR–DNA contacts, or GR can associate with other transcriptional regulators that are themselves bound to DNA.

Specific genomic occupancy by a TRF is typically highly context specific, with patterns of binding differing substantially in distinct cellular and physiological settings. Importantly, locus-specific GR binding is not a sufficient determinant of regulatory activity. In this section we will focus on outlining the direct and indirect interactions GR establishes with the genome (**Figure 2.2**); see Supplementary information S1 (**Table 2.S1**) for a summary of experimental techniques referred to in this section

Direct and indirect sequence-specific binding in vitro.

DNA binding domains of GR and other nuclear receptors contain two highly conserved subdomains, each with four cysteine residues coordinating a single Zn ion, followed by an amphipathic helix and a peptide loop. The helix of the first subdomain contains the "P box", bearing three residues that make base-specific contacts in the major groove of the binding sequence. The second subdomain helix makes nonspecific contacts with DNA helix backbone and minor groove, whereas the peptide loop provides "D box" residues important for GR dimerization[16,17].

The DBD positions GR at specific genomic sites by at least three classes of GR–DNA interactions, and at least one class of GR–protein interaction (**Figure 2.2A**). The best characterized GR–DNA interaction is through the canonical GR binding sequence (GBS) (**Figure 2.2Aa,B,D,F**), which is composed of two pseudo-palindromic hexameric AGAACA repeats, separated by a three base pair spacer[17]. The 'DNA reading helix' utilizes the side chains of Arg447, Lys442, and Val443, to make three base-specific contacts within the major groove of each GBS half site (**Figure 2.2B, D**). GR binding in this head-to-head fashion creates interactions between two sister GR-DBDs, which promotes GR–GR and GR–DNA

interactions, creating positive cooperativity[18,19]. Five amino acids within the D box provide critical protein–protein contacts, stabilizing the GR-DBD dimer on DNA. Mutational disruption of the D-loop conformation can affect GR's regulatory activity[20]. For example, Ala458 makes a hydrogen bond with Ile483 on the dimer partner, and mutation of the Ala to Thr has been shown to alter GR activity in a gene-specific manner[20–22].This mutation, though initially believed to be devoid of dimerization potential, still forms dimers on DNA but with diminished cooperativity[19]. In addition, a relatively weak LBD–LBD interaction (1.5 μM using GR-LBD fragments in the presence of ligand and coregulator peptide)[23] may contribute to GR dimerization, but the biological significance of this contact has not been established.

The second class of GR–DNA interaction is a recently discovered Inverted Repeat GBS (IR-GBS) (**Figure 2.2Ab,C,E**), characterized as an alternative GR-binding motif containing $CTCC(N)_{0–2}GGAGA$[24]. Structural studies of the GR-DBD–IR-GBS revealed that GR molecules bound these non-identical sites on opposite sides of DNA, separated by a one base pair spacer in a head-to-tail fashion[25] (**Figure 2.2C**). One monomer makes three contacts within a high-affinity binding site mediated by Lys442 and Val443, which are the same side chains that participate in recognizing the DNA in the GR-DBD–GBS structure. Arg447, however, establishes hydrogen bonds with a guanine in canonical GBS structures, but in the IR-GBS structure is prevented from making these base-specific interactions due to steric clashes with a thymine[25]. The other monomer uses only Arg447 to make one base-specific contact to a guanine. As the two GR monomers bind to opposite sides of the DNA at IR-GBSs, they are not in direct contact through the DBD dimerization interface.

Biochemical analyses reveal that GR binds IR-GBS elements with negative cooperativity whereby binding of one monomer dramatically reduces the propensity of a

second monomer to bind. This contrasts strongly with the cooperative binding seen at canonical GBS[25]. Furthermore, nuclear magnetic resonance (NMR) studies indicate that the dimerization loop residues display significant changes in chemical environment, consistent with dimerization when bound to a canonical GBS, whereas they are unaffected upon binding to an IR-GBS[26]. As with canonical GBSs, it has not been examined whether LBD–LBD interactions participate in GR binding to IR-GBSs. Taken together, we infer from these *in vitro* studies that monomeric GR most likely binds these elements *in vivo*. Other 3-keto steroid receptors, MR, PR, and AR, all bind a canonical GBS but only GR is able to bind a IR-GBS[26]. Even the MR-DBD, which shares 90% sequence identity with the GR-DBD, cannot bind or repress transcription from an IR-GBS due to epistatic mutations that occurred during the evolution of the steroid receptor family that limit this function in all steroid receptors except GR[26].

In a third class of GR–DNA interactions (**Figure 2.2Ac**), not yet characterized structurally, selective binding of GR to canonical half site DNA sequences (consensus AGAACA) has been reported[21,27]. This binding may be facilitated by secondary interactions between GR and other non-GR TRFs bound to DNA proximal to the GBS, although there is no evidence for enrichment of particular TRF motifs contiguous with these half site GBSs, the understanding of which TRFs are involved in the regulation of gene expression at such composite GREs is still very limited. Finally, GR can occupy specific genomic regions without directly binding DNA (**Figure 2.2Ad**). In this mechanism, known as tethering, the DBD makes protein–protein contacts with other TRFs, such as AP-1 or NF-κB, specifically bound at their cognate sequence motifs[28–31] (**Figure 2.2Ad**).

Context-specific genomic occupancy in vivo.

The four classes of sequence-specific GR–DNA interactions characterized *in vitro* (**Figure 2.2A**) occur at high frequency in mammalian genomes. Chromatin immunoprecipitation followed by deep sequencing (ChIP-seq)[32] is a robust, sensitive and fairly accurate (albeit not to single nucleotide resolution) method to identify genomic segments occupied by GR *in vivo* (for details of the technique see Supplementary information S1 (**Table 2.S1**) here termed GR occupied regions (GORs). It is clear from these approaches that GORs, while numerous, overlie only a small fraction of the potential sites predicted from *in vitro* studies, and that they are highly context specific. For example, only 0.5% of 11,666 GORs identified in mouse liver were found in common in four other mouse cell types examined[33,34], and 83% were unique to liver[35]. While differences in methodological and statistical approaches may complicate those meta-analyses dramatically different GORs among different cell types have been commonly noted[13].

GORs typically reside proximal both to GR regulated genes and to genes not regulated by GR[21,36], so mere proximity is uninformative either about regulatory function or about which genes are regulated by which functional GORs. What is clear is that GORs do not necessarily regulate the gene closest in linear proximity within the genome, and that non-GR-regulated genes commonly reside between a GOR and the nearest GR-regulated gene. Provisional evidence suggests that most GORs reside >10 kb from a GC-responsive gene, and that many GORs are 10-100 kb or more upstream, downstream, and within introns[37] relative to the transcription start sites of genes they are thought to regulate (**Figure 2.3a**). Such action-at-a distance is well-documented, heavily studied, and virtually not understood for metazoan transcriptional regulators. How particular GREs and promoters interact and

what drives specificity of one GRE–promoter interaction over others is a focus of current debate, but one or more of several proposed gene looping mechanisms (reviewed in[38]) are likely involved.

Clearly, the determinants of GOR context-specificity are not understood, but some positive and negative correlates have been uncovered. For example, DNase I hypersensitive sites (DHSs) (see Supplementary information S1 (**Table 2.S1**)), which are thought to identify "open chromatin" regions depleted of nucleosomes, but occupied by TRFs and other non-nucleosomal proteins, were found in whole genome analyses in multiple cell types to pre-exist at up to 95% of GORs formed upon hormone treatment[34]. DHSs, of course, vastly outnumber GORs (DHS ~2.5% of the genome, GORs 0.02 to 0.05% of the genome in the cell types tested)[34]; it is interesting to speculate that DHSs at eventual GORs reflect non-GR TRFs pre-bound at eventual composite GREs. Similar results were obtained using Formaldehyde-Assisted Isolation of Regulatory Elements and deep sequencing (FAIRE seq - See Supplementary information S1 (**Table 2.S1**)). In this case, pre-existing FAIRE seq signal indicative of open chromatin, increased upon glucocorticoid stimulation, suggesting that GR interacts with pre-existing open chromatin, and then further alters the chromatin environment[39]. In contrast, negative regulatory DNA sequence (NRS) motifs were identified proximal to a subset of GBs as anti-correlates to GR genomic occupancy. Here, GBSs that were not occupied by GR via ChIP-seq experiments were found to have an over-representation of proximal NRS motifs[40]. Interestingly, NRSs did not affect DNase I sensitivity. Rather, NRSs appeared to be enriched in para-speckle proteins that were speculated to repress transcription by blocking GR from binding to its canonical site.

GRE context and regulatory logic

The context specificity of GORs demonstrates that DHSs, recognition sequences (GBSs) identified *in vitro*, and non-GR TRFs with which GR interacts are insufficient to localize or predict localization of GR *in vivo.* A dramatic example is the IR-GBS-containing GORs, widespread and highly occupied in mouse fibroblasts[24], but detected sparingly in certain other cell types[41–43]. The relationship of a GOR to a functional GRE is similarly complicated by context. Transient reporter assays, in which a plasmid bearing a genomic fragment underlying a GOR is inserted adjacent to a minimal promoter and reporter (typically luciferase) gene, is transfected into a GR-expressing cell line (see Supplementary information (**Table 2.S1**)), have been widely used to assess GOR regulatory capacity, and therefore to infer GRE activity. However, the transient conditions for a newly introduced plasmid are surely different from the chromosomal and cellular settings of the endogenous GOR, and the strong context-specificity of GR-mediated gene regulation raises concerns about the validity of GRE activity inferred using that approach.

Genome editing using zinc finger nucleases or transcription activator-like effector nucleases (TALENs) provide, in principle, routes for assaying GRE activity *in situ.* Unfortunately, the technical complexity of these methods together with incomplete appreciation of the overriding importance of context, has left those approaches largely unused. As a result, only one single GRE has been validated at its endogenous locus *in vivo*. In that case, a short deletion introduced into the first intron of the mouse circadian clock gene *Per2* fortuitously covered a GOR ~25kb downstream from the transcription start site, and produced allele-specific loss of glucocorticoid-mediated induction of *Per2* expression in mesenchymal stem cells[44]. With the emergence of CRISPR-Cas technologies[45], precise

genome editing will rapidly become the gold standard for validation of GRE activity in any chosen cell type- or physiological-context. Interestingly, underscoring the importance of context, the *Per2* GRE identified in mouse mesenchymal stem cells appears not to be functional in human lung carcinoma cells, as assessed by CRISPR-driven deletion (K. Ehmsen, pers. comm.).

As mentioned above, the genes and gene networks regulated by GR within different cell types are distinct. For example, **Figure 2.3b** shows the vastly different sets of target genes regulated by the synthetic glucocorticoid in two different cancer cell lines. Most GR target genes are unique to each cell line, demonstrating striking cell-context specificity. Indeed, 62 genes regulated in both cells lines are differentially responsive (activated in one cell line, repressed in the other) showing that even genes regulated by GR in both cell lines can be controlled by distinct mechanisms.

Genetic strategies have also been used to identify GR surfaces that are engaged in the regulation of gene expression in different contexts. For instance, a triple point mutant, denoted 30iiB (E198K/F199L/W213R) abrogated activation by the AF1 domain, and single point mutants within the DBD dimerization domain (A458T), and AF2 (E755R) knocked out activation by those domains individually for genes tested in U2OS cells. These three "surfaces", together with a naturally occurring splice variant, GRγ, which inserts an a single amino acid (R452) into the lever arm region of the DBD, were employed in U2OS cells in gene-specific patterns. For example, GR-mediated activation of one set of genes depended solely on AF2, whereas another set employed a combination of AF1, DBD and AF2[21,46,47]. As each altered GR region contains protein interaction domains, it is reasonable to assume

that the different patterns of domain utilization reflect assembly of distinct regulatory complexes at the GREs controlling the different sets of genes.

Beginning with context determinants defined in the genome and GR itself, what can be inferred about the "regulatory logic" that confers such specificity, yet permits facile plasticity when conditions are altered? By manipulating ligand dose, duration of treatment, and other parameters that affect relative levels of GR activity, one study uncovered "toggle switch"-like behavior in U2OS cell GREs in which a set of genes was activated at low levels of GR activity, and underwent a dramatic shift to repression at a transition point as GR activity increased[48]. This stereotyped regulatory mode has been denoted as incoherent feed forward loop type 1 (I1-FFL) logic[49], and suggests that GREs that regulate these differentially responsive genes initiate two arms of a regulatory circuit, yielding net activation or repression as the level of GR activity is altered by various context determinants. Different feed forward loop network motifs initiated by GR have been described in macrophages[50]. These findings imply that some, perhaps many, GREs may contain molecular switches that readily confer plasticity in a highly context-sensitive manner. Thus, the GRE appears to emerge as the regulatory logic module, driven by multiple signals and conferring specificity while enabling plasticity. Below we suggest a conceptual and mechanistic basis for such a 'regulatory logic' for GR and likely other eukaryotic TRFs.

Allosteric effectors of GR

If, as outlined above, regulatory specificity is determined by assembly of distinct regulatory complexes, each attuned to a particular gene-, cell- and physiologic-context, what are the molecular determinants of context, and how are their effects communicated to GR?

Extending a few examples into a general conclusion, current data suggest that four classes of cellular signals interact with GR, allosterically altering its conformation. The first two classes, the hormonal ligands and covalent PTMs, represent the endpoints of various signal transduction pathways that communicate physiologic context to the receptor. A third signal class is the extensive array of DNA sequences bound specifically by GR; these differ at different GREs, and thus these signals impart gene-specific context to the receptor. Finally, the particular cell-specific repertoire of transcription factors, including the variability in levels of their expression and activities, are reflections of cell type. Thus, the differences in the expression of transcription factors that interact with GR convey cell-specific, allosteric signals. As described above, GR:TRF interactions occur at composite GREs, and their special subset, the tethering GREs.

DNA binding sequences.

High-throughput studies have identified thousands of proteins that interact with DNA[51] and TRFs are the most common group with sequence-specific DNA binding activity. Specific DNA sequences serve not only as platforms for binding, but appear also to act as direct allosteric effectors of transcription factors[18,19,52].

Although the oligomeric state of GR at different points in signalling *in vivo* (monomeric, dimeric or higher order) is a matter of current debate[42,43,53,54], it is established that full-length ligand-bound GR is a monomer in solution *in vitro*, even at high concentrations[55,56]. At canonical GBSs, DNA binding increases the local concentration and favorable orientation of protein–protein interactions between two DBDs resulting in productive dimerization[17]. Indeed, GR not only makes sequence specific interactions with

each hexameric half site of the DNA, but the 3bp spacer between the half sites, at least at certain GBSs, drives cooperative DNA binding and dimerization[18]; moreover, allosteric changes provoked by one half site sequence can be transduced intermolecularly across the dimer interface, conformationally altering the dimer partner[19]. Spacer sequences can alter DNA shape, resulting in conformational changes that originate from the DNA-reading helix, allosterically propagate through the 'lever arm', and alter the conformation of the receptor's D box, ultimately affecting GR transcriptional activity[19]. Such conformational distinctions can be seen in comparisons of different GR-DBD–GBS structures **(Figure 2.2F)**. Finally, recent work has shown that sequence at the +8 and -8 positions flanking the GBS that alter DNA conformation also affect GR-DBD structure, as assessed by NMR analysis *in vitro,* and *in vivo* using a zinc finger nuclease generated, genomically integrated, GBS reporter system and endogenous GR[57]. Collectively, these studies suggest that DNA sequence-specific conformational states of GR result in the generation or stabilization of distinct patterns of GR surfaces, which serve as interaction platforms, driving alternative transcriptional outcomes.

Additionally, NMR analyses revealed allosteric communications between the dimerized GR monomers at canonical GBSs[19], and molecular dynamics simulations (for further details of the method see Supplementary information **(Table 2.S1)**) suggested allosteric regulation between monomers at IR-GBSs (although this might not be relevant *in vivo*, as IR-GBS elements may be occupied by only monomeric GR in the *in vivo* scenario[26]). In both cases, DNA acts as a ligand to impart allosteric changes that could affect affinity for coregulators and ultimately regulatory outcomes, implying that the allosteric transitions extend into the N-terminal domain and LBD. So far, structural studies of GR in the context of allosteric modulation by DNA have been limited to isolated domains, but work

with full-length vitamin D receptor (VDR) and retinoid X receptor (RXR) heterodimer confirm allosteric communication throughout NR complexes upon DNA binding[58]. Indeed, hydrogen deuterium exchange mass spectrometry (HDX-MS) on the full-length liganded VDR–RXR–DNA complex revealed conformation changes upon binding ligands, DNA, and coregulators. Changes within the VDR DNA binding sequence (VBS) read through the VDR-DBD had far-reaching intramolecular allosteric effects that altered solvent accessibility of regions within the sister LBD of the complexed RXR molecule[58]. Future work with full-length GR will be essential to fully describe the allosteric consequences of differential GBS binding.

LBD-binding ligands.

The NR3C family members: GR, MR, PR, and AR evolved from a common ancestral gene and share high sequence similarity within their LBDs; however, key architectural differences of each LBD produce strict ligand selectivity[59]. Differences in the aromatization of the A ring of the ligand (See **Figure 2.4a** for ligand ring naming and carbon numbering), driven in part by the 3-oxo group interaction with polar residues within NR3 family LBDs distinguish endogenous NR3C family member 3-keto ligands from related estrogen receptor (ER) 3-hydroxy specific ligand recognition[59,60] (**Figure 2.4a**). The first crystal structure of the GR-LBD complexed with a ligand (dexamethasone) displayed an intricate network of polar and non-polar interactions defines ligand selectivity - every polar atom within dexamethasone directly interacts with the GR-LBD[23].

Despite strong binding specificity, dexamethasone occupies only ~65% of the GR ligand binding pocket (leaving >200 $Å^3$ excess volume within the 590 $Å^3$ binding pocket[23]);

similarly, the endogenous ligand cortisol binds specifically, but fails to fill the binding pocket[61] (**Figure 2.4b**). The additional volume within the binding pocket offers potential space for interaction with alternative modulatory ligands. Moreover, LBD structures bound to cortisol and an alternate GR ligand, RU-486, (**Figure 2.4c,d**) reveal extensive structural malleability within the LBD that appears to enable interaction with a wide range of potential ligands[62] that could confer different allosteric changes resulting in ligand-specific alterations in regulatory outcomes. The capacities of functional GR ligands to fill only a portion of the binding pocket or to alter the shape of the pocket challenge and, at the same time, liberate the concept and practice of ligand design. Using the dexamethasone-occupied pocket as a guide, arylpyrazole compounds were developed as non-steroidal GR ligands and shown indeed to differ from dexamethasone in their phenotypic and molecular actions in several target cell types[63]. Notably, "selective GR modulators" (SGRMs) that preserve the anti-inflammatory and immunosuppressive actions of standard glucocorticoids, but do not show adverse effects that accompany chronic glucocorticoid therapeutic regimens have been long sought without success. Hence, discovery of molecules that achieve pre-selected context-specific regulatory outcomes has proven difficult; we suggest in our concluding remarks a targeted approach.

In summary, GR ligands are critical physiologic or pharmacologic context inputs, conferring allosteric transitions[64] that affect GR-associated molecular chaperone affinities or functions, and activate GR nuclear localization signals and DNA binding activity. Even subtle modifications of GR ligand chemistry or dose can affect gene-specific GOR formation[48,63], or regulatory complex composition and/or function (*e.g.,* altered HAT activity) without affecting GOR formation itself[63].

Post-translational modifications.

Covalent PTMs of TRFs are conferred as the endpoints of cell signalling pathways that provide physiologic context information distinct from that provided by noncovalently associated hormonal ligands. PTMs can confer allosteric transitions, create or inactivate protein interaction surfaces, affect protein localization, stability, DNA binding, ligand response and regulatory activity of TRFs in a context-specific manner. GR can be modified by several PTMs at distinct sites (**Figure 2.5**)[65–69]. Here, we summarize some of those PTMs and their effects on GR action.

Phosphorylation. Phosphorylation (generally on Ser, Thr, or Tyr) of nuclear receptors plays important roles in ligand binding, nuclear localization, DNA binding, and modulating interactions with coregulators[68,70]. GR maintains a basal level of phosphorylation, but additional sites are phosphorylated upon ligand treatment[71–73]. To date, there are seven experimentally confirmed phosphorylation sites on GR clustered within the N-terminal domain, including Ser113, Ser134, Ser141, Ser203, Ser211, Ser226, and Ser404[65]. These residues are conserved among human, mice, and rats[73]. Although the enzymes responsible *in vivo* are uncertain, these sites can be modified *in vitro* by cyclin-dependent kinases (CDK), mitogen protein kinases (MAPK), c-Jun N-terminal kinases (JNK), and glycogen synthase kinase-3 (GSK-3)[74], implying that multiple signaling pathways communicate with GR. Mutational analysis of particular sites led to mixed reports of effects of phosphorylation on transcriptional regulatory activities of GR[75,76], consistent with strong gene-, cell-, and physiology-specific context dependence, as expected.

In general, phosphorylation of GR increases the protein half-life, and mutations to alanine of several different phosphorylation sites, one site at a time, each results in rapid protein degradation[67,76]. Phosphorylation of sites Ser203, Ser211, and Ser226, located within AF1, are predicted to affect exposure of protein surfaces critical for cofactor interactions[77,78]. Phosphorylation of Ser211 results in increased recruitment of GR to GORs and subsequent regulation of GR target genes[79]. Mutation of Ser203 prevents phosphorylation at Ser226, suggesting interdependence of those modifications[72,80]. Upon ligand binding, Ser203 is phosphorylated and GR is selectively partitioned to the nucleus. Mutation of that site precludes nuclear accumulation and subsequent GR dependent gene regulation[80], whereas S226 phosphorylation, presumably via JNK signalling pathways, results in increased nuclear export, thus decreasing gene regulation through GR[68]. Ser203 is phosphorylated *in vitro* by CDK and MAPK[74,81]. This apparent integration across different signalling pathways may contribute to the differences in transcriptional outcomes observed in different cellular and physiological contexts[70,80]. Another site, Ser404, phosphorylated *in vitro* by GSK-3, appears to be hypo-phosphorylated in nuclear fractions, but loss of S404 phosphorylation changes the conformation of GR, alters cofactor recruitment and transcriptional responses, and increases glucocorticoid-induced apoptosis[82,83].

Ubiquitylation. The 8.5 kDa ubiquitin polypeptide is covalently attached to Lys residues of a target protein and promotes protein turnover by targeting them to the proteasome for degradation. Cells treated with proteasome inhibitors showed enhanced GR regulatory activity[67,84] and increased *in vivo* DNA occupancy time, indicating that ubiquitylation plays a role in regulating GR stability[11,12]. Furthermore, ubiquitylation of GR at Lys419[85] was shown to stimulate GR nuclear export and subsequent degradation[86].

SUMOylation. The small ubiquitin-related modifier-1 (SUMO-1) polypeptide, when covalently linked to Lys residues of target proteins can alter their stability, localization or transcriptional regulatory activity[67,87]. GR can be SUMOylated at three sites: Lys277 and Lys293 within the NTD and Lys703 within the LBD; these modifications have been shown to have effects on GR activity that are highly context dependent[66,88]. Genome-wide profiling of GR SUMOylation mutants revealed enhanced GR recruitment to DNA, specifically at genes involved in cell growth, proliferation, and survival[89].

Although its functions are context dependent, SUMOylation has most commonly been linked to transcriptional repression[90,91]. One recent example concerns the repression of inflammatory genes[92,93] bearing IR-GBS-containing GORs. Specifically, SUMOylation at Lys293 within the NTD is required for IR-GBS-mediated repression, whereas mutations of the other SUMO sites have no effect. SUMOylation of Lys293 also promotes the recruitment of coregulators SMRT and NCoR, but does not affect GR-mediated activation from certain genes linked to canonical GBS-containing GORs[92]. It appears that GR SUMOylation not only facilitates assembly of repressive regulatory complexes but assists in the binding of GR to weaker associated sites, at least at the IR-GBS-containing GORs examined. Furthermore, loss of this SUMOylaton site results in diminished GR-mediated repression at certain AP-1/NF-κB tethering sites[93], apparently due to inhibition of regulatory complex assembly.

Acetylation. GR is acetylated at Lys494 and Lys495 within the hinge region at a common acetylation motif, KXKK, where X is any amino acid. This motif is conserved among the 3-keto steroid receptors, suggesting that this acetylation may be important for certain general functions of these receptors. The histone acetyltransferase (HAT) proteins CLOCK and BMAL1 were inferred from cell-based assays to acetylate GR at these Lys

residues[94]. KXKK acetylation reduced the affinity of GR binding to canonical GBSs *in vitro* and also reduced its ability to regulate transcription in transfection assays. The regulation of this modification by CLOCK is critical for circadian rhythm maintenance[95]. GR is deacetylated *in vitro* by histone deacetylase 2 (HDAC2) which appears to be important for repression of NF-κB-regulated genes[96].

Nitrosylation. S-Nitrosylation involves covalent attachment of a nitric oxide to a thiol group on a Cys residue. Nuclear receptors contain two zinc finger domains with four Cys residues each. Nitric oxide can target these residues and cause the release of bound Zn^{2+}. For GR, this modification has been shown to inhibit ligand binding[69]. Exposure of COS-7 cells to exogenous nitric oxide sources has been shown to inhibit DNA binding and dimerization of nuclear receptor complexes such as VDR-RXR[97], but to date this has not been studied for GR.

Composite GRE-bound non-GR TRFs.

The fourth class of allosteric effectors that alter GR activity are non-GR TRFs bound at composite GREs. Correlative studies, suggest that composite GREs are 0.5-2kb genomic segments containing one or more of the four classes of GR binding motifs described above, clustered with binding sequences for non-GR TRFs[36]. In the first described composite GRE, GR interacted both with GBS DNA and with an AP-1 factor bound contiguously, and either activated or repressed transcription depending on the subunit composition of the AP-1 factor[28,98,99]. A simple interpretation is that particular combinations of non-GR TRFs bound at composite GREs could bias GR occupancy at those elements, and by directly interacting with GR confer allosteric effects, modulating GR activity. Occupancy of the non-GR TRF

binding sites at a given composite GRE will differ in different cell types, owing to cell context-specific differential absolute and relative expression and activities of the respective TRFs, thereby providing a context-dependent signal for transcriptional regulation mediated by GR.

Coregulators as GR signalling readers

TRFs such as GR nucleate assembly of large (~10^2 polypeptides and non-coding RNAs) transcriptional regulatory complexes, typically including other TRFs and distinct arrays of coregulatory factors, which confer structural or functional changes on the transcription machinery or chromatin, thereby positively or negatively modulating target gene mRNA production. Roughly 300 coregulators have been identified, many of which are themselves multifactor complexes, and shown to exhibit a wide range of functions[100,101]. Which coregulators interact with GR depends not only on coregulator availability in a given cell type, but also on the integrated effects of the four classes of signals that communicate context information to GR discussed above. Hence, coregulators can be viewed as readers of GR-mediated signaling - it is the coregulators that convert the integrated signal-driven allosteric transitions at distinct receptor surfaces into context-specific transcriptional regulatory actions. Originally classified as coactivators and corepressors, many coregulators have been recognized to both activate and repress genes in a context specific manner[102]. Here, we shall classify coregulators based on their mechanisms of action rather than regulatory outcome in any particular context, and will discuss those classes shown to date to interact with GR (**Table 2.1**).

Basis of GR-Coregulator interactions.

Structural analysis has provided deep insight into the interaction of coregulators with regulatory domain AF2 near the GR C-terminus, a highly conserved 12-α helix and 4-β sheet motif that folds into a three layer helical bundle[103]. AF2 of nuclear receptors is minimally comprised of helices 3, 4 and 12[104,105] with which NR coregulators interact through highly conserved LXXLL motifs (so called NR boxes)[106,107] or (L/I)XX(I/V)I or LXXX(I/L)XXX(I/L) motifs (referred to as CoRNR boxes)[108] (X=any amino acid) (**Figure 2.4c,d**). In contrast, the entire NTD of GR, which includes the AF1 domain, is highly disordered. Interaction of coregulator SRC-2 and TATA-box binding protein with AF1 stabilizes it and increases its alpha-helical content[109,110] (similar disordered regions are commonly directed to fold into well-ordered functional domains in other TRFs and other proteins[111]); little else is known of the structure and dynamics of the functional domain, or of the AF1–coregulator interfaces. Still less is known structurally about the Tau2 domain within the GR hinge domain (**Figure 2.1a**), which interacts with coregulator Hic-5[112,113] (**Table 2.1**). Importantly, because systematic mutagenesis and mapping in multiple contexts has not been carried out, we are far from understanding even the number of functional surfaces that potentially might form within any given GR domain.

Functional classes of GR coregulators.

Table1 displays five functional classes of coregulators reported to interact with GR, together with specific examples from each class (for a more complete description of coregulators see [102,114,115]). As this is an active field of research, additional GR-associating coregulator classes are likely yet to be discovered. These coregulators typically interact with

GR bound by one of the "standard" cortisol-like glucocorticoid ligands, such as cortisol itself or dexamethasone. One class of coregulators, histone deacetylase complexes (HDACs), binds GR at canonical GBS-containing sites only when bound by exogenous drug RU486, thus calling into question the physiological significance of these interactions. In contrast, at IR-GBS-containing sites, HDACs interact with GR bound by standard ligands. Below we consider briefly examples from each functional class.

Structural and Enzymatic Complexes.

p160 Steroid Receptor Coregulator (SRC) Family. The p160 family is composed of three members; SRC-1 (also known as NCoA-1), SRC-2 (also known as NCoA-2, TIF-2, GRIP-1), and SRC-3 (also known as NCoA-3, pCIP, AIB1, ACTR, TRAM1)[100]. These proteins function as scaffolds and can associate with between six and ten other proteins[116], including other coregulators, such as CoCoA, CCAR1, histone acetyltransferases (CBP/p300 and pCAF), and histone methyltransferases (CARM1 and PRMT1)[117–121]. Knockdown of individual SRC proteins showed context-specific effects on GR transcriptional regulation[122,123].

The SRC proteins contain three functional domains: the N-terminal basic helix-loop-helix-Per/ARNT/Sim (bHLH-PAS) domain, which binds to AF1 domain of GR increasing its stability and α-helical content[109]; the central receptor interaction domain (RID), which contains three LXXLL motifs and two transcriptional activation domains (AD1 and 2) and associates with the AF2 domain of GR within the LBD[124,125] (see also **Figure 2.4C**); and the C-terminal activation domains, which interact with HATs and histone methyltransferases[118].

GR interacts preferentially with SRC-2 over other p160 members[126–128]. SRC-2 can form distinct foci within the nucleus that also contain p300, PCAF, and nuclear receptors GR, AR, ER, as well as others. Dexamethasone-bound but not RU486-bound GR was observed to localize to these substructures[129]. SRC-2 can also be phosphorylated in a GR-interaction-dependent manner, and mutations to key phosphorylation sites result in reduced expression from selected target genes[130]. Also functional in repression, SRC-2was the first coregulator shown to deploy distinct regulatory domains[131] for up- or down-regulation of GR-dependent transcription in different contexts[29,132–134].

Mediator and other structural and enzyme-interacting complexes. The ~30 protein, 1.2 MDa Mediator complex, which forms a physical link between TRFs, such as GR, and the general transcription machinery, and regulates transcription through affect RNA polymerase II and TFIIH activity[135]. Structural studies suggest that Mediator activity is controlled allosterically[136–139]. For example, interaction with LBDs of nuclear receptors as well as other TRFs induces conformational changes resulting in the formation of a Mediator "pocket domain" that enables Mediator–RNA polymerase II interaction[137,140]. Core Mediator components also stably interact with kinase modules including CDK8 or CDK19 which in turn recruit other enzymes, such as, the HAT GCN5L[141]. GR binds to two distinct Mediator subunits. MED1 interacts with the GR-LBD in a ligand-dependent manner via LXXLL motifs, whereas another mediator subunit, MED14 establishes interactions with AF1 domain of GR independent of ligand[142]. GR target gene regulation appears to be differentially dependent upon MED1 or MED14[143]. Other coregulators such as Hic-5, CoCoA, and CCAR1 also interact either directly or indirectly with GR in the formation of multi-subunit context-specific transcriptional regulatory complexes[112,113,119–121].

Chromatin Remodelling Complexes.

Human cells contain an extensive family of evolutionarily conserved multi-protein SWI/SNF-related ATPases that catalyze various transformations of the regular nucleosome packaging of chromatin[144,145]. Predominant among them are the closely related BRM and BRG1 complexes, which interact context-specifically with GR[146,147], repositioning nucleosomes to facilitate the accessibility of DNA to transcription factors and transcriptional machinery[148–150]. The recruitment of SWI/SNF complexes by GR to facilitate transcription[151] served as the first report of TRF coregulators. GR has been shown to interact directly with multiple subunits of the BRM and BRG1 complexes[152,153], with binding reported to DBD, LBD and Tau1/AF1 domains of GR, likely in a context-specific manner.

Methyltransferases.

Protein arginine methyltrasferases (PRMTs) such as PRMT4 (also known as CARM1) and lysine methyl transferase including G9a methylate histone and other proteins. They interact with GR directly or with GR-bound coregulators p160 or p300, and either activate or repress GR target genes in a context-dependent manner[154,155].

Histone Acetyltransferases (HATs).

HATs modify histones and other proteins, forming ε-*N*-acetyllysine at selected lysine residures[156]. Among the various HAT families, CREB-binding protein (CBP), p300 and pCAF (ADA/SAGA) interact with GR directly through interactions with the AF1 domain , or indirectly through p160 coregulators associated with the AF2 domain of GR to modulate transcription[100,124,157–159]. HATs target histones as well as non-histone proteins (including nuclear receptors) in a context-specific manner[117,160], and although generally associated with

transcriptional activation[161], mechanisms of action as well as regulatory outcomes appear also to be context-dependent[162].

Histone deacetylases (HDACs).

Nuclear receptor corepressor (NCoR) and silencing mediator of retinoic and thyroid receptors (SMRT) form multi-protein complexes that include histone deacetylases (HDACs)[163]. NCoR and SMRT contain extended helical motifs, (L/I)XX(I/V)I or LXXX(I/L)XXX(I/L), termed CoRNR boxes, which can interact with the LBD of GR. This binding occurs preferentially when GR is bound to RU-486 over GR bound to a standard glucocorticoid[128], as helix 12 of the GR-LBD in the RU-486-bound form is optimally positioned to permit CoRNR association[164] (**Figure 2.4d**). While that GR interaction may not be physiologically significant, NCoR/SMRT co-occupy IR-GBS-containing or NF-κB- or AP-1-tethered GORs at which standard GC-bound GR is SUMOylated within the NTD[92,93].

Precision and plasticity via allostery

Because it is expressed ubiquitously, GR provides a dramatic example of perhaps the most critical property of eukaryotic TRFs – their context-specificity. That is, GR governs networks of genes that are precisely determined in a given setting, yet differ dramatically as a function of cell type and physiologic state. This context-driven plasticity likely reflects the integration of four classes of context-specific signals outlined above — hormonal ligands, PTMs, GR binding DNA sequences and adjacent binding sequence motifs for non-receptor TRFs. The hormones and PTMs provide physiologic context information to GR, the two classes of DNA binding sequences, which together comprise composite GREs, provide gene

context, and the array of TRFs available to occupy non-GR binding sequences provides cell context.

We and others have shown that certain GR signals, namely hormonal ligands and DNA binding sequences, are allosteric effectors, conferring specific alterations on GR conformation. Our model assumes that all four classes of signals impacting GR transcriptional regulatory activity operate allosterically, and that their effects integrate to produce context-specific patterns of GR surfaces. These patterns of surfaces are recognized and bound by specific coregulator complexes, which typically possess enzymatic activities that confer structural and/or functional changes on the transcription machinery and/or the chromatin. Coregulators are generally not themselves cell specific; rather, we propose that they assemble in unique combinations at each GRE, based on the context-specific interactions established at these sites.

A provocative extension of these ideas is that GR and other TRFs may typically lack intrinsic regulatory activities, instead serving merely as molecular scaffolds, patterns of surfaces produced by signaling, to which coregulators, the actual regulatory machinery, combinatorially associate (**Figure 2.6**). This accounts readily for the common observation that TRFs can activate transcription in one context and repress it in another. In this sense, GR *activities* reflect its molecular conformations, which emerge owing to context-specific cues, whereas GR *functions* are integrated regulatory outcomes of coregulatory enzyme actions that associate with those various conformations. Thus, we suggest that GR structure, together with its determinants, are keys to understanding both its regulatory precision and plasticity. These ideas likely extend, at least conceptually, to other eukaryotic transcriptional regulatory factors[52], all of which face the same precision/plasticity challenge.

Conclusions and Perspectives

Biology is marked by astounding specificity of time and place, whether exemplified by butterfly migration, embryo patterning, neurite outgrowth or chromosome segregation. Here, we have considered such specificity from the perspective of a single polypeptide that regulates vertebrate gene expression. GR (which is ~90 kDa) associates non-covalently with a simple ligand (362 Da for cortisol), and controls transcription within networks comprising thousands of genes, distinct to cell type and condition. What explains such specificity?

We suggest here that GR's actions are expanded, refined and directed through interactions with multiple partner TRFs, dozens of signalling pathway-induced PTMs, hundreds of coregulators, tens of thousands of potential genomic binding sites, hundreds of cell types each with unique patterns of chromatin structure, and countless physiological states. This culminates in exquisitely complex regulation of gene transcription characterized by remarkable precision in any particular contextual setting, yet facile plasticity to adapt when that context is altered. We propose that allosteric regulation of GR by these various inputs lies at the basis of context specificity of gene expression.

This capacity of GR to integrate signalling inputs allosterically to produce distinct transcriptional outputs can be expanded further by the fact that multiple, distinct regulatory complexes established at individual GORs may collaborate to control any single GR-regulated gene. In addition, ligand chemistry and concentration readily alter GR-regulated gene network transcription, opening the door to speculation that endogenous ligands for GR may not be limited to cortisol. With this appreciation of the daunting task to understand this complexity sufficiently to render it predictive, what is a strategy to gain detailed insight into

signal- and allostery-determined regulation of transcription? Clearly, neither systems nor reductionist approaches alone will suffice[41].

In our view, an essential step in obtaining holistic understanding of the regulation of metazoan gene transcription is to identify "causative primary regulated genes", defined as genes directly regulated by a given TRF, whose regulation by that factor is essential for a given phenotypic change. Next, functional response elements for such genes must be identified. Finally, allosteric, compositional and enzymatic changes in their corresponding regulatory complexes upon change of context could reveal properties and mechanisms, such as a distinct pattern of functional, allosterically-specified surfaces, that correspond to the particular phenotype. Importantly, identification of molecular features that can serve as surrogates for complex physiologic or pathologic outcomes could form a basis for predicting different combinations of contexts or signals that produce or preclude a given outcome, as well as frame a new approach for screening and assaying therapeutic candidates.

Acknowledgements

We thank the members of the Yamamoto lab for critical reading of the manuscript, with special note to Elaine Kirschke for insightful discussions, Samantha Cooper, Sheng-Hong Chen, and Benjamin Schiller for use of unpublished data, and Kirk Ehmsen for use of unpublished data and assistance with Figure 3. E.R.W. is supported by a National Institutes of Health (NIH) predoctoral NRSA (1G31GM113397-01A1) from the National Institute of General Medical Sciences (NIGMS). M.T.K. is supported by NIH postdoctoral NRSA (5T32HL007731-20) from the National Heart, Lung, and Blood Institute (NHLBI) and by NIH R01CA020535 from the National Cancer Institute. E.A.O. is supported by NIH grant R01DK095750 from the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK), by an American Heart Association (AHA) grant 14GRNT20460124, and a W.M. Keck Foundation Medical Research Grant. K.R.Y. is supported by NIH R01CA020535 from the National Cancer Institute, NIH R21ES026068 from the National Institute of Environmental Health Sciences, and MCB-1615826 from the National Science Foundation.

Figures

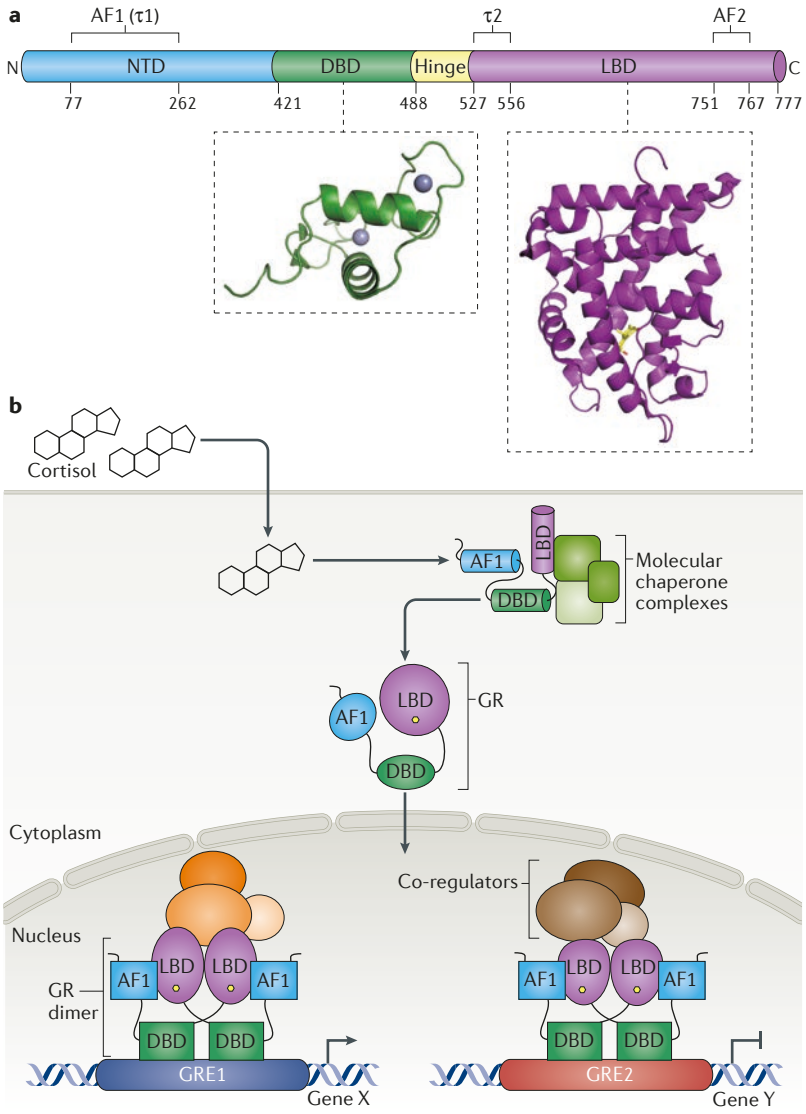


Figure 2.1: Glucocorticoid receptor signaling and DNA binding.

a) Linear domain structure of glucocorticoid receptor (GR). GR comprises: the amino-terminal domain (NTD), DNA binding domain (DBD), hinge and ligand binding domain (LBD). Embedded in these domains are segments that participate, context-specifically, in transcriptional regulation: activation function domain 1 (AF1 also known as trans-activating domain 1 / Tau1 / τ1), trans-activating domain 2 (Tau2 / τ2), and activation function

domain 2 (AF2). Insets: Crystal structures are shown for a single DBD (adapted from crystal structure Protein Data Bank accession number (PDB) 1R4R[17]) and coordinated Zn are shown as grey spheres; LBD (purple) liganded with cortisol (yellow) and complexed with SRC-2/TIF2 peptide (not shown) (PDB:4P6X[61]). b) Overview of signalling mediated by natural GR ligand, cortisol. Activating ligand interacts with monomeric GR associated with molecular chaperone-containing complexes in the cytosol. This induces local and remote allosteric changes that potentiate nuclear transport and other activities. Within the nucleus, GR nucleates multi-component transcriptional regulatory complexes, comprising various other Transcriptional regulatory factors (TRFs) and transcriptional coregulators at different glucocorticoid response elements (GREs) to activate or repress transcription at particular target genes. GRE1 and GRE2 represent distinct GREs within the genome, GeneX and GeneY represent the genes under the control of GRE1 and GRE2 respectively.

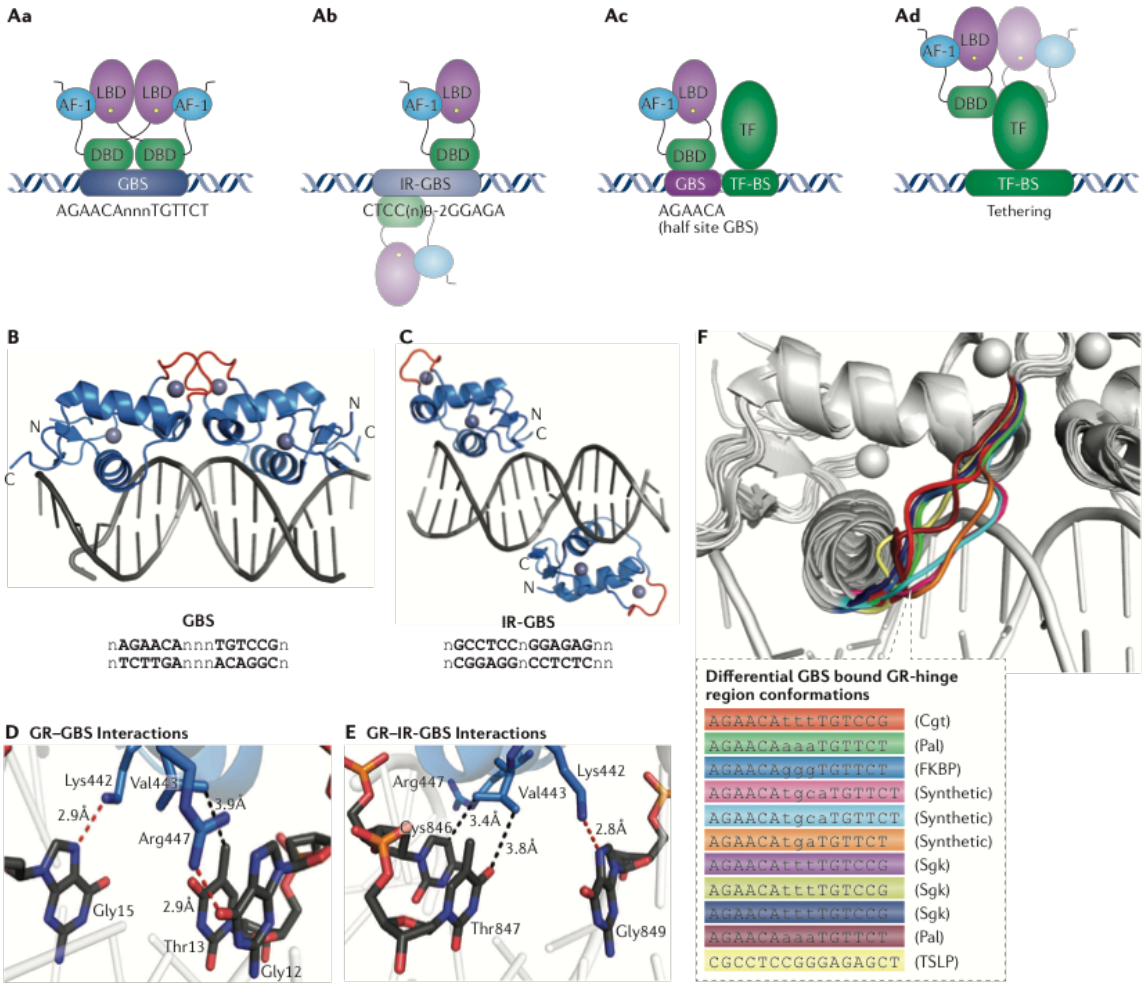


Figure 2.2: Modes of site-specific glucocorticoid receptor–genome interactions.

(A) Glucocorticoid receptor (GR) associates with specific genomic sites in multiple ways. *[a]* Two GR monomers bind a canonical GR binding sequence (GBS) in a head-to-head fashion; dimerization is achieved through interactions in the sister DBDs. *[b]* GR binds to inverted repeat (IR)-GBSs. The crystal structure of this interaction (see also C) shows two GR monomers bound to opposite sides of the DNA in a head-to-tail fashion; however, negative cooperativity argues that GR may bind as a monomer to IR-GBSs *in vivo* (indicated by shading of the second GR monomer). *[c]* GR can interact with half-site GBSs; these elements typically contain a

hexamer related to the consensus sequence that is palindromic in the full GBS and may operate in conjunction with proximal non-GR transcriptional regulatory factors (TRFs), although there is no evidence for enrichment of particular TRF binding site (TRF-BS) motifs contiguous with the GBSs.*[d]* GR can interact at specific genomic sites without directly binding DNA. Here, GR tethers onto a non-GR TRF through protein–protein interactions. The faded GR monomers depict uncertainty whether monomeric or dimeric GR binds at tethering elements. (B) Overall crystal structure of the DNA binding domain (DBD) of GR with the canonical GBS (PDB: 3FYL[18]). (C) Overall crystal structure of the complex between DBD of GR and the IR-GBS (PDB: 4HN5[25]). (D) GR-DBD makes three base-specific contacts within the major groove of the DNA on GBS. This interaction is mediated by hydrogen bonds (red) and van der Waals interactions (black). (E) GR-DBD makes similar contacts to an IR-GBS as in the case of the canonical GBS, with the exception of Arg447, which does not make any contacts with the DNA. (F) Superposition of all deposited structures of the interactions between DBD and DNA reveals especially striking conformational differences within the GR "lever arm" (residues 469-474, a loop connecting the DNA reading helix with the dimerization region), demonstrating allosteric modulation of GR by DNA sequence. DNA sequences from each structure are listed and color coded, the gene the sequence is derived from is listed in parenthesis (crystal structure Protein Data Bank accession numbers (PDBs) listed in order of sequences above: 3FYL[18], 3G99[18], 3G6U[18], 1GLU[17], 1R4O[17], 1R4R[17], 3G9P[18], 3G9O[18], 3G9M[18], 3G9J[18], and 4HN5[25]).

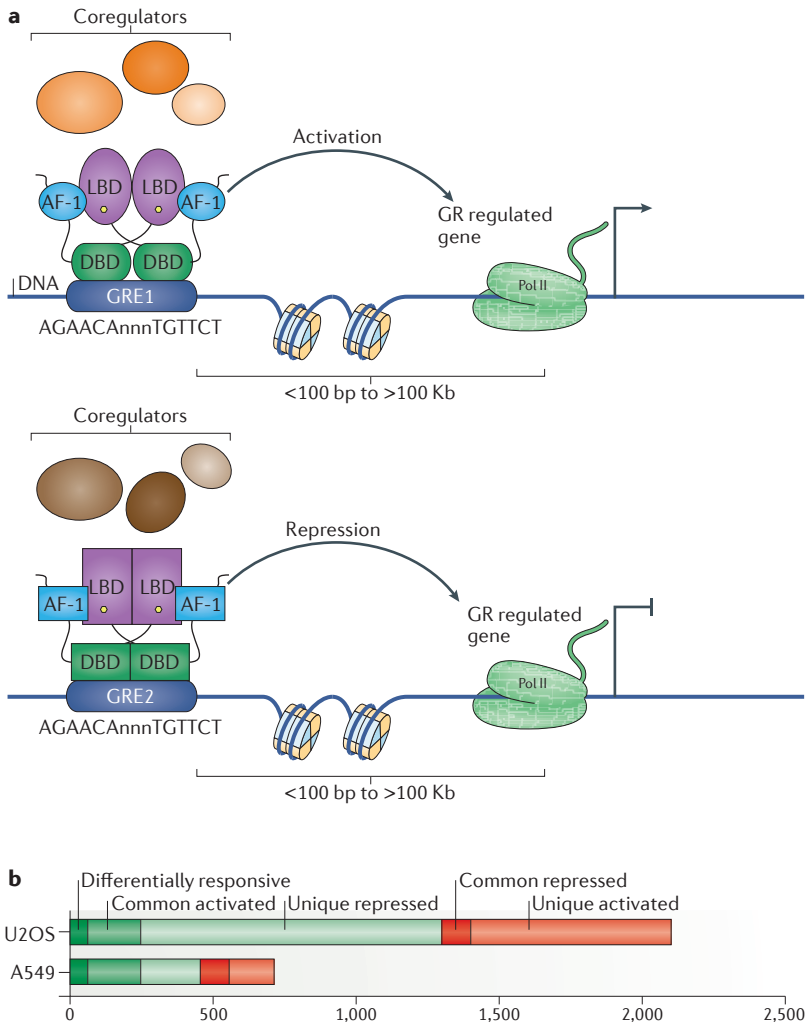


Figure 2.3: Context-specific glucocorticoid receptor occupancy and gene regulation.

(a) GR regulated genes commonly are linked to multiple glucocorticoid receptor occupied regions (GORs); mostly >10 kb from the transcription start site of the regulated gene, one or more of which may be a functional glucocorticoid response element (GRE) for that gene. As shown here, GREs may be close (<100 bp), or far (>>100 kb) from their target genes. Abbreviations: General Transcription factors (GTFs), RNA polymerase II (Pol II). (b) Illumina Human Ref8 beadchip analysis of glucocorticoid regulated genes. Genes regulated by 4 hour, 100 nM dexamethasone treatment in lung carcinoma (A549) (Yamamoto lab unpublished data) and osteosarcoma (U2OS)[21] cells. Differentially responsive genes are up-

regulated in one cell line and down-regulated in the other. Common genes are similarly regulated in both cell lines. Unique genes are regulated in one cell line and not regulated in the other. Abundance of genes within the unique and differentially regulated classes demonstrate cell-context specific regulation by GR. Differentially responsive genes demonstrate distinct mechanisms of regulation of target genes common to the two cell types.

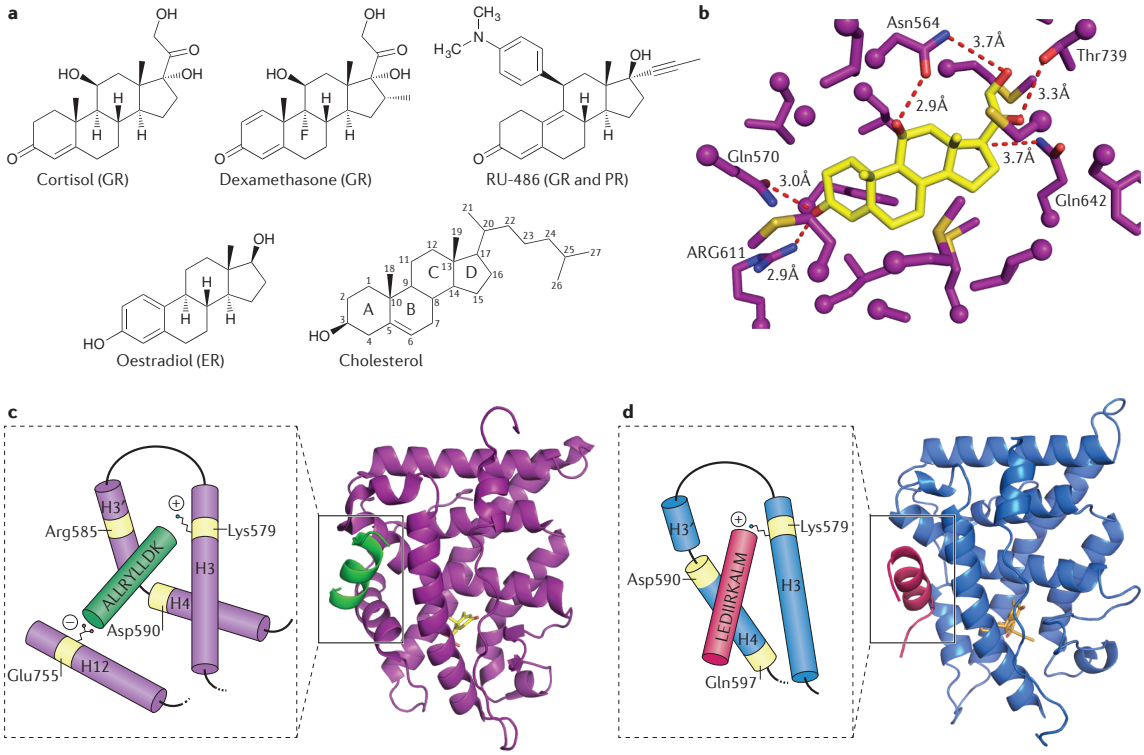


Figure 2.4: Glucocorticoid receptor–ligand interactions.

(a) Cortisol, dexamethasone and RU-486 are three glucocorticoid receptor (GR) ligands; RU-486 also binds to progesterone receptor (PR); oestradiol, which binds estrogen receptor, but not GR is shown for comparison; cholesterol is shown to provide the sterol carbon numbering convention. (b) Ligand binding pocket (purple) of the ligand binding domain (LBD) of GR bound to its endogenous ligand, cortisol (yellow). Residues within 4.2 Å are shown and hydrogen bonds are depicted in red. This structure highlights the intricate network of interactions between GR and ligand as well as the amount of unoccupied space within the ligand binding pocket with ligand bound. (c) Overall structure of the GR-LBD bound to cortisol and the SRC-2 coregulator peptide (green) (PDB: 4P6X[61]). The inset cartoon represents how the LXXLL motif (the so called Nuclear Receptor Box) of the coregulator peptide interacts with GR. The peptide is held in place by a conserved charge clamp

interaction, the positive and negatively charged residues of Glu755 (-) and Lys579 (+) on helix 3 (H3) and helix 12 (H12) mediate this interaction. There is an additional charge clamp that occurs through the residues Arg585 and Asp590. This is referred to as the active conformation associated with specific coregulator binding. (d) Overall structure of the GR-LBD bound to RU-486 and the NCoR coregulator peptide (hot pink) (PDB: 3H52[164]). The inset cartoon represents how the extended CoRNR boxes of the coregulator peptide interact with GR. The extended peptide only makes one of the conserved interactions, through Lys579, and helix 12 (H12) is displaced. This is considered an inactive form of the LBD. Differences in GR conformations presented in parts c and d indicate that different ligands can promote formation of alternative protein surfaces on GR that in turn differentially affect coregulator binding.

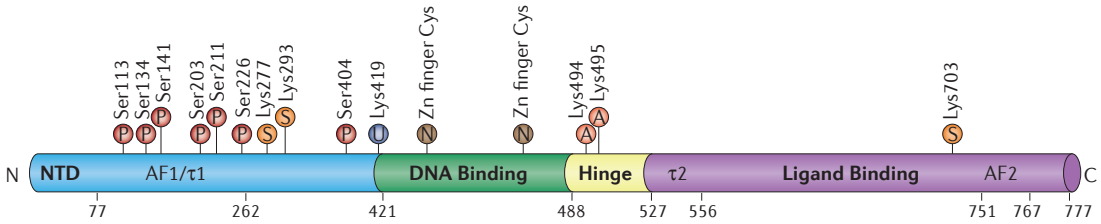


Figure 2.5: Sites of glucocorticoid receptor post-translational modifications.

Major reported modifications, including phosphorylation (P), SUMOylation (S), ubiquitylation (U), acetylation (A), and nitrosylation (N) are mapped onto the glucocorticoid receptor domain schematic.

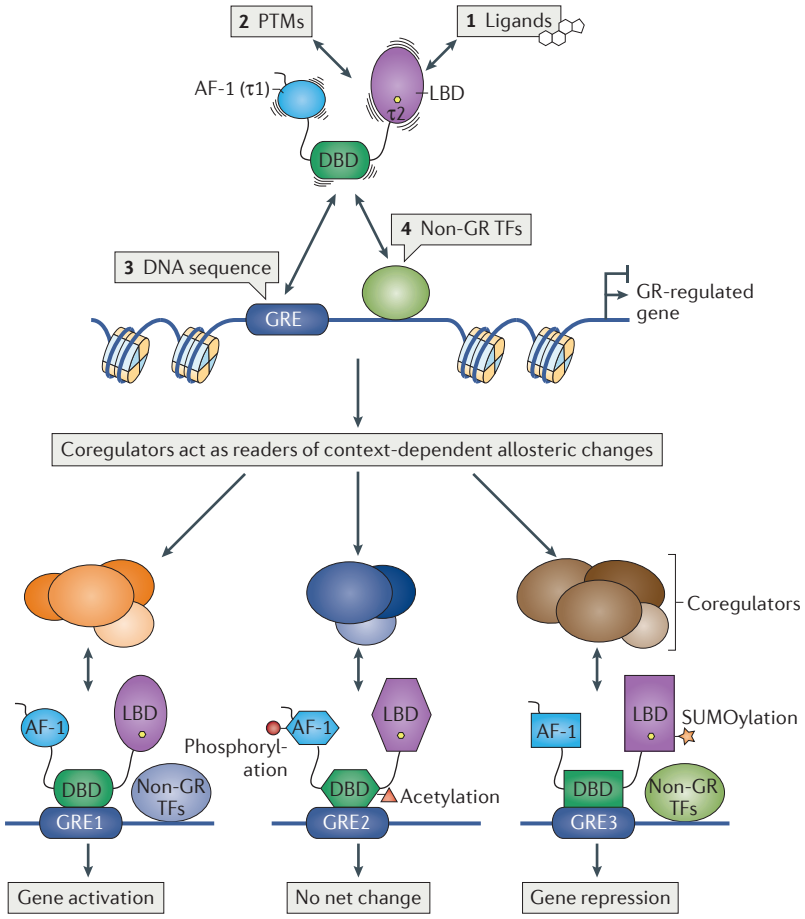


Figure 2.6: A model for transcriptional regulation - Precision and plasticity of TRF function achieved via allostery.

In our model for metazoan transcriptional regulation, TRFs such as glucocorticoid receptor (GR) may lack intrinsic regulatory activities, serving instead as molecular scaffolds that can assume different conformations in response to modification by different combinations of specific signalling inputs. Four classes of signalling inputs are described: (1) ligands and (2) post-translational modifications (PTMs) provide physiological context, (3) DNA binding sequences provide gene context, and (4) non-GR TRFs provide cell context. Each class of signals confers distinct allosteric effects, and their integrated actions can produce a vast array of GR conformations, which induce, expose or stabilize context-specific protein surfaces that

are read by coregulators. These coregulators, generally large multi-component enzyme-containing complexes, interact in distinct combinations with patterns of cognate GR surfaces and enzymatically modify the general transcriptional machinery and/or surrounding chromatin in and around glucocorticoid response elements (GREs) and target promoters and genes. Upper panel - lines around GR domains depict direct and allosteric conformational alterations imposed by the signaling inputs. Note that lack of net regulation (neither activation nor repression) could reflect balanced actions of one or more GREs acting on a single gene (not shown). AF1, activation function domain 1 (also known as τ1); - DBD, DNA binding domain; GBS, GR binding sequence; LBD, ligand binding domain; τ2, transactivation domain 2; TF, transcription factor.

Table 1 | **Five functional classes of co-regulators reported to interact with GR**

Functional class	Examples	GR-interaction surfaces	Targets and regulatory outcomes	Refs
Structural and enzyme-interacting	p160 SRC family: SRC-1 (also known as NCoA-1)	LXXLL motif interacts with AF2 bHLH-PAS–AF1 interaction	• Transcription activation best characterized; SRC-2, at least, can repress in appropriate contexts • Interaction increases α-helical content of AF1, increasing its structural stability	29,100,109, 116–134
	SRC-2 (also known as NCoA-2, TIF2 and GRIP1)			
	SRC-3 (also known as NCoA-3, CIP, AIB1 and TRAM1)			
	Mediator	LXXLL motif within MED1 interacts with the LBD of GR, and MED14 interacts with AF1	• Recruitment and allosteric regulation of transcriptional machinery, including RNA polymerase II and TFIIH • Recruitment of multiple enzymatic activities, including phosphorylation and acetylation (at Ser10 and Lys14, respectively) of histone H3 through associated kinase CDK8 and HAT GCN5L	135–138, 140–143
	HIC-5 (also known as TGFB1I1)	HIC-5 binds to the tau2 domain within the GR hinge	• Transcription complex assembly and Mediator recruitment • May block chromatin remodelling • Differently affects DNA binding and transcription regulatory activity of different gene classes	112,113
	COCOA	Indirectly via interaction with the amino terminus of p160	• Acts synergistically with p160	119,120
	CCAR1	Directly binds t GR or is indirectly recruited by COCOA	• Recruits Mediator and p160 to response elements	121
Chromatin-remodelling	BRG1 (also known as SMARCA4) and BRM SWI/SNF-related ATPases	Multiple BAF subunits interact with the DBD, LBD and AF1 in different contexts	• ATP-dependent repositioning of nucleosomes • Relieves repressive state of chromatin • Multiple context-specific effects of BRM knockdown on GR-dependent genes	146–153
Methyltransferases	CARM1 (also known as PRMT4)	Indirectly via interaction with carboxy-terminal p160	• Methylates histone H3 and non-histone proteins *in vitro* • Activates SRC-2-dependent genes	120,154
	Lys methyltransferase G9a (also known as EHMT2)	N-terminal portion of G9a binds to the GR NTD	• Positive regulation by enhanced recruitment of CARM1 and p300 to GR target genes • Facilitates methylation of Lys9 on H3, which is associated with transcription repression	155
HATs	CBP/p300	Directly to AF1 or indirectly through partner p160–AF2	• Acetylation of histones • Cell-type-specific transcriptional outcomes	100,117,124, 157,158,160, 165
	PCAF (in ADA and SAGA complexes)	Directly to AF1 or indirectly through either partner p160–AF2 or CBP/p300–p160–AF2	• Acetylation of histones • Less well defined, but contains TAF subunits, which have been postulated to aid in the recruitment of general transcription factors to DNA	100,159,166
HDACs	NCoR	(L/I)XX(I/V)I or LXXX(I/L)XXX(I/L) CoRNR motif interacts with AF2 when GR is bound to RU-486	• Transcription repression correlated with histone deacetylation in most cases	24,92,93, 128,163,164, 167,168
	SMRT	NTD interaction in GR–IR-GBS or GR-tethered when GR is bound to a standard ligand		

Table 2.1: Five functional classes of coregulators reported to interact with GR.

Table of known GR coregulators, how they interact with GR, and the regulatory outcomes of this interaction.

Method	Type	Utility/Application	Process	Pitfalls	Reference
ChIP-seq (Chromatin immunoprecipitation followed by deep sequencing)	*in vivo*	Identify GR occupied regions (GORs) within the genome *in vivo.* Particularly useful for determining changes in occupancy in different cellular conditions (*e.g.* different cell types)	Cellular chromatin is crosslinked and fragmented by mechanical or chemical cleavage. Next an antibody bound resin is used to precipitate the target factor together with crosslinked genomic fragments, the crosslinks are reversed, and the co-precipitated DNA fragments are prepared into a library and sequenced.	Relies on the specificity of the antibody-antigen recognition as well as the context-specific exposure of the epitope, efficiency of crosslinking and reverse crosslinking, accessibility of chromatin for pulldown, and proper library preparation and normalization to cellular chromatin input. Care should be taken when interpreting ChIP-seq results, especially when making quantitative interpretations comparing ChIP-seq peak signals done in different conditions	1–6
ChIP-exo (Chromatin immunoprecipitation followed by exonuclease digestion and deep sequencing)	*in vivo*	Identify GR occupied regions (GORs) within the genome *in* vivo with base pair resolution on a genome-wide scale	Similar general protocol as ChIP-seq except it uses an endonuclease to degrade accessible DNA before the crosslinking is reversed. Protein-bound DNA is protected from cleavage and, upon sequencing, reveals genomic occupied sites at base pair resolution.	The pitfalls associated with ChIP-seq apply here. Additionally, differences in exonuclease properties can lead to nuclease-specific artifacts and altered ChIP-exo "footprint."	4,7,8
DNase-seq (DNaseI hypersensitive site (DHS) sequencing)	*in vivo*	Identify chromatin regions that are most accessible to nuclease cleavage by DNaseI throughout the genome. These regions, referred to as "nucleosome-depleted," are thought to be "open chromatin" often important in regulation and occupied by TRFs and other non-nucleosomal proteins	Low concentrations of DNaseI are added to permeabilized cells or isolated nuclei. Open, or "nucleosome-depleted" regions of chromatin are more sensitive to cleavage by the enzyme. These can be detected genome-wide by deep sequencing.	Relies on permeabilization of cells or isolation of nuclei - both of which are inefficient steps that can create bias. Although less biased than other nucleases, DNaseI may have some sequence specificity that could influence DNA cleavage.	6,9–13
DNaseI footprinting	*in vitro*	Determine *in vitro* protein–DNA binding affinity and sequence specificity	High affinity protein–DNA interactions typically protect DNA from cleavage by DNaseI, resulting in a protein-specific "footprint" of protected DNA with an intensity of protection	Normally performed *in vitro* on unchromatinized DNA. For GR, predominantly limited to DBD, however, recently done on full-length purified GR in a limited	14–16

			roughly proportional to the fractional occupancy of the protein at the binding site. Labeled PCR amplified DNA is incubated with variable amounts of purified protein. The complexes are digested with limiting amounts of DNaseI and the footprint is visualized via PAGE.	number of conditions.	
FAIR-seq (Formaldehyde-Assisted Isolation of Regulatory Elements and deep sequencing)	*in vivo*	Alternate method to identify "nucleosome-depleted" regions of chromatin	This procedure identifies "open chromatin" based on the observation that nucleosome rich regions of the genome are more efficiently crosslinked by formaldehyde than nucleosome depleted regions. Briefly, genomic DNA is crosslinked, the DNA is then fragmented and phenol chloroform extracted to segregate nucleosome-bound (organic phase) from unbound (aqueous phase). "Nucleosome-depleted" DNA is identified via deep sequencing.	Limited use with GR in different cell types and under different physiological conditions with some notable exceptions.	17,18
X-ray Crystallography	*in vitro*	Structural analysis of GR–DNA Interactions	Obtain three-dimensional structure from exposing a crystal of protein–DNA complex to an x-ray beam. The diffraction pattern intensities obtained can be used to determine structure factors and calculate electron density maps from which structures can be derived.	For GR, work has been done with isolated domains and not with full-length receptor. X-ray crystallography is not optimal for intrinsically disordered proteins and almost half of GR is disordered. This causes issues with obtaining large quantities of purified full-length receptor required for crystallography. Additionally, crystallization can stabilize non-physiological conformations; hence, derived structures, especially protein–protein interaction surfaces, should be considered as models requiring validation through other tests.	19−25
NMR (Nuclear magnetic	*in vitro*	Probing of the specific chemical environment experienced by	NMR makes use of the particular magnetic properties of atomic nuclei to allow for the study of	For GR, NMR analysis has been conducted with isolated DBD bound to different GR binding	26−30

resonance)		specific atoms within a protein–DNA complex.	dynamic features of the protein-DNA interaction. Different labeling strategies can be used, either the protein or DNA is labeled with a heavy isotope, such as ^{15}N or ^{13}C. Spectra of individual residues or DNA bases can be used to compare protein alone to DNA bound, and vice-versa.	sequences. High concentrations of protein necessary for analysis precludes study of full-length GR.	
Molecular dynamics	*in silico*	Monitor the computer-simulated movement of atoms within a macromolecule in different states and ask how a structural model behaves under different perturbations.	Computer-modeled movements of atoms and molecules within a macromolecule are constrained by computed inter-particle forces and potential energies, interatomic potentials, and molecular mechanics force fields. Simulations occur over short, fixed time intervals and give information about dynamics within a macromolecule.	Molecular dynamics has been used to monitor how the GR-DBD interacts with different DNA sequences. This has been conducted with isolated domains as these are the only structural models available.	30–34
HDX-MS (Hydrogen deuterium exchange mass spectrometry)	*in vitro*	Used to identify changes in surface exposed regions of GR under different signaling contexts. These changes in solvent accessibility can be used to infer changes in conformation upon change of signaling context (*e.g.* ligand or DNA binding).	Deuterium exchanges more rapidly with solvent exposed amide hydrogens and slower with regions buried within the core of the protein or covered upon interaction with a partner. In this method, protein is incubated in a deuterated environment to allow amide hydrogens to exchange; then the reaction is quenched, the protein is digested with the acid protease pepsin, and is subjected to liquid chromatography mass spectrometry (LC-MS) to determine the amount of deuterium uptake, and thus solvent accessibility, of each proteolytic cleavage product.	HDX with GR-DNA interactions has been conducted with isolated domains. Magnitude of changes in solvent accessibility do not scale directly with changes in conformation or dynamics, for example small changes in solvent accessibility can be due to large changes in conformation. As GR forms a homodimer, it is difficult to determine the degree to which each sister subunit is undergoing a change is solvent accessibility.	35–37
FP (Fluorescence Polarization)	*in vitro*	GR–DNA Binding Assay	Monitor binding of proteins to fluorescently labeled oligonucleotide probe. A fluorophore, generally attached to the smaller of the two reactants (a short DNA fragment in this case), is excited with polarized light. Upon binding to protein, the combined mass of the complex increases, slowing the tumbling of the fluorophore. This decrease in	For GR–DNA interactions, FP has been conducted with full length GR and with isolated domains using purified proteins or extracts. Fluorescent label may affect the labeled DNA conformation or influence protein–DNA binding.	38,39

			tumbling rate is measured as a change in the intensity of fluorescence emission at a particular angle relative to the initial polarized excitation.		
EMSA (Electrophoretic Mobility Shift Assay)	*in vitro*	GR–DNA Binding Assay	Monitor binding of proteins to labeled an oligonucleotide probe. A protein–DNA complex will migrate more slowly on a non-denaturing polyacrylamide gel. Antibodies can also be used to target the protein of interest to generate a supershifted protein-antibody-DNA band.	*In vitro* assay that is useful for determining apparent equilibrium binding affinity, but can be difficult to quantify the kinetics of the protein–DNA complex. Has been conducted with full-length GR and with isolated domains using purified proteins or extracts.	21,27,28,40
Luciferase Reporter Assays	*in vivo*	Monitor transcriptional activity change in response to GR–genome interactions	Generally, a plasmid bearing a GOR-containing fragment cloned upstream of a minimal promoter and the luciferase reporter gene is transfected into cells where endogenous or overexpressed GR can bind to the GOR fragment and potentially stimulate or repress transcription of the luciferase reporter gene. The amount of luciferase reporter made can be measured by adding the substrate luciferin to cells. The luciferase enzyme will then catalyze a reaction that produces oxyluciferin and light, which is used to infer transcriptional activity.	Uses highly abundant exogenous DNA that does not reflect native chromatin states, endogenous nucleosomal packing, histone modifications, *etc*. Vulnerable to numerous artifacts upon TRF protein overexpression and gene dosage. Typically tests a single GOR fragment in conjunction with a non-native promoter. Does not account for post-RNA polymerase II initiation events, which have been shown to be highly regulated in endogenous contexts.	21,28,37,40,41
Precise genome editing (CRISPR-Cas, TALEN, and ZFN)	*in vivo*	Deletion, or single base pair resolution manipulation of potential GREs at endogenous loci. Used in conjunction with ChIP, qPCR, RNA-seq, transcriptomics, or other *in vivo* assays, can validate GRE activity and identify GRE target gene.	Each of these technologies uses a DNA sequence specific nuclease directed to a particular genomic site (e.g. a potential GRE). After cleavage of the endogenous locus, endogenous cellular machinery repairs the DNA break. Non-homologous end joining can introduce small insertions and deletions mutating the endogenous sequence. Alternatively, when a repair template is provided, the cells can be steered towards using homology driven repair to introduce predetermined	Although powerful methods, TALEN and ZFN technologies have proven to be time consuming and expensive routes to obtaining desired precisely edited genomic elements. CRISPR-Cas systems have shown great promise to precisely edit potential GREs as well as introduce affinity or fluorescent protein tags, and inducible degradation signaling sequences to endogenous	2,42

			sequences with single base pair precision.	TRF gene bodies. All of these technologies have low but finite rates of off-target editing.	
Fluorescence Microscopy	*in vivo*	Monitor GR-DNA Interactions in cells	In general, fluorescently tagged GR is expressed in a cell line of choice and its localization is monitored. This basic technique has been combined with Number and Brightness methodologies to infer the oligomeric status of GR on chromatin. This technique has also been combined with fluorescence recovery after photobleaching (FRAP) to analyze GR-chromatin dynamics.	Adding a large fluorescent tag to a protein may inhibit natural function, or have stress inducing / toxic effects on cells. Tagged proteins are overexpressed and strongly altered stoichiometries can produce various artifacts. Inference of oligomerization states seems particularly vulnerable to such artifacts.	43–48

Supplemental Table 2.1: Methods to probe glucocorticoid receptor-DNA Interactions.

Table of methods commonly used in the field to measure GR-DNA interactions, both *in vitro* and *in vivo*.

References

1. Bridgham, Jamie T., Geeta N. Eick, Claire Larroux, Kirti Deshpande, Michael J. Harms, and others. 2010. 'Protein Evolution by Molecular Tinkering: Diversification of the Nuclear Receptor Superfamily from a Ligand-Dependent Ancestor', *PLoS Biology*, 8.10, ed. by Laurence D. Hurst: e1000497.

2. Revollo, Javier R., and John A. Cidlowski. 2009. 'Mechanisms Generating Diversity in Glucocorticoid Receptor Signaling', *Annals of the New York Academy of Sciences*, 1179: 167–78.

3. Kadmiel, Mahita, and John A. Cidlowski. 2013. 'Glucocorticoid Receptor Signaling in Health and Disease', *Trends in Pharmacological Sciences*, 34.9 (Elsevier Ltd): 518–30.

4. Lewis-Tuffin, Laura J, Christine M Jewell, Rachelle J Bienstock, Jennifer B Collins, and John A Cidlowski. 2007. 'Human Glucocorticoid Receptor Binds RU-486 and Is Transcriptionally Active', *Molecular and Cellular Biology*, 27.6: 2266–82.

5. Picard, Didier, Bushra Khursheed, Michael J Garabedian, Marc G Fortin, Susan Lindquist, and others. 1990. 'Reduced Levels of hsp90 Compromise Steroid Receptor Action in Vivo', *Nature*, 348.6297: 166–68.

6. Chandler, V L, B a Maler, and K R Yamamoto. 1983. 'DNA Sequences Bound Specifically by Glucocorticoid Receptor in Vitro Render a Heterologous Promoter Hormone Responsive in Vivo.', *Cell*, 33.2: 489–99 7. Yamamoto, K. R. Steroid Receptor Regulated Transcription of Specific Genes and Gene Networks. *Annu. Rev. Genet.* **19,** 209–252 (1985).

8. Yamamoto, K R, B D Darimont, R L Wagner, and J A Iñiguez-Lluhí. 1998. 'Building Transcriptional Regulatory Complexes: Signals and Surfaces.', *Cold Spring Harbor Symposia on Quantitative Biology*, 63: 587–98.

9. McNally, James G. 2000. 'The Glucocorticoid Receptor: Rapid Exchange with Regulatory Sites in Living Cells', *Science*, 287.5456: 1262–65.

10. Becker, Matthias. 2002. 'Dynamic Behavior of Transcription Factors on a Natural Promoter in Living Cells', *EMBO Reports*, 3.12: 1188–94.

11. Stavreva, Diana A., W. G. Muller, Gordon L Hager, Carolyn L Smith, and James G McNally. 2004. 'Rapid Glucocorticoid Receptor Exchange at a Promoter Is Coupled to Transcription and Regulated by Chaperones and Proteasomes', *Molecular and Cellular Biology*, 24.7: 2682–97.

12. Meijsing, Sebastiaan H., Cem Elbi, Hans F. Luecke, Gordon L. Hager, and Keith R. Yamamoto. 2007. 'The Ligand Binding Domain Controls Glucocorticoid Receptor Dynamics Independent of Ligand Release.', *Molecular and Cellular Biology*, 27.7: 2442–51.

13. Sacta, Maria A, Yurii Chinenov, and Inez Rogatsky. 2016. 'Glucocorticoid Signaling: An Update from a Genomic Perspective', *Annual Review of Physiology*, 78.1: 155–80.

14. Wyllie, A H. 1980. 'Glucocorticoid-Induced Thymocyte Apoptosis Is Associated with Endogenous Endonuclease Activation', *Nature*, 284.5756: 555–56.

15. Patel, Rucha, Jasmine Williams-Dautovich, and Carolyn L Cummins. 2014. 'Minireview: New Molecular Mediators of Glucocorticoid Receptor Activity in Metabolic Tissues',

Molecular Endocrinology, 28.7: 999–1011.

16. Kumar, R, and E B Thompson. 1999. 'The Structure of the Nuclear Hormone Receptors.', *Steroids*, 64.5: 310–19.

17. Luisi, B. F., W. X. Xu, Z. Otwinowski, L. P. Freedman, K. R. Yamamoto, and others. 1991. 'Crystallographic Analysis of the Interaction of the Glucocorticoid Receptor with DNA.', *Nature*, pp. 497–50.

18. Meijsing, Sebastiaan H, Miles A Pufall, Alex Y So, Darren L Bates, Lin Chen, and others. 2009. 'DNA Binding Site Sequence Directs Glucocorticoid Receptor Structure and Activity', *Science*, 324.5925: 407–10.

19. Watson, Lisa C, Kristopher M Kuchenbecker, Benjamin J Schiller, John D Gross, Miles A Pufall, and others. 2013. 'The Glucocorticoid Receptor Dimer Interface Allosterically Transmits Sequence-Specific DNA Signals', *Nature Structural & Molecular Biology*, 20.7: 876–83.

20. Heck, S, M Kullmann, A Gast, H Ponta, H J Rahmsdorf, and others. 1994. 'A Distinct Modulating Domain in Glucocorticoid Receptor Monomers in the Repression of Activity of the Transcription Factor AP-1', *The EMBO Journal*, 13.17: 4087–95.

21. Schiller, Benjamin J, Rajas Chodankar, Lisa C Watson, Michael R Stallcup, and Keith R Yamamoto. 2014. 'Glucocorticoid Receptor Binds Half Sites as a Monomer and Regulates Specific Target Genes', *Genome Biology*, 15.8: 418.

22. Reichardt, Holger M., Klaus H. Kaestner, Jan Tuckermann, Oliver Kretz, Oliver Wessely, and others. 1998. 'DNA Binding of the Glucocorticoid Receptor Is Not

Essential for Survival', *Cell*, 93.4: 531–41.

23. Bledsoe, Randy K., Valerie G. Montana, Thomas B. Stanley, Chris J. Delves, Christopher J. Apolito, and others. 2002. 'Crystal Structure of the Glucocorticoid Receptor Ligand Binding Domain Reveals a Novel Mode of Receptor Dimerization and Coactivator Recognition', *Cell*, 110: 93–105.

24. Surjit, Milan, Krishna Priya Ganti, Atish Mukherji, Tao Ye, Guoqiang Hua, and others. 2011. 'Widespread Negative Response Elements Mediate Direct Repression by Agonist- Liganded Glucocorticoid Receptor', *Cell*, 145.2 (Elsevier Inc.): 224–41.

25. Hudson, William H, Christine Youn, and Eric A Ortlund. 2013. 'The Structural Basis of Direct Glucocorticoid-Mediated Transrepression', *Nature Structural & Molecular Biology*, 20.1: 53–58.

26. Hudson, William H, Bradley R Kossmann, Ian Mitchelle S. de Vera, Shih-wei Chuo, Emily R Weikum, and others. 2016. 'Distal Substitutions Drive Divergent DNA Specificity among Paralogous Transcription Factors through Subdivision of Conformational Space', *Proceedings of the National Academy of Sciences of the United States of America*, 113.2: 326–31.

27. Lim, Hee-woong, N Henriette Uhlenhaut, Alexander Rauch, Juliane Weiner, Sabine Hübner, and others. 2015. 'Genomic Redistribution of GR Monomers and Dimers Mediates Transcriptional Response to Exogenous Glucocorticoid in Vivo.', *Genome Research*, 25.6: 836–44.

28. Miner, J. N., and K. R. Yamamoto. 1992. 'The Basic Region of AP-1 Specifies

Glucocorticoid Receptor Activity at a Composite Response Element.', *Genes & Development*, 6.12B: 2491–2501.

29. De Bosscher, Karolien, Wim Vanden Berghe, and Guy Haegeman. 2001. 'Glucocorticoid Repression of AP-1 Is Not Mediated by Competition for Nuclear Coactivators', *Molecular Endocrinology*, 15.2: 219–27.

30. Luecke, Hans F, and Keith R Yamamoto. 2005. 'The Glucocorticoid Receptor Blocks P-TEFb Recruitment by NFkappaB to Effect Promoter-Specific Transcriptional Repression.', *Genes & Development*, 19.9: 1116–27.

31. De Bosscher, Karolien, Wim Vanden Berghe, and Guy Haegeman. 2003. 'The Interplay between the Glucocorticoid Receptor and Nuclear Factor-κB or Activator Protein-1: Molecular Mechanisms for Gene Repression', *Endocrine Reviews*, 24.4: 488–52.

32. Landt, Stephen G, Georgi K Marinov, Anshul Kundaje, Pouya Kheradpour, Florencia Pauli, and others. 2012. 'ChIP-Seq Guidelines and Practices of the ENCODE and modENCODE Consortia.', *Genome Research*, 22.9: 1813–31.

33. Steger, David J., Gregory R. Grant, Michael Schupp, Takuya Tomaru, Martina I. Lefterova, and others. 2010. 'Propagation of Adipogenic Signals through an Epigenomic Transition State', *Genes & Development*, 24.10: 1035–44.

34. John, Sam, Peter J Sabo, Robert E Thurman, Myong-Hee Sung, Simon C Biddie, and others. 2011. 'Chromatin Accessibility Pre-Determines Glucocorticoid Receptor Binding Patterns', *Nature Genetics*, 43.3: 264–68.

35. Grøntved, Lars, Sam John, Songjoon Baek, Ying Liu, John R Buckley, and others. 2013.

'C/EBP Maintains Chromatin Accessibility in Liver and Facilitates Glucocorticoid Receptor Recruitment to Steroid Response Elements', *The EMBO Journal*, 32.11: 1568–83.

36. So, A. Y. L., Chaivorapol, C., Bolton, E. C., Li, H. & Yamamoto, K. R. Determinants of cell- and gene-specific transcriptional regulation by the glucocorticoid receptor. *PLoS Genet.* **3,** e94 (2007).

37. Reddy, T. E. *et al.* Genomic determination of the glucocorticoid response reveals unexpected mechanisms of gene regulation. *Genome Res.* **19,** 2163–2171 (2009).

38. Merkenschlager, M. & Nora, E. P. CTCF and Cohesin in Genome Folding and Transcriptional Gene Regulation. *Annu. Rev. Genomics Hum. Genet.* **17,** 17–43 (2016).

39. Burd, C. J. *et al.* Analysis of chromatin dynamics during glucocorticoid receptor activation. *Mol. Cell. Biol.* **32,** 1805–1817 (2012).

40. Telorac, J. *et al.* Identification and characterization of DNA sequences that prevent glucocorticoid receptor binding to nearby response elements. *Nucleic Acids Res.* **44,** 6142–6156 (2016).

41. Uhlenhaut, N. H. *et al.* Insights into Negative Regulation by the Glucocorticoid Receptor from Genome-wide Profiling of Inflammatory Cistromes. *Mol. Cell* **49,** 158–171 (2013).

42. Presman, D. M. *et al.* Live Cell Imaging Unveils Multiple Domain Requirements for In Vivo Dimerization of the Glucocorticoid Receptor. *PLoS Biol.* **12,** e1001813 (2014).

43. Starick, S. R. *et al.* ChIP-exo signal associated with DNA-binding motifs provides insight into the genomic binding of the glucocorticoid receptor and cooperating transcription factors. *Genome Res.* **25,** 825–835 (2015).

44. So, A. Y.-L., Teresita U Bernal, Marlisa L Pillsbury, Keith R Yamamoto, and Brian J Feldman. 2009. 'Glucocorticoid Regulation of the Circadian Clock Modulates Glucose Homeostasis', *Proceedings of the National Academy of Sciences*, 106.41: 17582–87.

45. Ran, F Ann, Patrick D Hsu, Jason Wright, Vineeta Agarwala, David A Scott, and others. 2013. 'Genome Engineering Using the CRISPR-Cas9 System', *Nature Protocols*, 8.11: 2281–2308.

46. Rogatsky, Inez, J.-C. Wang, Mika K Derynck, Daisuke F Nonaka, Daniel B Khodabakhsh, and others. 2003. 'Target-Specific Utilization of Transcriptional Regulatory Surfaces by the Glucocorticoid Receptor', *Proceedings of the National Academy of Sciences*, 100.24: 13845–50.

47. Thomas-Chollier, Morgane, Lisa C Watson, Samantha B Cooper, Miles A Pufall, Jennifer S Liu, and others. 2013. 'A Naturally Occuring Insertion of a Single Amino Acid Rewires Transcriptional Regulation by Glucocorticoid Receptor Isoforms', *Proceedings of the National Academy of Sciences*, 110.44: 17826–31.

48. Chen, Sheng-Hong, Kiriko Masuno, Samantha B Cooper, and Keith R Yamamoto. 2013. 'Incoherent Feed-Forward Regulatory Logic Underpinning Glucocorticoid Receptor Action', *Proceedings of the National Academy of Sciences*, 110.5: 1964–69.

49. Mangan, S, and U Alon. 2003. 'Structure and Function of the Feed-Forward Loop

Network Motif', *Proceedings of the National Academy of Sciences*, 100.21: 11980–85.

50. Chinenov, Yurii, Maddalena Coppo, Rebecca Gupte, Maria A Sacta, and Inez Rogatsky. 2014. 'Glucocorticoid Receptor Coordinates Transcription Factor-Dominated Regulatory Network in Macrophages', *BMC Genomics*, 15.1: 656.

51. Hudson, William H, and Eric A Ortlund. 2014. 'The Structure, Function and Evolution of Proteins That Bind DNA and RNA', *Nature Reviews Molecular Cell Biology*, 15.11 (Nature Publishing Group): 749–60.

52. Lefstin, Jeffrey A, and Keith R Yamamoto. 1998. 'Allosteric Effects of DNA on Transcriptional Regulators.', *Nature*, 392.6679: 885–88.

53. Presman, Diego M, Sourav Ganguly, R Louis Schiltz, Thomas A Johnson, Tatiana S Karpova, and others. 2016. 'DNA Binding Triggers Tetramerization of the Glucocorticoid Receptor in Live Cells', *Proceedings of the National Academy of Sciences*, 113.29: 8236–41.

54. Gebhardt, J Christof M, David M Suter, Rahul Roy, Ziqing W Zhao, Alec R Chapman, and others. 2013. 'Single-Molecule Imaging of Transcription Factor Binding to DNA in Live Mammalian Cells.', *Nature Methods*, 10.5: 421–26.

55. Robblee, James P., Michael T. Miura, and David L. Bain. 2012. 'Glucocorticoid Receptor–Promoter Interactions: Energetic Dissection Suggests a Framework for the Specificity of Steroid Receptor-Mediated Gene Regulation', *Biochemistry*, 51.22: 4463–72.

56. Bain, David L., Qin Yang, Keith D. Connaghan, James P. Robblee, Michael T. Miura,

and others. 2012. ‘Glucocorticoid Receptor-DNA Interactions: Binding Energetics Are the Primary Determinant of Sequence-Specific Transcriptional Activity’, *Journal of Molecular Biology*, 422.1 (Elsevier Ltd): 18–32.

57. Schöne, Stefanie, Marcel Jurk, Mahdi Bagherpoor Helabad, Iris Dror, Isabelle Lebars, and others. 2016. ‘Sequences Flanking the Core-Binding Site Modulate Glucocorticoid Receptor Structure and Activity’, *Nature Communications*, 7: 12621.

58. Zhang, Jun, Michael J Chalmers, Keith R Stayrook, Lorri L Burris, Yongjun Wang, and others. 2011. ‘DNA Binding Alters Coactivator Interaction Surfaces of the Intact VDR–RXR Complex’, *Nature Structural & Molecular Biology*, 18.5: 556–63.

59. Thornton, J W. 2001. ‘Evolution of Vertebrate Steroid Receptors from an Ancestral Estrogen Receptor by Ligand Exploitation and Serial Genome Expansions’, *Proceedings of the National Academy of Sciences*, 98.10: 5671–76.

60. Eick, Geeta N., Jennifer K. Colucci, Michael J. Harms, Eric A. Ortlund, and Joseph W. Thornton. 2012. ‘Evolution of Minimal Specificity and Promiscuity in Steroid Hormone Receptors’, *PLoS Genetics*, 8.11, ed. by Jianzhi Zhang: e1003072.

61. He, Yuanzheng, Wei Yi, Kelly Suino-Powell, X Edward Zhou, W David Tolbert, and others. 2014. ‘Structures and Mechanism for the Design of Highly Potent Glucocorticoids.’, *Cell Research*, 24.6 (Nature Publishing Group): 713–26.

62. Kauppi, Björn, Clarissa Jakob, Mathias Färnegårdh, Jie Yang, Harri Ahola, and others. 2003. ‘The Three-Dimensional Structures of Antagonistic and Agonistic Forms of the Glucocorticoid Receptor Ligand-Binding Domain: RU-486 Induces a

Transconformation That Leads to Active Antagonism.', *The Journal of Biological Chemistry*, 278.25: 22748–54.

63. Wang, Jen-Chywan, Nilesh Shah, Carlos Pantoja, Sebastiaan H. Meijsing, Joseph D. Ho, and others. 2006. 'Novel Arylpyrazole Compounds Selectively Modulate Glucocorticoid Receptor Regulatory Activity.', *Genes & Development*, 20.6: 689–99.

64. Ricketson, D., U. Hostick, L. Fang, K.R. Yamamoto, and B.D. Darimont. 2007. 'A Conformational Switch in the Ligand-Binding Domain Regulates the Dependence of the Glucocorticoid Receptor on Hsp90', *Journal of Molecular Biology*, 368.3: 729–41.

65. Ismaili, Naima, and Michael J Garabedian. 2004. 'Modulation of Glucocorticoid Receptor Function via Phosphorylation.', *Annals of the New York Academy of Sciences*, 1024.1: 86–101.

66. Tian, Sha, Hetti Poukka, Jorma J Palvimo, and Olli A Jänne. 2002. 'Small Ubiquitin-Related Modifier-1 (SUMO-1) Modification of the Glucocorticoid Receptor.', *The Biochemical Journal*, 367.Pt 3: 907–11.

67. Wallace, Andrew D., and J A Cidlowski. 2001. 'Proteasome-Mediated Glucocorticoid Receptor Degradation Restricts Transcriptional Signaling by Glucocorticoids.', *The Journal of Biological Chemistry*, 276.46: 42714–21.

68. Itoh, M, M Adachi, H Yasui, M Takekawa, H Tanaka, and others. 2002. 'Nuclear Export of Glucocorticoid Receptor Is Enhanced by c-Jun N-Terminal Kinase-Mediated Phosphorylation', *Molecular Endocrinology*, 16.10: 2382–92.

69. Galigniana, M D, G Piwien-Pilipuk, and J Assreuy. 1999. 'Inhibition of Glucocorticoid

Receptor Binding by Nitric Oxide.', *Molecular Pharmacology*, 55.2: 317–23.

70. Ward, Robert D., and Nancy L. Weigel. 2009. 'Steroid Receptor Phosphorylation: Assigning Function to Site-Specific Phosphorylation', *BioFactors*, 35.6: 528–36.

71. Housley, Paul R, and William B Pratt. 1983. 'Direct Demonstration of Glucocorticoid Receptor Phosphorylation by Intact L-Cells.', *The Journal of Biological Chemistry*, 258.7: 4630–35.

72. Wang, Zhen, Weiwei Chen, Evelyn Kono, Thoa Dang, and Michael J Garabedian. 2007. 'Modulation of Glucocorticoid Receptor Phosphorylation and Transcriptional Activity by a C-Terminal-Associated Protein Phosphatase', *Molecular Endocrinology*, 21.3: 625–34.

73. Bodwell, Jack E, Jeffrey C Webster, Christine M Jewell, John A Cidlowski, Jiong-ming Hu, and others. 1998. 'Glucocorticoid Receptor Phosphorylation: Overview, Function and Cell Cycle-Dependence', *The Journal of Steroid Biochemistry and Molecular Biology*, 65.1–6: 91–99.

74. Krstic, M D, I Rogatsky, K R Yamamoto, and M J Garabedian. 1997. 'Mitogen-Activated and Cyclin-Dependent Protein Kinases Selectively and Differentially Modulate Transcriptional Enhancement by the Glucocorticoid Receptor.', *Molecular and Cellular Biology*, 17.7: 3947–54.

75. Mason, Sheryl A, and P R Housley. 1993. 'Site-Directed Mutagenesis of the Phosphorylation Sites in the Mouse Glucocorticoid Receptor.', *The Journal of Biological Chemistry*, 268.29: 21501–4.

76. Jewell, Christine M. 1997. 'Mouse Glucocorticoid Receptor Phosphorylation Status Influences Multiple Functions of the Receptor Protein', *Journal of Biological Chemistry*, 272.14: 9287–93.

77. Chen, Weiwei, Thoa Dang, Raymond D Blind, Zhen Wang, Claudio N Cavasotto, and others. 2008. 'Glucocorticoid Receptor Phosphorylation Differentially Affects Target Gene Expression', *Molecular Endocrinology*, 22.8: 1754–66.

78. Garza, Anna M S, Shagufta H Khan, and Raj Kumar. 2010. 'Site-Specific Phosphorylation Induces Functionally Active Conformation in the Intrinsically Disordered N-Terminal Activation Function (AF1) Domain of the Glucocorticoid Receptor', *Molecular and Cellular Biology*, 30.1: 220–30.

79. Miller, Aaron L, M Scott Webb, Alicja J Copik, Yongxin Wang, Betty H Johnson, and others. 2005. 'p38 Mitogen-Activated Protein Kinase (MAPK) Is a Key Mediator in Glucocorticoid-Induced Apoptosis of Lymphoid Cells: Correlation between p38 MAPK Activation and Site-Specific Phosphorylation of the Human Glucocorticoid Receptor at Serine 211.', *Molecular Endocrinology (Baltimore, Md.)*, 19.6: 1569–83.

80. Wang, Zhen, Jeremy Frederick, and Michael J. Garabedian. 2002. 'Deciphering the Phosphorylation "Code" of the Glucocorticoid Receptor in Vivo', *Journal of Biological Chemistry*, 277.29: 26573–80.

81. King, K L, and J A Cidlowski. 1998. 'Cell Cycle Regulation and Apoptosis.', *Annual Review of Physiology*, 60.1: 601–17.

82. Galliher-Beckley, Amy J., and John A. Cidlowski. 2009. 'Emerging Roles of

Glucocorticoid Receptor Phosphorylation in Modulating Glucocorticoid Hormone Action in Health and Disease', *IUBMB Life*, 61.10: 979–86.

83. Galliher-Beckley, Amy Jo, Jason Grant Williams, Jennifer Brady Collins, and John Anthony Cidlowski. 2008. 'Glycogen Synthase Kinase 3 -Mediated Serine Phosphorylation of the Human Glucocorticoid Receptor Redirects Gene Expression Profiles', *Molecular and Cellular Biology*, 28.24: 7309–22.

84. Deroo, Bonnie J, Claudia Rentsch, Sowmini Sampath, Janel Young, Donald B DeFranco, and others. 2002. 'Proteasomal Inhibition Enhances Glucocorticoid Receptor Transactivation and Alters Its Subnuclear Trafficking.', *Molecular and Cellular Biology*, 22.12: 4113–23.

85. Wallace, Andrew D., Yan Cao, Sindhu Chandramouleeswaran, and John A. Cidlowski. 2010. 'Lysine 419 Targets Human Glucocorticoid Receptor for Proteasomal Degradation', *Steroids*, 75.12 (Elsevier Inc.): 1016–23.

86. Kino, Tomoshige, Szu-Heng Liou, Evangelia Charmandari, and George P Chrousos. 2006. 'Glucocorticoid Receptor Mutants Demonstrate Increased Motility inside the Nucleus of Living Cells: Time of Fluorescence Recovery after Photobleaching (FRAP) Is an Integrated Measure of Receptor Function.', *Molecular Medicine (Cambridge, Mass.)*, 10.7–12: 80–88.

87. Gill, Grace. 2005. 'Something about SUMO Inhibits Transcription', *Current Opinion in Genetics & Development*, 15.5: 536–41.

88. Le Drean, Yves, Nathalie Mincheneau, Pascale Le Goff, and Denis Michel. 2002.

'Potentiation of Glucocorticoid Receptor Transcriptional Activity by Sumoylation', *Endocrinology*, 143.9: 3482–89.

89. Paakinaho, Ville, Sanna Kaikkonen, Harri Makkonen, Vladimir Benes, and Jorma J. Palvimo. 2014. 'SUMOylation Regulates the Chromatin Occupancy and Anti-Proliferative Gene Programs of Glucocorticoid Receptor.', *Nucleic Acids Research*, 42.3: 1575–92.

90. Treuter, Eckardt, and Nicolas Venteclef. 2011. 'Transcriptional Control of Metabolic and Inflammatory Pathways by Nuclear Receptor SUMOylation', *Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease*, 1812.8 (Elsevier B.V.): 909–18.

91. Flotho, Annette, and Frauke Melchior. 2013. 'Sumoylation: A Regulatory Protein Modification in Health and Disease', *Annual Review of Biochemistry*, 82.1: 357–85.

92. Hua, Guoqiang, Laetitia Paulen, and Pierre Chambon. 2016b. 'GR SUMOylation and Formation of an SUMO-SMRT/NCoR1-HDAC3 Repressing Complex Is Mandatory for GC-Induced IR nGRE-Mediated Transrepression', *Proceedings of the National Academy of Sciences*, 113.5: E626–34.

93. Hua, Guoqiang, Krishna Priya Ganti, and Pierre Chambon. 2016a. 'Glucocorticoid-Induced Tethered Transrepression Requires SUMOylation of GR and Formation of a SUMO-SMRT/NCoR1-HDAC3 Repressing Complex.', *Proceedings of the National Academy of Sciences of the United States of America*, 113.5: E635–43.

94. Nader, Nancy, George P Chrousos, and Tomoshige Kino. 2009. 'Circadian Rhythm Transcription Factor CLOCK Regulates the Transcriptional Activity of the

Glucocorticoid Receptor by Acetylating Its Hinge Region Lysine Cluster: Potential Physiological Implications', *The FASEB Journal*, 23.5: 1572–83.

95. Kino, Tomoshige, and George P. Chrousos. 2011. 'Acetylation-Mediated Epigenetic Regulation of Glucocorticoid Receptor Activity: Circadian Rhythm-Associated Alterations of Glucocorticoid Actions in Target Tissues', *Molecular and Cellular Endocrinology*, 336.1–2 (Elsevier Ireland Ltd): 23–30.

96. Ito, Kazuhiro, Satoshi Yamamura, Sarah Essilfie-Quaye, Borja Cosio, Misako Ito, and others. 2006. 'Histone Deacetylase 2–mediated Deacetylation of the Glucocorticoid Receptor Enables NF-κB Suppression', *The Journal of Experimental Medicine*, 203.1: 7–13.

97. Kröncke, K D, and C Carlberg. 2000. 'Inactivation of Zinc Finger Transcription Factors Provides a Mechanism for a Gene Regulatory Role of Nitric Oxide.', *FASEB Journal : Official Publication of the Federation of American Societies for Experimental Biology*, 14.1: 166–73.

98. Diamond, M., J. Miner, S. Yoshinaga, and K. Yamamoto. 1990. 'Transcription Factor Interactions: Selectors of Positive or Negative Regulation from a Single DNA Element', *Science*, 249.4974: 1266–72.

99. Miner, J N, M I Diamond, and K R Yamamoto. 1991. 'Joints in the Regulatory Lattice: Composite Regulation by Steroid Receptor-AP1 Complexes.', *Cell Growth & Differentiation : The Molecular Biology Journal of the American Association for Cancer Research*, 2.10: 525–30.

100. Jenkins, Bethany D, Christian B Pullen, and Beatrice D Darimont. 2001. 'Novel Glucocorticoid Receptor Coactivator Effector Mechanisms', *Trends in Endocrinology & Metabolism*, 12.3: 122–26.

101. Bannister, Andrew J, and Tony Kouzarides. 2011. 'Regulation of Chromatin by Histone Modifications', *Cell Research*, 21.3 (Nature Publishing Group): 381–95.

102. Millard, Christopher J., Peter J. Watson, Louise Fairall, and John W R Schwabe. 2013. 'An Evolving Understanding of Nuclear Receptor Coregulator Proteins', *Journal of Molecular Endocrinology*, 51.3: T23–36.

103. Vandevyver, Sofie, Lien Dejager, and Claude Libert. 2014. 'Comprehensive Overview of the Structure and Regulation of the Glucocorticoid Receptor.', *Endocrine Reviews*, 35.4: 671–93.

104. Parker, M G, and R White. 1996. 'Nuclear Receptors Spring into Action.', *Nature Structural Biology*, 3.2: 113–15.

105. Wang, Zhulun, Gérard Benoit, Jinsong Liu, Srividya Prasad, Piia Aarnisalo, and others. 2003. 'Structure and Function of Nurr1 Identifies a Class of Ligand-Independent Nuclear Receptors', *Nature*, 423.6939: 555–60.

106. Torchia, J, D W Rose, J Inostroza, Y Kamei, S Westin, and others. 1997. 'The Transcriptional Co-Activator p/CIP Binds CBP and Mediates Nuclear-Receptor Function.', *Nature*, 387.6634: 677–84.

107. Heery, D M, E Kalkhoven, S Hoare, and M G Parker. 1997. 'A Signature Motif in Transcriptional Co-Activators Mediates Binding to Nuclear Receptors.', *Nature*,

387.6634 (England): 733–36.

108. Hu, Xiao, and Mitchell A Lazar. 1999. 'The CoRNR Motif Controls the Recruitment of Corepressors by Nuclear Hormone Receptors.', *Nature*, 402.6757: 93–96.

109. 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).

110. Khan, Shagufta H., Jun Ling, and Raj Kumar. 2011. 'TBP Binding-Induced Folding of the Glucocorticoid Receptor AF1 Domain Facilitates Its Interaction with Steroid Receptor Coactivator-1', *PLoS ONE*, 6.7, ed. by Austin John Cooney: e21939.

111. Dahlman-Wright, Karin, T Almlöf, Iain J McEwan, J A Gustafsson, and A P Wright. 1994. 'Delineation of a Small Region within the Major Transactivation Domain of the Human Glucocorticoid Receptor That Mediates Transactivation of Gene Expression.', *Proceedings of the National Academy of Sciences of the United States of America*, 91.5: 1619–23.

112. Yang, L, J Guerrero, H Hong, D B DeFranco, and M R Stallcup. 2000. 'Interaction of the tau2 Transcriptional Activation Domain of Glucocorticoid Receptor with a Novel Steroid Receptor Coactivator, Hic-5, Which Localizes to Both Focal Adhesions and the Nuclear Matrix.', *Molecular Biology of the Cell*, 11.6: 2007–18.

113. Chodankar, Rajas, D.-Y. Wu, Benjamin J Schiller, Keith R Yamamoto, and Michael R Stallcup. 2014. 'Hic-5 Is a Transcription Coregulator That Acts before And/or after Glucocorticoid Receptor Genome Occupancy in a Gene-Selective Manner',

Proceedings of the National Academy of Sciences, 111.11: 4007–12.

114. Dasgupta, Subhamoy, David M Lonard, and Bert W O'Malley. 2014. 'Nuclear Receptor Coactivators: Master Regulators of Human Health and Disease.', *Annual Review of Medicine*, 65.3: 279–92.

115. Perissi, V, and M G Rosenfeld. 2005. 'Controlling Nuclear Receptors: The Circular Logic of Cofactor Cycles', *Nature Reviews: Molecular Cell Biology*, 6.7: 542–54.

116. Lonard, David M., and Bert W. O'Malley. 2007. 'Nuclear Receptor Coregulators: Judges, Juries, and Executioners of Cellular Regulation', *Molecular Cell*, 27.5: 691–700.

117. Fonte, Cosima, Julien Grenier, Amalia Trousson, Anne Chauchereau, Olivier Lahuna, and others. 2005. 'Involvement of Beta-Catenin and Unusual Behavior of CBP and p300 in Glucocorticosteroid Signaling in Schwann Cells.', *Proceedings of the National Academy of Sciences of the United States of America*, 102.40: 14260–65.

118. Xu, Jianming, Ray-Chang Wu, and Bert W O'Malley. 2009. 'Normal and Cancer-Related Functions of the p160 Steroid Receptor Co-Activator (SRC) Family', *Nature Reviews Cancer*, 9.9 (Nature Publishing Group): 615–30.

119. Kim, Jeong Hoon, Hongwei Li, and Michael R. Stallcup. 2003. 'CoCoA, a Nuclear Receptor Coactivator Which Acts through an N-Terminal Activation Domain of p160 Coactivators', *Molecular Cell*, 12.6: 1537–49.

120. Stallcup, Michael R, Jeong Hoon Kim, Catherine Teyssier, Young-Ho Lee, Han Ma, and others. 2003. 'The Roles of Protein–protein Interactions and Protein Methylation

in Transcriptional Activation by Nuclear Receptors and Their Coactivators', *The Journal of Steroid Biochemistry and Molecular Biology*, 85.2–5: 139–45.

121. Kim, Jeong Hoon, Catherine K. Yang, Kyu Heo, Robert G. Roeder, Woojin An, and others. 2008. 'CCAR1, a Key Regulator of Mediator Complex Recruitment to Nuclear Receptor Transcription Complexes', *Molecular Cell*, 31.4: 510–19.

122. Szapary, Daniele, Ying Huang, and S Stoney Simons. 1999. 'Opposing Effects of Corepressor and Coactivators in Determining the Dose-Response Curve of Agonists, and Residual Agonist Activity of Antagonists, for Glucocorticoid Receptor-Regulated Gene Expression', *Molecular Endocrinology*, 13.12: 2108–21.

123. Trousson, Amalia, Julien Grenier, Cosima Fonte, Liliane Massaad-Massade, Michael Schumacher, and others. 2007. 'Recruitment of the p160 Coactivators by the Glucocorticoid Receptor: Dependence on the Promoter Context and Cell Type but Not Hypoxic Conditions', *The Journal of Steroid Biochemistry and Molecular Biology*, 104.3–5: 305–11.

124. Voegel, Johannes J., M J Heine, M Tini, V Vivat, P Chambon, and others. 1998. 'The Coactivator TIF2 Contains Three Nuclear Receptor-Binding Motifs and Mediates Transactivation through CBP Binding-Dependent and -Independent Pathways.', *The EMBO Journal*, 17.2: 507–19.

125. Darimont, Beatrice D, Richard L Wagner, James W Apriletti, Michael R Stallcup, Peter J Kushner, and others. 1998. 'Structure and Specificity of Nuclear Receptor-Coactivator Interactions', *Genes & Development*, 12.21: 3343–56.

126. Li, Xiaotao, Jiemin Wong, Sophia Y Tsai, Ming-jer Tsai, and Bert W O'Malley. 2003. 'Progesterone and Glucocorticoid Receptors Recruit Distinct Coactivator Complexes and Promote Distinct Patterns of Local Chromatin Modification.', *Molecular and Cellular Biology*, 23.11: 3763–73.

127. Kurihara, Isao, Hirotaka Shibata, Toshihiko Suzuki, Takashi Ando, Sakiko Kobayashi, and others. 2002. 'Expression and Regulation of Nuclear Receptor Coactivators in Glucocorticoid Action', *Molecular and Cellular Endocrinology*, 189.1–2: 181–89.

128. Ronacher, Katharina, Katie Hadley, Chanel Avenant, Elisabeth Stubsrud, S. Stoney Simons, and others. 2009. 'Ligand-Selective Transactivation and Transrepression via the Glucocorticoid Receptor: Role of Cofactor Interaction', *Molecular and Cellular Endocrinology*, 299.2: 219–31.

129. Ogawa, Hidesato, Ruth T. Yu, Tokuko Haraguchi, Yasushi Hiraoka, Yoshihiro Nakatani, and others. 2004. 'Nuclear Structure-Associated TIF2 Recruits Glucocorticoid Receptor and Its Target DNA', *Biochemical and Biophysical Research Communications*, 320.1: 218–25.

130. Dobrovolna, Jana, Yurii Chinenov, Megan a Kennedy, Bill Liu, and Inez Rogatsky. 2012. 'Glucocorticoid-Dependent Phosphorylation of the Transcriptional Coregulator GRIP1', *Molecular and Cellular Biology*, 32.4: 730–39.

131. Rogatsky, Inez, Hans F Luecke, Dale C Leitman, and Keith R Yamamoto. 2002. 'Alternate Surfaces of Transcriptional Coregulator GRIP1 Function in Different Glucocorticoid Receptor Activation and Repression Contexts', *Proceedings of the National Academy of Sciences*, 99.26: 16701–6.

132. Kamei, Y, L Xu, T Heinzel, J Torchia, R Kurokawa, and others. 1996. 'A CBP Integrator Complex Mediates Transcriptional Activation and AP-1 Inhibition by Nuclear Receptors.', *Cell*, 85.3: 403–14.

133. Sheppard, K.-A., K. M. Phelps, A. J. Williams, D. Thanos, C. K. Glass, and others. 1998. 'Nuclear Integration of Glucocorticoid Receptor and Nuclear Factor-κB Signaling by CREB-Binding Protein and Steroid Receptor Coactivator-1', *Journal of Biological Chemistry*, 273.45: 29291–94.

134. De Bosscher, K., W. Vanden Berghe, L. Vermeulen, S. Plaisance, E. Boone, and others. 2000. 'Glucocorticoids Repress NF-Kappa B-Driven Genes by Disturbing the Interaction of p65 with the Basal Transcription Machinery, Irrespective of Coactivator Levels in the Cell', *Proceedings of the National Academy of Sciences*, 97.8: 3919–24.

135. Allen, Benjamin L, and Dylan J Taatjes. 2015. 'The Mediator Complex: A Central Integrator of Transcription', *Nature Reviews Molecular Cell Biology*, 16.3 (Nature Publishing Group): 155–66.

136. Knuesel, Matthew T., and Dylan J. Taatjes. 2011. 'Mediator and Post-Recruitment Regulation of RNA Polymerase II', *Transcription*, 2.1: 28–31.

137. Meyer, Krista D, Shih-Chieh Lin, Carrie Bernecky, Yuefeng Gao, and Dylan J Taatjes. 2010. 'p53 Activates Transcription by Directing Structural Shifts in Mediator', *Nature Structural & Molecular Biology*, 17.6 (Nature Publishing Group): 753–60.

138. Taatjes, Dylan J, Anders M Näär, Frank Andel, Eva Nogales, and Robert Tjian. 2002. 'Structure, Function, and Activator-Induced Conformations of the CRSP Coactivator.',

Science (New York, N.Y.), 295.5557: 1058–62.

139. Knuesel, Matthew T., Krista D. Meyer, Carrie Bernecky, and Dylan J. Taatjes. 2009. 'The Human CDK8 Subcomplex Is a Molecular Switch That Controls Mediator Coactivator Function', *Genes & Development*, 23.4: 439–51.

140. Bernecky, Carrie, Patricia Grob, Christopher C. Ebmeier, Eva Nogales, and Dylan J. Taatjes. 2011. 'Molecular Architecture of the Human Mediator–RNA Polymerase II–TFIIF Assembly', *PLoS Biology*, 9.3, ed. by Gregory A. Petsko: e1000603.

141. Meyer, Krista D, Aaron J Donner, Matthew T Knuesel, Autumn G York, Joaquín M Espinosa, and others. 2008. 'Cooperative Activity of cdk8 and GCN5L within Mediator Directs Tandem Phosphoacetylation of Histone H3', *The EMBO Journal*, 27.10: 1447–57.

142. Hittelman, A B, D Burakov, J A Iñiguez-Lluhí, L P Freedman, and M J Garabedian. 1999. 'Differential Regulation of Glucocorticoid Receptor Transcriptional Activation via AF-1-Associated Proteins.', *The EMBO Journal*, 18.19: 5380–88.

143. Chen, Weiwei, Inez Rogatsky, and Michael J Garabedian. 2006. 'MED14 and MED1 Differentially Regulate Target-Specific Gene Activation by the Glucocorticoid Receptor.', *Molecular Endocrinology (Baltimore, Md.)*, 20.3: 560–72.

144. Narlikar, Geeta J., Ramasubramanian Sundaramoorthy, and Tom Owen-Hughes. 2013. 'Mechanisms and Functions of ATP-Dependent Chromatin-Remodeling Enzymes', *Cell*, 154.3 (The Authors): 490–503.

145. Pazin, M J, and J T Kadonaga. 1997. 'SWI2/SNF2 and Related Proteins: ATP-Driven

Motors That Disrupt Protein-DNA Interactions?', *Cell*, 88.6: 737–40).

146. Fryer, Christy J., and Trevor K. Archer. 1998. 'Chromatin Remodelling by the Glucocorticoid Receptor Requires the BRG1 Complex.', *Nature*, 393.6680: 88–91.

147. Engel, Karin B, and Keith R Yamamoto. 2011. 'The Glucocorticoid Receptor and the Coregulator Brm Selectively Modulate Each Other's Occupancy and Activity in a Gene-Specific Manner', *Molecular and Cellular Biology*, 31.16: 3267–76.

148. Ostlund Farrants, A K, P Blomquist, H Kwon, and O Wrange. 1997. 'Glucocorticoid Receptor-Glucocorticoid Response Element Binding Stimulates Nucleosome Disruption by the SWI/SNF Complex.', *Molecular and Cellular Biology*, 17.2: 895–905.

149. Collingwood, T N, F D Urnov, and A P Wolffe. 1999. 'Nuclear Receptors: Coactivators, Corepressors and Chromatin Remodeling in the Control of Transcription.', *Journal of Molecular Endocrinology*, 23.3: 255–75.

150. King, Heather A., Kevin W. Trotter, and Trevor K. Archer. 2012. 'Chromatin Remodeling during Glucocorticoid Receptor Regulated Transactivation', *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms*, 1819.7: 716–26.

151. Yoshinaga, S., C. Peterson, I Herskowitz, and K. R. Yamamoto. 1992. 'Roles of SWI1, SWI2, and SWI3 Proteins for Transcriptional Enhancement by Steroid Receptors', *Science*, 258.5088: 1598–1604.

152. Wallberg, Annika E, Kristen E Neely, Ahmed H Hassan, J A Gustafsson, Jerry L Workman, and others. 2000. 'Recruitment of the SWI-SNF Chromatin Remodeling

Complex as a Mechanism of Gene Activation by the Glucocorticoid Receptor tau1 Activation Domain.', *Molecular and Cellular Biology*, 20.6: 2004–13.

153. Muratcioglu, Serena, Diego M. Presman, John R. Pooley, Lars Grøntved, Gordon L. Hager, and others. 2015. 'Structural Modeling of GR Interactions with the SWI/SNF Chromatin Remodeling Complex and C/EBP.', *Biophysical Journal*, 109.6 (Biophysical Society): 1227–39.

154. Chen, D., H Ma, H Hong, S S Koh, S M Huang, and others. 1999. 'Regulation of Transcription by a Protein Methyltransferase.', *Science (New York, N.Y.)*, 284.5423: 2174–77.

155. Bittencourt, Danielle, D.-Y. Wu, Kwang Won Jeong, Daniel S Gerke, Laurie Herviou, and others. 2012. 'G9a Functions as a Molecular Scaffold for Assembly of Transcriptional Coactivators on a Subset of Glucocorticoid Receptor Target Genes', *Proceedings of the National Academy of Sciences*, 109.48: 19673–78.

156. Lee, Kenneth K, and Jerry L Workman. 2007. 'Histone Acetyltransferase Complexes: One Size Doesn't Fit All', *Nature Reviews Molecular Cell Biology*, 8.4: 284–95.

157. Almlöf, Tova, Annika E. Wallberg, Jan-Åke Gustafsson, and Anthony P. H. Wright. 1998. 'Role of Important Hydrophobic Amino Acids in the Interaction between the Glucocorticoid Receptor τ1-Core Activation Domain and Target Factors', *Biochemistry*, 37.26: 9586–94.

158. Yao, T P, G Ku, N Zhou, R Scully, and D M Livingston. 1996. 'The Nuclear Hormone Receptor Coactivator SRC-1 Is a Specific Target of p300.', *Proceedings of the*

National Academy of Sciences, 93.20: 10626–31.

159. Wallberg, A E, K E Neely, J A Gustafsson, J L Workman, A P Wright, and others. 1999. 'Histone Acetyltransferase Complexes Can Mediate Transcriptional Activation by the Major Glucocorticoid Receptor Activation Domain.', *Molecular and Cellular Biology*, 19.9: 5952–59.

160. Fonte, Cosima, Amalia Trousson, Julien Grenier, Michael Schumacher, and Charbel Massaad. 2007. 'Opposite Effects of CBP and p300 in Glucocorticoid Signaling in Astrocytes', *The Journal of Steroid Biochemistry and Molecular Biology*, 104.3–5: 220–27.

161. Verdin, Eric, and Melanie Ott. 2015. '50 Years of Protein Acetylation: From Gene Regulation to Epigenetics, Metabolism and Beyond.', *Nature Reviews. Molecular Cell Biology*, 16.4 (Nature Publishing Group): 258–64.

162. Shahbazian, Mona D, and Michael Grunstein. 2007. 'Functions of Site-Specific Histone Acetylation and Deacetylation.', *Annual Review of Biochemistry*, 76: 75–100.

163. Stewart, M. David, and Jiemin Wong. 2009. 'Nuclear Receptor Repression: Regulatory Mechanisms and Physiological Implications.', *Progress in Molecular Biology and Translational Science*, 87.C (Elsevier Inc.): 235–59.

164. Schoch, Guillaume A., Brigitte D'Arcy, Martine Stihle, Dominique Burger, Dominik Bär, and others. 2010. 'Molecular Switch in the Glucocorticoid Receptor: Active and Passive Antagonist Conformations', *Journal of Molecular Biology*, 395.3 (Elsevier Ltd): 568–77.

165. Kuznetsova, Tatyana, Shuang-Yin Wang, Nagesha A. Rao, Amit Mandoli, Joost H. A. Martens, and others. 2015. 'Glucocorticoid Receptor and Nuclear Factor Kappa-B Affect Three-Dimensional Chromatin Organization', *Genome Biology*, 16.1 (Genome Biology): 264.

166. Ogryzko, Vasily V., Tomohiro Kotani, Xiaolong Zhang, R.Louis Schiltz, Tazuko Howard, and others. 1998. 'Histone-like TAFs within the PCAF Histone Acetylase Complex', *Cell*, 94.1: 35–44.

167. Guenther, Matthew G., Orr Barak, and Mitchell A. Lazar. 2001. 'The SMRT and N-CoR Corepressors Are Activating Cofactors for Histone Deacetylase 3.', *Molecular and Cellular Biology*, 21.18: 6091–6101.

168. Chen, J D, and R M Evans. 1995. 'A Transcriptional Co-Repressor That Interacts with Nuclear Hormone Receptors.', *Nature*, pp. 454–57.

Supplemental References

1. Landt, Stephen G, Georgi K Marinov, Anshul Kundaje, Pouya Kheradpour, Florencia Pauli, and others. 2012. 'ChIP-Seq Guidelines and Practices of the ENCODE and modENCODE Consortia.', *Genome Research*, 22.9: 1813–31.

2. Telorac, Jonas, Sergey V. Prykhozhij, Stefanie Schöne, David Meierhofer, Sascha Sauer, and others. 2016. 'Identification and Characterization of DNA Sequences That Prevent Glucocorticoid Receptor Binding to Nearby Response Elements', *Nucleic Acids Research*, 44.13: 6142–56.

3. Uhlenhaut, N. Henriette, Grant D. Barish, Ruth T. Yu, Michael Downes, Malith

Karunasiri, and others. 2013. ‘Insights into Negative Regulation by the Glucocorticoid Receptor from Genome-Wide Profiling of Inflammatory Cistromes’, *Molecular Cell*, 49.1: 158–71.

4. Starick, Stephan R, Jonas Ibn-Salem, Marcel Jurk, Céline Hernandez, Michael I Love, and others. 2015. ‘ChIP-Exo Signal Associated with DNA-Binding Motifs Provides Insight into the Genomic Binding of the Glucocorticoid Receptor and Cooperating Transcription Factors’, *Genome Research*, 25.6: 825–35.

5. Reddy, Timothy E, Florencia Pauli, Rebekka O Sprouse, Norma F Neff, Kimberly M Newberry, and others. 2009. ‘Genomic Determination of the Glucocorticoid Response Reveals Unexpected Mechanisms of Gene Regulation’, *Genome Research*, 19.12: 2163–71.

6. John, Sam, Peter J Sabo, Robert E Thurman, Myong-Hee Sung, Simon C Biddie, and others. 2011. ‘Chromatin Accessibility Pre-Determines Glucocorticoid Receptor Binding Patterns’, *Nature Genetics*, 43.3: 264–68.

7. Rhee, Ho Sung, and B. Franklin Pugh. 2011. ‘Comprehensive Genome-Wide Protein-DNA Interactions Detected at Single-Nucleotide Resolution.’, *Cell*, 147.6 (Elsevier Inc.): 1408–19.

8. Lim, Hee-woong, N Henriette Uhlenhaut, Alexander Rauch, Juliane Weiner, Sabine Hübner, and others. 2015. ‘Genomic Redistribution of GR Monomers and Dimers Mediates Transcriptional Response to Exogenous Glucocorticoid in Vivo.’, *Genome Research*, 25.6: 836–44.

9. Gross, D S, and W T Garrard. 1988. 'Nuclease Hypersensitive Sites in Chromatin.', *Annual Review of Biochemistry*, 57: 159–97.

10. Crawford, Gregory E. 2005. 'Genome-Wide Mapping of DNase Hypersensitive Sites Using Massively Parallel Signature Sequencing (MPSS)', *Genome Research*, 16.1: 123–31.

11. Burd, Craig J, James M Ward, Valerie J Crusselle-Davis, Grace E Kissling, Dhiral Phadke, and others. 2012. 'Analysis of Chromatin Dynamics during Glucocorticoid Receptor Activation.', *Molecular and Cellular Biology*, 32.10: 1805–17.

12. John, Sam, Peter J. Sabo, Thomas A. Johnson, Myong-Hee Sung, Simon C. Biddie, and others. 2008. 'Interaction of the Glucocorticoid Receptor with the Chromatin Landscape.', *Molecular Cell*, 29.5: 611–24.

13. Koohy, Hashem, Thomas A. Down, and Tim J. Hubbard. 2013. 'Chromatin Accessibility Data Sets Show Bias due to Sequence Specificity of the DNase I Enzyme.', *PloS One*, 8.7: e69853.

14. Bain, David L., Qin Yang, Keith D. Connaghan, James P. Robblee, Michael T. Miura, and others. 2012. 'Glucocorticoid Receptor-DNA Interactions: Binding Energetics Are the Primary Determinant of Sequence-Specific Transcriptional Activity', *Journal of Molecular Biology*, 422.1 (Elsevier Ltd): 18–32.

15. Robblee, James P., Michael T. Miura, and David L. Bain. 2012. 'Glucocorticoid Receptor–Promoter Interactions: Energetic Dissection Suggests a Framework for the Specificity of Steroid Receptor-Mediated Gene Regulation', *Biochemistry*, 51.22:

4463–72.

16. De Angelis, Rolando W., Nasib K. Maluf, Qin Yang, James R. Lambert, and David L. Bain. 2015. 'Glucocorticoid Receptor-DNA Dissociation Kinetics Measured in Vitro Reveal Exchange on the Second Time Scale.', *Biochemistry*, 54.34: 5306–14.

17. Giresi, Paul G, Jonghwan Kim, Ryan M McDaniell, Vishwanath R Iyer, and Jason D Lieb. 2007. 'FAIRE (Formaldehyde-Assisted Isolation of Regulatory Elements) Isolates Active Regulatory Elements from Human Chromatin.', *Genome Research*, 17.6: 877–85.

18. Burd, Craig J., and Trevor K. Archer. 2013. 'Chromatin Architecture Defines the Glucocorticoid Response.', *Molecular and Cellular Endocrinology*, 380.1–2 (Elsevier Ireland Ltd): 25–31.

19. Shi, Yigong. 2014. 'A Glimpse of Structural Biology through X-Ray Crystallography', *Cell*, 159.5 (Elsevier Inc.): 995–1014.

20. Luisi, B. F., W. X. Xu, Z. Otwinowski, L. P. Freedman, K. R. Yamamoto, and others. 1991. 'Crystallographic Analysis of the Interaction of the Glucocorticoid Receptor with DNA.', *Nature*, pp. 497–505.

21. Meijsing, Sebastiaan H, Miles A Pufall, Alex Y So, Darren L Bates, Lin Chen, and others. 2009. 'DNA Binding Site Sequence Directs Glucocorticoid Receptor Structure and Activity', *Science*, 324.5925: 407–10.

22. Hudson, William H, Christine Youn, and Eric A Ortlund. 2013. 'The Structural Basis of Direct Glucocorticoid-Mediated Transrepression', *Nature Structural & Molecular*

Biology, 20.1: 53–58.

23. Bledsoe, Randy K., Valerie G. Montana, Thomas B. Stanley, Chris J. Delves, Christopher J. Apolito, and others. 2002. 'Crystal Structure of the Glucocorticoid Receptor Ligand Binding Domain Reveals a Novel Mode of Receptor Dimerization and Coactivator Recognition', *Cell*, 110: 93–105.

24. He, Yuanzheng, Wei Yi, Kelly Suino-Powell, X Edward Zhou, W David Tolbert, and others. 2014. 'Structures and Mechanism for the Design of Highly Potent Glucocorticoids.', *Cell Research*, 24.6 (Nature Publishing Group): 713–26.

25. Schoch, Guillaume A., Brigitte D'Arcy, Martine Stihle, Dominique Burger, Dominik Bär, and others. 2010. 'Molecular Switch in the Glucocorticoid Receptor: Active and Passive Antagonist Conformations', *Journal of Molecular Biology*, 395.3 (Elsevier Ltd): 568–77.

26. Campagne, S, V Gervais, and A Milon. 2011. 'Nuclear Magnetic Resonance Analysis of Protein-DNA Interactions', *Journal of The Royal Society Interface*, 8.61: 1065–78.

27. Watson, Lisa C, Kristopher M Kuchenbecker, Benjamin J Schiller, John D Gross, Miles A Pufall, and others. 2013. 'The Glucocorticoid Receptor Dimer Interface Allosterically Transmits Sequence-Specific DNA Signals', *Nature Structural & Molecular Biology*, 20.7: 876–83.

28. Thomas-Chollier, Morgane, Lisa C Watson, Samantha B Cooper, Miles A Pufall, Jennifer S Liu, and others. 2013. 'A Naturally Occuring Insertion of a Single Amino Acid Rewires Transcriptional Regulation by Glucocorticoid Receptor Isoforms',

Proceedings of the National Academy of Sciences, 110.44: 17826–31.

29. Schöne, Stefanie, Marcel Jurk, Mahdi Bagherpoor Helabad, Iris Dror, Isabelle Lebars, and others. 2016. 'Sequences Flanking the Core-Binding Site Modulate Glucocorticoid Receptor Structure and Activity', *Nature Communications*, 7: 12621.

30. Hudson, William H, Bradley R Kossmann, Ian Mitchelle S. de Vera, Shih-wei Chuo, Emily R Weikum, and others. 2016. 'Distal Substitutions Drive Divergent DNA Specificity among Paralogous Transcription Factors through Subdivision of Conformational Space', *Proceedings of the National Academy of Sciences of the United States of America*, 113.2: 326–31.

31. Karplus, Martin, and J Andrew McCammon. 2002. 'Molecular Dynamics Simulations of Biomolecules.', *Nature Structural Biology*, 9.9: 646–52.

32. Bredenberg, Johan, and Lennart Nilsson. 2002. 'Conformational States of the Glucocorticoid Receptor DNA-Binding Domain from Molecular Dynamics Simulations', *Proteins: Structure, Function, and Genetics*, 49.1: 24–36.

33. Eriksson, M. A., T. Härd, and L. Nilsson. 1995. 'Molecular Dynamics Simulations of the Glucocorticoid Receptor DNA-Binding Domain in Complex with DNA and Free in Solution.', *Biophysical Journal*, 68.2: 402–26.

34. Stockner, Thomas, Heinz Sterk, Robert Kaptein, and Alexandre M J J Bonvin. 2003. 'Molecular Dynamics Studies of a Molecular Switch in the Glucocorticoid Receptor.', *Journal of Molecular Biology*, 328.2: 325–34.

35. Konermann, Lars, Jingxi Pan, and Yu-Hong Liu. 2011. 'Hydrogen Exchange Mass

Spectrometry for Studying Protein Structure and Dynamics.', *Chemical Society Reviews*, 40.3: 1224–34.

36. Kirschke, Elaine, Devrishi Goswami, Daniel Southworth, Patrick R. Griffin, and David A. Agard. 2014. 'Glucocorticoid Receptor Function Regulated by Coordinated Action of the Hsp90 and Hsp70 Chaperone Cycles', *Cell*, 157.7 (Elsevier Inc.): 1685–97.

37. Khan, Shagufta H., Smita Awasthi, Chunhua Guo, Devrishi Goswami, Jun Ling, and others. 2012. 'Binding of the N-Terminal Region of Coactivator TIF2 to the Intrinsically Disordered AF1 Domain of the Glucocorticoid Receptor Is Accompanied by Conformational Reorganizations', *Journal of Biological Chemistry*, 287.53: 44546–60.

38. Zhang, Hairong, Qian Wu, and Mikhail Y Berezin. 2015. 'Fluorescence Anisotropy (Polarization): From Drug Screening to Precision Medicine.', *Expert Opinion on Drug Discovery*, 10.11 (Informa UK, Ltd.): 1145–61.

39. Shah, Nilesh, and Thomas S Scanlan. 2004. 'Design and Evaluation of Novel Nonsteroidal Dissociating Glucocorticoid Receptor Ligands', *Bioorganic & Medicinal Chemistry Letters*, 14.20: 5199–5203.

40. Surjit, Milan, Krishna Priya Ganti, Atish Mukherji, Tao Ye, Guoqiang Hua, and others. 2011. 'Widespread Negative Response Elements Mediate Direct Repression by Agonist- Liganded Glucocorticoid Receptor', *Cell*, 145.2 (Elsevier Inc.): 224–41.

41. Hua, Guoqiang, Laetitia Paulen, and Pierre Chambon. 2016. 'GR SUMOylation and Formation of an SUMO-SMRT/NCoR1-HDAC3 Repressing Complex Is Mandatory

for GC-Induced IR nGRE-Mediated Transrepression', *Proceedings of the National Academy of Sciences*, 113.5: E626–34.

42. Ran, F Ann, Patrick D Hsu, Jason Wright, Vineeta Agarwala, David A Scott, and others. 2013. 'Genome Engineering Using the CRISPR-Cas9 System', *Nature Protocols*, 8.11: 2281–2308.

43. Unruh, Jay R., and Enrico Gratton. 2008. 'Analysis of Molecular Concentration and Brightness from Fluorescence Fluctuation Data with an Electron Multiplied CCD Camera.', *Biophysical Journal*, 95.11 (Elsevier): 5385–98.

44. McNally, James G. 2000. 'The Glucocorticoid Receptor: Rapid Exchange with Regulatory Sites in Living Cells', *Science*, 287.5456: 1262–65.

45. Presman, Diego M, Sourav Ganguly, R Louis Schiltz, Thomas A Johnson, Tatiana S Karpova, and others. 2016. 'DNA Binding Triggers Tetramerization of the Glucocorticoid Receptor in Live Cells', *Proceedings of the National Academy of Sciences*, 113.29: 8236–41.

46. Presman, Diego M., M. Florencia Ogara, Martín Stortz, Lautaro D. Alvarez, John R. Pooley, and others. 2014. 'Live Cell Imaging Unveils Multiple Domain Requirements for In Vivo Dimerization of the Glucocorticoid Receptor', *PLoS Biology*, 12.3, ed. by Keith R. Yamamoto: e1001813.

47. Stavreva, Diana A., W. G. Muller, Gordon L Hager, Carolyn L Smith, and James G McNally. 2004. 'Rapid Glucocorticoid Receptor Exchange at a Promoter Is Coupled to Transcription and Regulated by Chaperones and Proteasomes', *Molecular and*

Cellular Biology, 24.7: 2682–97.

48. Kino, Tomoshige, Szu-Heng Liou, Evangelia Charmandari, and George P Chrousos. 2006. 'Glucocorticoid Receptor Mutants Demonstrate Increased Motility inside the Nucleus of Living Cells: Time of Fluorescence Recovery after Photobleaching (FRAP) Is an Integrated Measure of Receptor Function.', *Molecular Medicine (Cambridge, Mass.)*, 10.7–12: 80–88.

CHAPTER 3: THE GLUCOCORTICOID RECEPTOR BINDS DIRECTLY TO A GR HALF-SITE SEQUENCE EMBEDDED WITHIN AP-1 RECOGNITION MOTIFS

Emily R. Weikum [1], Ian Mitchelle S. de Vera[2], Jerome C. Nwachukwu [3], Kendall W. Nettles[3], Douglas J. Kojetin[2], Eric A. Ortlund[12]

1) Department of Biochemistry, Emory University School of Medicine, Atlanta, GA 30322 USA

2)Department of Integrative Structural and Computational Biology, The Scripps Research Institute, Jupiter, FL 33458 USA

3)Department of Molecular Medicine, The Scripps Research Institute, Jupiter, FL 33458 USA

This manuscript describes a novel mechanism for the glucocorticoid receptor in the control of transcription. We use structural biology and biochemistry to show, contrary to current models, that the glucocorticoid receptor binds directly to an AP-1 DNA response element. This mechanism represents a paradigm shift in our current understanding of glucocorticoid receptor biology. This work is under review for publication.

---

[a] Conceived and designed the experiments: ERW IMSdV JCN KWN DJK EAO. Performed the experiments: ERW IMSdV JCN KWN DJK EAO. Analyzed the data: ERW IMSdV JCN KWN DJK EAO. Wrote the paper: ERW IMSdV JCN KWN DJK EAO.

Abstract

The glucocorticoid receptor (GR) is a ligand-regulated transcription factor that controls the expression of extensive gene networks, driving both up- and down-regulation. GR utilizes multiple DNA-binding-dependent and -independent mechanisms to achieve context-specific transcriptional outcomes. The DNA-binding-independent mechanism involves tethering of GR to the pro-inflammatory transcription factor AP-1 through protein-protein interactions. This mechanism has served as the predominant model of GR-mediated transrepression of inflammatory genes. However, ChIP-seq data have consistently shown GR to occupy AP-1 response elements (TREs), even in the absence of AP-1. Therefore, the current model is insufficient to explain GR action at these sites. Here, we show that GR regulates a subset of inflammatory genes in a DNA-binding-dependent manner. Using structural biology and biochemical approaches, we show that GR binds directly to TREs via sequence-specific contacts to a GR-binding sequence (GBS) half-site found embedded within the TRE motif. Furthermore, we show that GR-mediated transrepression observed at TRE sites to be DNA-binding-dependent. This work expands our understanding of this multifaceted transcription factor, showing that GR uses multiple mechanisms to control gene expression and suppress inflammatory gene expression in the absence of AP-1.

Introduction

Glucocorticoid receptor (GR) is a ligand-regulated transcription factor in the nuclear receptor superfamily which activates and represses the expression of thousands of genes[1]. GR contains a modular domain architecture common to the superfamily: an unstructured N-terminal domain (NTD) which contains the activation function-1 (AF-1) region that interacts with coregulator proteins; a zinc finger DNA binding domain (DBD); a hinge region; and a ligand-binding domain (LBD) which contains the ligand-sensitive AF-2 surface that also enables interaction with coregulators[2,3]. In humans, GR activity is regulated by the steroid hormone cortisol, or by exogenous glucocorticoids (GCs) such as dexamethasone, or DEX[1]. Most unliganded GR resides in the cytoplasm bound to chaperone proteins. Upon binding ligand, GR undergoes a conformational change exposing nuclear localization signals and subsequently translocates to the nucleus (4,5). In the nucleus, GR interacts with genomic response elements via multiple DNA-binding-dependent and -independent mechanisms[6-8].

GR binds directly to DNA at canonical GR binding sequence (GBS) composed of two pseudo-palindromic hexameric AGAACA repeats separated by a three-base pair spacer[9,10] (**Figure 3.1a*i***). GR binding to GBS sequences occurs in a head-to-head fashion through a protein-protein interaction between two GR DBD proteins[11]. GR also binds to a newly characterized inverted repeat-GBS (IR-GBS/nGREs), which shows monomeric GR binding to a CTCC$(N)_{0\text{-}2}$GGAGA motif[12,13] (**Figure 3.1a*ii***). Unlike binding to a canonical GBS, GR binding to IR-GBS sequences occurs in a tail-to-tail fashion where GR DBD proteins do not interact[13,14]. Lastly, GR binds degenerate GBSs that are found in conjunction with other transcription factor (TF) binding sites or composite elements[15,16] (**Figure 3.1a*iii*)**. However,

in contrast to these DNA-binding-dependent mechanisms, GR also represses transcription without direct interaction with DNA[17,18].

GR-mediated transrepression involves binding of GR to other TFs, such as activator protein-1 (AP-1) and nuclear factor-kappa beta (NF-κB), through protein-protein interactions[18,19] (**Figure 3.1a*iv*)**. This mechanism, referred to as tethering, has been the accepted model for GR-mediated transrepression of inflammatory genes[20,21]. GR is a potent repressor of AP-1 activity, driving its popularity as a sought-after drug target for anti-inflammatory therapies[22,23]. GR-targeted therapeutics, collectively known as glucocorticoids (GCs), are the predominant treatment for chronic inflammatory diseases such as arthritis and asthma[24]. However, continued administration of GCs results in numerous side effects, including diabetes, muscle wasting, and Cushing's syndrome, which diminish the effectiveness of treatment[25,26]. These side effects have driven extensive efforts to develop dissociative GCs that separate the beneficial anti-inflammatory properties from side effects[23]. Unfortunately, efforts to make these selective modulators have been unsuccessful, encouraging a reexamination into the mechanisms of GR-mediated transrepression of inflammation.

With advances in genome-wide sequencing, studies using chromatin immunoprecipitation-sequencing (ChIP-seq) have consistently found GR to occupy AP-1 response elements (TREs) in the absence of AP-1 as a tethering factor, suggesting the current tethering model is insufficient to explain GR occupancy of TRE motifs[27-31]. Additionally, the discovery of DNA-binding-dependent repression at IR-GBSs revealed GR can interact directly with genomic DNA to repress gene transcription[12,13]. Therefore, we asked whether DNA binding was required for transrepression at TRE-containing inflammatory genes. Using

a combination of structural and biochemical techniques, we show GR binds to a GBS half-site located within TRE motifs in a sequence-specific fashion. We found that direct interaction of GR with the GBS DNA half-site within TREs is vital for transcriptional repression and monomeric GR is preferred at these sites, a mechanism similar to repressive GR-IR:GBS interactions[13,14]. Taken together, our findings suggest that in addition to tethering, GR is able to regulate a subset of AP-1 driven inflammatory genes through a DNA-binding-dependent mechanism.

Results

GR is recruited to AP-1 target genes in a DNA-binding-dependent manner

ChIP-seq data have found GR to be highly enriched at TRE motifs[27-31]. AP-1 subunits are largely cytoplasmic until treatment with an activator, such as lipopolysaccharide (LPS) or tumor necrosis factor alpha (TNF-α); therefore, AP-1 would not be present in the nucleus or at these GR occupied sites in the absence of an AP-1 activator. However, ChIP-seq for GR in DEX-treated mouse primary macrophages without an AP-1 activator showed the TRE motif was the second most enriched site compared to a canonical GBS[27]. Furthermore, even with AP-1 activation the ChIP-seq profiles revealed that 50.3% of the GR occupancy is without a tethering factor[27].

Based on these findings that indicate GR can occupy TREs in the absence of a tethering factor, we hypothesized that GR may bind to some TRE sequences directly. To test GR recruitment to TREs, we performed ChIP using a tetracycline-inducible system in HEK293T cells, which contain very low levels of endogenous GR. The exogenous receptors were detected using their N-terminal hemagglutinin (HA) epitope tag (**Figure 3.1b**). We tested GR occupancy on two TRE-containing promoters; interleukin-11 (*IL11*) and vascular cell adhesion molecule-1 (*VCAM1*). In the absence of an AP-1 activator, HA-tagged WT GR occupies the TRE motif from both promoters in the presence of a GR agonist, DEX (**Figure 3.1c**). GR occupancy is maintained with cellular pretreatment of an AP-1 activator, TNF-α. As both promoters contain a TRE and a NF-κB response element (κBRE), it is likely that DNA fragments generated for qPCR could encompass both response element motifs (**Figure 3.1d**). As a control, WT GR occupies several canonical GBS motifs that show comparable

levels of GR recruitment (**Figure 3.1e**). We have observed a DNA-binding-dependent mechanism is also possible at κBREs (*Ortlund Lab, unpublished*). These data, when considered with genome-wide ChIP analysis[27,31], suggests that tethering alone is inadequate to explain GR recruitment to these sites.

GR Ser425Gly mutation cannot be used to distinguish DNA-binding-dependent from -independent mechanisms

Compared to other steroid receptors, GR is unique in its ability to counter AP-1 activity[32,33]. Despite sharing 90% sequence identity in the DBD and complete conservation of DNA-contacting residues, the mineralocorticoid (MR), which is able to transactivate from a canonical GBS, is unable to repress a TRE-containing reporter[32-35] (**Figure 3.1f**). Swapping the GR DBD into the full-length MR protein partially restores transrepression, suggesting that the GR DBD is necessary to repress inflammatory genes[32].

To determine the specific residues within GR that allow for this transrepressive function, the GR and MR DBDs were aligned, which revealed seven amino acid differences—including Ser425 in GR, which is a Gly in MR. When this GR residue is mutated to the equivalent site in MR (GR Ser425Gly), the mutation renders GR incapable of repressing select inflammatory genes but maintains the ability to bind to a canonical GBS[32]. This mutation did not affect binding and transactivation from canonical sites, and it was therefore concluded that GR did not need to directly bind DNA in order to repress, solidifying support for a tethering hypothesis. However, this conclusion was made before DNA-binding-dependent repression at IR-GBSs was identified[12-14]. At IR-GBS sites, the GR Ser425Gly mutation reduced the affinity for IR-GBS DNA sequences and was deficient in

transrepression of the *TSLP* IR-GBS in cells[13,14]. Based on these findings, the GR Ser425Gly mutation cannot be used to discriminate DNA-binding-dependent from independent mechanisms.

Despite conflicting reports on the effect this mutation has on GR-mediated transrepression at TREs, we tested whether GR Ser425Gly had an effect on GR recruitment to TREs in cells[36]. Compared to WT GR, the Ser425Gly mutant showed a reduced occupancy on TREs (**Figure 3.1c**). This reduced occupancy is in line with the effect observed at IR-GBSs[13,14], suggesting mechanisms of transrepression might be similar. Conversely, both WT GR and the Ser425Gly mutant are similarly recruited to canonical GBS sites in the promoters of*, GILZ,* and *SGK1,* but reduced occupancy at *FKBP5* (**Figure 3.1d**).

GR represses AP-1 target genes in the absence of tethering factors

Since GR can occupy genomic TREs without tethering factors, we hypothesized that GR might also repress transcription without tethering factors. Canonical AP-1 target genes are expressed at low levels in the absence of pro-inflammatory signaling, making the study of DNA-dependent GR-mediated transrepression difficult[19,37,38]. Stimulation of AP-1 and its target genes would confound our aims of delineating DNA-binding-dependent from -independent transrepression. To circumvent this issue, we constructed reporter plasmids that measure transcriptional repression without prior activation of AP-1. The reporters contain a strong SV40 enhancer and promoter that constitutively expresses luciferase; and in between the enhancer and promoter regions we inserted ~150 bases of the promoters from *IL11*, *VCAM1,* and interleukin-6 (*IL6*), all of which contain TREs and are known to be upregulated by AP-1[39-42]. With co-transfection of full-length GR or mutants, a loss of luciferase signal is

a readout for gene repression[12]. When these reporter plasmids were transfected with WT GR into U-2 OS cells, which do not express endogenous GR, only in the presence of GR and upon DEX treatment is GR able to repress transcription of the reporter, indicating that GR alone is sufficient to repress these TRE-containing inflammatory genes (**Figure 3.2a**).

To test the requirement of DNA-binding for the observed repression, we generated a new mutant GR Lys442Ala/Arg447Ala, termed "DNA-Dead" GR, which prevents binding of GR to multiple DNA response elements due to a loss of sequence-specific contacts[9,13] (**Figure 3.2b**). ChIP experiments revealed that GR Lys442Ala/Arg447Ala is not recruited to TREs or to canonical GBS, and this mutant is also unable to repress transcription (**Figure 3.2b,c**), indicating the importance of DNA binding.

GR binds directly to TREs

Since ChIP and transrepression analysis suggested DNA-dependent interactions were critical for GR action at inflammatory genes, we hypothesized that GR could bind directly to TREs. To test this, we used fluorescence polarization assays to monitor full-length GR binding to *IL11* TRE and *SGK* GBS DNA sequences[45]. Surprisingly, GR bound to the *IL11* TRE with similar affinity (42 nM) to the canonical *SGK* GBS (34 nM) **(Figure 3.3a)**. Additionally, we found that the GR DNA binding domain (DBD) bound to TREs from the *IL6, IL11, VCAM1, CSF1, IL1α,* and *MMP13* promoters (**Figure 3.3b,c**) with similar affinities as IR-GBS (13). Full-length GR has a higher affinity for TREs, suggesting that domains other than the DBD contribute to DNA binding, potentially by increased non-sequence specific interactions, pre-ordering the DBD, or by making additional contacts with DNA[32,43,44].

The crystal structure of the GR DBD:IR-GBS complex revealed one monomer of GR bound to opposite sides of the DNA in an everted fashion[13,14]. NMR analysis of the GR DBD:IR-GBS complex confirmed that residues critical for dimerization on GBS are not perturbed, suggesting monomeric GR is likely sufficient at these elements[14]. To test whether the GR DBD adopts a canonical head-to-head dimeric or IR-GBS-like monomeric conformation on TREs, we performed 2D [^{1}H,^{15}N]-HSQC NMR on the GR DBD:*IL11* TRE complex. Binding of *IL11* TRE to ^{15}N-labeled GR DBD causes large NMR chemical shift perturbations for residues within the DNA reading helix, such as Cys441 and Val488 (**Figure 3.4a**). However, in contrast to the dimeric GR-GBS complex, residues within the GR DBD dimerization loop, such as Ala458 and Gly459, showed no change upon binding *IL11* TRE (**Figure 3.4b-d**) (9,11), indicating GR binds as a monomer to the *IL11* TRE.

To determine whether dimerization is required for transrepression, we performed transcriptional reporter assays using two well-characterized full-length GR dimerization mutants, GRdim (Ala458Thr) and GRmon (Ala458Thr/Iso634Ala)[20,46]. Both mutants repressed the constitutively active SV40 TRE luciferase reporters in the presence of GR agonist, suggesting monomeric GR is preferred at these sites (**Figure 3.4e**). It is not surprising that GRdim maintained transrepressive function as this mutant represses inflammation *in vivo*[20]. However, because GRdim can form dimers on canonical GBSs[11], we also tested GRmon, which was shown to be predominantly monomeric in cells[46]. GRmon is better than GRdim at repressing inflammatory gene expression (**Figure 3.4e**), suggesting that monomeric GR is responsible for GR-mediated repression of this subset of genes.

GR recognizes TREs in a sequence specific manner

To determine how GR recognizes TRE sequences, we determined crystal structures of GR DBD bound to the *IL11* and *VCAM1* TREs (**Figure 3.5**). The GR DBD: *IL11* TRE complex crystallized in the $P2_12_12_1$ space group and data were collected to 2.15Å **(Figure 3.5a, Table 3.1)**. The GR DBD: *VCAM1* TRE complex crystallized in the $P2_12_12_1$ space group and data were collected to 2.29 Å (**Figure 3.5b, Table 3.1**). Both structures show two GR monomers bound to opposite sides of the TRE DNA sequences in an everted fashion, similar to the GR DBD:IR-GBS structure[13]. However, in the TRE structures, one of the GR monomers straddles the end-stacking junction where the DNA makes a pseudo-continuous helix; this GR monomer does not make base-specific interactions and only contacts the DNA backbone. Based on our cellular transrepression and NMR footprinting analysis above, it is likely this GR monomer is only important for efficient crystal packing and may not be biologically relevant *in vivo*.

In both structures, GR recognizes a hexameric TGA(G/C)TC sequence; though the third base differs, our structures show that GR does not directly contact this base. Analysis of the GR-DNA structural interfaces using PISA[47] revealed a favorable free energy gained with the GR DBD (Monomer 1)-DNA interaction. The free energy gain upon complex formation is -9.7 kcal/mol for *IL11* TRE and -8.9 kcal/mol for *VCAM1* TRE, values similar to GR DBD:IR-GBS complex formation (13). Monomer 1 of the GR DBD:*IL11* TRE complex positions the DNA reading helix in the major groove (**Figure 3.5a**). Three side chains, Arg447, Lys442, and Val443, participate in base-specific interactions with the DNA and are consistently used by the GR DBD to recognize DNA sequences (**Figure 3.5c**) (9,13). Arg447 makes hydrogen bonds to the N7 position and the terminal oxygen on guanine 107, and van

der Waals contacts to the methyl group on thymine 106. Lys442 makes hydrogen bonds to the N7 position on adenine 91, and Val443 makes van der Waals contacts to guanine 107 (**Figure 3.5b)**. Monomer 1 of the GR DBD:*VCAM1* TRE complex only uses two side chains to make base-specific contacts with the DNA (**Figure 3.5d**). Arg447 makes hydrogen bonds to the N7 position on guanine 523, and Val443 makes van der Waals contacts to the same base. Arg447 also makes van der Waals contacts to the methyl group on thymine 522. In both structures, additional side chains participate in backbone interactions, marking the boundaries of the GR binding footprint.

In all structures, GR recognizes similar DNA bases with consistent spacing. In the *IL11* structure, Arg447 hydrogen bonds to a guanine but also makes side-on hydrophobic contacts with the methyl group on a neighboring thymine base. Additionally, Lys442 makes hydrogen bonds to a pyrimidine base. We do not observe a direct interaction between GR and this base in the *VCAM1* structure; instead, Lys442 interacts with the DNA backbone at the same position. In the GBS structure, Val443 has an additional base contact though van der Waals forces with a thymine; TREs have a pyrimidine base in this position, and therefore the Val residue is shifted to make hydrophobic contacts with the guanine base instead. Our structures show that GR recognizes a GBS half-site sequence embedded within the TRE.

GR and AP-1 likely compete for the same binding site

Structural alignment of the GR DBD:*IL11* TRE structure with the AP-1:TRE structure reveals that GR and AP-1 would likely compete for the same DNA binding site (**Figure 3.6a**). To test if GR competes with AP-1 for binding to the TRE, we performed a time resolved-fluorescence resonance energy transfer (TR-FRET) competition assay by

monitoring the concentration-dependent effect of GR DBD on the interaction between AP-1 and *IL11* TRE DNA. Titration of AP-1 resulted in an increase in TR-FRET signal, showing binding of AP-1 to *IL11* TRE. Notably, when increasing amounts of GR DBD is added, TR-FRET signal is reduced indicating that GR competes with AP-1 binding to the *IL11* TRE (**Figure 3.6b**).

Discussion

GR is a ligand-regulated transcription factor that controls distinct gene networks across numerous biological processes including development, metabolism, and inflammation[1]. Current models suggest that GR uses distinct mechanisms to recognize promoters with different DNA sequences (**Figure 3.1a**). GR can cooperatively homodimerize at GBSs or bind as a monomer to IR-GBSs and composite elements[9,13]. Conversely, GR can tether onto other DNA-bound TFs, such as AP-1, but not make direct contact with DNA[19]. However, new genome-wide methodologies have revealed a new complexity of GR-genome interactions and suggest there may be alternative mechanisms beyond the current view of GR signaling[8,27,48]. Here, we propose a new mechanism by which GR binds directly to a GBS-like half-site located within a canonical AP-1 TRE, such as sequences found in the promoter regions of *IL11* and *VCAM1*.

As detailed above, genome-wide ChIP-seq studies find that GR occupies the TRE motif across numerous cell types[27-31] (**Figure 3.1b**). These experiments are conducted with GR agonist alone, which does not alter the subcellular localization of AP-1 subunits, and in the absence of AP-1 activation—suggesting GR occupies these sites without AP-1[27]. However, loss of AP-1 results in a significant reduction of GR DNA occupancy[29]. This could suggest that GR requires tethering with AP-1 in order to bind DNA, which could also be explained by increased chromatin accessibility gained by AP-1 activation and subsequent GR binding to TREs[30,49]. GR, like most nuclear receptors, predominately binds to accessible chromatin[50]. While GR and AP-1 have been shown to interact directly[43], other studies were unable to validate this interaction[44,51]. Additionally, structural overlays of GR and AP-1 bound to a TRE sequence suggests they likely compete for the same DNA binding site on

some elements (**Figure 3.6a**), and we confirmed that GR and AP-1 indeed compete for binding to the TRE sequence (**Figure 3.6b**), consistent with previous studies[43,44].

Further support for a GR DNA-binding-dependent repressive mechanism comes from single molecule tracking experiments that revealed only 3% of cellular GR is likely to be tethered to other TFs, indicating that DNA-binding-dependent mechanisms represent the majority of GR-chromatin interactions[27,52,53]. We therefore hypothesized that GR might be able to bind directly to TREs. We show here that full-length GR is able to bind the *IL11* TRE with an affinity nearly identical to canonical GBS (**Figure 3.3a**). Furthermore, we demonstrate that GR alone is sufficient to transrepress. Ablation of GR DNA binding results in an attenuation of GR occupancy at TREs as well as GR-mediated transrepression (**Figure 3.1b, 2b**). The latter data is in line with other reports that show the importance of the GR DNA binding domain for repression of inflammatory genes[32,36,43,44,54]. Another striking example is that GR has already been shown to bind directly to a TRE-like sequence in the rat tyrosine hydroxylase (TH) gene promoter[55]. Though the promoter contains a TRE and a canonical GBS, GR was shown to regulate transcription through the TRE sequence by directly binding to the TRE-like site (TGACTAA). This sequence is almost identical to the GR footprint identified by our structural analysis (**Figure 3.5**). Binding at this site is conserved in humans, suggesting GR DNA-binding-dependent mechanisms at TREs may represent an evolutionary conserved model[56].

Initial support for the tethering hypothesis stemmed from studies of dimerization deficient GR mutants, which generated complex and often conflicting interpretations[57]. The GRdim mutant, designed with the aim of breaking intramolecular protein-protein interactions at the GR homodimerization interface[9], was the main driver for a DNA-binding-independent

mechanism at inflammatory genes. GRdim was shown to not bind DNA and displays a reduced ability to transactivate genes, whereas transrepression was unaffected[32,58]. Furthermore, whereas complete GR knock-out mice die quickly after birth, GRdim knock-in mice lived and maintained the ability to combat inflammatory challenge[20,59]. These results drove the conclusions that dimerization of GR is required for DNA binding and that dimerization is not necessary for GR to repress inflammatory gene expression[21]. However, it was later shown that GRdim does not affect GR stoichiometry on DNA and still forms dimers *in vitro* and in cells; instead, GRdim affects cooperative binding to DNA[11,46]. These results suggest that the GRdim mutant cannot be used to rule out a DNA-binding-dependent mechanism at TREs.

The GR Ser425Gly mutant was previously used to show that DNA binding was dispensable at TREs; this mutant binds to canonical GBSs but is unable to repress inflammatory genes[32]. However, these interpretations were made before the DNA-binding-dependent GR mechanisms at IR-GBS were identified. It has since been shown that the GR Ser425Gly mutant not only has diminished binding to an IR-GBS but also affects GR transrepression from IR-GBS sites[14]. In this work, we further show that GR Ser425Gly mutant is poorly recruited to TREs (**Figure 3.1b**). Taken together, our findings suggest that GR repression of inflammatory genes likely occurs using a similar mechanism to repression of IR-GBSs through a preference for monomeric GR.

Our findings represent a shift in our understanding of GR-mediated repression of inflammation. Our data support a DNA-binding-dependent mechanism for GR repression at TREs, however a variety of mechanisms are likely involved[17,23]. What drives the selection between DNA-binding-dependent repression and tethering/transrepression remains unclear,

but this work adds yet another layer of complexity to the role of GR in regulating transcription.

Materials and Methods

Reporter Gene Assays

The U-2 OS human osteosarcoma cells were obtained from the American Type Culture Collection (ATCC) and cultured at 37 °C with 5% CO_2 in Dulbecco's modified Eagle's medium (DMEM), 10% stripped fetal bovine serum (FBS), and 5% penicillin/streptomycin without phenol red. A 150 bp region of the *IL11, IL6,* and *VCAM1* TRE-containing inflammatory gene promoters was cloned in the pGL3 plasmid in between the SV40 enhancer and promoter with a Firefly luciferase reporter gene downstream, as described previously[12,13]. The sequences cloned into the pGL3 vectors are as follows: *IL11*, -17 to -167 upstream of the transcription start site; *VCAM1*, -385 to -585 upstream of the transcription start site; and *IL6*, -938 to –1131 upstream of the transcription start site. All numbers correspond to the human GRCh38.p7 genome, accessed through *Ensembl*[61]. U-2 OS cells were transfected with 50 ng of the indicated reporter; 10 ng of wild type (WT) or mutant full-length GR, or full length MR in a pcDNA3 vector; and 1 ng of constitutively active *Renilla* luciferase under the control of pRL-TK promoter; with FuGene HD in OptiMEM (Invitrogen) according to the manufacturer's protocol. Twenty-four hours after transfection, cells were treated with media alone, 100 nM dexamethasone (DEX), or 100 nM aldosterone for MR experiments. Twenty-four hours post treatment, firefly and *Renilla* luciferase activity was detected using the Duel-Glo Luciferase Assay system (Promega) and read on a Biotek Synergy 4 plate-reader. Data were plotted as firefly luciferase activity divided by *Renilla* luciferase activity, normalized to control for each well and plotted using GraphPad Prism (v7). Data are representative of three independent biological replicates, and normalized values were compared using Tukey's multiple comparison test.

ChIP-PCR

WT and mutant GR were expressed in HEK293T cells using the Tet-On® Advanced inducible gene expression system (Takara Bio USA, Inc., Mountain View, CA). Cells in 6-well plates were co-transfected with the TransIT®-LT1 transfection reagent (Mirus Bio LLC, Madison, WI) and 1.25 μg/well of pTet-On Advanced reverse tetracycline-controlled transactivator (rtTA), and 1.25 μg/well of pTight-FRT-Hygro2-HA-GR-WT, -Ser425Gly, or - Lys442Ala/Arg447Ala expression plasmids. Control cells were transfected with empty pTight vector instead of the GR expression plasmid. After 24 h, the media were replaced with phenol red-free DMEM + 10% csFBS, with or without 1 ug/ml Doxycycline (Dox). The next day, the cells were treated with 100 nM Dex alone or in combination with 10 ng/ml TNFα for 1 h.

Quantitative ChIP assay was performed as previously described with some modification[62]. Cells were fixed in 11% formaldehyde for 15 min, quenched with 0.125 M glycine for 10 min, and rinsed with cold 1x PBS. The cells were disrupted in lysis buffer[62], incubated at 4 °C for 1 h and then sonicated. The lysates were then incubated with 100 μl Dynabeads protein G (Invitrogen, Thermo Fisher Scientific Inc., Waltham, MA) and 10 μL pre-immune rabbit IgG for 1 h at 4 °C, and centrifuged at 12,000 rpm for 15 min at 4 °C. 100 μl of the pre-cleared supernatant was mixed with an HA antibody (Y-11, Santa Cruz Biotechnology Inc., Dallas, TX), 25 μl Dynabeads protein G (Invitrogen, Thermo Fisher Scientific Inc.), and lysis buffer to make a 200 μl IP mixture that was rotated overnight at 4 °C. The precipitates were sequentially washed in low salt, high salt, and LiCl buffers, and twice in 1x TE buffer. The crosslinks were then reversed at 65 °C for 3 h. DNA fragments were isolated using QIAquick PCR purification kit (QIAGEN, Hilden, Germany), and

analyzed by qPCR using TaqMan® 2x master mix and the following custom TaqMan® real-time PCR assays (Applied Biosystems, Thermo Fisher Scientific Inc.):

IL11

Fwd 5’ - AACTTTTCCTTCCGTGCCCT - 3’

Rev 5’ - TGACACATCCTGACTCACCC - 3’

Hyb 5’ - TGAATGGAAAAGGCAGGCAG - 3’

VCAM1

Fwd 5’ - CCAGGACAGAGAGAGGAGCT - 3’

Rev 5’ - AGTTTAACAGACACCCAGCCA - 3’

Hyb 5’ - TCAGCAGTGAGAGCAACTGA - 3’

FKBP5

Fwd 5’ - AGGGTGTTCTGTGCTCTTCAA - 3’

Rev 5’ - CGAGCTGCAAAACATCACTT - 3’

Hyb 5’ - CTGCCCTAGAGCAATTTTGTT - 3’

GILZ

Fwd 5’ - CCGTTGCTGCTCTGCTATTG - 3’

Rev 5’ - TTCCCTGTCAGAGCAAGCAC - 3’

Hyb 5’ - GCTGTTGCCAGACATCCAAT - 3’

SGK1

Fwd 5’ - TGTCAGCGTCCATCCAAATG - 3’

Rev 5’ - ACAGCATGATTGATCCTCAGC - 3’

Hyb 5’ - TGGGCACAGTGAGATGACTC - 3’

Protein Expression and Purification

GR DBD (residues 417-506) was expressed and purified as described previously[13] as a 6X-Histidine tag fusion protein using the pMCSG7 vector. GR DBD was expressed in BL21 (DE3)pLysS E. *coli* and induced with 0.3 mM IPTG and grown for 4 h at 32 °C. Cells were lysed in 20 mM Tris-HCl (pH 7.4), 1 M NaCl, 25 mM imidazole, and 5% glycerol via sonication. Protein was purified using affinity chromatography (His-Trap) followed by gel filtration chromatography. Protein was then concentrated to 3-4 mg/ml in 20 mM Tris-HCl (pH 7.4), 150 mM NaCl, and 5% glycerol, flash frozen in liquid N_2, and stored at -80 °C. Full-length GR was expressed in baculovirus-infected SF9 cells treated with 1 μM triamcinolone acetonide (TA) for 24 hours. Full-length GR was a gift from Prof. David Bain and purified as described previously[63]. ^{15}N-GR DBD was expressed in E. coli BL21(DE3) pLysS cells using a standard minimal media protocol with $^{15}NH_4Cl$ as the sole nitrogen source and purified as described above. The 6X-His tag was cleaved with TEV protease overnight at 4 °C, passed through an Ni-NTA column, and the flow through containing purified ^{15}N-GR DBD was collected and verified to be >99% pure by SDS-PAGE.

Nucleic Acid Binding Assays

Synthesized FAM-labeled nucleic acid duplexes (Integrated DNA Technologies) of various TREs were annealing by heating to 90 °C followed by slow cooling to room temperature. Fluorescence polarization assays were performed by adding increasing concentrations of WT or mutant GR DBD (1 nM- 50 μM) and 10 nM of the FAM-labeled

DNA. Reactions were performed in 20 mM Tris-HCL (pH 7.4), 150 mM NaCl, and 5% glycerol. Polarization was monitored on a Biotek Synery 4 plate-reader at an excitation/emission wavelength of 485/528 nm. Three technical replicates and three biological replicates were conducted. Data plotted using GraphPad Prism (v7) are a compilation of all data collected and error bars represent standard error of measurement (SEM). Binding data was analyzed with an F-test to compare a two-site binding event to a one-site binding event with Hill slope; this test generated an F-statistic and p-value for a two-site binding model, which are represented in **Figure 3.3** along with dissociation values (K_d) and coefficient of determination (r^2).

Sequences of DNA constructs used for fluorescence polarization assays were: *IL6*: 5'-(FAM) CAAGTGCTGAGTCACTAATAA – 3', 5' – TTATTAAGACTCAGCACGTTG – 3'; *IL11* 5' –(FAM) GGGTGAGTCAGGATGTGTCAGG – 3', 5' – CCTGACACATCCTGACTCACCC – 3'; *VCAM1* 5' –(FAM) TTCCGGCTGACTCATCAAGCG – 3', 5' – CGCTTGATGAGTCAGCCGGAA – 3'; *IL1(F*5' –(FAM) GAAGAAGACTGACTCTCAGGCTTAAGC – 3', 5' – GCTTAAGCCTGAGAGTCAGTCTTCTTC – 3'; *MMP13* 5' –(FAM) ATAAGTGATGACTCACCATTGCA – 3', 5' – TGCAATGGTGAGTCATCACTTAT – 3', In all cases, (FAM) indicates the position of 6-FAM (fluorescein).

NMR analysis

NMR data were collected on a Bruker 700 MHz NMR instrument equipped with a QCI cryoprobe. For NMR studies to monitor binding to DNA, the 19-nt *IL11* TRE DNA duplex was reconstituted in 20 mM phosphate (pH 6.7), 100 mM NaCl, 1 mM TCEP, 10%

D_2O buffer to a final concentration of 437 μM; subsequently annealed by denaturing at 95 °C for 3 minutes and then equilibrated to room temperature (20-23 °C) overnight. Two-dimensional [^{1}H,^{15}N]-HSQC spectra were collected at 25 °C for free ^{15}N-GR DBD in the absence or presence of 0.44:1 or 2.3:1 of *IL11* TRE DNA duplex; or 1.5x GBS consensus DNA sequence. Data were processed using Bruker Topspin (v3.2) and analyzed using NMRViewJ (OneMoon Scientific, Inc.). Chemical shift perturbations were calculated using previously published GR DBD NMR chemical shifts[35] using the minimum chemical shift perturbation procedure[63].

Structure Determination of GR DBD-TRE Complexes

Crystals of the GR DBD:*IL11* complex were grown by hanging drop vapor diffusion in 0.2 M lithium nitratie, 15 % PEG 3350, 1 % glycerol with a 2:1 protein:DNA molar ratio. Crystals were cryo-protected with 30 % PEG 3350 and 15 % glycerol and flash cooled in liquid N2. Crystals of the GR DBD:*VCAM1* complex were grown by hanging drop vapor diffusion in 0.05 M Na cacodylate (pH 7), 5 mM $MgCl_2$, 1 mM spermine, and 2 % t-butanol with a 2:1 protein:DNA molar ratio. Crystals were cryo-protected with 50 % glycerol and flash cooled in liquid N_2. Data were collected at 1.00 Å wavelength at the 22-ID beamline (Advanced Photon Source, Argonne, IL) and processed using the HKL-2000 software[64]. The structures were phased using a previously solved structure of GR DBD:IR-GBS complex (PDB 4hn5) in Phenix[13,65]. Structure refinement and validation was performed using PHENIX refine software and model building was performed in COOT[65,66]. PDB Redo was used iteratively to optimize refinement parameters and geometry[67]. PyMOL v1.8.2 was used to visualize structures and generate figures (Schrödinger, LLC).

Generation of GR Lys442Ala/Arg447Ala "DNA Dead" and GR A458T/I634A (GRmon) Mutants

The QuikChange Site-directed Mutagenesis kit (Stratagene) was used to generate the GR Lys442Ala/Arg447Ala and GR Ala458Thr/Iso634Ala mutants. The "DNA Dead" mutation was made in both the GR DBD only for bacterial expression and in the full-length GR PCDNA3 vector for reporter gene assays. The GRmon mutant was also made in the full-length GR PCDNA3 vector for reporter gene assays.

TR-FRET Competition Assays

Lyophilized Lumi4-Tb Cryptate anti-HIS6 antibody (Cisbio) was reconstituted in 250 μL of distilled water to create a working stock per the manufacturer instructions[68] and subsequently diluted 1:100 in reaction buffer containing 20 mM Tris-HCL (pH 7.4), 150 mM NaCl, and 5% glycerol. The Lanthascreen TR-FRET-based assay was performed by adding increasing concentrations of HIS-tagged AP-1 (1.5 nM- 50 nM), 10 nM of the FAM-labeled *IL11* TRE DNA, and 5 μL of diluted antibody. Competition experiments were conducted by adding increasing concentrations of GR DBD (50 nM – 1 μM) lacking the 6xHIS tag. TR-FRET was measured on a Biotek Neo (Winooski, VT) plate-reader at an excitation and emission wavelength of 340 and 520 nm, respectively. Two technical replicates and three biological replicates were conducted. Plots generated using GraphPad Prism (v7) are a compilation of all data collected; data are shown as a ratio of acceptor to donor values; and error bars represent standard error of the mean (SEM).

Acknowledgements

We thank David Bain, PhD for providing full-length GR used for binding studies. We thank Sam Hong, PhD for providing AP-1 used for competition assays. We thank Oskar Laur of the Emory Cloning Core for assistance with generating reporter constructs. X-ray data were collected at Southeast Regional Collaborative Access Team (SER-CAT) 22-ID beamline at the Advanced Photon Source, Argonne National Laboratory. Supporting institutions may be found at www.ser-cat.org/members/html. Use of the Advanced Photon source was supported by the U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences, under Contract No. W-31-109-Eng-38.

Figures

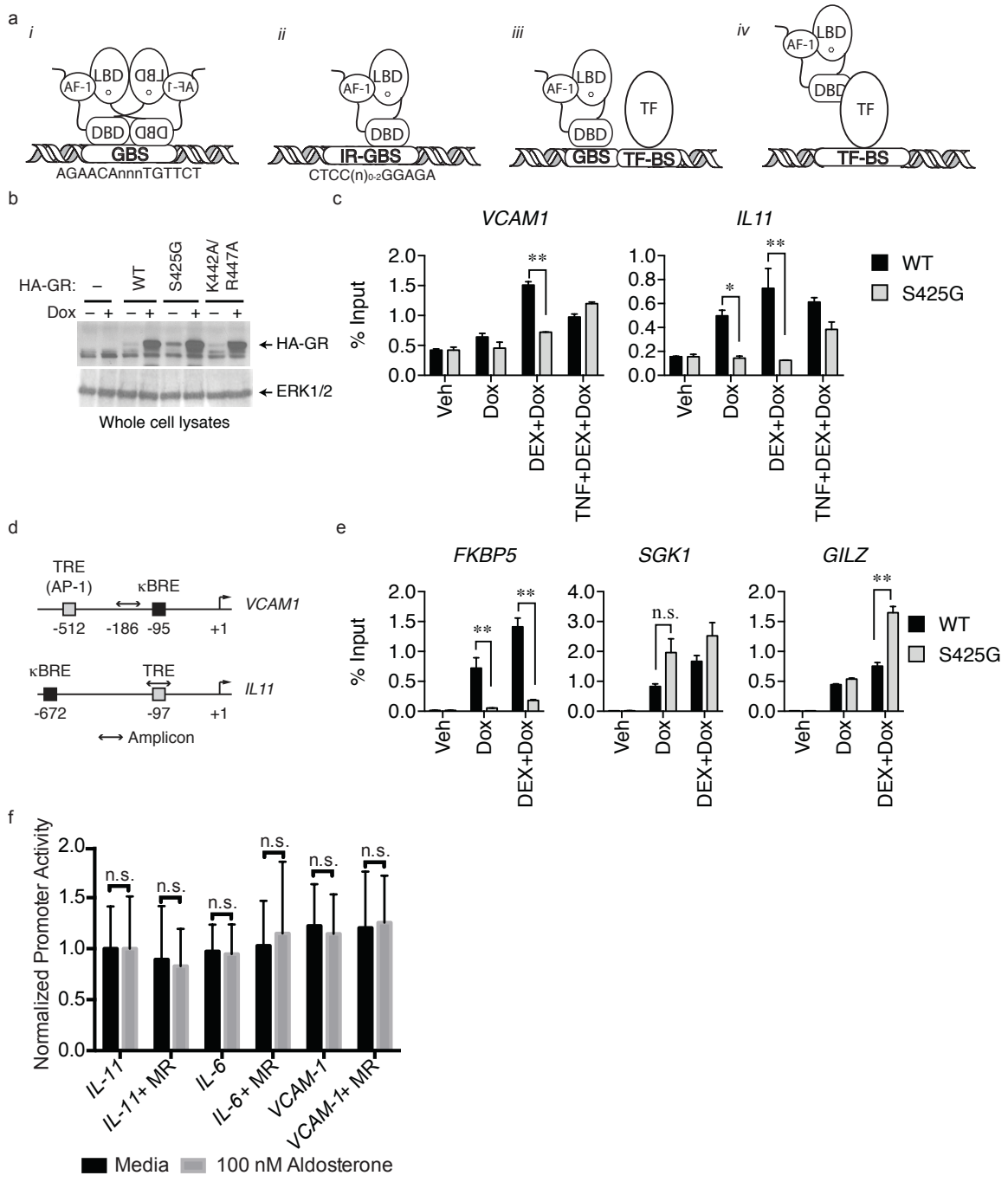


Figure 3.1: WT GR, but not GR Ser425Gly, is recruited to TREs in the absence of tethering factors

(a) Cartoon representation of GR-DNA interactions. (b-e) MCF-7 cells were transfected and steroid-deprived with or without 1 μg/ml Dox as described in methods, stimulated for 1 hr with 100 nM dexamethasone (DEX) alone or in combination with 10 ng/ml TNFα, and analyzed by ChIP assay using anti-HA antibody. HA-GR occupancies at the *IL11* and *VCAM1* promoters were determined by ChIP-PCR and shown as percent input (mean ± s.e.m) relative to un-stimulated transfectants. (b) MCF7 whole cell lysates analyzed by western blot using anti-HA and anti-ERK1/2 antibodies. (c) WT GR occupies TRE target genes where GR SER425GLY has reduced occupancy at these sites (d) Schematic of *IL11* and *VCAM1* promoters. Both promoters contain a TRE and NF-κB (κBRE) recognition element. (e) WT GR and GR Ser425Gly are similarly recruited to canonical GBS containing promoters from *FKBP5, GILZ,* and *SGK1* (f) MR is unable to transrepress a constitutively active luciferase reporter containing a portion of the *IL11*, *IL6*, and *VCAM1* promoters upon treatment with 100 nM aldosterone, a MR agonist.

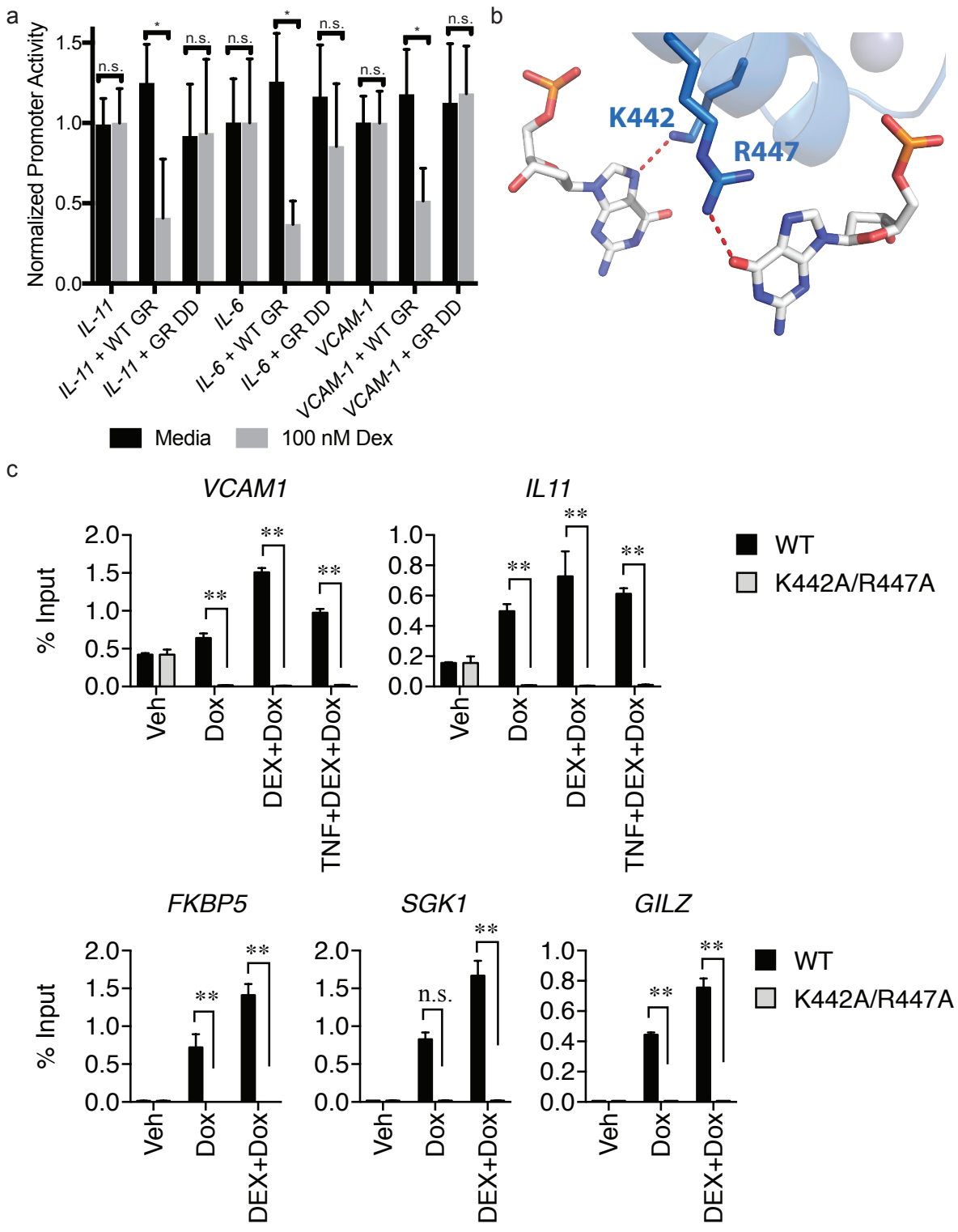
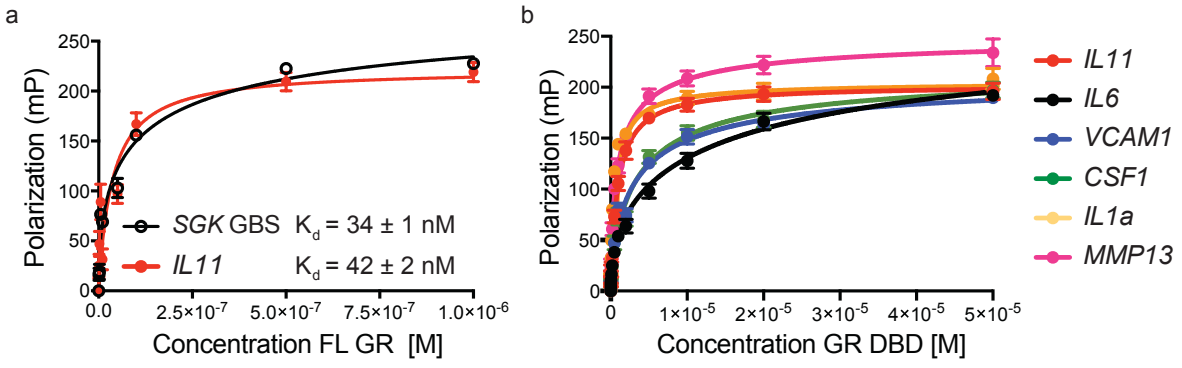


Figure 3.2: WT GR, but not GR Lys442Ala/Arg447Ala, is able to transrepress inflammatory genes.

(a) WT GR transrepresses constitutively active luciferase reporters containing portions of the *IL11*, *IL6*, and *VCAM1* promoters upon treatment with 100 nM DEX but not media alone; GR Lys442Ala/Arg447Ala (GR DNA Dead "DD") is unable to repress inflammatory gene reporters. (b) GR uses K442 and R447 to make base specific interactions on a canonical GBS (PDB: 3FYL), these side chains were mutated to Ala to disrupt sequence-specific DNA interactions. (c) Compared to WT GR, GR Lys442Ala/Arg447Ala does not occupy TREs or to canonical GBS from *FKBP5, GILZ,* and *SGK1*.



c

	High site K_d	Low site K_d	*p* for 2-site binding	r^2
IL11	12 ± 2 nM	1 ± 0.1 μM	0.0137	0.99
IL6	140 ± 51 nM	10 ± 1 μM	< 0.0001	0.99
VCAM1	402 ± 31 nM	7 ± 5 μM	0.0126	0.99
CSF1	461 ± 45 nM	6 ± 4 μM	0.0318	0.99
IL1a	157 ± 20 nM	12 ± 7 μM	< 0.001	0.99
MMP13	68 ± 6 nM	1 ± 0.1 μM	0.0003	0.99

Figure 3.3: **GR binds a variety of TRE sequences.**

(a) Fluorescence polarization binding assays monitored the ability of full-length GR to bind to the *IL11* TRE and a canonical GBS from the *SGK* promoter. (b) GR DBD binding to TREs from the *IL11*, *IL6*, *VCAM1*, *CSF1*, *IL-1a*, and *MMP13* promoters. (c) Values of fits shown in (b).

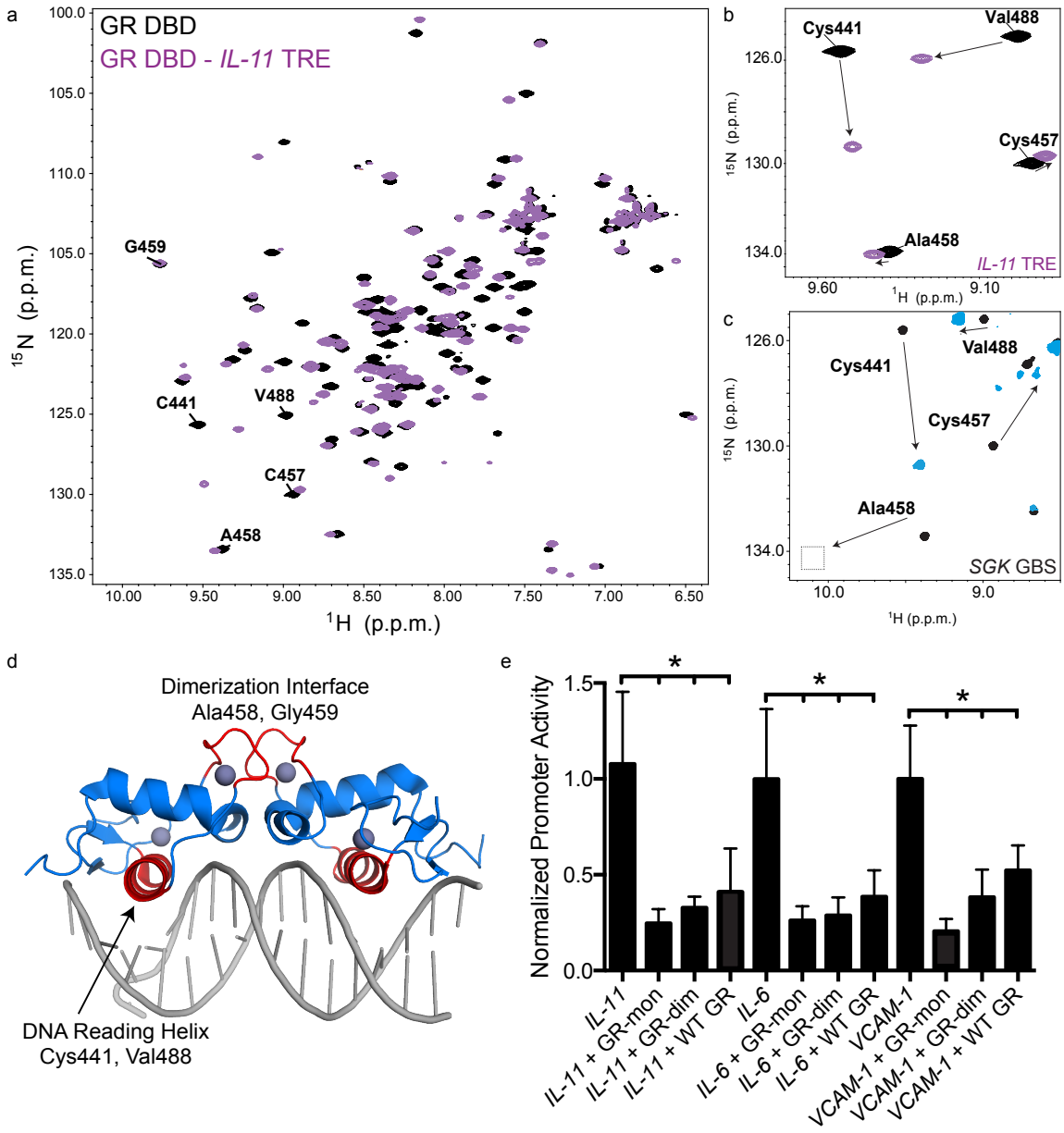


Figure 3.4: Monomeric GR is preferred to repress inflammatory genes.

(a) 2D [^{1}H,^{15}N]-HSQC NMR analysis of ^{15}N-GR DBD binding to *IL11* TRE DNA; data for GR DBD alone is shown in black and the GR DBD:*IL11* complex shown in purple. (b) Zoom-in view of the 2D NMR data show that residues contacting DNA in the complex, including Cys441 and Val488, show significant chemical shift perturbations upon binding DNA. Residues in the dimerization loop (D loop) do not show perturbations, suggesting GR

binds as a monomer to *IL11* TRE. (c) D loop residues show perturbations when GR binds as a dimer on a canonical GBS. (d) GR DBD-GBS crystal structure (PDB: 3FYL) with the DNA reading helix and D loop highlighted in red. Residues highlighted in panels a-c are located within these two regions. (e) Dimerization deficient mutants GRdim (Ala458Thr) and GRmon (Ala458Thr/Iso634Ala) cause more repression of the constitutively active TRE luciferase reporters compared to WT GR.

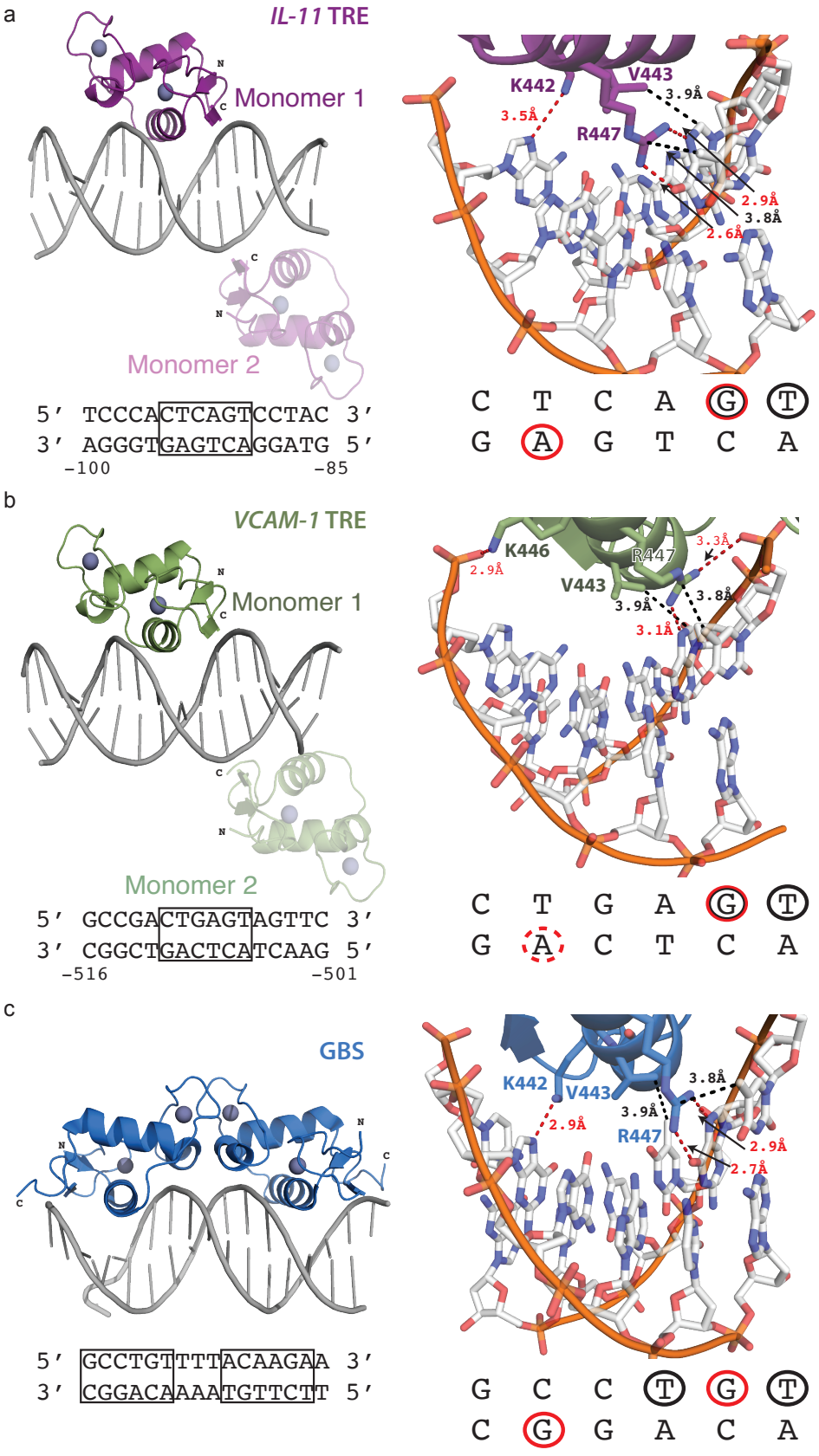

a
IL-11 TRE
Monomer 1
Monomer 2
5′ TCCCACTCAGTCCTAC 3′
3′ AGGGTGAGTCAGGATG 5′
−100
−85
K442
V443
R447
3.5Å
3.9Å
2.9Å
3.8Å
2.6Å
C T C A G T
G A G T C A
b
VCAM-1 TRE
Monomer 1
Monomer 2
5′ GCCGACTGAGTAGTTC 3′
3′ CGGCTGACTCATCAAG 5′
−516
−501
K446
R447
V443
2.9Å
3.3Å
3.9Å
3.8Å
3.1Å
C T G A G T
G A C T C A
c
GBS
5′ GCCTGTTTTACAAGAA 3′
3′ CGGACAAAATGTTCTT 5′
K442
V443
R447
2.9Å
3.8Å
3.9Å
2.9Å
2.7Å
G C C T G T
C G G A C A

Figure 3.5: Crystal structures of GR DBD bound to the IL11 TRE, VCAM1 TRE, and a GBS.

(a,b) Structures of the GR DBD bound to the (a) *IL11* (purple) and (b) *VCAM1* (green) TRE sequences. In both structures, two GR monomers were bound to opposite sides of the DNA sequence. The darker colored monomer (#1) sits over the GR binding footprint, highlighted in red, and the faded monomer (#2) sits over an end-stacking junction without making direct DNA contacts. (c) In contrast, the structure of GR DBD bound to a canonical GBS shows GR binds as a homodimer. For all structures, GR base-specific contacts are shown to the right. GR contacts the DNA through hydrogen bonds (red) and Van der Waals contacts (black) made between Arg447, Lys442, and Val443. These contacts are highlighted on the sequence below, where red circles represent contacts made through hydrogen bonds and black circles are bases contacted by Van der Waals interactions. The dotted red circle in the *VCAM1* footprint indicates contacts being made by GR with the backbone of the adenine base but not a base-specific interaction. This data shows that GR makes similar contacts on all sites.

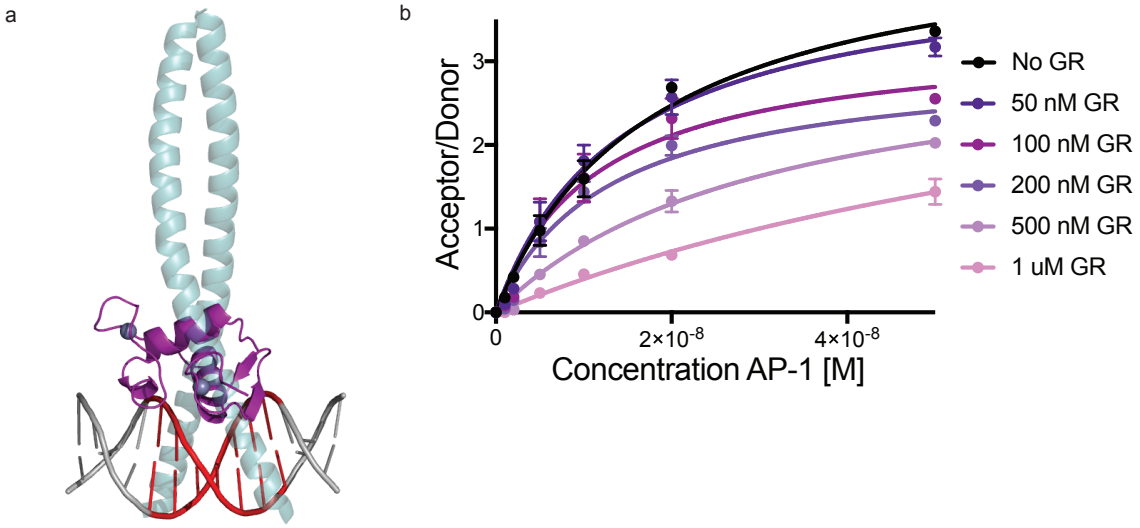


Figure 3.6: GR and AP-1 Compete for the same DNA binding site.

(a) Overlay of GR DBD-*IL11* TRE complex and AP-1 TRE complex (GR, purple; AP-1 green; TRE, red) (PDB Code: **1FOS**) (b) TR-FRET assay showing GR competes with AP-1 for binding to FAM-labeled *IL11* TRE DNA via decreased acceptor/donor signal.

	GR DBD – *IL11*	GR DBD – *VCAM1*
Data Collection		
Space Group	$P2_12_12_1$	$P2_12_12_1$
Cell Dimension	A=39.1 b=96.9 c=104.6	a=39.4, b=96.3, c=105.2
Resolution (Å)	2.15 (2.23 – 2.15)*	2.29 (2.37 - 2.29)*
R_{sym}	7.7 (47.5)	11.9 (58.2)
I/σ	8.9 (1.7)	9.7 (2.1)
Completeness	98.3 (82.9)	94.4 (98.1)
Redundancy	5.3 (3.5)	7.0 (6.6)
Refinement		
Resolution	2.15	2.29
No. Reflections	21901	17582
R_{work}/R_{free}	20.1/22.7	19.4/22.8
No. Atoms:		
Protein	1082	1094
DNA	650	650
Water	47	42
B-factors:		
Protein	47.3	42.3
DNA	64.7	50.9
Water	48.2	40.5
R.m.s. deviations:		
Bond lengths (Å)	0.009	0.008
Bond angles (°)	1.06	1.88
PDB Code	5VA7	5VA0

*Data collected from a single crystal; values in parentheses are for the highest-resolution shell.

Table 3.1: Data Collection and Refinement Statistics

Data collection and refinement statistics for the GR DBD in complex with the TRE from the *IL11* and *VCAM1* promoter.

References

1. Kadmiel, M. and J. A. Cidlowski (2013). "Glucocorticoid receptor signaling in health and disease." Trends Pharmacol Sci **34**(9): 518-530.
2. Carson-Jurica, M. A., W. T. Schrader and B. W. O'Malley (1990). "Steroid receptor family: structure and functions." Endocr Rev **11**(2): 201-220.
3. Kumar, R. and E. B. Thompson (1999). "The structure of the nuclear hormone receptors." Steroids **64**(5): 310-319.
4. Pratt, W. B. and D. O. Toft (1997). "Steroid receptor interactions with heat shock protein and immunophilin chaperones." Endocr Rev **18**(3): 306-360.
5. Picard, D., B. Khursheed, M. J. Garabedian, M. G. Fortin, S. Lindquist and K. R.
6. Yamamoto, K. R. (1985). "Steroid receptor regulated transcription of specific genes and gene networks." Annu Rev Genet **19**: 209-252.
7. Meijsing, S. H. (2015). "Mechanisms of Glucocorticoid-Regulated Gene Transcription." Adv Exp Med Biol **872**: 59-81
8. Weikum, E. R., M. T. Knuesel, E. A. Ortlund and K. R. Yamamoto (2017). "Glucocorticoid receptor control of transcription: precision and plasticity via allostery." Nat Rev Mol Cell Biol.
9. Luisi, B. F., W. X. Xu, Z. Otwinowski, L. P. Freedman, K. R. Yamamoto and P. B. Sigler (1991). "Crystallographic analysis of the interaction of the glucocorticoid receptor with DNA." Nature **352**(6335): 497-505.
10. Meijsing, S. H., M. A. Pufall, A. Y. So, D. L. Bates, L. Chen and K. R. Yamamoto (2009). "DNA binding site sequence directs glucocorticoid receptor structure and activity." Science **324**(5925): 407-410.

11. Watson, L. C., K. M. Kuchenbecker, B. J. Schiller, J. D. Gross, M. A. Pufall and K. R. Yamamoto (2013). "The glucocorticoid receptor dimer interface allosterically transmits sequence-specific DNA signals." Nat Struct Mol Biol **20**(7): 876-883.

12. Surjit, M., K. P. Ganti, A. Mukherji, T. Ye, G. Hua, D. Metzger, M. Li and P. Chambon (2011). "Widespread negative response elements mediate direct repression by agonist-liganded glucocorticoid receptor." Cell **145**(2): 224-241.

13. Hudson, W. H., C. Youn and E. A. Ortlund (2013). "The structural basis of direct glucocorticoid-mediated transrepression." Nat Struct Mol Biol **20**(1): 53-58.

14. Hudson, W. H., B. R. Kossmann, I. M. de Vera, S. W. Chuo, E. R. Weikum, G. N. Eick, J. W. Thornton, I. N. Ivanov, D. J. Kojetin and E. A. Ortlund (2016). "Distal substitutions drive divergent DNA specificity among paralogous transcription factors through subdivision of conformational space." Proc Natl Acad Sci U S A **113**(2): 326-331.

15. Miner, J. N. and K. R. Yamamoto (1992). "The basic region of AP-1 specifies glucocorticoid receptor activity at a composite response element." Genes Dev **6**(12B): 2491-2501.

16. Schiller, B. J., R. Chodankar, L. C. Watson, M. R. Stallcup and K. R. Yamamoto (2014). "Glucocorticoid receptor binds half sites as a monomer and regulates specific target genes." Genome Biol **15**(7): 418.

17. Ratman, D., W. Vanden Berghe, L. Dejager, C. Libert, J. Tavernier, I. M. Beck and K. De Bosscher (2013). "How glucocorticoid receptors modulate the activity of other transcription factors: a scope beyond tethering." Mol Cell Endocrinol **380**(1-2): 41-54.

18. De Bosscher, K., W. Vanden Berghe and G. Haegeman (2003). "The interplay between the glucocorticoid receptor and nuclear factor-kappaB or activator protein-1: molecular mechanisms for gene repression." Endocr Rev **24**(4): 488-522.
19. Herrlich, P. (2001). "Cross-talk between glucocorticoid receptor and AP-1." Oncogene **20**(19): 2465-2475.
20. Reichardt, H. M., K. H. Kaestner, J. Tuckermann, O. Kretz, O. Wessely, R. Bock, P. Gass, W. Schmid, P. Herrlich, P. Angel and G. Schutz (1998). "DNA binding of the glucocorticoid receptor is not essential for survival." Cell **93**(4): 531-541.
21. Karin, M. (1998). "New twists in gene regulation by glucocorticoid receptor: is DNA binding dispensable?" Cell **93**(4): 487-490.
22. De Bosscher, K., W. Vanden Berghe and G. Haegeman (2000). "Mechanisms of anti-inflammatory action and of immunosuppression by glucocorticoids: negative interference of activated glucocorticoid receptor with transcription factors." J Neuroimmunol **109**(1): 16-22.
23. Clark, A. R. and M. G. Belvisi (2012). "Maps and legends: the quest for dissociated ligands of the glucocorticoid receptor." Pharmacol Ther **134**(1): 54-67.
24. Cato, A. C. and E. Wade (1996). "Molecular mechanisms of anti-inflammatory action of glucocorticoids." Bioessays **18**(5): 371-378.
25. Kleiman, A. and J. P. Tuckermann (2007). "Glucocorticoid receptor action in beneficial and side effects of steroid therapy: lessons from conditional knockout mice." Mol Cell Endocrinol **275**(1-2): 98-108.
26. Schacke, H., W. D. Docke and K. Asadullah (2002). "Mechanisms involved in the side effects of glucocorticoids." Pharmacol Ther **96**(1): 23-43.

27. Uhlenhaut, N. H., G. D. Barish, R. T. Yu, M. Downes, M. Karunasiri, C. Liddle, P. Schwalie, N. Hubner and R. M. Evans (2013). "Insights into negative regulation by the glucocorticoid receptor from genome-wide profiling of inflammatory cistromes." Mol Cell **49**(1): 158-171.

28. Lim, H. W., N. H. Uhlenhaut, A. Rauch, J. Weiner, S. Hubner, N. Hubner, K. J. Won, M. A. Lazar, J. Tuckermann and D. J. Steger (2015). "Genomic redistribution of GR monomers and dimers mediates transcriptional response to exogenous glucocorticoid in vivo." Genome Res **25**(6): 836-844.

29. Biddie, S. C., S. John, P. J. Sabo, R. E. Thurman, T. A. Johnson, R. L. Schiltz, T. B. Miranda, M. H. Sung, S. Trump, S. L. Lightman, C. Vinson, J. A. Stamatoyannopoulos and G. L. Hager (2011). "Transcription factor AP1 potentiates chromatin accessibility and glucocorticoid receptor binding." Mol Cell **43**(1): 145-155.

30. John, S., P. J. Sabo, R. E. Thurman, M. H. Sung, S. C. Biddie, T. A. Johnson, G. L. Hager and J. A. Stamatoyannopoulos (2011). "Chromatin accessibility pre-determines glucocorticoid receptor binding patterns." Nat Genet **43**(3): 264-268.

31. Reddy, T. E., F. Pauli, R. O. Sprouse, N. F. Neff, K. M. Newberry, M. J. Garabedian and R. M. Myers (2009). "Genomic determination of the glucocorticoid response reveals unexpected mechanisms of gene regulation." Genome Res **19**(12): 2163-2171.

32. Heck, S., M. Kullmann, A. Gast, H. Ponta, H. J. Rahmsdorf, P. Herrlich and A. C. Cato (1994). "A distinct modulating domain in glucocorticoid receptor monomers in the repression of activity of the transcription factor AP-1." EMBO J **13**(17): 4087-4095.

33. Cato, A. C., H. Konig, H. Ponta and P. Herrlich (1992). "Steroids and growth promoting factors in the regulation of expression of genes and gene networks." J Steroid Biochem Mol Biol **43**(1-3): 63-68.

34. Pearce, D. and K. R. Yamamoto (1993). "Mineralocorticoid and glucocorticoid receptor activities distinguished by nonreceptor factors at a composite response element." Science **259**(5098): 1161-1165.

35. Hudson, W. H., C. Youn and E. A. Ortlund (2014). "Crystal structure of the mineralocorticoid receptor DNA binding domain in complex with DNA." PLoS One **9**(9): e107000.

36. Wei, P., N. Inamdar and W. V. Vedeckis (1998). "Transrepression of c-jun gene expression by the glucocorticoid receptor requires both AP-1 sites in the c-jun promoter." Mol Endocrinol **12**(9): 1322-1333.

37. Angel, P., M. Imagawa, R. Chiu, B. Stein, R. J. Imbra, H. J. Rahmsdorf, C. Jonat, P. Herrlich and M. Karin (1987). "Phorbol ester-inducible genes contain a common cis element recognized by a TPA-modulated trans-acting factor." Cell **49**(6): 729-739.

38. Karin, M., Z. Liu and E. Zandi (1997). "AP-1 function and regulation." Curr Opin Cell Biol **9**(2): 240-246.

39. Ahmad, M., P. Theofanidis and R. M. Medford (1998). "Role of activating protein-1 in the regulation of the vascular cell adhesion molecule-1 gene expression by tumor necrosis factor-alpha." J Biol Chem **273**(8): 4616-4621.

40. Sawa, Y., T. Ueki, M. Hata, K. Iwasawa, E. Tsuruga, H. Kojima, H. Ishikawa and S. Yoshida (2008). "LPS-induced IL-6, IL-8, VCAM-1, and ICAM-1 expression in human lymphatic endothelium." J Histochem Cytochem **56**(2): 97-109.

41. Tohjima, E., D. Inoue, N. Yamamoto, S. Kido, Y. Ito, S. Kato, Y. Takeuchi, S. Fukumoto and T. Matsumoto (2003). "Decreased AP-1 activity and interleukin-11 expression by bone marrow stromal cells may be associated with impaired bone formation in aged mice." J Bone Miner Res **18**(8): 1461-1470.

42. Tang, W., L. Yang, Y. C. Yang, S. X. Leng and J. A. Elias (1998). "Transforming growth factor-beta stimulates interleukin-11 transcription via complex activating protein-1-dependent pathways." J Biol Chem **273**(10): 5506-5513.

43. Yang-Yen, H. F., J. C. Chambard, Y. L. Sun, T. Smeal, T. J. Schmidt, J. Drouin and M. Karin (1990). "Transcriptional interference between c-Jun and the glucocorticoid receptor: mutual inhibition of DNA binding due to direct protein-protein interaction." Cell **62**(6): 1205-1215.

44. Schule, R., P. Rangarajan, S. Kliewer, L. J. Ransone, J. Bolado, N. Yang, I. M. Verma and R. M. Evans (1990). "Functional antagonism between oncoprotein c-Jun and the glucocorticoid receptor." Cell **62**(6): 1217-1226.

45. Bain, D. L., Q. Yang, K. D. Connaghan, J. P. Robblee, M. T. Miura, G. D. Degala, J. R. Lambert and N. K. Maluf (2012). "Glucocorticoid receptor-DNA interactions: binding energetics are the primary determinant of sequence-specific transcriptional activity." J Mol Biol **422**(1): 18-32.

46. Presman, D. M., M. F. Ogara, M. Stortz, L. D. Alvarez, J. R. Pooley, R. L. Schiltz, L. Grontved, T. A. Johnson, P. R. Mittelstadt, J. D. Ashwell, S. Ganesan, G. Burton, V. Levi, G. L. Hager and A. Pecci (2014). "Live cell imaging unveils multiple domain requirements for in vivo dimerization of the glucocorticoid receptor." PLoS Biol **12**(3): e1001813.

47. Krissinel, E. and K. Henrick (2007). "Inference of macromolecular assemblies from crystalline state." J Mol Biol **372**(3): 774-797.

48. Sacta, M. A., Y. Chinenov and I. Rogatsky (2016). "Glucocorticoid Signaling: An Update from a Genomic Perspective." Annu Rev Physiol **78**: 155-180.

49. John, S., P. J. Sabo, T. A. Johnson, M. H. Sung, S. C. Biddie, S. L. Lightman, T. C. Voss, S. R. Davis, P. S. Meltzer, J. A. Stamatoyannopoulos and G. L. Hager (2008). "Interaction of the glucocorticoid receptor with the chromatin landscape." Mol Cell **29**(5): 611-624.

50. Ray, A. and K. E. Prefontaine (1994). "Physical association and functional antagonism between the p65 subunit of transcription factor NF-kappa B and the glucocorticoid receptor." Proc Natl Acad Sci U S A **91**(2): 752-756.

51. Lucibello, F. C., E. P. Slater, K. U. Jooss, M. Beato and R. Muller (1990). "Mutual transrepression of Fos and the glucocorticoid receptor: involvement of a functional domain in Fos which is absent in FosB." EMBO J **9**(9): 2827-2834.

52. Baschant, U., S. Culemann and J. Tuckermann (2013). "Molecular determinants of glucocorticoid actions in inflammatory joint diseases." Mol Cell Endocrinol **380**(1-2): 108-118.

53. Gebhardt, J. C., D. M. Suter, R. Roy, Z. W. Zhao, A. R. Chapman, S. Basu, T. Maniatis and X. S. Xie (2013). "Single-molecule imaging of transcription factor binding to DNA in live mammalian cells." Nat Methods **10**(5): 421-426.

54. Oro, A. E., S. M. Hollenberg and R. M. Evans (1988). "Transcriptional inhibition by a glucocorticoid receptor-beta-galactosidase fusion protein." Cell **55**(6): 1109-1114.

55. Rani, C. S., N. Elango, S. S. Wang, K. Kobayashi and R. Strong (2009). "Identification of an activator protein-1-like sequence as the glucocorticoid response element in the rat tyrosine hydroxylase gene." Mol Pharmacol **75**(3): 589-598.

56. Sheela Rani, C. S., A. Soto-Pina, L. Iacovitti and R. Strong (2013). "Evolutionary conservation of an atypical glucocorticoid-responsive element in the human tyrosine hydroxylase gene." J Neurochem **126**(1): 19-28.

57. Beck, I. M., K. De Bosscher and G. Haegeman (2011). "Glucocorticoid receptor mutants: man-made tools for functional research." Trends Endocrinol Metab **22**(8): 295-310.

58. Tuckermann, J. P., H. M. Reichardt, R. Arribas, K. H. Richter, G. Schutz and P. Angel (1999). "The DNA binding-independent function of the glucocorticoid receptor mediates repression of AP-1-dependent genes in skin." J Cell Biol **147**(7): 1365-1370.

59. Cole, T. J., J. A. Blendy, A. P. Monaghan, K. Krieglstein, W. Schmid, A. Aguzzi, G. Fantuzzi, E. Hummler, K. Unsicker and G. Schutz (1995). "Targeted disruption of the glucocorticoid receptor gene blocks adrenergic chromaffin cell development and severely retards lung maturation." Genes Dev **9**(13): 1608-1621.

60. Jonat, C., Rahmsdorf, H.J., Park, K.K., Cato, A.C., Gebel, S., Ponta, H. and Herrlich, P. (1990). "Antitumor promotion and antiinflammation: down-modulation of AP-1 (Fos/Jun) activity by glucocorticoid hormone." **Cell 62**: 1189-1204.

61. Flicek, P., I. Ahmed, M. R. Amode, D. Barrell, K. Beal, S. Brent, D. Carvalho-Silva, P. Clapham, G. Coates, S. Fairley, S. Fitzgerald, L. Gil, C. Garcia-Giron, L. Gordon, T. Hourlier, S. Hunt, T. Juettemann, A. K. Kahari, S. Keenan, M. Komorowska, E. Kulesha, I. Longden, T. Maurel, W. M. McLaren, M. Muffato, R. Nag, B. Overduin,

M. Pignatelli, B. Pritchard, E. Pritchard, H. S. Riat, G. R. Ritchie, M. Ruffier, M. Schuster, D. Sheppard, D. Sobral, K. Taylor, A. Thormann, S. Trevanion, S. White, S. P. Wilder, B. L. Aken, E. Birney, F. Cunningham, I. Dunham, J. Harrow, J. Herrero, T. J. Hubbard, N. Johnson, R. Kinsella, A. Parker, G. Spudich, A. Yates, A. Zadissa and S. M. Searle (2013). "Ensembl 2013." Nucleic Acids Res **41**(Database issue): D48-55.

62. Nwachukwu, J. C., S. Srinivasan, N. E. Bruno, A. A. Parent, T. S. Hughes, J. A. Pollock, O. Gjyshi, V. Cavett, J. Nowak, R. D. Garcia-Ordonez, R. Houtman, P. R. Griffin, D. J. Kojetin, J. A. Katzenellenbogen, M. D. Conkright and K. W. Nettles (2014). "Resveratrol modulates the inflammatory response via an estrogen receptor-signal integration network." Elife **3**: e02057.
63. Robblee, J. P., M. T. Miura and D. L. Bain (2012). "Glucocorticoid receptor-promoter interactions: energetic dissection suggests a framework for the specificity of steroid receptor-mediated gene regulation." Biochemistry **51**(22): 4463-4472.
64. Otwinowski, Z. and W. Minor (1997). "Processing of X-ray diffraction data collected in oscillation mode." Methods Enzymol **276**: 307-326.
65. Adams, P. D., P. V. Afonine, G. Bunkoczi, V. B. Chen, I. W. Davis, N. Echols, J. J. Headd, L. W. Hung, G. J. Kapral, R. W. Grosse-Kunstleve, A. J. McCoy, N. W. Moriarty, R. Oeffner, R. J. Read, D. C. Richardson, J. S. Richardson, T. C. Terwilliger and P. H. Zwart (2010). "PHENIX: a comprehensive Python-based system for macromolecular structure solution." Acta Crystallogr D Biol Crystallogr **66**(Pt 2): 213-221.

66. Emsley, P., B. Lohkamp, W. G. Scott and K. Cowtan (2010). "Features and development of Coot." Acta Crystallogr D Biol Crystallogr **66**(Pt 4): 486-501.

67. Joosten, R. P., J. Salzemann, V. Bloch, H. Stockinger, A. C. Berglund, C. Blanchet, E. Bongcam-Rudloff, C. Combet, A. L. Da Costa, G. Deleage, M. Diarena, R. Fabbretti, G. Fettahi, V. Flegel, A. Gisel, V. Kasam, T. Kervinen, E. Korpelainen, K. Mattila, M. Pagni, M. Reichstadt, V. Breton, I. J. Tickle and G. Vriend (2009). "PDB_REDO: automated re-refinement of X-ray structure models in the PDB." J Appl Crystallogr **42**(Pt 3): 376-384.

68. Mathis, G. (1995). "Probing molecular interactions with homogeneous techniques based on rare earth cryptates and fluorescence energy transfer." Clin Chem **41**(9): 1391-1397.

CHAPTER 4: STRUCTURAL ANALYSIS OF THE GLUCOCORTICOID RECEPTOR LIGAND-BINDING DOMAIN IN COMPLEX WITH TRIAMCINOLONE ACETONIDE AND A FRAGMENT OF THE ATYPICAL COREGULATOR, SMALL HETERODIMER PARTNER

Emily R. Weikum[1], Denise C. Okafor[1], Emma H. D'Agostino[1], Jennifer K. Colucci[1], Eric A. Ortlund[13]

1) Department of Biochemistry, Emory University School of Medicine, Atlanta GA 30322 USA

This manuscript describes the crystal structure of the glucocorticoid receptor ligand binding domain in complex with a glucocorticoid ligand widely used in the clinic and laboratory, triamcinolone acetonide, and a fragment from the atypical coregulator, small heterodimer partner. In order to investigate how ligands interact with the receptor, we rely on these x-ray crystal structures. Identifying the structural mechanisms that guide ligand-receptor recognition is of particular interest to the glucocorticoid receptor field and the pharmaceutical industry as a way to guide future drug design. This work was accepted for publication in Molecular Pharmacology for a special issue on Structural Basis for Receptor-Ligand Interactions in April 2017.

---

[a] Conceived and designed the experiments: ERW DCO EAO. Performed the experiments: ERW DCO EHD EAO. Analyzed the data: ERW DCO EHD EAO. Wrote the paper: ERW DCO JKC EAO.

Abstract

The synthetic glucocorticoids (GCs) dexamethasone, mometasone furoate, and triamcinolone acetonide have been pharmaceutical mainstays to treat chronic inflammatory diseases. These drugs bind to the glucocorticoid receptor (GR), a ligand-activated transcription factor and member of the nuclear receptor superfamily. GR is widely recognized as a therapeutic target for its ability to counter pro-inflammatory signaling. Despite the popularity of GCs in the clinic, long-term use leads to numerous side effects, driving the need for new and improved drugs with less off-target pharmacology. X-ray crystal structures have played an important role in the drug-design process, permitting the characterization of robust structure-function relationships. However, steroid receptor ligand-binding domains (LBDs) are inherently unstable and their crystallization has required extensive mutagenesis to enhance expression and crystallization. Here, we utilize an ancestral variant of GR as a tool to generate a high-resolution crystal structure of GR in complex with the potent glucocorticoid triamcinolone acetonide (TA) and a fragment of the small heterodimer partner (SHP). Using structural analysis, molecular dynamics and biochemistry, we show that TA increases intramolecular contacts within the LBD to drive affinity and enhance stability of the receptor-ligand complex. These data support the emerging theme that ligand-induced receptor conformational dynamics at the mouth of the pocket play a major role in steroid receptor activation. This work also represents the first GR structure in complex with SHP, which has been suggested to play a role in modulating hepatic GR function.

Introduction

The nuclear receptor superfamily is comprised of a family of ligand-regulated transcription factors that are critical for maintaining specific gene expression profiles across a number of biological processes[1]. The glucocorticoid receptor (GR), the founding member of the nuclear receptor superfamily, is ubiquitously expressed and both up- and down-regulates thousands of genes involved in immunity, metabolism, and inflammation[2]. GR has a modular architecture consisting of five primary domains: an N-terminal region that contains an activator function (AF-1) surface that interacts with coregulators; a highly conserved zinc finger DNA-binding domain (DBD); a flexible hinge region; and a ligand-binding domain (LBD), which binds ligands that modulate receptor activity. Ligands drive conformational changes within the LBD that modulate a second activation function surface (AF-2), which in turn enables selective interaction with coregulators[1,3-5]. Unliganded GR resides in the cytosol bound to chaperone proteins. Upon binding to glucocorticoids, GR undergoes a conformational change resulting in an exchange of chaperones and translocation to the nucleus[6,7]. In the nucleus, GR interacts with the genome via multiple mechanisms to then recruit transcriptional coregulator proteins, which serve to promote or repress transcription[8,9]. The simplistic view of GR signaling suggests that GR will bind directly to DNA at canonical GR binding sequences (GBS) as a homodimer to activate transcription or can bind as a monomer to inverted repeat GBS (IR-GBS/nGREs) to repress transcription[10,11]. GR can also regulate transcription in a DNA-independent manner by interacting directly with the pro-inflammatory transcription factors activator protein-1 (AP-1) and nuclear factor kappa-beta (NF-κB) through protein-protein interactions to block their activity[12-14]. This mechanism,

known as tethering, is the major mechanism targeted by pharmaceutical companies for anti-inflammatory therapies[15].

Since their discovery in 1948, glucocorticoids (GCs) have been the most efficacious treatment for chronic inflammatory diseases such as rheumatoid arthritis and asthma[16]. GC-bound GR is unmatched in its anti-inflammatory action, resulting in its prevalence in the clinic and worldwide sales that surpass 10 billion dollars per year[17]. Yet despite the effectiveness of GCs, long-term exposure leads to numerous debilitating side effects including weight gain, muscle wasting, and development of Cushing's syndrome[18]. These adverse effects have generally been attributed to GR's role in transactivation[18], spurring studies focused on the molecular mechanisms of GR-ligand interactions to support the development of more selective GCs.

Triamcinolone Acetonide (TA) is a highly potent synthetic glucocorticoid that has been FDA-approved to treat allergic rhinitis as a nasal spray (Nasacort®; Sanofi, Bridgewater, New Jersey)[18,19]. It is also used to treat macular edema as a result of diabetes (KENALOG) and as an over-the-counter cream to treat skin lesions (Bristol-Myers Squibb Company; Princeton, New Jersey)[19]. In addition to its prolific clinical uses, TA has been widely used in the laboratory setting to probe GR biology since it shows a higher affinity than dexamethasone (Dex) and promotes the soluble expression and purification of the full length receptor for *in vitro* biochemical characterization[1,20-22].

Soluble expression and purification of human GR LBD is challenging and requires mutation at several sites to promote solubility and purification[23]. These mutations, designed via sequence alignments to other steroid receptor LBDs, enable expression in *E. coli* and

greatly enhance crystallization. However, their effect on ligand selectivity, conformational dynamics and activity remains unknown, resulting in uncertainty when constructing robust structure-function relationships. Recently, resurrected ancestral proteins have been used as a tool not only to study evolutionary biology, but also to generate x-ray crystal structures of novel complexes[24-26]. Here, we utilize AncGR2, the ancestral precursor to the modern GR, which recapitulates both the structure and function of its extant counterpart while exhibiting better expression, purification, and crystallization in the laboratory setting along with a higher tolerance to mutagenic analysis, which is critical for probing structure-function relationships[24,25,27,28]. AncGR2 and hGR LBDs share 78.6 % sequence identity with 100 % of the residues within the ligand binding pocket conserved. This makes it an ideal surrogate for hGR structure-function studies, and allowed us to generate the first x-ray crystal structure of the GR LBD complexed with TA and with a peptide from the atypical coregulator, small heterodimer partner (SHP).

Materials and Methods

Protein Expression and Purification

AncGR2 LBD (GenBank accession number EF631976.1) was cloned into a pMALCH10T vector, transformed into *Escherichia coli* strain BL21(pLysS), and expressed as a maltose-binding protein (MBP) with a hexahistidine tag. Cultures (9 L in Terrific Broth) were grown at 37 °C until an OD_{600} of 0.6. Protein expression was induced with the addition of 0.5 mM isopropyl-1-thio-D-galactopyranoside (IPTG) and 25 μM TA or 50 μM dexamethasone and grown for 4 hours at 32 °C. Cells were lysed in a buffer containing 20 mM Tris-HCl (pH 7.4), 300 mM NaCl, 25 mM imidazole, and 5 % glycerol via sonication on ice. Fusion protein was purified by affinity chromatography (GE Healthcare His-Trap FF) and the MBP tag was removed by Tobbaco Etch Virus (TEV) protease cleavage with an additional two passes over HisTrap FF media (GE Healthcare). After purification, protein was dialyzed overnight at 4 °C into buffer containing 20 mM HEPES (pH 8.0), 200 mM NaCl, 10 % glycerol, and 50 [illegible]CHAPS. Protein was concentrated to 3-5 mg/mL, flash frozen in liquid N_2, and stored at -80 °C.

Protein Crystallization, Data Collection, and Structure Determination

AncGR2 LBD-TA was concentrated to 4.75 mg/mL and incubated with a peptide derived from SHP (NH2-QGASRPAILYALLSSSLK-OH) at two-fold molar excess. Crystals were grown by hanging drop vapor diffusion at 18 °C in drops containing 1 μL of AncGR2-TA-SHP and 1 μL of 100 mM HEPES pH 7.5, 1.5 % glycerol, and 25 % PEG 300. Crystals were flash frozen in liquid nitrogen, using a cryopreservative consisting of 100 mM HEPES pH 7.5, 10 % glycerol, and 40 % PEG 300. Data were collected remotely from the South East Regional Collaborative Access Team (SER-CAT) at the Advanced Photon Source (APS), 22ID beamline (Argonne National Laboratories, Chicago, IL). Data were processed and scaled using HKL-2000[28] and phased by molecular replacement using Phaser-MR[29]. The structure was phased using the previously solved AncGR2 – dexamethasone – nuclear

receptor coactivator 2 (TIF2) peptide complex as a search model (3GN8)[24]. Structure refinement and validation was performed using PHENIX (v1.11.1) and model building was performed in COOT[29,30]. PDB Redo was used iteratively to optimize refinement parameters and geometry (Joosten et al., 2009). PyMOL (v1.8.2) was used to visualize structures and generate figures (Schrödinger, LLC).

$R_{factors}$ for the final model are 21.5 % and 25.2 % for R_{work} and R_{free}, respectively. MolProbity was used for model validation, indicating that 98 % of the residues fall in the most favored regions of the Ramachandran plot with none in disallowed regions[29,31]. The overall MolProbity score was 0.70, placing the structure in the 100th percentile for overall geometric quality among protein crystal structures of comparable resolution. The remaining data collection and refinement statistics can be found in **Table 5.1**. Coordinates and structure factors have been deposited in the Protein Data Bank (PDB) with the accession number, 5UFS.

Ligand Binding & Competition Assays

Hexahistidine-tagged MBP-fused AncGR2 LBD-Dex was dialyzed overnight into buffer containing 150 mM NaCl, 10 mM HEPES pH 7.4, 3 mM EDTA, 5 mM DTT, and 0.005 % Tween-20. All binding experiments were performed using this buffer. Binding affinity for dexamethasone-fluorescein was measured with 12 nM dexamethasone-fluorescein and protein concentrations from 10^{-10} to 10^{-5} M. Polarization was monitored on a Biotek Neo plate-reader at an excitation/emission wavelength of 485/528 nm (Winooski, VT). Three technical replicates and three biological replicates were conducted and graphs are a compilation of all data collected. Binding data were fit with a one-site binding curve in GraphPad Prism v7 (GraphPad, Inc). Competition assays were performed at a protein

concentration 1.2 times the binding affinity for dexamethasone and in the presence of 12 nM dexamethasone-fluorescein and 10^{-10} to 10^{-5} M of competing ligand. Three technical replicates and three biological replicates were conducted and graphs are a compilation of all data collected. GraphPad Prism v7 was used to analyze data using a one-site, fit K_i curve.

Differential Scanning Fluorimetry (DSF)

DSF assays for protein-ligand complexes (**Figure 5.2d**) were performed with 5 μM of AncGR2 LBD (expressed and purified with varying ligands), buffer containing 20 mM Tris pH 7.4, 300 mM NaCl, and 5 % glycerol, and a final 1:1000 dilution of SYPRO® orange dye (Sigma) to a final volume of 20 μL. Reactions with coregulator peptides (**Figure X.5e**) consisted of 5 μM protein, 5 μM peptide, buffer, and SYPRO dye as above. The peptide sequences used were as follows: SHP NR Box 1 (NH2-QGASRPAILYALLSSSLK-OH) and TIF2 NR Box 3 (NH2-KENALLRYLLDKDD-OH). For all DSF experiments reactions were heated from 25 °C to 95 °C in 0.5 °C increments every minute and fluorescence was monitored using the ROX filter (602 nm) with a StepOne Plus Real Time PCR System (ThermoFisher). Two technical replicates and three biological replicates were conducted and graphs are a compilation of all data collected. Data was normalized by subtracting each data point from a dye-buffer only control reaction. Data was fit using a Boltzmann sigmoidal curve to determine the melting temperature (T_m), defined as 50 % unfolding.

ProSMART Analysis

ProSMART alignment tool provides a conformation-independent structural comparison of two proteins[32]. Pairwise comparisons were conducted between AncGR2 LBD-TA-SHP and AncGR2 LBD-Dex-TIF2 (PDB 3GN8); AncGR2 LBD-TA-SHP and AncGR2 LBD-MF-TIF2 (PDB 4E2J); AncGR2 LBD-Dex-TIF2 and AncGR2 LBD-MF-TIF2[24,25]. The

comparisons generate a Procrustes score, which is mapped onto the LBD structures. This score is the r.m.s.d. of the central residue of two corresponding structural fragments of length *n*, where *n* is an odd number of amino acids.

Molecular Dynamics Simulations

Three complexes were prepared for molecular dynamics simulations using AncGR2 LBD: i) TA-TIF2 ii) Dex-TIF2 (PDB 3GN8) and iii) MF-TIF2 (PDB 4E2J). The structure for TA-TIF2 was created by replacing SHP peptide with TIF2 and an all-atom minimization of the structure. The complexes were solvated in an octahedral box of TIP3P[33] water with a 10 Å buffer around the protein complex. Na^+ and Cl^- ions were added to neutralize the protein and achieve physiological conditions, to a final concentration of 150 mM. All systems were set up using xleap in AmberTools[34] with the parm99-bsc0 forcefield[35]. Parameters for TA, MF and Dex were obtained using Antechamber[36] in AmberTools. All minimizations and simulations were performed with Amber14[37]. Systems were minimized with 5000 steps of steepest decent followed by 5000 steps of conjugate gradient minimization with 500 kcal/ mol·$Å^2$ restraints on all atoms. Restraints were removed from all atoms excluding the atoms in both the ligand and the TIF2 peptide, and the previous minimization was repeated. The systems were heated from 0 to 300 K using a 100-ps run with constant volume periodic boundaries and 5 kcal/mol·$Å^2$ restraints on all protein and ligand atoms. Twelve ns of MD equilibration was performed with 10 kcal/ mol·$Å^2$ restraints on protein and ligand atoms using the NPT ensemble. Restraints were reduced to 1 kcal/ mol·$Å^2$ for an additional 10 ns of MD equilibration. Then restraints were removed and 500 ns production simulations were performed for each system in the NPT ensemble. A 2 fs timestep was used and all bonds between heavy atoms and hydrogens were fixed with the

SHAKE algorithm[38]. A cut-off distance of 10 Å was used to evaluate long-range electrostatics with Particle Mesh Ewald (PME) and for van der Waals forces. 25,000 evenly spaced frames were taken from each simulation for analysis.

MD trajectories were analyzed with various tools. Structural averaging and analysis were performed with the CPPTRAJ module[39] of AmberTools (v14). Hydrogen bonds were identified using HBPLUS[40] with default criteria (minimum angles: D-H-A 90.0, H-A-AA 90.0, D-A-AA 90; maximum distances D-A 3.9Å, H-A 2.5Å, D = Donor, A = acceptor, AA= atom attached to acceptor). Root mean square fluctuations (RMSF) analysis was performed on Cα atoms of protein residues. RMSF was computed relative to the reference structure (crystal structure) for each frame in the trajectory. CPPTRAJ was used to calculate distances between Cα atoms or hydrogen bonding atoms over trajectories. All distances between residue pairs were calculated using Cα atoms. The MMTSB toolset was used to perform a cluster analysis with a 2 Å RMSD cutoff[41].

Results

Structural analysis of AncGR2-Triamcinolone/SHP complex

To understand how TA interacts with GR, we determined the x-ray crystal structure of AncGR2 LBD – TA complex bound to a fragment of the coregulator SHP. Crystals formed in the C2 space group with two AncGR2 LBD – TA – SHP complexes in the asymmetric unit. Data sets were collected to 2.1 Å with 98.6 % completeness. Full data collection and refinement statistics are summarized in **Table 4.1**.

The AncGR2 LBD is bound by one molecule of TA and the SHP coregulator peptide (**Figure 4.1a**). The LBD adopts a canonical fold with 11 α-helices and 4 β-strands that fold into three layers of a helical sandwich bundle (**Figure 4.1a**)[5]. This folding creates a hydrophobic ligand binding pocket (LBP) encompassing TA (**Figure 4.1b**), which is supported by unambiguous electron density (**Figure 4.1c**).

TA is coordinated via extensive hydrophobic contacts and a series of specific hydrogen-bonding interactions (**Figure 4.1d**). This hydrogen bonding network includes a water molecule, which is a common feature for SR LBD-ligand complexes and is seen in all AncGR2 LBD-ligand structures to date[24,25]. This water hydrogen bonds with the terminal oxygen of the A-ring. The amide of Gln39 and guanidinium group of Arg80 not only hydrogen bond to the water molecule but also form hydrogen bonds to the A-ring 3-keto oxygen (**Figure 4.1d,2a**). Molecular dynamics simulations (MD) of ligand-hydrogen bonding interactions show that over a 500 ns simulation, these hydrogen bonds alternate between being direct or water-mediated (**Figure 4.2a**). In addition, there is a water-mediated hydrogen bond that occurs between the Arg80 and Gln39 residues that persists for 47% of

the simulation. Both of these hydrogen-bonding interactions have a joint effect in keeping TA in the correct orientation within the LBP over the duration of the simulation. Asn33 hydrogen bonds to the C-ring 11-hydroxyl. Finally, Asn33 and Thr208 hydrogen bond to the 21-hydroxyl located off the C-17 position on the D-ring. These hydrogen bonds are well supported via MD and persist throughout the entirety of the simulations (**Figure 4.2b**).

The C-17 acetonide moiety on TA increases the affinity and stability for GR LBD over Dex.

Synthetic GCs typically show increased affinity and selectivity for the receptor over GR's endogenous ligand, cortisol. Corticosteroids, like most cholesterol-derived steroid hormones, contain three 6-carbon rings (A, B, C) and one 5-carbon ring (D) (**Figure 4.3a**). Dex varies from cortisol by the addition of a C1-C2 double bond in ring A, a C-16 α-methyl, and a C-9 α-fluoro group. TA varies from Dex by the addition of the 16,17 acetonide moiety. MF varies from Dex by the presence of a chloro group at C-9 instead of a fluoro group, a chloro group instead of a 21-hydroxyl off the C-17 D-ring, and the large furoate moiety off the C-17 position. The C-17 furoate moiety is more polar and labile than the TA acetonide moiety and is cleaved *in vivo*. The active metabolite, mometasone, contains a at hydroxyl at the C17 position due to loss of the fuorate moiety; thus, conclusions based on analysis of the GR-MF complex cannot be reliably correlated with its *in vivo* potency. However, MF has proven a useful tool to probe GR LBD biochemistry *in vitro* and is therefore included in our analysis. For synthetic GCs, the steroid backbone halogen substitutions are thought to increase potency and the substitutions at position 17 account for the largest functional differences[21,42]. These substitutions on TA and MF drive the increased affinity and reported potencies over Dex[43,44].

Similar to previous reports, we show that TA displaces bound fluorescent Dex (FL-Dex) from the AncGR2 LBD with a K_i of 3.2 nM (K_d for FL-Dex = 38 nM) (**Figure 4.3b**)[43,45,46]. Furthermore, we show there is an 8 °C increase in thermal stability of the LBD when bound by TA or MF over Dex (**Figure 4.3c**). To understand how TA has such a high affinity and potency for GR, we compared our structure to the previously solved AncGR2-Dex-TIF2 and AncGR2-MF-TIF2 complexes (PDB 3GN8, 4E2J) (**Figure 4.4**)[24,25].

Superposition of the different ligand-bound structures revealed only subtle differences in the RMSD's of main chain carbon alphas. The difference between AncGR2-TA and AncGR2-Dex is 0.346 Å, and between AncGR2-TA and AncGR2-MF is 0.349 Å. Within the ligand-binding pocket, the same side chains participate in hydrogen bonding interactions, including a water molecule in all structures (**Figure 4.4a**). This is not surprising when comparing Dex and TA, as the ligands are almost identical except for the acetonide moiety, which is contacted by hydrophobic interactions. However, MF-bound GR makes one fewer hydrogen bonding interaction; the C17 position contains a chloro group rather than a hydroxyl, altering the interaction with Thr208. Yet, as with MF, the acetonide C-17 addition is oriented almost 90° from the steroid backbone. This allows for the ligand to expand and fill the binding pocket as well as generate additional hydrophobic contacts between GR and the ligand. These large substitutions on TA and MF cause a 100 $Å^3$ (1.1-fold) and 200 $Å^3$ (1.3-fold) increase, respectively, in pocket volume compared to Dex (**Figure 4.4b**). Therefore, the differences in potency and affinity must be due to changes in hydrophobic contacts and long-range allosteric changes generated by the ligands. To investigate this possibility, we combined in-depth structural analysis with MD simulations.

TA increases intramolecular contacts within the LBD to drive affinity and stability

To identify how differential ligands affect local structure, we performed pairwise comparisons of the three AncGR2 ligand structures using ProSMART[32] (**Figure 4.5a**). Each chain is compared separately and the final models are colored by their Procrustes score, which defines the similarity of an aligned fragment according to the legend provided in Figure 4. The largest differences are seen in the mouth of the ligand binding pocket, encompassing the loops after H1, H3, and H6, and the transition between H9 and H10. Differences in the H9-H10 loop are not surprising, as this loop is often difficult to model and is not supported by strong electron density in any structure. However, there are meaningful differences in local residue conformation at the entrance to the ligand binding pocket, where TA constricts the opening of the pocket to 4.6 Å, compared to 6.3 Å for Dex and 10.1 Å for MF as measured by the distance between Ser25 on H3 and Pro106 on H6. (**Figure 4.5b**). This TA-driven pocket constriction is also observed in MD simulations. The average Ser25-Pro106 distances in the simulations are 6.4 Å, 10.5 Å and 11.3 Å respectively for TA, Dex and MF. To link differential contacts with ligand-induced differences in receptor motion, we used MD to calculate the root mean square fluctuations between the key LBD-ligand complexes over 500 ns simulations. To focus only on ligand-induced dynamic changes, we first replaced SHP with TIF2 to mitigate peptide-driven effects. The MD trajectories showed less motion at residues 101-109 (Helix 6-7) in the TA complex compared to both Dex and MF, suggesting TA introduces a stabilizing effect (**Figure 4.5c,d**).

To obtain the structure most representative of the dominant conformation sampled in each trajectory, clustering was performed with the MMTSB tool. For each AncGR2 complex, the structure with the lowest RMSD in the most populated cluster was used as a

representative for the complex. Comparison of these structures shows a rearrangement of the residues in the binding pocket of the TA bound structure as a result of stabilizing, hydrophobic interactions between the acetonide group and Met29 and Met108 (**Figure 4.5e**). The interaction causes a shift of other residues (e.g. Met103) and shifts in the overall positions of H6 and H7, bringing them closer (**Figure 4.5f,g**). This movement is responsible for the closer distances observed between Ser25 and Pro106 in the TA-bound structure. Similar effects are observed in AncGR2 LBD-TA-SHP and LBD-TA-TIF2 structures, indicating that the TA ligand drives this effect.

Recognition of the SHP NR Box 1 LXXLL Motif

Coregulator proteins interact with SRs via an α-helix containing a short LXXLL motif (L- leucine, X- any amino acid)[47,48]. These coregulators bind to the same hydrophobic groove on the surface of SR LBDs, the AF-2[49,50]. The AF-2 is comprised of helices 3, 4, and 12 and is generally held in place by a charge clamp formed by a lysine on H3 and a glutamate on H12 that interacts with the helix dipole[44,50]. GR is known to interact with a wide variety of coregulator proteins, yet structural analysis has been limited to the typical coregulators of the p160 SRC family, specifically SRC-2 (TIF2, GRIP-1). Here, we show GR complexed with 11 residues of the SHP NR Box 1 peptide. This is the first GR LBD structure with this coregulator.

SHP is an atypical orphan nuclear receptor comprised only of an LBD that co-folds with the transcriptional corepressor EID1[51-53]. SHP has been shown to act as a potent corepressor for numerous nuclear receptors including ER, RXR, LRH-1, and GR[51,54-57]. To modulate receptor activity, SHP utilizes two canonical NR Box motifs to bind to the AF-2

surface, competing directly with coactivators[57]. SHP's repression of GR activity has been postulated to have important biological implications in the liver and pancreas[54].

The overall AncGR2-TA-SHP structure shows the AF-H in the active conformation and the SHP peptide bound at the AF surface (**Figure 4.6a**). The peptide inserts leucine side chains into the surface and is further stabilized by a charge clamp interaction, which is conserved across NRs. Lys48 on H3 hydrogen bonds with the backbone carbonyl of Leu24, and Glu224 hydrogen bonds to the free amide nitrogen of Ile20 (**Figure 4.6b**). The terminus of the peptide is stabilized by a hydrogen bond involving Lys48 interacting directly with the terminal Ser27 residue. The peptide is further stabilized by Met62, which makes van der Waals contacts with Tyr22. As in previous NR LBD-SHP structures, this interaction holds the tyrosine residue in the center of the peptide helix, which is further stabilized between the aromatic face and the Ser26 sidechain. The NR Box 1 peptide contains an arginine residue at the first position, but there was not strong enough density to model in this side chain; therefore it is modeled as an alanine residue. The SHP NR Box 1 and 2 peptides look most similar to the LXXLL motif seen in the coactivator PGC-1α (**Figure 4.6d**). It has been postulated that SHP antagonizes PGC-1α activation of GR. This repression of GR was shown to inhibit PEPCK, implying a role for SHP in modulating GR function within the liver[54].

The SHP peptide in our structure interacts with the same AF-2 surface seen in the previously solved AncGR2 LBD-TIF2 NR Box 3 structure (**Figure 4.6c**)[25]. Since SHP interacts with GR via a classical NR coactivator motif, similar interactions are expected. In both GR-peptide complexes, the charge clamp formed by Lys48 and Glu224 is conserved. Despite these peptides having sequence differences, both peptides stabilize AncGR2 LBD to similar levels (**Figure 4.6d,e**).

Discussion

In the 1940s, GR's endogenous ligand, cortisol, was identified as a potent suppressor of inflammation[58]. Since then, synthetic GCs with dramatically improved affinity, potency and selectivity have become the most widely used treatment for anti-inflammatory therapies[17]. Targeting GR is not challenging, however, current GCs drive the activation of metabolic, homeostatic and growth pathways in addition to immunosuppression[15,16,18]. X-ray crystal structures are required to build robust structure-function relationships; however, this has proved difficult due to the instability of recombinantly expressed, purified human GR LBD (hGR). To circumvent this issue, we utilized the AncGR2 LBD derived from the phylogenetically reconstructed GR present in the ancestor of bony vertebrates[24-26]. This ancestral receptor has been extensively used to understand the evolution of corticoid selectivity in modern GR; it displays the same ligand selectivity and agonist response as the human receptor but shows enhanced expression, solubility, crystallizability, and tolerance to mutation[24,25,27]. Therefore, AncGR2 represents a powerful tool to explore novel GR-ligand complexes that would otherwise be difficult to probe[24-26].

The synthetic glucocorticoid TA is used in the clinic to treat allergic rhinitis, macular edema, and skin lesions[19,59]. TA is more hydrophobic than other synthetic GCs, making it ideal for topical use and allowing for a prolonged duration of action[18,19]. Furthermore, TA has gained popularity in the laboratory due to its increased affinity for GR over other synthetic compounds (**Figure 4.3c**) such as dexamethasone[46,59], and for its ability to promote the expression and solubility of the intact receptor. These improvements in affinity, potency, and bioavailabity over other GCs makes TA a critical ligand to study. We show that TA significantly increases the thermal stability of the LBD relative to dexamethasone, (**Figure**

4.3d) causing GR to assume a more compact structure with smaller conformational fluctuations near the pocket (**Figure 4.5**). Driving these effects with TA is the bulky hydrophobic C17 acetonide moiety, which generates additional contacts relative to other ligands and repositions the H6-H7 loop to constrict the pocket (**Figure 4.5d-e**). The bound TA maintains interactions with the hydrogen bond network critical for GR specificity and transactivation (**Figure 4.1d, 2, 4a**), but enables greater intramolecular contacts, which likely explains the increased affinity and stability (**Figure 4.5**). Ligand-driven perturbation of intramolecular contacts and dynamics at the mouth of the ligand binding pocket has been shown to be essential for selective modulation of other nuclear receptors such as ERα, PPARγ, and LRH-1[60-63]. Taken together, we have elucidated the structural mechanisms driving TA's enhanced affinity and ability to stabilize the GR LBD, which will inform future glucocorticoid design.

This work also offers the first visualization of GR's interaction with the atypical NR, SHP, which is part of a transcriptional corepressor complex that targets NRs in the active conformation[53,54,64,65]. SHP accomplishes this by utilizing two LXXLL motifs to mimic coactivators and bind directly to the AF surface. SHP is expressed at high levels in the liver and has been shown to interact directly with GR to modulate its function[54]. We demonstrate that indeed, SHP binds to the AF surface on GR and makes the conserved charge clamp interaction, similar to the previously solved structure of AncGR2 in complex with the coregulator TIF2 (**Figure 4.6b**). Furthermore, the GR LBD–TA complex was stabilized to similar levels whether in complex with SHP or TIF2 (**Figure 4.6e**). SHP has been suggested to play a role in modulating hepatic GR function and thereby metabolism[54]. It is likely that increased levels of SHP permit direct competition with coregulators for the GR-agonist-DNA

complex[53,64,64,65]. However, further studies are needed to test this mechanism and to determine its physiological and clinical relevance for hepatic GR action.

Developing improved GCs with less off target side effects, will require linking ligand-induced receptor motions with selective coregulator interactions to drive GR to specific promoters[65,66]. Obtaining structural and dynamic information, as presented here, is vital to this effort.

Acknowledgements

X-ray data were collected at Southeast Regional Collaborative Access Team (SER-CAT) 22-ID beamline at the Advanced Photon Source, Argonne National Laboratory. Supporting institutions may be found at www.ser-cat.org/members/html. Use of the Advanced Photon source was supported by the U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences, under Contract No. W-31-109-Eng-38.

Figures

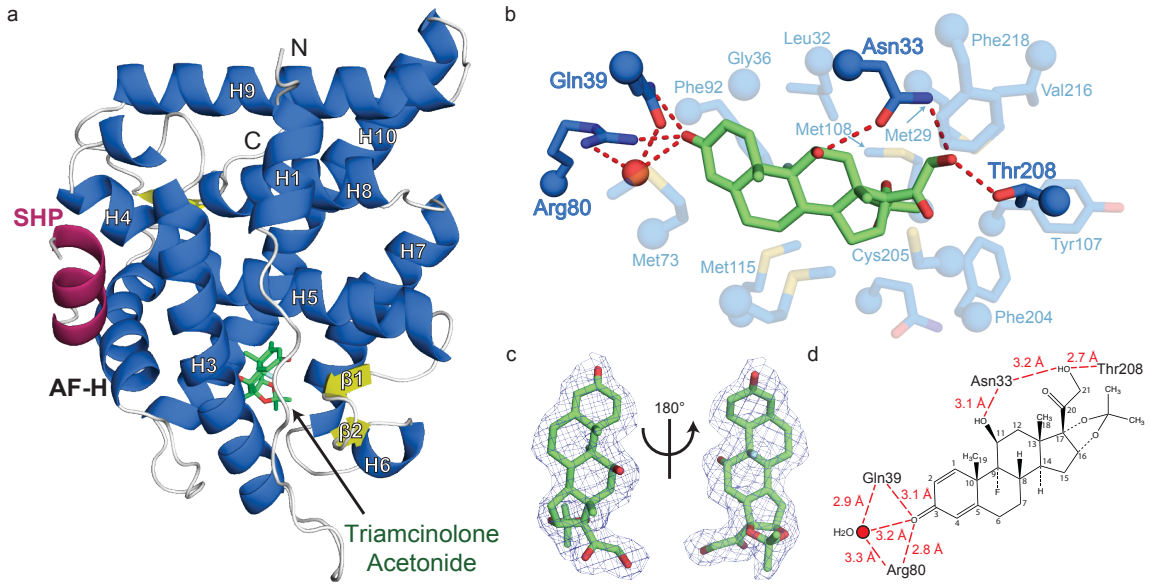


Figure 4.1: Structure of the AncGR2-TA-SHP NR Box 1 Complex.

(a) Cartoon representation of the overall structure of AncGR2 LBD (blue) in complex with TA (Green) and SHP (Pink) (b) Close-up view of ligand binding pocket (side chains are shown as sticks with α-carbons shown as spheres) Residues that participate in hydrophobic contacts are shown in faded blue and side chains that make hydrogen bonding interactions are shown darker blue with Hydrogen bonds denoted by red dashes. (c) Omit map (F_o-F_c) contoured at 2.5 σ around the ligand. (d) Schematic of hydrogen bonding network. Hydrogen bonds are shown in red.

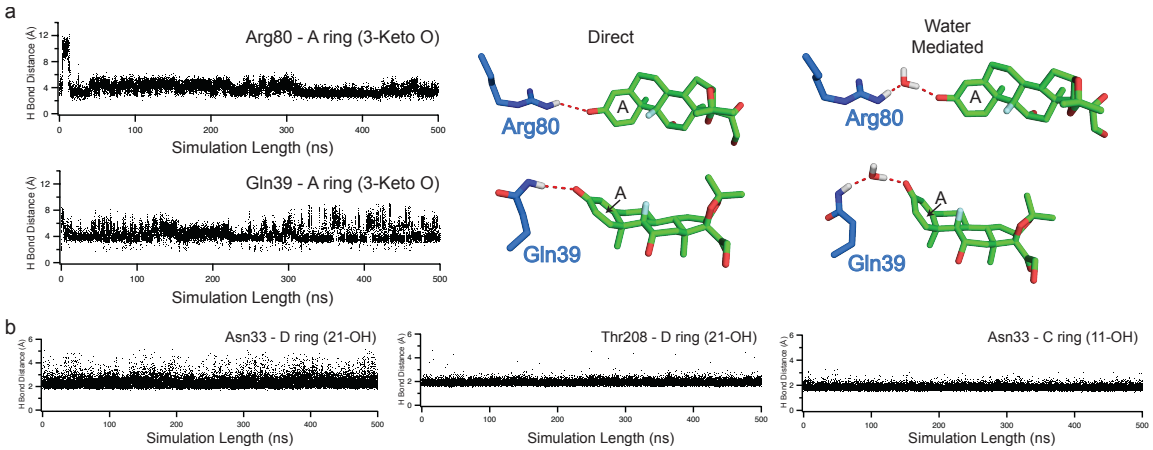


Figure 4.2: MD Simulations Support LBP-Hydrogen Bonding Interactions.

(a) Graph of hydrogen bonding distances between the guanidinium hydrogen atom on Arg80 and the amide hydrogen atom on Gln39 with the 3-keto oxygen on the A ring of TA during the 500 ns MD simulation. These hydrogen atoms either make contacts with the water molecule within the pocket or make direct interactions with the ligand. The A ring of TA is labeled in each view of the ligand (b) MD analysis supports the hydrogen bonding interactions seen between Asn33 and Thr308 with the ligand.

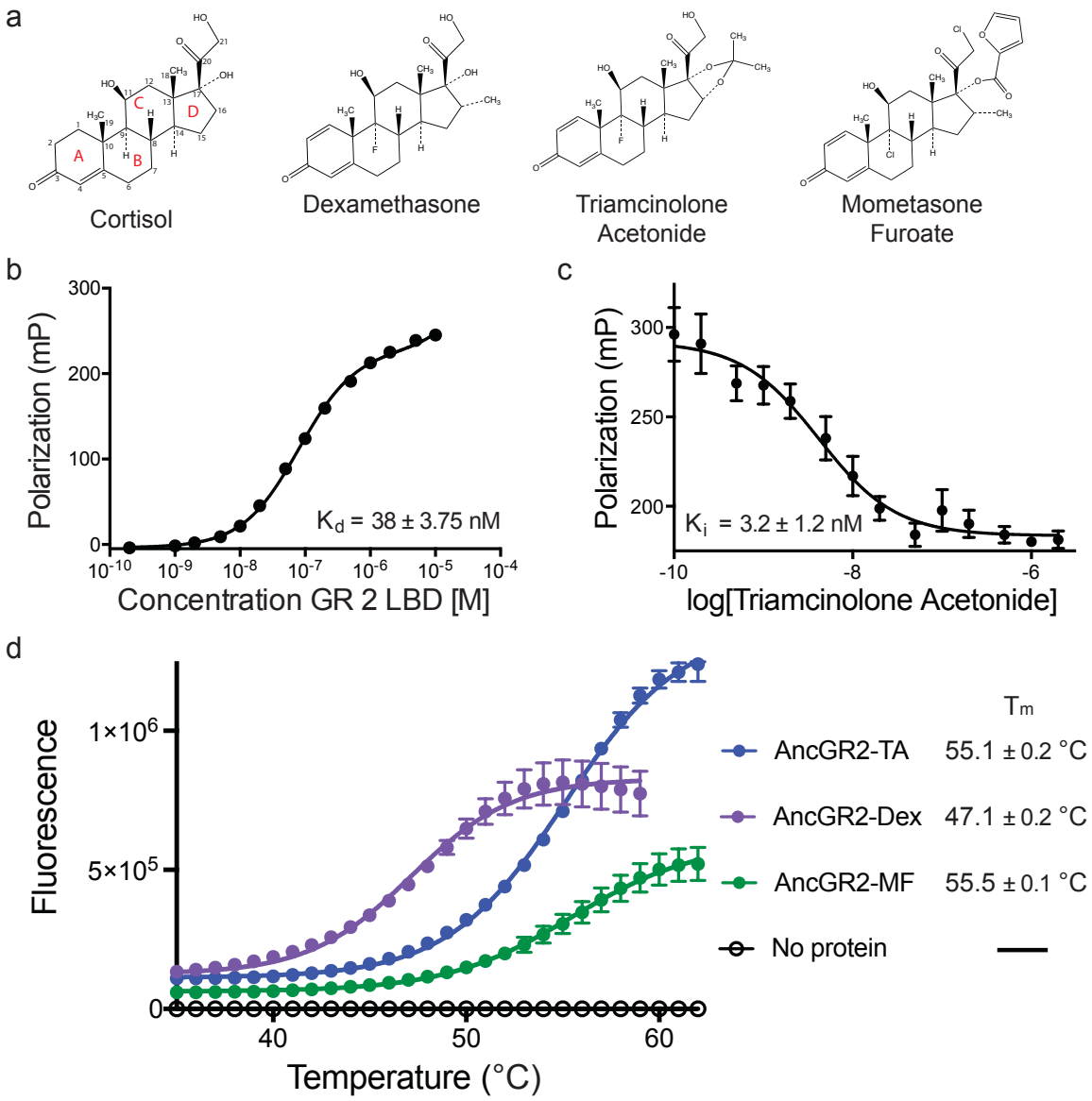


Figure 4.3: TA readily competes Dex out of the binding pocket and highly stabilizes the LBD.

(a) GR's endogenous ligand cortisol and a myriad of synthetic glucocorticoids including Dex, TA, and MF. (b) Fluorescent Dex (FL-Dex) binds AncGR2 LBD with a K_d of 38 nM. Binding was measured via fluorescence polarization and graphs are fit using a one site binding curve to calculate K_d values, error bars indicate standard error of the mean (SEM;

n=9). (c) TA competes FL-Dex out of the ligand binding pocket with a K_i of 3.2 nM. Graphs are fit using a one site, fit K_i equation to calculate K_i values, error bars indicate SEM (n=9). (d) DSF monitors the thermal stability of different AncGR2 ligand complexes. Both TA and MF-bound LBD increased the thermal stability by 8 °C. Graphs are fit using the Boltzmann sigmoidal equation, which calculates T_m as 50 % unfolding. Error bars error bars indicate SEM (n=6).

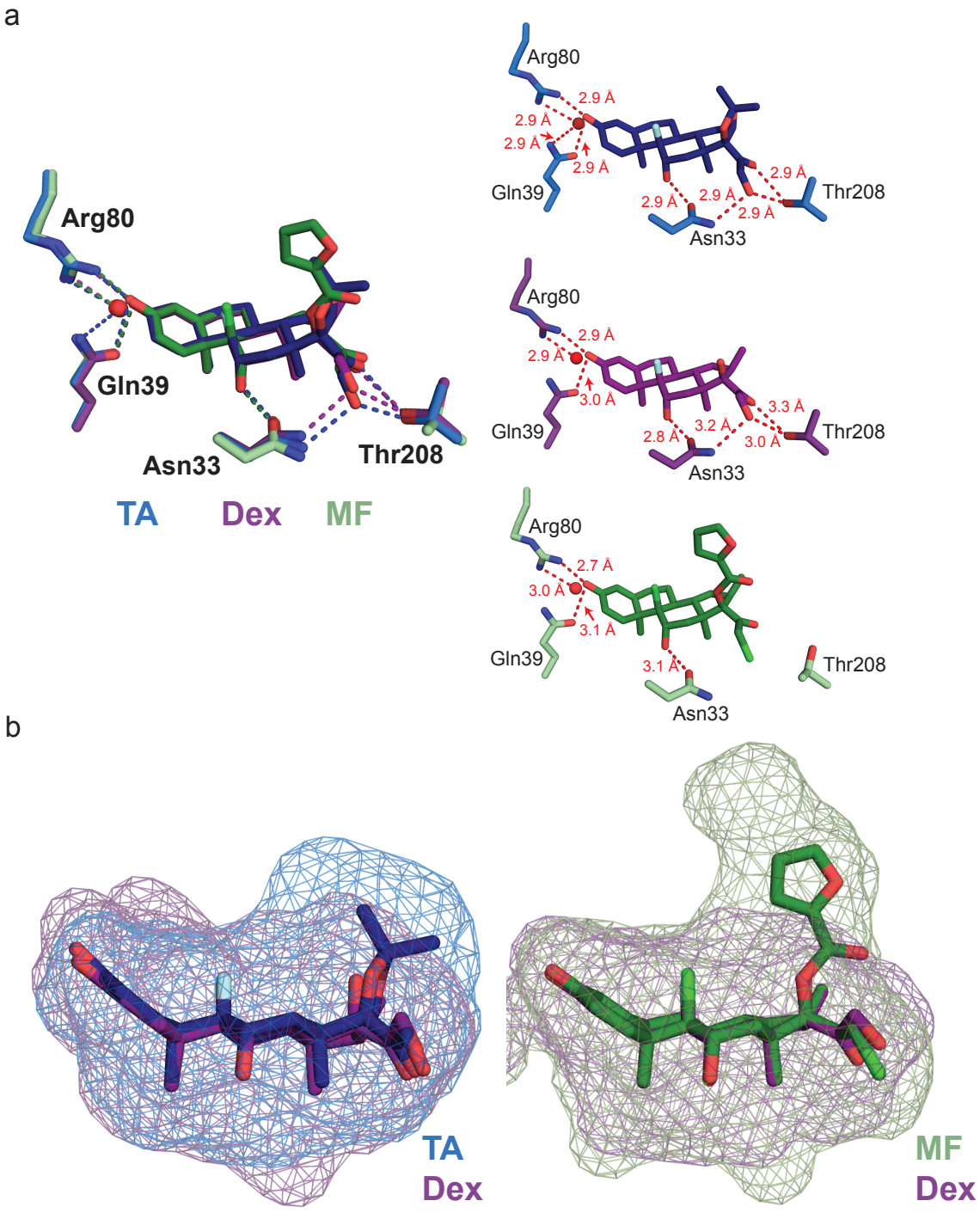


Figure 4.4: Comparison of Glucocorticoid Ligands bound to AncGR2 LBD.

(a) Overlay of TA (blue), Dex (purple), MF (green) within the pocket. Insets to the right show TA (blue), Dex (purple) and MF (green) hydrogen bonding networks. Hydrogen bonds

are shown in red. (b) Mesh depicting ligand pocket volume induced by different glucocorticoid ligands. TA and MF have large additions at the C-17 position, which the GR LBP expands to accommodate.

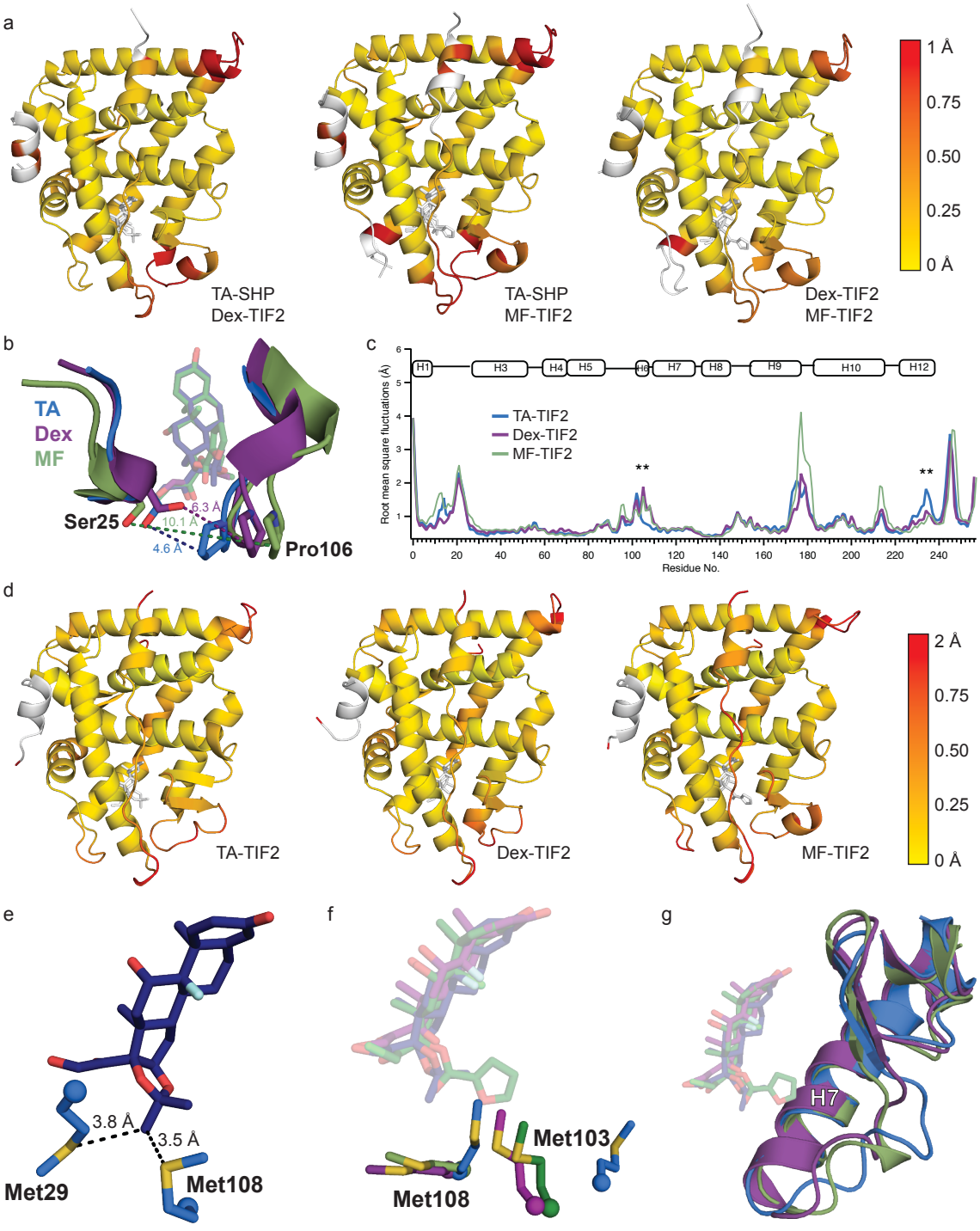


Figure 4.5: Structural comparison of different GR-ligand complexes.

(a) ProSMART analysis of pairwise comparisons of different structures of GR-ligand complexes. Areas shown in white were not used in the comparison. Structures are colored by Procrustes score of the central residue of an aligned fragment pair according to the legend shown on the right. The following comparisons were made: AncGR2 LBD-TA-SHP *versus* AncGR2 LBD-Dex-TIF2 (PDB 3GN8); AncGR2 LBD-TA-SHP *versus* AncGR2 LBD-MF-TIF2 (PDB 4E2J); AncGR2 LBD-Dex-TIF2 *versus* AncGR2 LBD-MF-TIF2. (b) Structural overlay of the mouth of the ligand binding pocket, which was one of the areas with the largest variance from the analysis done in (a). TA induces a narrowing of the pocket entrance, indicated by a 4.6 Å distance between Ser25 on H3 and Pro106 on H6. Dex has a 6.3 Å and MF has a 10.1 Å distance across the same area. (c) 500 ns molecular dynamics simulations looking at RMSFs between various GR-ligand complexes. ** indicate areas of significant differences in RMSF that are not obvious due to flexible loops, where higher RMSF values are to be expected. (d) The AncGR2 LBD-TA structure shows the acetonide group on TA makes hydrophobic contacts (black) with Met29 and Met108 (e) The hydrophobic contacts seen in d causes a shift in the position of other side chains within the ligand binding pocket, such as Met103. Structural overlays of AncGR2-ligand complexes show this is unique to the TA bound complex. (f) The shift in side chains causes an overall rearrangement of H6/H7, explaining the constricted pocket entrance seen in the TA structure.

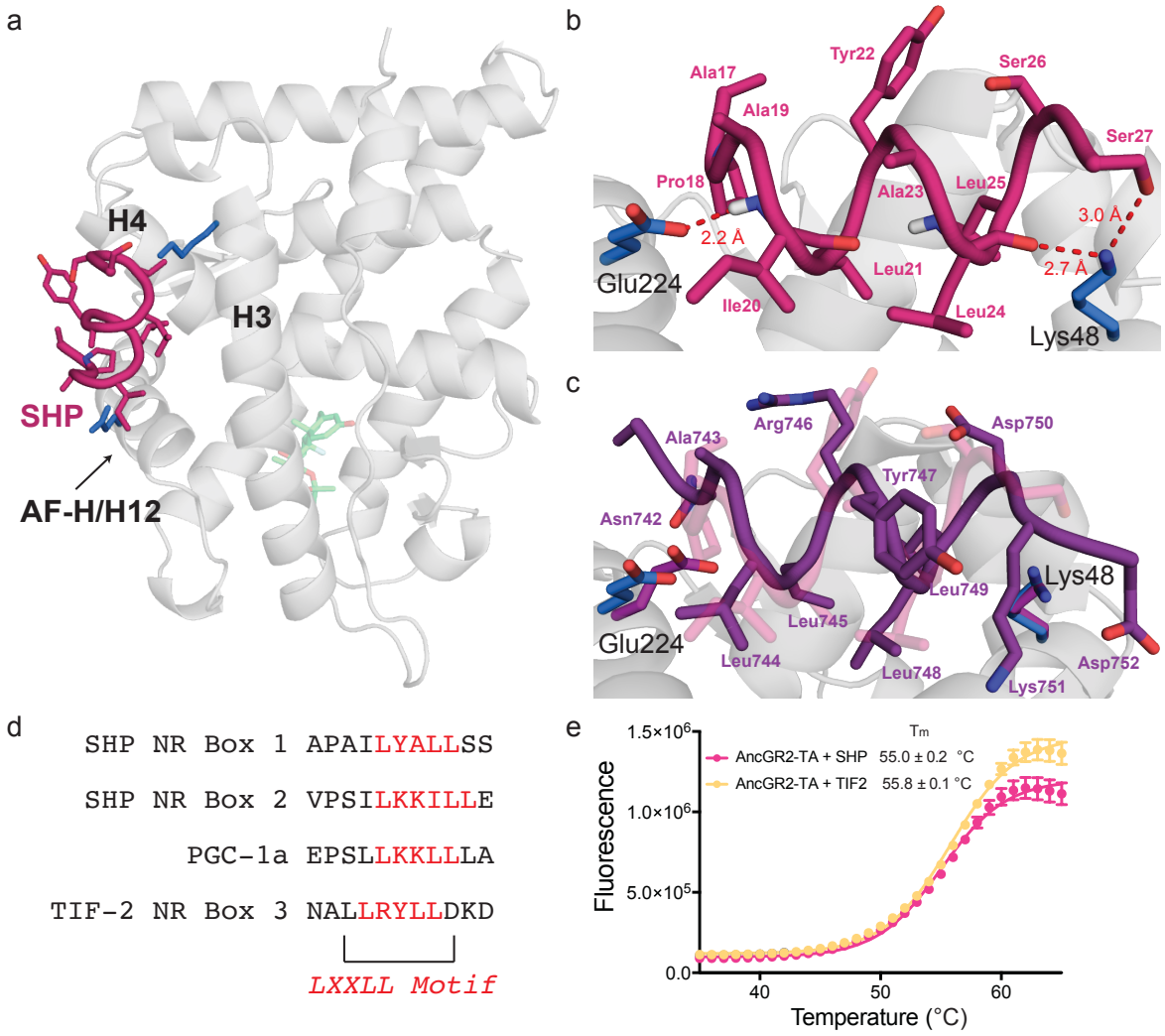


Figure 4.6: AncGR2-TA complex is bound by atypical coregulator, SHP.

(a) Overall structure of AncGR2 with SHP represented as sticks and ribbon (pink). The peptide is bound in the canonical AF-2 surface made by H3, H4, and the AF-H/H12. The peptide is held in place by charge clamp formed by Glu224 and Lys48, shown in sticks (blue). (b) Zoom in of SHP peptide. (c) Overlay of SHP (faded pink) with TIF2 peptide (purple) from the AncGR2-Dex (PDB 3GN8) structure. (d) Comparison of LXXLL motifs found in various coregulators. (e) DSF monitors the thermal stability AncGR2 LBD-TA in complex with the SHP NR Box 1 and TIF-2 NR Box 3 peptide. There is little difference between the two complexes.

Table 1: Data Collection and Refinement Statistics

	AncGR2 LBD-TA + SHP
Data Collection	
Space Group	C2
Cell Dimension	a=87.0, b=52.8, c=126.0
Resolution (Å)	2.11 (2.19-2.11)
R_{pim}	6.2 (37.0)
I/σ	2.67
Completeness	98.6 (93.2)
Redundancy	3.6 (3.0)
Refinement	
Resolution	2.11
No. Reflections	31942
R_{work}/R_{free}	21.5/25.2
No. Atoms	
Protein	4158
Ligands	62
Water	119
B-factors	
Protein	44.9
Ligands	31.8
Water	41.0
R.m.s. deviations	
Bond lengths (Å)	0.002
Bond angles (°)	0.52
PDB Code	5UFS

*Data collected from a single crystal; values in parentheses are for the highest-resolution shell.

Table 4.1: Data Collection and Refinement Statistics

Data collection and refinement statistics for the AncGR2 LBD in complex with TA and the SHP coregulator peptide.

References

1. Nagy, L. and J. W. Schwabe (2004). "Mechanism of the nuclear receptor molecular switch." Trends Biochem Sci **29**(6): 317-324.
2. Kadmiel, M. and J. A. Cidlowski (2013). "Glucocorticoid receptor signaling in health and disease." Trends Pharmacol Sci **34**(9): 518-530.
3. Carson-Jurica, M. A., W. T. Schrader and B. W. O'Malley (1990). "Steroid receptor family: structure and functions." Endocr Rev **11**(2): 201-220.
4. Gronemeyer, H. and D. Moras (1995). "Nuclear receptors. How to finger DNA." Nature **375**(6528): 190-191.
5. Kumar, R. and E. B. Thompson (1999). "The structure of the nuclear hormone receptors." Steroids **64**(5): 310-319.
6. O'Malley, B. W. and M. J. Tsai (1992). "Molecular pathways of steroid receptor action." Biol Reprod **46**(2): 163-167.
7. Picard, D., B. Khursheed, M. J. Garabedian, M. G. Fortin, S. Lindquist and K. R. Yamamoto (1990). "Reduced levels of hsp90 compromise steroid receptor action in vivo." Nature **348**(6297): 166-168.
8. Darimont, B. D., R. L. Wagner, J. W. Apriletti, M. R. Stallcup, P. J. Kushner, J. D. Baxter, R. J. Fletterick and K. R. Yamamoto (1998). "Structure and specificity of nuclear receptor-coactivator interactions." Genes Dev **12**(21): 3343-3356.
9. Hu, X. and M. A. Lazar (1999). "The CoRNR motif controls the recruitment of corepressors by nuclear hormone receptors." Nature **402**(6757): 93-96.
10. Hudson, W. H., C. Youn and E. A. Ortlund (2013). "The structural basis of direct glucocorticoid-mediated transrepression." Nat Struct Mol Biol **20**(1): 53-58.

11. Luisi, B. F., W. X. Xu, Z. Otwinowski, L. P. Freedman, K. R. Yamamoto and P. B. Sigler (1991). "Crystallographic analysis of the interaction of the glucocorticoid receptor with DNA." Nature **352**(6335): 497-505.

12. Mordacq, J. C. and D. I. Linzer (1989). "Co-localization of elements required for phorbol ester stimulation and glucocorticoid repression of proliferin gene expression." Genes Dev **3**(6): 760-769.

13. Schule, R., P. Rangarajan, S. Kliewer, L. J. Ransone, J. Bolado, N. Yang, I. M. Verma and R. M. Evans (1990). "Functional antagonism between oncoprotein c-Jun and the glucocorticoid receptor." Cell **62**(6): 1217-1226.

14. Yang-Yen, H. F., J. C. Chambard, Y. L. Sun, T. Smeal, T. J. Schmidt, J. Drouin and M. Karin (1990). "Transcriptional interference between c-Jun and the glucocorticoid receptor: mutual inhibition of DNA binding due to direct protein-protein interaction." Cell **62**(6): 1205-1215.

15. De Bosscher, K., W. Vanden Berghe and G. Haegeman (2000). "Mechanisms of anti-inflammatory action and of immunosuppression by glucocorticoids: negative interference of activated glucocorticoid receptor with transcription factors." J Neuroimmunol **109**(1): 16-22.

16. Cato, A. C. and E. Wade (1996). "Molecular mechanisms of anti-inflammatory action of glucocorticoids." Bioessays **18**(5): 371-378.

17. Clark, A. R. and M. G. Belvisi (2012). "Maps and legends: the quest for dissociated ligands of the glucocorticoid receptor." Pharmacol Ther **134**(1): 54-67.

18. Schacke, H., W. D. Docke and K. Asadullah (2002). "Mechanisms involved in the side effects of glucocorticoids." Pharmacol Ther **96**(1): 23-43.

19. Fredman, R. and M. Tenenhaus (2013). "Cushing's syndrome after intralesional triamcinolone acetonide: a systematic review of the literature and multinational survey." Burns **39**(4): 549-557.

20. Bain, D. L., Q. Yang, K. D. Connaghan, J. P. Robblee, M. T. Miura, G. D. Degala, J. R. Lambert and N. K. Maluf (2012). "Glucocorticoid receptor-DNA interactions: binding energetics are the primary determinant of sequence-specific transcriptional activity." J Mol Biol **422**(1): 18-32.

21. Derendorf, H. and E. O. Meltzer (2008). "Molecular and clinical pharmacology of intranasal corticosteroids: clinical and therapeutic implications." Allergy **63**(10): 1292-1300.

22. Robblee, J. P., M. T. Miura and D. L. Bain (2012). "Glucocorticoid receptor-promoter interactions: energetic dissection suggests a framework for the specificity of steroid receptor-mediated gene regulation." Biochemistry **51**(22): 4463-4472.

23. Gebhardt, J. C., D. M. Suter, R. Roy, Z. W. Zhao, A. R. Chapman, S. Basu, T. Maniatis and X. S. Xie (2013). "Single-molecule imaging of transcription factor binding to DNA in live mammalian cells." Nat Methods **10**(5): 421-426.

24. Bridgham, J. T., E. A. Ortlund and J. W. Thornton (2009). "An epistatic ratchet constrains the direction of glucocorticoid receptor evolution." Nature **461**(7263): 515-519.

25. Kohn, J. A., K. Deshpande and E. A. Ortlund (2012). "Deciphering modern glucocorticoid cross-pharmacology using ancestral corticosteroid receptors." J Biol Chem **287**(20): 16267-16275.

26. Colucci, J. K. and E. A. Ortlund (2013). "X-ray crystal structure of the ancestral 3-ketosteroid receptor-progesterone-mifepristone complex shows mifepristone bound at the coactivator binding interface." PLoS One **8**(11): e80761.

27. Ortlund, E. A., J. T. Bridgham, M. R. Redinbo and J. W. Thornton (2007). "Crystal structure of an ancient protein: evolution by conformational epistasis." Science **317**(5844): 1544-1548.

28. Otwinowski, Z. and W. Minor (1997). "Processing of X-ray diffraction data collected in oscillation mode." Methods Enzymol **276**: 307-326.

29. Adams, P. D., P. V. Afonine, G. Bunkoczi, V. B. Chen, I. W. Davis, N. Echols, J. J. Headd, L. W. Hung, G. J. Kapral, R. W. Grosse-Kunstleve, A. J. McCoy, N. W. Moriarty, R. Oeffner, R. J. Read, D. C. Richardson, J. S. Richardson, T. C. Terwilliger and P. H. Zwart (2010). "PHENIX: a comprehensive Python-based system for macromolecular structure solution." Acta Crystallogr D Biol Crystallogr **66**(Pt 2): 213-221.

30. Emsley, P., B. Lohkamp, W. G. Scott and K. Cowtan (2010). "Features and development of Coot." Acta Crystallogr D Biol Crystallogr **66**(Pt 4): 486-501.

31. Davis, I. W., A. Leaver-Fay, V. B. Chen, J. N. Block, G. J. Kapral, X. Wang, L. W. Murray, W. B. Arendall, 3rd, J. Snoeyink, J. S. Richardson and D. C. Richardson (2007). "MolProbity: all-atom contacts and structure validation for proteins and nucleic acids." Nucleic Acids Res **35**(Web Server issue): W375-383.

32. Nicholls, R. A., M. Fischer, S. McNicholas and G. N. Murshudov (2014). "Conformation-independent structural comparison of macromolecules with ProSMART." Acta Crystallogr D Biol Crystallogr **70**(Pt 9): 2487-2499.

33. Jorgensen, W. L., J. Chandrasekhar, J. D. Madura, R. W. Impey and M. L. Klein (1983). "Comparison of simple potential functions for simulating liquid water." The J of Chem phy **79**(2): 926-935.

34. Case, D., V. Babin, J. Berryman, R. Betz, Q. Cai, D. Cerutti, T. Cheatham Iii, T. Darden, R. Duke and H. Gohlke (2014). "Amber 14."Pérez, A., I. Marchán, D. Svozil, J. Sponer, T. E. Cheatham, C. A. Laughton and M. Orozco (2007). "Refinement of the AMBER force field for nucleic acids: improving the description of α/γ conformers." Biophysical journal **92**(11): 3817-3829.

35. Pérez, A., I. Marchán, D. Svozil, J. Sponer, T. E. Cheatham, C. A. Laughton and M. Orozco (2007). "Refinement of the AMBER force field for nucleic acids: improving the description of α/γ conformers." Biophysical journal **92**(11): 3817-3829.

36. Wang, H., X. Wang, X. Xu, M. Kyba and A. J. Cooney (2016). "Germ Cell Nuclear Factor (GCNF) Represses Oct4 Expression and Globally Modulates Gene Expression in Human Embryonic Stem (hES) Cells." J Biol Chem **291**(16): 8644-8652.

37. Case, D. and P. Kollman (2012). "AmberTools 12." University of California, San Francisco.

38. Ryckaert, J.-P., G. Ciccotti and H. J. Berendsen (1977). "Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes." Journal of Computational Physics **23**(3): 327-341.

39. Roe, D. R. and T. E. Cheatham III (2013). "PTRAJ and CPPTRAJ: software for processing and analysis of molecular dynamics trajectory data." J of Chem Theory and Comp **9**(7): 3084-3095.

40. Thiruchelvam, P. T., C. F. Lai, H. Hua, R. S. Thomas, A. Hurtado, W. Hudson, A. R. Bayly, F. J. Kyle, M. Periyasamy, A. Photiou, A. C. Spivey, E. A. Ortlund, R. J. Whitby, J. S. Carroll, R. C. Coombes, L. Buluwela and S. Ali (2011). "The liver receptor homolog-1 regulates estrogen receptor expression in breast cancer cells." Breast Cancer Res Treat **127**(2): 385-396.

41. Feig, M., J. Karanicolas and C. L. Brooks (2004). "MMTSB Tool Set: enhanced sampling and multiscale modeling methods for applications in structural biology." J Mol Graphs and Model **22**(5): 377-395.

42. Bikowski, J., R. Pillai and B. Shroot (2006). "The position not the presence of the halogen in corticosteroids influences potency and side effects." J Drugs Dermatol **5**(2): 125-130.

43. Grossmann, C., T. Scholz, M. Rochel, C. Bumke-Vogt, W. Oelkers, A. F. Pfeiffer, S. Diederich and V. Bahr (2004). "Transactivation via the human glucocorticoid and mineralocorticoid receptor by therapeutically used steroids in CV-1 cells: a comparison of their glucocorticoid and mineralocorticoid properties." Eur J Endocrinol **151**(3): 397-406.

44. Nehme, A., E. K. Lobenhofer, W. D. Stamer and J. L. Edelman (2009). "Glucocorticoids with different chemical structures but similar glucocorticoid receptor potency regulate subsets of common and unique genes in human trabecular meshwork cells." BMC Med Genomics **2**: 58.

45. Bledsoe, R. K., V. G. Montana, T. B. Stanley, C. J. Delves, C. J. Apolito, D. D. McKee, T. G. Consler, D. J. Parks, E. L. Stewart, T. M. Willson, M. H. Lambert, J. T. Moore, K. H. Pearce and H. E. Xu (2002). "Crystal structure of the glucocorticoid

receptor ligand binding domain reveals a novel mode of receptor dimerization and coactivator recognition." Cell **110**(1): 93-105.

46. Yoneda, Y., D. Han, K. Ogita and A. Watanabe (1995). "Distinction between binding of [3H]triamcinolone acetonide to a ligand binding domain on the glucocorticoid receptor complex in cytosol fractions of brain and liver from the rat with intact adrenals." Brain Res **685**(1-2): 105-116.
47. Jenkins, B. D., C. B. Pullen and B. D. Darimont (2001). "Novel glucocorticoid receptor coactivator effector mechanisms." Trends Endocrinol Metab **12**(3): 122-126.
48. Heery, D. M., E. Kalkhoven, S. Hoare and M. G. Parker (1997). "A signature motif in transcriptional co-activators mediates binding to nuclear receptors." Nature **387**(6634): 733-736.
49. Vandevyver, S., L. Dejager and C. Libert (2014). "Comprehensive overview of the structure and regulation of the glucocorticoid receptor." Endocr Rev **35**(4): 671-693.
50. Lee, Y. S., D. Chanda, J. Sim, Y. Y. Park and H. S. Choi (2007). "Structure and function of the atypical orphan nuclear receptor small heterodimer partner." Int Rev Cytol **261**: 117-158.
51. Seol, W., H. S. Choi and D. D. Moore (1996). "An orphan nuclear hormone receptor that lacks a DNA binding domain and heterodimerizes with other receptors." Science **272**(5266): 1336-1339.
52. Zhi, X., X. E. Zhou, Y. He, C. Zechner, K. M. Suino-Powell, S. A. Kliewer, K. Melcher, D. J. Mangelsdorf and H. E. Xu (2014). "Structural insights into gene repression by the orphan nuclear receptor SHP." Proc Natl Acad Sci U S A **111**(2): 839-844.

53. Borgius, L. J., K. R. Steffensen, J. A. Gustafsson and E. Treuter (2002). "Glucocorticoid signaling is perturbed by the atypical orphan receptor and corepressor SHP." J Biol Chem **277**(51): 49761-49766.

54. Johansson, L., A. Bavner, J. S. Thomsen, M. Farnegardh, J. A. Gustafsson and E. Treuter (2000). "The orphan nuclear receptor SHP utilizes conserved LXXLL-related motifs for interactions with ligand-activated estrogen receptors." Mol Cell Biol **20**(4): 1124-1133.

55. Johansson, L., J. S. Thomsen, A. E. Damdimopoulos, G. Spyrou, J. A. Gustafsson and E. Treuter (1999). "The orphan nuclear receptor SHP inhibits agonist-dependent transcriptional activity of estrogen receptors ERalpha and ERbeta." J Biol Chem **274**(1): 345-353.

56. Ortlund, E. A., Y. Lee, I. H. Solomon, J. M. Hager, R. Safi, Y. Choi, Z. Guan, A. Tripathy, C. R. Raetz, D. P. McDonnell, D. D. Moore and M. R. Redinbo (2005). "Modulation of human nuclear receptor LRH-1 activity by phospholipids and SHP." Nat Struct Mol Biol **12**(4): 357-363.

57. Lee, Y. K., H. Dell, D. H. Dowhan, M. Hadzopoulou-Cladaras and D. D. Moore (2000). "The orphan nuclear receptor SHP inhibits hepatocyte nuclear factor 4 and retinoid X receptor transactivation: two mechanisms for repression." Mol Cell Biol **20**(1): 187-195.

58. Kendall, E. C. (1951). "The development of cortisone as a therapeutic agent." Antibiot Chemother (Northfield) **1**(1): 7-15.

59. Gee, A. C. and J. A. Katzenellenbogen (2001). "Probing conformational changes in the estrogen receptor: evidence for a partially unfolded intermediate facilitating ligand binding and release." Mol Endocrinol **15**(3): 421-428.

60. Kojetin, D. J. and T. P. Burris (2013). "Small molecule modulation of nuclear receptor conformational dynamics: implications for function and drug discovery." Mol Pharmacol **83**(1): 1-8.

61. Musille, P. M., M. C. Pathak, J. L. Lauer, W. H. Hudson, P. R. Griffin and E. A. Ortlund (2012). "Antidiabetic phospholipid-nuclear receptor complex reveals the mechanism for phospholipid-driven gene regulation." Nat Struct Mol Biol **19**(5): 532-537, S531-532.

62. Srinivasan, S., J. C. Nwachukwu, A. A. Parent, V. Cavett, J. Nowak, T. S. Hughes, D. J. Kojetin, J. A. Katzenellenbogen and K. W. Nettles (2013). "Ligand-binding dynamics rewire cellular signaling via estrogen receptor-alpha." Nat Chem Biol **9**(5): 326-332.

63. Bavner, A., L. Johansson, G. Toresson, J. A. Gustafsson and E. Treuter (2002). "A transcriptional inhibitor targeted by the atypical orphan nuclear receptor SHP." EMBO Rep **3**(5): 478-484.

64. Macchiarulo, A., G. Rizzo, G. Costantino, S. Fiorucci and R. Pellicciari (2006). "Unveiling hidden features of orphan nuclear receptors: the case of the small heterodimer partner (SHP)." J Mol Graph Model **24**(5): 362-372.

65. Sacta, M. A., Y. Chinenov and I. Rogatsky (2016). "Glucocorticoid Signaling: An Update from a Genomic Perspective." Annu Rev Physiol **78**: 155-180.

66. Weikum, E. R., M. T. Knuesel, E. A. Ortlund and K. R. Yamamoto (2017). "Glucocorticoid receptor control of transcription: precision and plasticity via allostery." Nat Rev Mol Cell Biol **18**(3): 159-174.

CHAPTER 5: A STRUCTURAL INVESTIGATION INTO *OCT4* REGULATION BY ORPHAN NUCLEAR RECEPTORS, GERM CELL NUCLEAR FACTOR (GCNF) AND LIVER RECEPTOR HOMOLOG-1 (LRH-1)

Emily R. Weikum[1], Micheal L. Tuntland[1], Michael N. Murphy[1], Eric A. Ortlund[14]

1) Department of Biochemistry, Emory University School of Medicine, Atlanta GA 30322 USA

Weikum, ER, Tuntland ML, Murphy MN, Ortlund EA (2016) A structural investigation into *Oct4* regulation by orphan nuclear receptors, germ cell nuclear factor (GCNF) and liver receptor homolog-1 (LRH-1). Journal of Molecular Biology, Issue 24(4981-4992).

Two nuclear receptors are key regulators of *Oct4* expression during development. In order to understand the structural basis for this regulation, it is crucial to solve x-ray crystal structures of each receptor in complex with their DNA recognition element. In this work, I crystallized GCNF bound to the *Oct4* direct repeat sequence. This structure was directly compared to the structure of LRH-1 bound to the same DNA sequence. This work allowed direct comparison of the sequence specific contacts that guide *Oct4* regulation. This work was accepted for publication in Journal Molecular Biology in November 2016.

---

[a] Conceived and designed the experiments: ERW MLT EAO. Performed the experiments: ERW MLT MNM EAO. Analyzed the data: ERW MLT EAO. Wrote the paper: ERW MLT EAO.

Abstract

Oct4 is a transcription factor required for maintaining pluripotency and self-renewal in stem cells. Prior to differentiation, *Oct4* must be silenced to allow for the development of the three germ layers in the developing embryo. This fine-tuning is controlled by the nuclear receptors, liver receptor homolog-1 and germ cell nuclear factor. Liver receptor homolog-1 is responsible for driving the expression of *Oct4* where germ cell nuclear factor represses its expression upon differentiation. Both receptors bind to a DR0 motif located within the *Oct4* promoter. Here, we present the first structure of mouse germ cell nuclear factor DNA binding domain in complex with the *Oct4* DR0. The overall structure revealed two molecules bound in a head-to-tail fashion on opposite sides of the DNA. Additionally, we solved the structure of the human liver receptor homolog-1 DNA binding domain bound to the same element. We explore the structural elements that govern *Oct4* recognition by these two nuclear receptors.

Introduction

The pluripotency of embryonic stem cells is maintained by a specific group of factors including leukemia inhibitor factor, and transcription factors Oct4, Sox2, and Nanog[1]. All of these proteins are critical for maintaining precise gene expression profiles in ES cells. Oct4, a member of the POU-domain family is widely recognized as the gatekeeper, preventing embryonic stem cell differentiation by maintaining pluripotent gene expression and inhibiting lineage-determining factors[2]. During this time, *Oct4* expression must be tightly regulated to ensure proper expression levels. Upon exposure to differentiation cues, such as the presence of retinoic acid, Oct4 and the other pluripotency factors are repressed in a temporal and spatial manner to ensure proper development of the three germ layers. Members of the nuclear receptor superfamily of ligand-regulated transcription factors are responsible for ensuring this specific pattern of *Oct4* expression[3,4].

Nuclear receptors (NR) play key roles in diverse biological processes, including maintaining homeostasis, metabolism, development, and many others[4,5]. These receptors all share the same core domain structure: a N-terminal transactivation domain (NTD), a DNA binding domain (DBD) with two highly conserved zinc fingers, a flexible hinge region, and a ligand binding domain (LBD) that contains an activator function-2 helix that is critical for binding coregulators **(**Fig. X.2A**)**. These receptors bind to palindromic DNA sequences as a monomer, homodimer, or heterodimer to regulate transcription from their target genes[6]. A number of these receptors including, liver receptor homolog-1 (LRH-1, *NR5A2*), germ cell nuclear factor (GCNF, *NR6A1*), steroidogenic factor (SF-1, *NR5A1*), and retinoic acid receptor (RAR, *NR1B1*) are known to regulate of *Oct4* expression by binding to response elements within the promoter regions[3,7,8,9].

LRH-1 regulates *Oct4* gene expression by binding DNA as a monomer to a 9 nucleotide recognition element comprised of a YCA followed by a NR half-site (AGGCCR) sequence (Y= pyrimidine, R= purine)[10]. LRH-1 contains a canonical NR DBD composed of two helical zinc fingers, which recognize the NR-halfsite. LRH-1, and all NR5A family members, contain a unique *fushi tarazu* factor 1 (Ftz-F1) domain located C-terminal to the DBD that has been shown to play a functionally important role as a protein interaction module required for the recruitment of other transcription factors[11-14]. This domain was structurally characterized in the LRH-1 DBD *CYP7A1* complex by X-ray crystallography then by NMR for SF-1[10,15], and assumes an α-helix fold that packs against body of the zinc-figure domain. Mutations that untether the Ftz-F1 helix from the body of the protein dramatically reduce both DNA binding and transactivation. LRH-1 makes structurally conserved contacts with its DNA half-site through both the canonical DNA-reading helix, located within the zinc finger domain, and the C-terminal extension (CTE), located between the zinc-finger domain and Ftz-F1 domain. The CTE-DNA interactions are influenced by the Ftz-F1 helix, which serves to orient the CTE in the DNA minor groove[10].

In early embryonic development, LRH-1 is highly expressed within the inner cell mass and primitive endoderm of the blastocyst, where other NRs are secluded to other cell types[3]. To maintain pluripotency, LRH-1 binds directly to DR0 sequences within the proximal promoter and proximal enhancer to drive expression of *Oct4*[16]. This regulation via LRH-1 is critical, as loss of LRH-1 results in embryonic lethality at day 6.5. LRH-1$^{-/-}$ embryos also exhibit low *Oct4* expression and die before liver development is established. Recently, it has even been shown that LRH-1 can replace Oct4 in the reprogramming of

somatic cells into pluripotent stem cells[17]. However, upon signals for differentiation, LRH-1 levels are dramatically reduced while GCNF is recruited to repress *Oct4*.

GCNF is an orphan nuclear receptor that was first identified from mouse heart tissue and shows high expression levels in developing germ cells, oocytes, and spermatogenic cells[18]. GCNF comprises its own unique NR superfamily subclass with a DBD that resembles retinoid X receptor (RXR) but a LBD more closely related to COUP transcription factor 2 (COUP-TF). Additionally, GCNF does not have the typical AF-2 helix within its LBD but instead contains a predicted β-sheet that interacts with the transcriptional corepressors nuclear receptor corepressor-1 (NCoR) and nuclear receptor corepressor-2 (NCoR2/SMRT)[7,19]. GCNF represses *Oct4* expression by binding directly to a DR0 element within the proximal promoter region, the same site as LRH-1[18-21]. This transrepression of *Oct4* is required for cell differentiation as loss of GCNF results in embryonic lethality at day E10.5 from cardiovascular complications and other severe developmental abnormalities[19,22]. DNA binding by GCNF is functionally critical as a deletion of the DBD phenocopies the $GCNF^{-/-}$ mice[23]. GCNF binds the *Oct4* DR0 element in the proximal promoter as a homodimer and recruits transcriptional corepressors and DNA methyltransferases in order to ultimately silence *Oct4* expression[24]. This promoter methylation is maintained well beyond GCNF expression in order to maintain *Oct4* silencing[25,26]. This process is depicted in **Figure 5.1**.

Here, we present the crystal structures of the mGCNF DBD-TA and hLRH-1 DBD-FtzF1 bound to the *mOct4* DR0 proximal promoter element. Throughout the rest of this manuscript mGCNF DBD-TA and hLRH-1 DBD-FtzF1 will be referred to as GCNF DBD and LRH-1 DBD, respectively. This work represents the first crystal structure of GCNF and

permitted the visualization of the sequence specific contacts that facilitate recognition of this element for both of these orphan NRs.

Results

GCNF and LRH-1 directly bind the mOct4 DR0

To characterize *in vitro* binding affinity and kinetics of GCNF and LRH-1 we monitored the ability of recombinant GCNF DBD, LRH-1 DBD, and full-length LRH-1 to bind a FAM labeled 16 bp *Oct4* DR0 fragment via fluorescence polarization (**Figure 5.2**). Intact, GCNF has been shown to bind the *Oct4* DR0 as a homodimer via electric mobility shift assay[18,19]. Here, GCNF bound the *Oct4* DR0 with a two-site binding mechanism with the K_d of the high affinity site at 170 nM and 2.2 μM for the low affinity site (**Figure 5.2c**). This apparent two-site binding, rather than cooperative binding, may be due to the lack of the LBD which presumably facilities dimerization[24]. The LRH-1 DBD displayed a one site binding mechanism with an apparent K_d of 60 nM (**Figure 5.2b**). The full-length LRH-1 construct also fit a one site binding curve with the 16 bp *Oct4* element with an apparent affinity of 30 nM. The data is summarized in a **Figure 5.2d**.

Structural analysis of GCNF and LRH-1 – Oct4 Complexes

To determine how these individual NRs recognize the same *Oct4* DR0, we solved crystal structures of each protein-DNA complex. The GCNF-*Oct4* 16 *bp* DR0 DNA complex crystallized in the $P2_12_12_1$ space group and data were collected to 2.1 Å with 96.2% completeness. The LRH-1-*Oct4* DR0 12 *bp* DNA complex crystallized in the space group $P4_3$ and data were collected to 2.2 Å with 99.8% completeness (**Table 5.1**).

The GCNF structure shows two molecules bound to opposite sides of the DNA in a head-to-tail fashion (**Figure 5.3a**). This orientation on DNA is very similar to the structure of the RXRα-DR1 complex (PDB ID: **1BY4**)[27]. The GCNF DBD adopts the canonical NR DBD

fold[6], and strong electron density allowed for modeling of the T and A box residues within the C-terminal extension (CTE) for molecule 2. One molecule, colored purple, positions the DNA reading helix into the major groove of the DNA at the first AGGTCA repeat (bases 106-111). This molecule makes three base-specific contacts (**Figure 5.3b**). The Lys96 side chain makes hydrogen bonds to the O6 and N7 positions on guanine 107. Arg101 make additional hydrogen bonds to the on guanine 97 at the N7 position on the other side of the DNA as well as to the backbone phosphate of a thymine 98. The second molecule, colored deep purple, makes similar contacts to the second AGGCTA sequence (bases 112-117) (**Figure 5.3b**). Here, Lys96 again makes hydrogen bonds to the O6 position of guanine 113. Arg101 makes hydrogen bonds to the N7 position on adenine 91 instead of a guanine and also makes additional hydrogen bonds the backbone of thymine 90. These interactions are supported by excellent electron density and additional side chains participating in backbone and water-mediated interactions are highlighted in **Figure 5.3c**.

The GCNF CTE, including the T/A box residues, was observed for molecule 2 (residues 141-156). There is strong electron density for these residues, which is shown in **Figure 5.5d**. The GCNF CTE dips into the minor groove at the TCAA sequence (bases 109-112) to make additional sequence specific contacts. Arg86 makes hydrogen bonds to the O2 positions on both thymine 109 and cytosine 110. Arg78 makes hydrogen bonds to the polypeptide backbone oxygen at Gly152, which may help lock the CTE into a more stable conformation for interaction with the DNA minor groove. In addition to DNA contacts, the CTE also makes a series of contacts with GNCF molecule 1 via a set of highly conserved residues (**Figure 5.4**). The largely negative CTE of molecule 2 rests within the minor groove but also makes contacts with the positive surface of molecule 1(**Figure 5.4a**). Namely,

Asp148 makes weak electrostatic contacts with Arg121.This intermolecular interface is also supported by a series of hydrophobic contacts driven largely by the conserved Met150, which interacts with Arg121 and Arg124 (**Figure 5.4b,c**).

The LRH-1 DBD crystallized as a monomer on a 12 bp *Oct4* fragment, with its DNA recognition helix resting in the major groove formed by the DR0 motif, flanked by two conserved zinc fingers (**Figure 5.5a**). This structure is similar to the previously solved LRH-1 – *hCYP7A1* complex (PDB ID: **2A66**)[10]. The DNA reading helix contains four residues that make distinct base-specific contacts with the *Oct4* DR0, including hydrogen bonds between Glu104 and cytosine 93, Lys107 and the O6 atom on guanine 113, Lys111 and the N7 atom on guanine 114, and Arg112 and the N7 atom on adenine 91 (**Figure 5.5b**). In addition to the core DBD there was also density to model the CTE and Ftz-F1 domains. The CTE of hLRH-1 wedges into the minor groove of the DNA duplex (**Figure 5.6c**). Base-specific contacts formed by the CTE include hydrogen bonds between Arg162 and thymine 95 and Arg165 and cytosine 110. Other stabilizing interactions with the phosphate backbone and the water network are shown in **Figure 5.5c**.

LRH-1 and GCNF differentially recognize the DR0 of the Oct4 Proximal Promoter

To understand how these distinct proteins recognize the same *Oct4* DR0 DNA, we compared both the primary amino acid sequence and structure (**Figure 5.6**). Alignment of mGCNF and hLRH-1 DBDs show 44.7% identity with 61.8% similarity. Though human LRH-1 DBD was used, sequence alignments show that human and mouse LRH-1 DBDs are 94.4 % identical, with the critical DNA binding residues being 100% conserved. Mouse GCNF (495 aa) and human GCNF (480 aa) are 95.4% identical, with both the DBD and CTE

100% identical. Furthermore, the core DR0 motif within the *Oct4* PP is conserved between mice and humans, suggesting these interactions would be maintained across species.

In order to visualize differences in the structure, we superimposed both to identify how each recognizes the same DNA element (**Figure 5.6**). Unlike, GCNF, LRH-1 binds to DNA as a monomer. While this is uncommon among NRs, other monomeric receptors include SF-1, ERR2, REV-ERB[15,28,29]. Perhaps not surprising, the LRH-1 DBD binds the *Oct4* DR0 element with greater affinity than the GCNF DBD, which is lacking the LBDs that are known to participate in receptor dimerization[24]. Superposition of the structures revealed LRH-1 makes four additional base-specific contacts within the major and minor groove. The root-mean-square deviation (r.m.s.d.) between the DBD core domains of LRH-1 (83-154) and GCNF (74-147) is 0.46 Å (63 Cα aligned). In the DNA reading helix, LRH-1 displays two additional base-specific contacts with Glu104 and Lys107 making hydrogen bonds to cytosine 93 and guanine 114, respectively (**Figure 5.5b**). The GCNF and LRH-1 CTE share 56.3% identity with 62.5% similarity and a similar backbone position within the DNA minor groove (r.m.s.d 0.51 Å; 13 Cα aligned). However, LRH-1 makes two additional H-bonding contacts vs GCNF via Arg160 and 162 (**Figure 5.5c,e**) where GCNF contains glycine and proline residues at those positions (**Figure 5.5d,e**). Mutation of the GCNF CTE glycine and proline to arginine residues (Gly149Arg and Pro151Arg), to mimic the LRH-1 CTE, increased the GCNF DBD affinity from 170 nM to 20 nM (**Figure 5.5f**). Conversely, simultaneous Arg160Gly and Arg162Pro mutations within the LRH-1 CTE reduced binding affinity from 60 nM to 750 nM (**Figure 5.5f**). Taken together, difference within the CTE drive differential affinity for the isolated DBDs for the Oct4 promoter. The difference in affinity between these isolated DBDs should not be used to predict in-cell kinetics; however,

it is clear that these proteins compete for *Oct4* proximal promoter binding when co-expressed *in vivo*.

Discussion

The orphan nuclear receptors, LRH-1 and GCNF play a critical role in regulating genes central to embryonic development[3,7]. Notably, LRH-1 and GCNF reciprocally regulate the pluripotency factor, Oct4, by binding to a DR0 response element within the promoter[3,7]. Though many have studied this mechanism, the structural basis for this regulation has yet to be explored. Here, we not only present the first structure of GCNF DBD but also LRH-1 DBD bound to the *Oct4* DR0 sequence and examine the sequence-specific contacts that guide *Oct4* regulation.

Understanding the DNA-binding properties of transcription factors is required to understand their biological function. Initial gel mobility shift assays show GCNF binds direct repeats of the AGGTCA sequence with no spacer or to extended half-sites of TCAAGGTCA sequence as a homodimer, though GCNF bound to the DR0 with higher affinity than the half-sites[18,19]. Deletion studies show that removal of the ligand-binding domain had no effect of DNA binding and *in vivo* removal of this domain did not affect transrepression[30,31]. Additionally, N-terminal domain deletions still bound DNA to the same levels of WT GCNF[20]. In contrast, removal of the DBD *in vivo* phenocopies the GCNF complete knockout mice[23]. Additionally, the DBD and the CTE are strictly required to bind DNA making them necessary to repress Oct4[20,30]. For this reason, we purified the minimal residues required for DNA recognition (**Figure 5.2A**).

We show that purified GCNF DBD binds to the *Oct4* DR0 via a two-site binding mechanism with a high and low affinity binding event (**Figure 5.2c**). In contrast, intact *in vitro* translated full-length protein binds DNA as a homodimer[15,20,30]. The overall structure

shows two GCNF molecules bound to opposite sides of the DNA and the sister DBDs only weakly interact (**Figure 5.3,4**). This structure is similar to retinoic X receptor bound to a DR1 sequence[27]. The GCNF DBD sequence has been shown to be most similar to the RXR DBD, which requires an intact LBD for receptor dimerization; therefore, it is likely that strong homodimerization of GCNF also requires the LBD[19,24]. Dimerization motifs have been proposed in the DBD, CTE, and LBD. Structural analysis reveals that the CTE makes a number of hydrophobic contacts with the other GCNF molecule (**Figure 5.4**). Additionally, the side chains that participate in homodimerization are very well conserved (**Figure 5.4c**). Therefore, structural studies of the GCNF LBD or full-length will be required complete the understanding of GCNF:DNA recognition[24,30].

LRH-1 is required to maintain *Oct4* expression[3]. Monomeric LRH-1 interacts with extended half-site sequences in both the proximal promoter and proximal enhancer to activate *Oct4* expression prior to differentiation[16]. The only reported structures of NR5A DBDs are the x-ray structure of the LRH-1-*hCYP7A1* complex (PDB ID: **2A66**) and the NMR solution structure of SF-1 in complex with a 9 bp fragment of *inhibin-α* (PDB ID: **2FF0**; 16 conformers) (*10,15*). While SF-1 and LRH-1 bind the same response elements, previous data show that SF-1 is not expressed at detectable levels in ES cells[3]. We show that purified LRH-1 DBD and full-length LRH-1 bind the *Oct4* DR0 with high affinity as a monomer positioned over the extended LRH-1 recognition sequence (**Figure 5.2b,d, 5.5**).

Comparing the GCNF and LRH-1 structures revealed that their respective response elements directly overlap in the DR0 of the PP. (**Figure 5.6**). LRH-1 is highly expressed in ES cells when *Oct4* expression is up regulated[3]. However, upon signals to differentiate, both LRH-1 and *Oct4* expression is rapidly decreased[3,28]. During this time GCNF expression is

high (**Figure 5.1**)[7]. LRH-1 has an apparent higher affinity *in vitro*, potentially due to additional side chain – DNA base interactions (**Figure 5.5,6**). Mutational analysis of GCNF CTE to mimic the LRH-1 CTE shows an increased affinity for the *Oct4* DR0 and the high LRH-1 affinity can be drastically reduced when the Arg residues of the CTE are mutated to the GCNF glycine and proline residues at the same positions (**Figure 5.5f**). Since binding was only being tested with isolated domains, this result is not surprising. Kinetic studies with intact GCNF may reveal similar affinities for the element. Yet, it is possible that GCNF will not have to compete for binding to the DR0 during embryonic development due to the inverse expression patterns of LRH-1 and GCNF (**Figure 5.6**)[3,18].

Mutational analyses of C-terminal extensions from numerous nuclear receptors such as SF-1, RXR, LRH-1, and GCNF have shown this region to be critical for DNA recognition and transcriptional regulation[10,24,32,33]. The GCNF structure shows electron density for the CTE within molecule 2 (**Figure 5.5d**). Removal of these residues from GCNF prevents DNA binding, highlighting their importance[19]. Furthermore, DNA sequence analysis revealed that the TCA sequence preceding the second direct repeat, that forms the minor groove CTE interaction, is most important for GCNF to bind DNA[19,20]. Here, GCNF uses an Arg residue to make two base-specific contacts within the minor groove (**Figure 5.5d**). PISA analysis of the GCNF structure reveals that the average gain on complex formation for molecule 2 and DNA is -9.1 kcal/mol (complex score 1.0) but with the CTE removed it is -5.2 kcal/mol (complex score 0.190) and has a much weaker complex score (1.0 being highly favorable complex)[34]. Like GCNF, the LRH-1 CTE also dips into the minor groove to make additional base-specific contacts (**Figure 5.5c**).

Structure alignments of GCNF, LRH-1, RXR, and SF-1 bound to DNA elements show high conservation among the overall DBD structures. The r.m.s.d. of LRH-1 and GCNF on the *Oct4* DR0 is 0.57 Å (76 Cα aligned), while RXRα and GCNF is 0.66 Å (66 Cα aligned) and SF-1 and LRH-1 is 1.12 Å (88 Cα aligned). Though GCNF is cited as most similar to RXR, the CTE regions are very different. When overlaid, these structures show that the RXRα CTE does not rest in the DNA minor groove but trails off away from the DNA. Furthermore, sequence alignments of $GCNF_{126\text{-}141}$ and $RXR\alpha_{201\text{-}216}$ only have 27% identity and 40% similarity. These differences also set GCNF apart in its own subclass of the nuclear receptor superfamily. $LRH\text{-}1_{152\text{-}167}$ and $SF\text{-}1_{79\text{-}94}$ on the other hand, are 100% identical. Aligned structures show that both of these proteins' CTEs rest in the minor groove of DNA leading into NR5A conserved Ftz-F1 helices. The NMR solutions of SF-1 show that, although the CTE adopts multiple conformations, all make contacts in the minor groove, as in both LRH-1 structures (PDB ID: **2A66** and **5L0M**). Interestingly, replacing the GCNF CTE with that from LRH-1/SF-1, induces a monomeric DNA binding preference likely by removing conserved GCNF intermolecular contacts (i.e. Met150) and increasing half-site affinity[24]. In agreement, the WT GCNF binds to the *Oct4* DR0 with an apparent two-site binding curve but the GCNF Gly142Arg/Pro151Arg mutant binds the same element using a one-site binding mechanism (**Figure 5.5f**). As these mutations now mimic the LRH-1 and SF-1 CTE, this result is not surprising. Structural studies using intact LRH-1 and GCNF would shed deep insight *Oct4* recognition; however, it is clear that the CTEs of these receptors play a critical role not only in sequence specificity but oligomierization on DNA[3].

Methods

Protein expression and purification

The DNA binding domain (DBD) of mouse GCNF (residues 69-180, UniProt Q64249) with a C104S mutation was cloned into a pMCGS7 vector with a 6X-Histidine tag. GCNF was expressed in BL21 (DE3) pLysS *E. coli* cells. The protein was grown at 37 °C for 2 hrs, then reduced to 20 °C and grown until an OD_{600} of 0.6, and then induced with 0.78 mM Isopropyl β-D-1-thiogalactopyranoside (IPTG). Cultures were grown overnight at 20 °C. Cells were lysed in 20 mM Tris-HCl (pH 7.4), 1M NaCl, 25 mM imidazole, and 5% glycerol via sonication. Protein was purified using affinity chromatography (His-Trap FF, GE Healthcare) followed by further purification via gel filtration chromatography. Protein was then concentrated to 2-3 mg/ml in 20 mM Tris-HCL (pH 7.4), 150 mM NaCl, and 5% glycerol, flash frozen in liquid N_2, and stored at -80 °C.

The DBD-FtzF1 domains of human LRH-1 (LRH-1 DBD) were expressed and purified similar to previous[10]. Briefly, cultures were grown in Terrific Broth (TB) at 37 °C to OD_{600} of 0.6, induced with 0.3 mM IPTG at 18 °C and grown overnight. Fusion protein was purified via affinity chromatography (His-Trap FF, GE Healthcare) and the MBP tag was removed following TEV protease cleavage with an additional pass over His-Trap FF resin. LRH-1 DBD was further purified using gel filtration (Superdex 75; GE Healthcare) equilibrated with 20 mM Tris-HCL (pH 7.4), 150 mM NaCl, and 5% glycerol. Eluted protein was concentrated to 11.5 mg/mL and either flash frozen in liquid N_2 and stored at -80 °C or used directly for crystallization experiments.

hLRH-1 has multiple full-length functional isoforms, the canonical sequence (isoform

2; 1-541) was used to number the residues of the crystal structure (*10*). Experimentally, isoform 1 (1-495) lacking residues 22-67 in the N-terminal domain, was used to generate a construct (amino acids 2-495) preceded by a TEV protease site, was cloned into the pE-SUMO-Amp vector (LifeSensors) and recombinantly expressed in *E. coli* BL21 DE3 cells. Cells were grown in LB at 37 °C to OD_{600} of 0.6, induced with 0.5 mM IPTG at 20 °C and grown overnight. Cells were lysed in 20 mM Tris-HCl (pH 8.5), 0.5 M NaCl, 25 mM imidazole, 2 mM CHAPS, 0.2% Triton-100X and pierce protease inhibitor tablets (Thermo Fisher Scientific) via sonication. Fusion protein was purified by affinity chromatography (His-Trap FF and HiTrap Heparin, GE Healthcare). The His-SUMO fusion tag was removed by incubation with TEV overnight and affinity chromatography used to isolate pure target protein.

Generation of GCNF DBD Gly149Arg/Pro151Arg and LRH-1 DBD Arg160Gly/Arg162Pro

Mutagenesis was performed following the MEGAWHOP protocol. Briefly, megaprimers were generated containing the desired mutations, purified and used for whole plasmid PCR[35]. Mutant proteins were prepared as wild-type.

Sequence Alignments and Analyses

The following sequences were obtained from UniProt: mGCNF (UniProt Q64249-1), hGCNF (Uniprot Q15406-1), mLRH-1 (UniProt P45448-1), hLRH-1 (UniProt O00482-1), hSF-1 (UniProt Q13285-1) and hRXRα (UniProt P19793-1) and aligned using Clustal Omega[36]. Jalview[37] was used for visualization and manipulation of the alignments. The Sequence Manipulation Suite (SMS) was used for percent identity and similarity calculations[38].

Nucleic acid binding assays

Synthesized FAM-labeled nucleic acid duplexes (Integrated DNA Technologies) of mouse *Oct4* DR0 (5' – [FAM] AGAGGTCAAGGCTAGA – 3') were annealed in a 1 L water bath heated to 90 °C then cooled slowly to room temperature. Fluorescence polarization assays were performed by adding increasing concentrations of GCNF, LRH-1 DBD, or FL LRH-1 (1 nM- 50 μM for the DBDs; 1 nM- 20 μM for FL hLRH-1) to 10 nM of the FAM-labeled DNA. Reactions were performed in 20 mM Tris-HCL (pH 7.4), 150 mM NaCl, and 5% glycerol. Polarization was monitored on a Biotek Synergy plate-reader at an excitation/emission wavelength of 485/528 nm. The program GraphPad Prism 6 was used to analyze binding data and generate graphs. Binding data were analyzed with an F-test to compare a two-site binding event to a one-site binding event with Hill slope. This test generated an F-statistic and p-value supporting a two-site binding model. These values are represented in Figure 2. Additionally, dissociation values (K_d) and coefficient of determination (r^r) are included.

Crystallization, data collection, and structure determination

Crystals of the GCNF-*mOct4* (16bp - 5' – AGAGGTCAAGGCTAGA – 3') complex were grown by hanging drop vapor diffusion in 0.1M Tris pH 8.5, 20% PEG 3350, 3% glycerol with a 2:1 protein:DNA molar ratio. Crystals were cryo-protected with 0.1M Tris pH 8.5, 30% PEG 3350, and 15% glycerol and flash cooled in liquid N_2. The hLRH-1 DBD-*mOct4* (12bp duplex - 5' – GGTCAAGGCTAG – 3') complex was formed by mixing 1:1.2 molar ratios of protein to DNA and incubating at 25 °C. Crystallization conditions were screened using a Phoenix nanolitre drop dispensing robot (Art Robbins) with a 1:1 protein (5-

6 mg/mL) to well solution ratio. Single well-formed crystals appeared overnight at 18 °C in 0.2 M calcium acetate and 20% (w/v) PEG 3350. Larger crystals were grown by hanging drop vapor diffusion in the same solution. Crystals were cryo-protected with an additional 20% (v/v) glycerol and flash cooled in liquid N_2.

Data were collected at the 22-ID beamline (Advanced Photon Source, Argonne, IL) and processed using HKL-2000[39]. The structures were phased using a low-resolution model GCNF-*Oct4*, previously generated in the lab or the hLRH-1 DBD-*CYP7a1* complex (PDB ID: **2A66**)[10]. Structure refinement and validation was performed using PHENIX refine software and model building was performed in COOT[42,41]. PDB Redo was used iteratively to optimize refinement parameters and geometry[42]. PyMOL v1.8.2 was used to visualize structures and generate figures (Schrödinger, LLC). Both structures showed good overall geometry with one Ramachandran outlier in the LRH-1 DBD – DNA complex and all other residues (93) in favored or allowed regions and zero Ramachandran outliers in the GCNF DBD – DNA complex and all in favored or allowed regions.

Accession Numbers

Coordinates and structure factors have been deposited in the Protein Data Bank with the accession numbers, 5KRB for GCNF:*Oct4* complex and 5L0M for LRH-1:*Oct4* complex.

Acknowledgements

X-ray data were collected at Southeast Regional Collaborative Access Team (SER-CAT) 22-ID beamline at the Advanced Photon Source, Argonne National Laboratory. Supporting institutions may be found at www.ser-cat.org/members/html. Use of the Advanced Photon source was supported by the U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences, under Contract No. W-31-109-Eng-38.

Figures

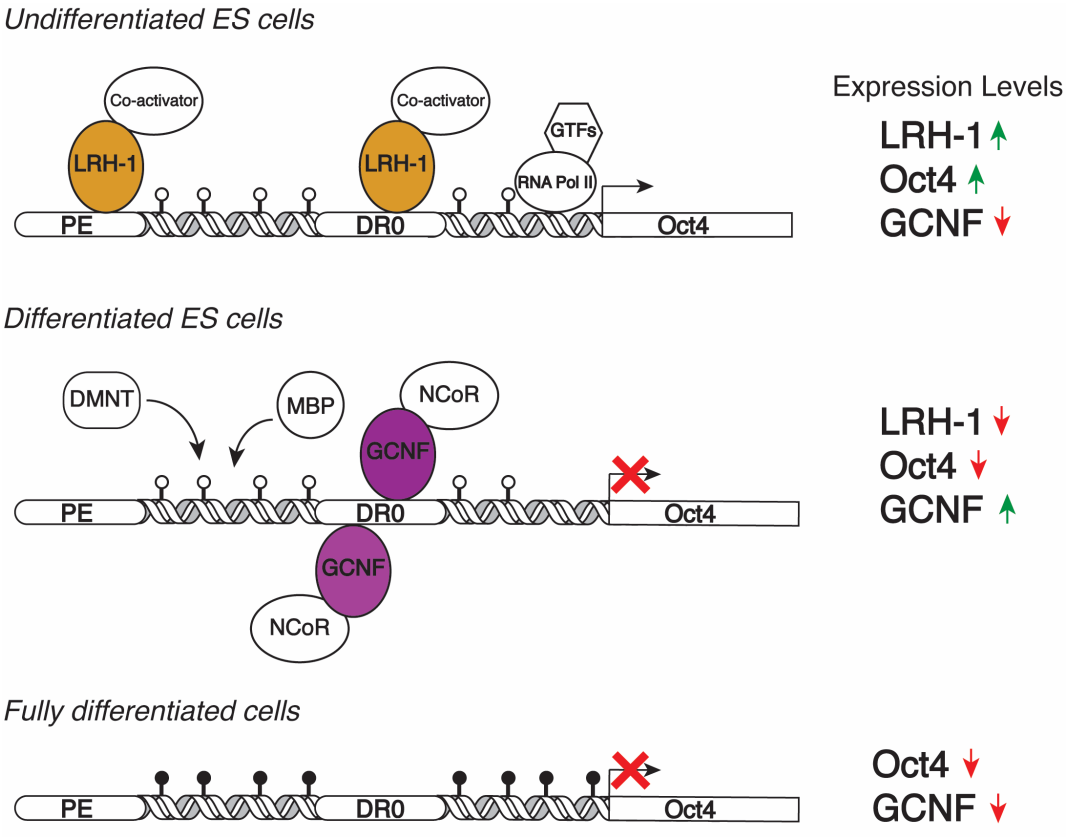
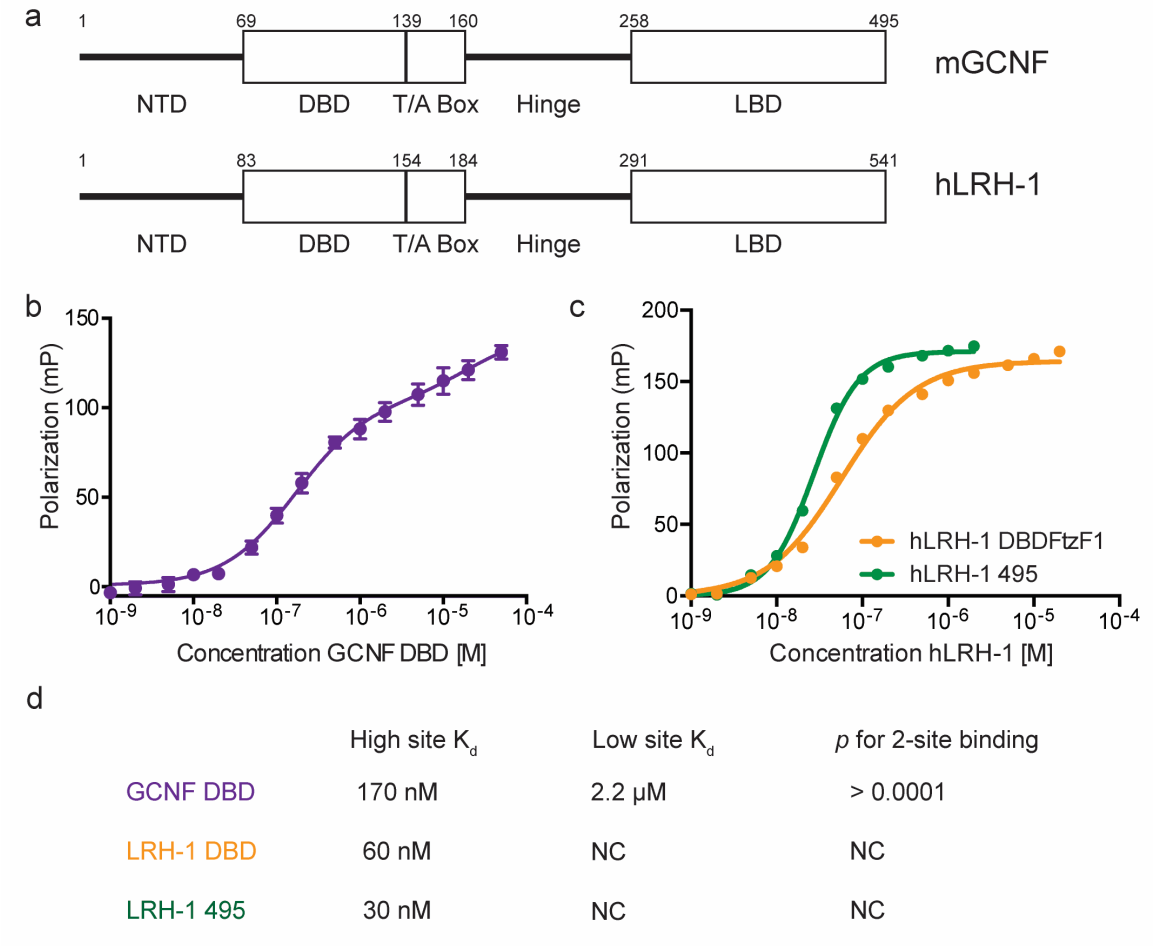


Figure 5.1: Schematic representation of differential regulation of Oct4 by LRH-1 and GCNF.

In undifferentiated ES cells, LRH-1 expression is high and drives *Oct4* expression by binding to DR0 sequences in the proximal enhancer and promoter. This binding recruits coactivators and the transcriptional machinery to drive gene expression. Upon signals to differentiate, LRH-1 expression is rapidly reduced while GCNF expression is high. GCNF then binds to the DR0 within the proximal promoter to repress Oct4 expression. Binding by GCNF recruits corepressors, such as NCoR, to block Oct4 expression. GCNF also recruits DNA methyltransferases (DMNT) and methyl-binding proteins (MBP) to methylate the Oct4 gene in order to efficiently shut off its expression.



d

	High site K_d	Low site K_d	*p* for 2-site binding
GCNF DBD	170 nM	2.2 μM	> 0.0001
LRH-1 DBD	60 nM	NC	NC
LRH-1 495	30 nM	NC	NC

Figure 5.2: GCNF and LRH-1 bind directly to the Oct4 DR0.

(a) Diagram of the GCNF and LRH-1 modular structure. In this study, GCNF DBD-TA (residues 69-180), LRH-1 DBD Ftz-F1 (residues 79-187), and full-length LRH-1 (residues 2-495) were used. (b) GCNF DBD-TA bound to the Oct4 DR0 in a two-site binding mechanism. (c) LRH-1 DBD Ftz-F1 (orange) and full-length LRH-1 (green) bind the Oct4 DR0 in a one-site binding mechanism. Binding data are represented as mean ± s.e.m from three replicates and from three independent fluorescence polarization experiments. (d) Summary of binding data.

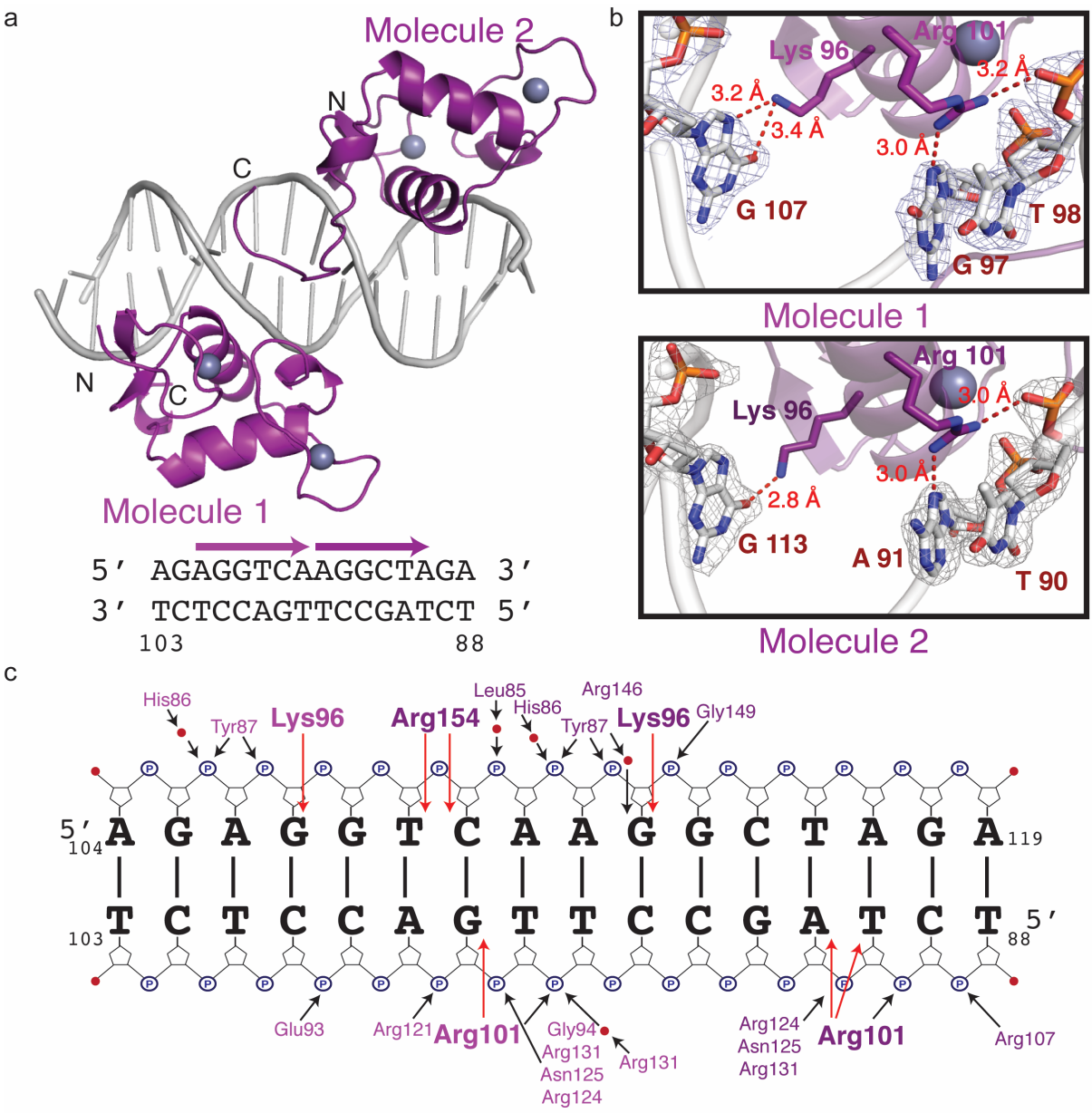


Figure 5.3: Structural Analysis of GCNF - mOct4 Complex.

(a) Overall structure of GCNF DBD (purple) bound to the Oct4 DR0 (gray). GCNF DBD and DNA shown as cartoon in purple and white, respectively and zinc atoms as spheres. Two molecules of GCNF bound to opposite sides of the DNA in a head-to-tail fashion. The DR0 sequence is shown below with arrows denoting the direction of GCNF over the sequence. (b) Each molecule of GCNF makes base-specific contacts with the DR0 (bases in white)

mediated by hydrogen bonds (red) made between Arg101 and Lys96. Mesh shows 2Fo-Fc electron density map contoured to 2σ around the DNA bases. (c) Schematic view of protein-DNA interactions. Larger, bold side chains denoted base-contacting side chains. Water molecules are indicated as red spheres.

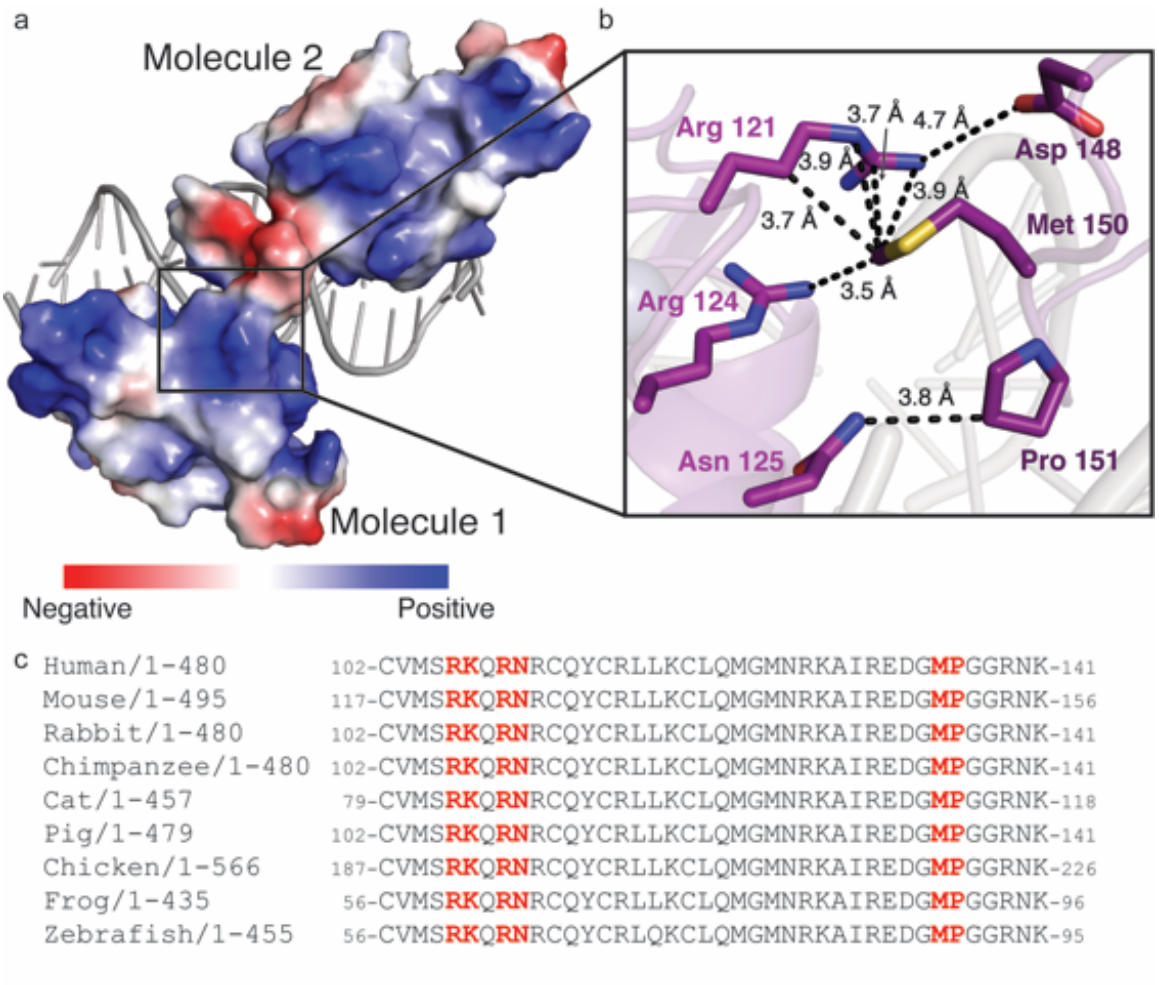


Figure 5.4: Interactions between GCNF molecules.

(a) Electrostatic surface of overall GCNF structure. (b) Zoom in of contacts (black) made between the two molecules. Side chains from molecule 1 are colored purple and dark purple from molecule 2. (c) Sequence alignments from numerous species reveal that the side chains that participate in homodimer formation are highly conserved.

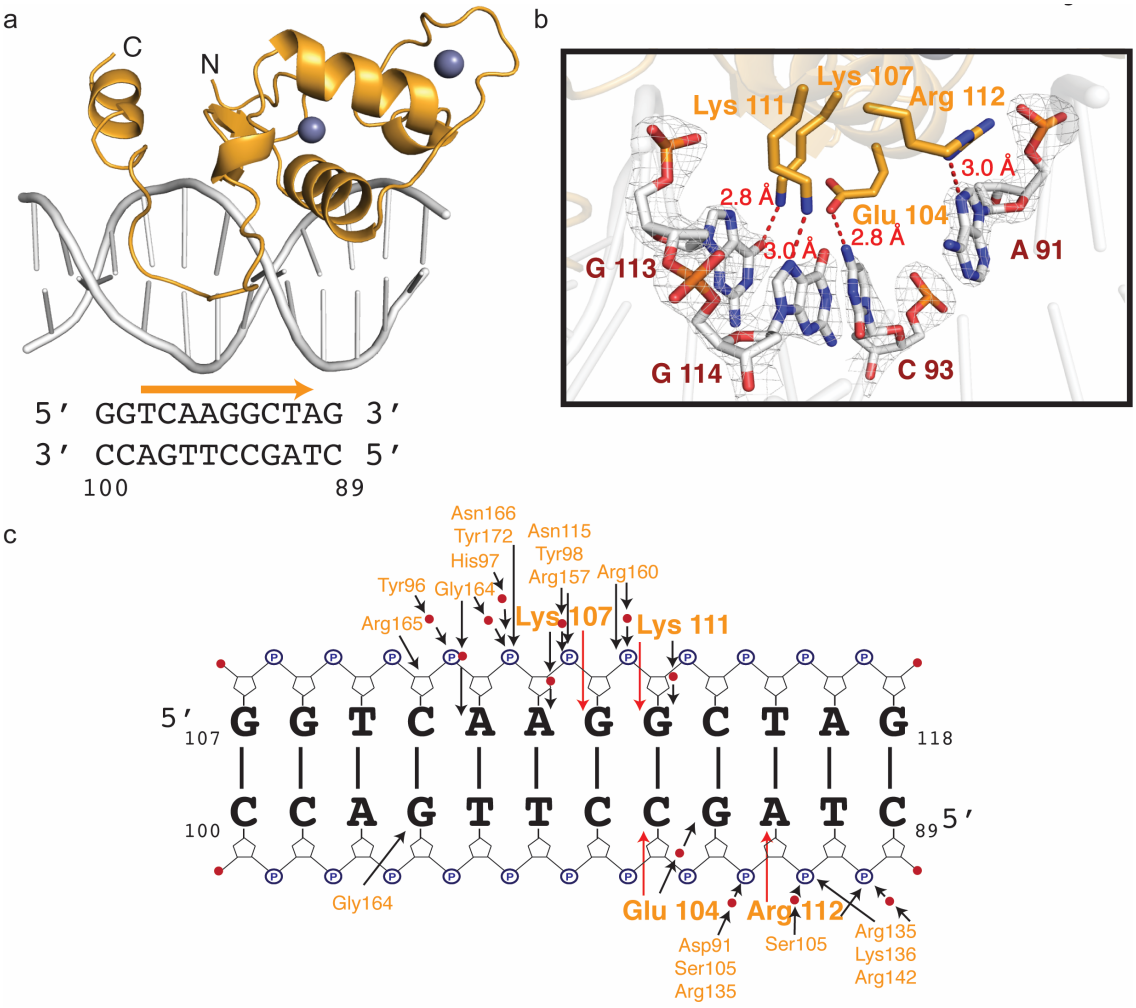


Figure 5.5: Structural Analysis of LRH-1 - mOct4 Complex.

(a) Overall structure of LRH-1 DBD (orange) bound to the 12 bp Oct4 DR0 (gray) with zinc atoms as spheres. The Oct4 DR0 sequence is shown below with arrows denoting the footprint and orientation of the LRH-1 binding site. (b) LRH-1 makes base-specific contacts with the DR0 (bases in white) mediated by hydrogen bonds (red) made between Glu104, Lys107, Lys 111, Arg112. Mesh shows 2Fo-Fc electron density map contoured to 2σ around the DNA bases. (c) Schematic view of protein-DNA interactions. Larger, bold side chains denoted base-contacting side chains. Water molecules are indicated as red spheres.

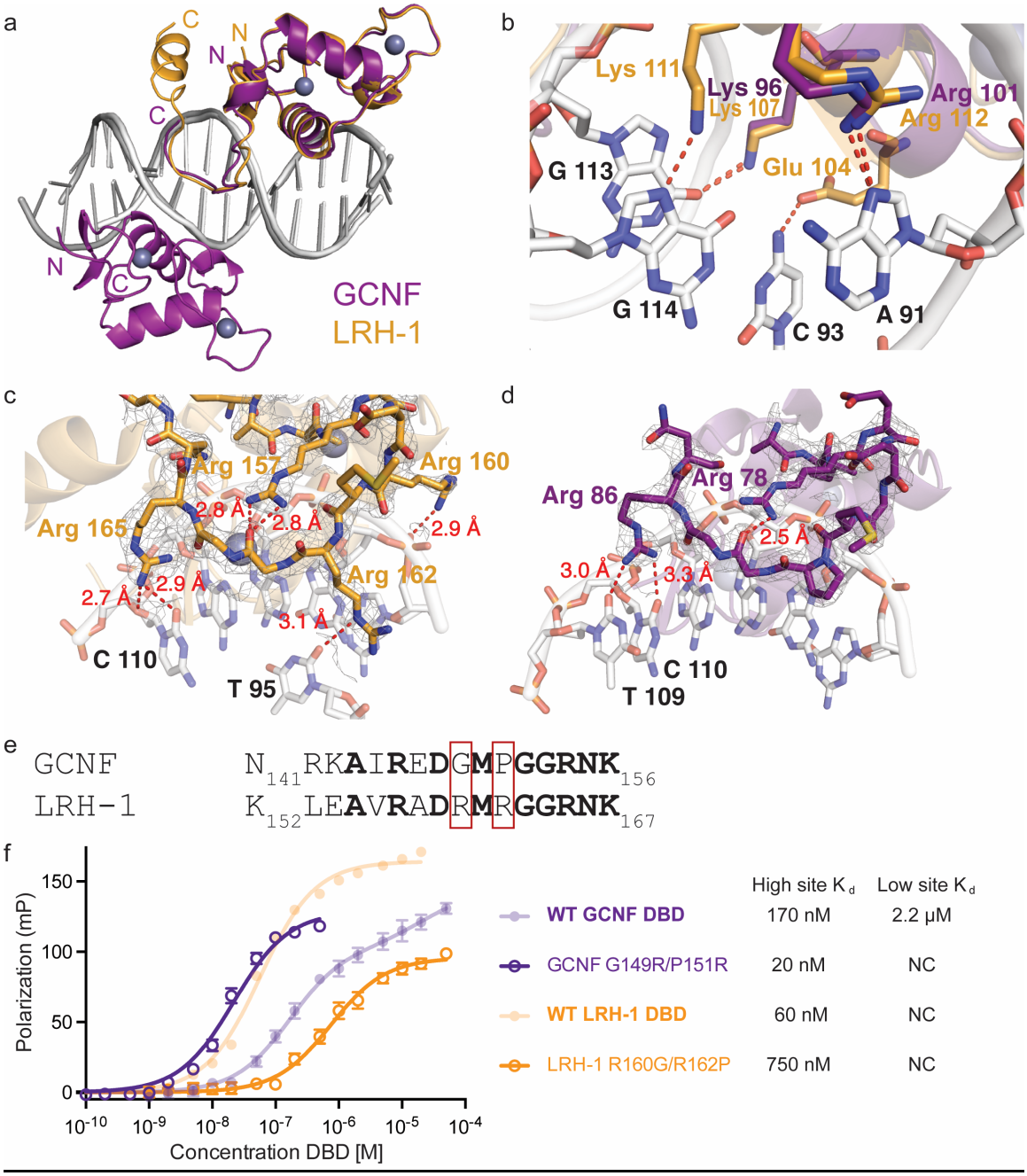


Figure 5.6: GCNF and LRH-1 comparison.

(a) Overlay of GCNF and LRH-1 structures. Molecule 2 of GCNF (Dark purple) sites directly on top of the LRH-1 (orange) recognition site. (b) Close up view of the base-specific

contacts mediated by these two receptors. LRH-1 makes an additional contact with Lys 107 making hydrogen bonds to guanine 114. (c) LRH-1 CTE has good electron density, mesh shows 2Fo-Fc electron density map contoured to 1σ around the residues. Arg162 and 165 make hydrogen bonds (red) to thymine 95 and cytosine 110. (d) GCNF CTE has good electron density, mesh shows 2Fo-Fc electron density map contoured to 2σ around the residues. Arg86 makes hydrogen bonds to thymine 109 and cytosine 110. Arg78 also folds into the CTE to make hydrogen bonds to Gly152. (e) Sequence alignment of GCNF and LRH-1 CTEs show LRH-1 to contain two additional Arg residues that are used for DNA binding. (f) Mutational analysis of CTE residues on *Oct4* binding: GCNF DBD Gly149Arg/Pro151Arg (open purple circles) bound to the *Oct4* DR0 with an affinity 20 nM, where WT GCNF DBD (faded closed purple circles) binds with an affinity of 170 nM. LRH-1 DBD Arg160Gly/Arg162Pro (open orange circles) bound with an affinity of 750 nM, where WT LRH-1 DBD (faded closed orange circles) bound with an affinity of 60 nM.

	GCNF DBD - *mOct4* (16bp)	**LRH-1 DBD - *mOct4* (12bp)**
Data Collection		
Space Group	$P2_12_12_1$	$P4_3$
Cell Dimension	a=53.6, b=69.5, c=84.5	a=40.9, b=40.9, c=105.1
Resolution (Å)	2.10 (2.18-2.10)	50-2.2 (2.28-2.20)
R_{pim}	6.9 (37.1)	3.7 (30.0)
I/σ	9.7 (2.0)	24.3 (2.1)
Completeness	96.2 (80.2)	99.8 (98.8)
Redundancy	5.2 (3.5)	7.2 (5.7)
Refinement		
Resolution	2.10	2.20
No. Reflections	18277	8789
R_{work}/R_{free}	21.6/26.8	15.4/19.7
No. Atoms		
Protein	1244	779
DNA	651	485
Water	43	50
B-factors		
Protein	56.0	59.1
DNA	51.8	52.6
Water	54.6	53.4
R.m.s. deviations		
Bond lengths (Å)	0.003	0.016
Bond angles (°)	0.57	1.65
PDB code	5KRB	5L0M

*Data collected from a single crystal; values in parentheses are for the highest-resolution shell

Table 5.1: Data Collection and Refinement Statistics

Data collection and refinement statistics for the GCNF DBD and LRH-1 DBD in complex with the *Oct4* DR0.

References

1. Paranjpe, S. S. & G. J. Veenstra. 2015. Establishing pluripotency in early development. Biochim Biophys Acta 1849.626-36.

2. Wu, G. & H. R. Scholer. 2014. Role of Oct4 in the early embryo development. Cell Regen (Lond) 3.7.

3. Gu, P., B. Goodwin, A. C. Chung, X. Xu, D. A. Wheeler, R. R. Price, C. Galardi, L. Peng, A. M. Latour, B. H. Koller, J. Gossen, S. A. Kliewer & A. J. Cooney. 2005. Orphan nuclear receptor LRH-1 is required to maintain Oct4 expression at the epiblast stage of embryonic development. Mol Cell Biol 25.3492-505.

4. Sylvester, I. & H. R. Scholer. 1994. Regulation of the Oct-4 gene by nuclear receptors. Nucleic Acids Res 22.901-11.

5. Mullen, E. M., P. Gu & A. J. Cooney. 2007. Nuclear Receptors in Regulation of Mouse ES Cell Pluripotency and Differentiation. PPAR Res 2007.61563.

6. Olefsky, J. M. 2001. Nuclear receptor minireview series. J Biol Chem 276.36863-4.

7. Fuhrmann, G., A. C. Chung, K. J. Jackson, G. Hummelke, A. Baniahmad, J. Sutter, I. Sylvester, H. R. Scholer & A. J. Cooney. 2001. Mouse germline restriction of Oct4 expression by germ cell nuclear factor. Dev Cell 1.377-87.

8. Palmieri, S. L., W. Peter, H. Hess & H. R. Scholer. 1994. Oct-4 transcription factor is differentially expressed in the mouse embryo during establishment of the first two extraembryonic cell lineages involved in implantation. Dev Biol 166.259-67.

9. Barnea, E. & Y. Bergman. 2000. Synergy of SF1 and RAR in activation of Oct-3/4 promoter. J Biol Chem 275.6608-19.

10. Solomon, I. H., J. M. Hager, R. Safi, D. P. McDonnell, M. R. Redinbo & E. A. Ortlund. 2005. Crystal structure of the human LRH-1 DBD-DNA complex reveals Ftz-F1 domain positioning is required for receptor activity. J Mol Biol 354.1091-102.

11. Yussa, M., U. Lohr, K. Su & L. Pick. 2001. The nuclear receptor Ftz-F1 and homeodomain protein Ftz interact through evolutionarily conserved protein domains. Mech Dev 107.39-53.

12. Pare, J. F., D. Malenfant, C. Courtemanche, M. Jacob-Wagner, S. Roy, D. Allard & L. Belanger. 2004. The fetoprotein transcription factor (FTF) gene is essential to embryogenesis and cholesterol homeostasis and is regulated by a DR4 element. J Biol Chem 279.21206-16.

13. Takemaru, Ki, F. Q. Li, H. Ueda & S. Hirose. 1997. Multiprotein bridging factor 1 (MBF1) is an evolutionarily conserved transcriptional coactivator that connects a regulatory factor and TATA element-binding protein. Proc Natl Acad Sci U S A 94.7251-6.

14. Cai, Y. N., Q. Zhou, Y. Y. Kong, M. Li, B. Viollet, Y. H. Xie & Y. Wang. 2003. LRH-1/hB1F and HNF1 synergistically up-regulate hepatitis B virus gene transcription and DNA replication. Cell Res 13.451-8.

15. Little, T. H., Y. Zhang, C. K. Matulis, J. Weck, Z. Zhang, A. Ramachandran, K. E. Mayo & I. Radhakrishnan. 2006. Sequence-specific deoxyribonucleic acid (DNA) recognition by steroidogenic factor 1: a helix at the carboxy terminus of the DNA binding domain is necessary for complex stability. Mol Endocrinol 20.831-43.

16. Sung, B., H. J. Do, S. W. Park, S. H. Huh, J. H. Oh, H. J. Chung, M. J. Kang, J. H. Kim, N. H. Kim & J. H. Kim. 2012. Regulation of OCT4 gene expression by liver receptor homolog-1 in human embryonic carcinoma cells. Biochem Biophys Res Commun 427.315-20.

17. Heng JC, Feng B, Han J, Jiang J, Kraus P, Ng JH, et al. The nuclear receptor Nr5a2 can replace Oct4 in the reprogramming of murine somatic cells to pluripotent cells. Cell Stem Cell. 2010;6:167-74.

18. Chen, F., A. J. Cooney, Y. Wang, S. W. Law & B. W. O'Malley. 1994. Cloning of a novel orphan receptor (GCNF) expressed during germ cell development. Mol Endocrinol 8.1434-44.

19. Yan, Z. H., A. Medvedev, T. Hirose, H. Gotoh & A. M. Jetten. 1997. Characterization of the response element and DNA binding properties of the nuclear orphan receptor germ cell nuclear factor/retinoid receptor-related testis-associated receptor. J Biol Chem 272.10565-72.

20. Cooney, A. J., G. C. Hummelke, T. Herman, F. Chen & K. J. Jackson. 1998. Germ cell nuclear factor is a response element-specific repressor of transcription. Biochem Biophys Res Commun 245.94-100.

21. Wang, H., X. Wang, X. Xu, M. Kyba & A. J. Cooney. 2016. Germ Cell Nuclear Factor (GCNF) Represses Oct4 Expression and Globally Modulates Gene Expression in Human Embryonic Stem (hES) Cells. J Biol Chem 291.8644-52..

22. Chung, A. C., D. Katz, F. A. Pereira, K. J. Jackson, F. J. DeMayo, A. J. Cooney & B. W. O'Malley. 2001. Loss of orphan receptor germ cell nuclear factor function results in ectopic development of the tail bud and a novel posterior truncation. Mol Cell Biol 21.663-77.

23. Lan, Z. J., A. C. Chung, X. Xu, F. J. DeMayo & A. J. Cooney. 2002. The embryonic function of germ cell nuclear factor is dependent on the DNA binding domain. J Biol Chem 277.50660-7.

24. Greschik, H., J. M. Wurtz, P. Hublitz, F. Kohler, D. Moras & R. Schule. 1999. Characterization of the DNA-binding and dimerization properties of the nuclear orphan receptor germ cell nuclear factor. Mol Cell Biol 19.690-703.

25. Gu, P., X. Xu, D. Le Menuet, A. C. Chung & A. J. Cooney. 2011. Differential recruitment of methyl CpG-binding domain factors and DNA methyltransferases by the orphan receptor germ cell nuclear factor initiates the repression and silencing of Oct4. Stem Cells 29.1041-51.

26. Sato, N., M. Kondo & K. Arai. 2006. The orphan nuclear receptor GCNF recruits DNA methyltransferase for Oct-3/4 silencing. Biochem Biophys Res Commun 344.845-51.

27. Zhao, Q., S. A. Chasse, S. Devarakonda, M. L. Sierk, B. Ahvazi & F. Rastinejad. 2000. Structural basis of RXR-DNA interactions. J Mol Biol 296.509-20.

28. Gearhart, M. D., S. M. Holmbeck, R. M. Evans, H. J. Dyson & P. E. Wright. 2003. Monomeric complex of human orphan estrogen related receptor-2 with DNA: a pseudo-dimer interface mediates extended half-site recognition. J Mol Biol 327.819-32.

29. Zhao, Q., S. Khorasanizadeh, Y. Miyoshi, M. A. Lazar & F. Rastinejad. 1998. Structural elements of an orphan nuclear receptor-DNA complex. Mol Cell 1.849-61..

30. Borgmeyer, U. 1997. Dimeric binding of the mouse germ cell nuclear factor. Eur J Biochem 244.120-7.

31. Okumura, L. M., B. J. Lesch & D. C. Page. 2013. The ligand binding domain of GCNF is not required for repression of pluripotency genes in mouse fetal ovarian germ cells. PLoS One 8.e66062.

32. Wilson, T. E., T. J. Fahrner & J. Milbrandt. 1993. The orphan receptors NGFI-B and steroidogenic factor 1 establish monomer binding as a third paradigm of nuclear receptor-DNA interaction. Mol Cell Biol 13.5794-804.

33. Melvin, V. S., S. C. Roemer, M. E. Churchill & D. P. Edwards. 2002. The C-terminal extension (CTE) of the nuclear hormone receptor DNA binding domain determines interactions and functional response to the HMGB-1/-2 co-regulatory proteins. J Biol Chem 277.25115-24.

34. Krissinel, E. & K. Henrick. 2007. Inference of macromolecular assemblies from crystalline state. J Mol Biol 372.774-97.

35. Miyazaki, K. 2011. MEGAWHOP cloning: a method of creating random mutagenesis libraries via megaprimer PCR of whole plasmids. Methods Enzymol 498.399-406.

36. Stothard, P. 2000. The sequence manipulation suite: JavaScript programs for analyzing and formatting protein and DNA sequences. Biotechniques 28.1102, 04.

37. Waterhouse, A. M., J. B. Procter, D. M. Martin, M. Clamp & G. J. Barton. 2009. Jalview Version 2--a multiple sequence alignment editor and analysis workbench. Bioinformatics 25.1189-91.

38. Sievers, F., A. Wilm, D. Dineen, T. J. Gibson, K. Karplus, W. Li, R. Lopez, H. McWilliam, M. Remmert, J. Soding, J. D. Thompson & D. G. Higgins. 2011. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol Syst Biol 7.539.

39. Otwinowski, Z., W. Minor & et al. 1997. Processing of X-ray diffraction data collected in oscillation mode. Methods Enzymol 276.307-26.

40. Adams, P. D., P. V. Afonine, G. Bunkoczi, V. B. Chen, I. W. Davis, N. Echols, J. J. Headd, L. W. Hung, G. J. Kapral, R. W. Grosse-Kunstleve, A. J. McCoy, N. W. Moriarty, R. Oeffner, R. J. Read, D. C. Richardson, J. S. Richardson, T. C. Terwilliger & P. H. Zwart. 2010. PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr D Biol Crystallogr 66.213-21.

41. Emsley, P., B. Lohkamp, W. G. Scott & K. Cowtan. 2010. Features and development of Coot. Acta Crystallogr D Biol Crystallogr 66.486-501.

42. Joosten, R. P., J. Salzemann, V. Bloch, H. Stockinger, A. C. Berglund, C. Blanchet, E. Bongcam-Rudloff, C. Combet, A. L. Da Costa, G. Deleage, M. Diarena, R. Fabbretti, G. Fettahi, V. Flegel, A. Gisel, V. Kasam, T. Kervinen, E. Korpelainen, K. Mattila, M. Pagni, M. Reichstadt, V. Breton, I. J. Tickle & G. Vriend. 2009. PDB_REDO: automated re-refinement of X-ray structure models in the PDB. J Appl Crystallogr 42.376-84.

CHAPTER 6: DISCUSSION

Conclusions

These collective works aim to better understand the mechanisms that guide nuclear receptor regulation. This regulatory specificity is determined by the assembly of distinct complexes, each attuned to a particular gene-, cell- and physiologic-context. This work expands the current simplistic views of NR signaling by utilizing structural biology and biochemistry to examine the allosteric effectors of NR function. These allosteric modifications result in different NR surfaces that are read by coregulator proteins, driving alternative transcriptional programs.

DNA as an allosteric modulator

How is DNA specificity achieved?

High-throughput studies have identified thousands of proteins that interact with DNA[1] and transcriptional regulatory factors (TRFs) are the most common group with sequence-specific DNA binding activity. Specific DNA sequences serve not only as platforms for binding, but appear also to act as direct allosteric effectors of transcription factors[2-4]. But how is this specificity achieved? As the human genomes contain approximately 3 billion base pairs and considering there is such a large number of DNA binding proteins, many of which have high sequence conservation, like NR DBDs, how are specific binding sequences distinguished?

There are two main mechanisms that account for specific binding: *direct* and *indirect* interactions between proteins and DNA. Direct recognition involves hydrogen bonding and steric interactions between protein side chains and bases within the major and minor groove

of DNA. Each DNA base has a unique pattern of hydrogen bond donors and acceptors in the major groove with less specific hydrogen bonding potential in the minor groove (**Figure 6.1**). DNA binding proteins commonly rely on alpha helices to position residues for direct DNA recognition. This structural motif is ideal to wedge within the major groove of DNA and present side chains that can "read" DNA base edges. This is the case for NR DBDs, which utilize a DNA-reading alpha helix to make base-specific contacts within the major groove. Within the major groove, the hydrogen bond donors and acceptors of T-A and G-C base pairs are exposed and easily contacted by protein side chains. Here, it is possible to distinguish a C-G base pair from G-C and vise versa with A-T. Yet, in terms of hydrogen bond donors, a C and a T are indistinguishable. In the minor groove, A and T are indistinguishable and G and C only differ slightly. Despite this lower information content, many DNA binding proteins also utilize conserved arginine and lysine residues to make base-specific contacts within the minor groove in addition to contacts to the major groove. Mutation of these minor groove-contacting residues shows a dramatic effect on binding affinity (**Figure 5.6**), suggesting both contacts are important for specificity.

Though initially it was believed that only direct interactions accounted for specificity, it was later shown that the indirect component is also important. Indirect recognition involves the identification of DNA shape, which can be altered depending on the sequence. The ability of these indirect interactions to affect transcriptional output was shown for the glucocorticoid receptor, where changes within the minor groove spacer region between binding sites altered DNA shape. Comparing the structures of GR bound to these different sequences revealed GR adopted slightly different conformations, resulting in changes to gene expression (**Figure 2.2F**).

How do different DNA sequences alter NR function?

As outlined in **Chapter 2**, the glucocorticoid receptor (GR) regulates transcription through a multitude of mechanisms (**Figure 2.1A**)[5]. GR binds directly to DNA at canonical GR binding sequence (GBS) composed of two pseudo-palindromic hexameric AGAACA repeats separated by a three-base pair spacer[6,7] (**Figure 2.1A*i*,B,D**). GR binding to GBS sequences occurs in a head-to-head fashion through a protein-protein interaction between two GR DBD proteins[3]. In addition to x-ray crystallographic analyses, NMR studies revealed allosteric communications occur between the dimerized GR monomers[8] (**Figure 3.4c**). At these sites GR's DNA reading helix makes sequence-specific contacts within each hexameric half site (**Figure 2.1D**). Structural studies have also revealed that the 3bp spacer between the half sites as well as the +8 and -8 positions flanking the GBS can alter GR conformations[3,9].

GR also binds to a newly characterized inverted repeat-GBS (IR-GBS/nGREs), which shows monomeric GR binding to a CTCC$(N)_{0\text{-}2}$GGAGA motif[12,13] (**Figure 2.1A*ii*,C,E**). Unlike binding to a canonical GBS, GR binding to IR-GBS sequences occurs in a tail-to-tail fashion where GR DBD proteins do not interact[13,14]. Molecular dynamics simulations suggested allosteric regulation between monomers at IR-GBSs (although this might not be relevant *in vivo*, as IR-GBS elements may be occupied by only monomeric GR in the *in vitro* scenario[26]).

However, in contrast to these DNA-binding-dependent mechanisms, GR also represses transcription without direct interaction with DNA[10,11]. DNA-independent transrepression involves binding of GR to other TFs, such as activator protein-1 (AP-1) and nuclear factor-kappa beta (NF-κB), through protein-protein interactions[12,13] (**Fig 1a*iv*)**. This

mechanism, referred to as tethering, has been the accepted model for GR-mediated transrepression of inflammatory genes[14,15]. However, many of these original papers in support of this DNA-independent mechanism failed to show a direct interaction between GR and AP-1 via immunoprecipitation experiments. Only one paper was able to detect an interaction and this was after crosslinking and days of exposure to the GR antibody. This was attributed to a weak or transient interaction. Yet, despite this result there has been continued overwhelming acceptance of this model and no one has gone on to fully characterize a GR-AP-1 complex. Additionally, much of the tethering model support came from GR mutants, which generated complex and often conflicting interpretations.

The GRdim mutant, designed with the aim of breaking intramolecular protein-protein interactions at the GR homodimerization interface, was another driver for a DNA-binding-independent mechanism at inflammatory genes. This mutant was shown to not bind DNA and displays a reduced ability to transactivate genes, whereas transrepression was unaffected. These results drove the conclusions that dimerization of GR is required for DNA binding and that dimerization is not necessary for GR to repress inflammatory gene expression. However, it was later shown that GRdim does not affect GR stoichiometry on DNA; GRdim can still forms dimers *in vitro* and in cells; instead, GRdim affects cooperative binding to DNA. These results suggest that the GRdim mutant cannot be used to rule out a DNA-binding-dependent mechanism at TREs. Another GR mutant, GR Ser425Gly mutant was used to show that DNA binding was dispensable at TREs; this mutant binds to canonical GBSs but is unable to repress inflammatory genes. However, these interpretations were made before the DNA-binding-dependent GR mechanisms at IR-GBS were identified. It has since been

shown that the GR Ser425Gly mutant not only has diminished binding to an IR-GBS but also affects GR transrepression from IR-GBS sites.

With this reexamination into the tethering model and the identification of DNA-binding-dependent mechanisms at IR-GBSs, we hypothesized that instead GR could bind directly to the DNA at these sites and compete with AP-1 for the same binding site. This work expands GR's role in regulating inflammatory genes by showing that GR can also directly bind the DNA at AP-1 response elements (TREs) (**Chapter 3**). We show that GR binds directly to TREs via sequence-specific contacts to a GR-binding sequence (GBS) half-site found embedded within the TRE motif (**Figure 3.3-5**). Furthermore, we show that GR-mediated transrepression observed at TRE sites to be DNA-binding-dependent (**Figure 3.1-2**). We also show that when we tested the GR mutants used to rule out a DNA-dependent mechanism that indeed their initial conclusions were incorrect (**Figure 3.1,4**). Our findings represent a dramatic shift in our understanding of GR-mediated repression of inflammation. What drives the selection between DNA-binding-dependent repression and tethering/ transrepression remains unclear, and yet to be explored. This work will be paramount as the field continues to seek selective gene modulators for the treatment of chronic inflammatory diseases. We not only proposed a new mechanism of repression but also show that monomeric GR is responsible for this repression (**Figure 3.4**). As we continue to discover characteristics that will distinguish transactivation from transrepression, these can potentially provide avenues for future therapeutic development.

This work also reports the first crystal structure of germ cell nuclear factor (GCNF) DBD and the liver receptor homolog-1 (LRH-1) DBD bound to the same sequence within the *Oct4* promoter at an overlapping site (**Chapter 5**)[16]. The overall structure of the GCNF-*Oct4*

DR0 shows two GCNF molecules bound to opposite sides of the DNA and the sister DBDs only weakly interact (**Figure 5.3,4**). The LRH-1 DBD crystallized as a monomer on a 12 bp *Oct4* fragment, with its DNA recognition helix resting in the major groove formed by the DR0 motif, flanked by two conserved zinc fingers (**Figure 5.5a**). Both receptors utilize a C-terminal sequence (CTE) to make sequence-specific contacts with the DNA in the minor groove. Mutational studies of this region revealed that addition of charged residues within the GCNF CTE enhanced DNA binding or removal of these resides within the LRH-1 DBD abolished DNA binding (**Figure 5.5f**). As these two proteins are required to differentially regulate *Oct4* during key stages in development this work allowed us to explore the structural mechanisms that govern this regulation.

Collectively, these studies suggest that DNA sequence-specific conformational states of NRs result in the generation or stabilization of distinct patterns of NR surfaces, which serve as interaction platforms, driving alternative transcriptional outcomes. In all cases, DNA acts as a ligand to impart these allosteric changes that likely extend into the N-terminal domain and LBD. So far, structural studies of GR, GCNF, and LRH-1 in the context of allosteric modulation by DNA have been limited to isolated domains, but work with full-length vitamin D receptor (VDR) and retinoid X receptor (RXR) heterodimer confirm allosteric communication throughout NR complexes upon DNA binding[17]. Indeed, hydrogen deuterium exchange mass spectrometry (HDX-MS) on the full-length liganded VDR–RXR–DNA complex revealed conformational changes upon binding ligands, DNA, and coregulators. Changes within the VDR DNA binding sequence (VBS) read through the VDR-DBD had far-reaching intramolecular allosteric effects that altered solvent accessibility of regions within the sister LBD of the complexed RXR molecule[17]. Future work with full-

length NRs will be essential to fully describe the allosteric consequences of differential DNA binding.

Ligands as allosteric modulators

The NR LBD is a complex allosteric signaling domain that upon ligand binding undergoes conformational changes to modulate coregulator binding surfaces[18,19]. The hydrophobic ligand binding pocket (LBP) is adapted to accept variety of small, lipophilic ligands, such as steroids, thyroid hormone, retinoids, and to exogenous pharmaceutical ligands (**Figure 1.4**)[20]. Guided by conserved hydrogen bonding and hydrophobic residues within the pocket, binding to these varied ligands drives the dramatic size differences within ligand binding pockets across NRs (**Figure 1.4e**). Structural studies with GR have shown that the LBD is highly dynamic and can accommodate binding to a variety of ligands. The endogenous ligand cortisol binds specifically, but fails to fill the binding pocket similarly, the exogenous glucocorticoid dexamethasone occupies only ~65% of the GR ligand binding pocket (leaving >200 $Å^3$ excess volume within the 590 $Å^3$ binding pocket[21])[22] **(Figure 2.4b)**. The additional volume within the binding pocket offers potential space for interaction with alternative modulatory ligands. Moreover, LBD structures bound to an alternate GR ligand, RU-486 **(Figure 2.4c,d)** reveal extensive structural malleability within the LBD that appears to enable interaction with a wide range of potential ligands[23]. In order to better understand how ligands allosterically modulate GR, we utilized x-ray crystallographic structures and molecular dynamics simulations of GR bound to a clinically prevalent GC, triaminacolone acetonide (TA). Binding by TA revealed increased intramolecular contacts within the pocket, which drove conformational changes within the mouth of the ligand binding domain. These changes in structure drove the stabilizing effect of this GR-ligand complex (**Chapter 4)**.

Moreover, these data support the emerging theme that ligand-induced receptor conformational dynamics at the mouth of the pocket play a major role in steroid receptor activation. Furthermore, this work also provided the first crystal structure of the GR LBD in complex with a peptide from the atypical coactivator SHP (**Figure 4.6a,b**). Most GR LBD structures have been solved in complex with a TIF2 coactivator peptide but attempts to make crystals of GR LBD-TA complex with TIF2 were unsuccessful. This suggests that different protein-ligand complexes present different surfaces that are better read by certain coregulators. However, as TIF2 and SHP utilize conserved residues to contact the AF surface and neither showed a difference in stabilizing the GR LBD-TA complex by thermal stability, we are unable to discern the forces that drove the preference for SHP in our structure of the GR LBD-TA complex (**Figure 4.6b-e**).

In summary, GR ligands are critical physiologic or pharmacologic context inputs that confer allosteric transitions[24] that affect gene-specific GOR formation[25,26], or regulatory complex composition and/or function (*e.g.,* altered HAT activity). To date, few GR LBD-ligand structures have been solved. Therefore, this work greatly expands our understanding of the structural mechanisms that guide high-affinity ligand interactions.

NRs as scaffolds

We present a model in which NRs act as scaffolds whose conformations are altered allosterically by signaling inputs (**Figure 2.6**). This model accounts for the precision and plasticity of eukaryotic transcriptional regulation.

Our model assumes that the signals impacting NR transcriptional regulatory activity operate allosterically, and that their effects integrate to produce context-specific patterns of

NR surfaces. These patterns of surfaces are recognized and bound by specific coregulator complexes, which typically possess enzymatic activities that confer structural and/or functional changes on the transcription machinery and/or the chromatin. Coregulators are generally not themselves cell specific; rather, we propose that they assemble in unique combinations at each DNA element, based on the context-specific interactions established at these sites.

A provocative extension of these ideas is that GR and other TRFs may typically lack intrinsic regulatory activities, instead serving merely as molecular scaffolds, patterns of surfaces produced by signaling, to which coregulators, the actual regulatory machinery, combinatorially associate. This accounts readily for the common observation that TRFs can activate transcription in one context and repress it in another. In this sense, NR *activities* reflect its molecular conformations, which emerge owing to context-specific cues, whereas NR *functions* are integrated regulatory outcomes of coregulatory enzyme actions that associate with those various conformations. Thus, we suggest that NR structure, together with its determinants, are keys to understanding both its regulatory precision and plasticity. These ideas likely extend, at least conceptually, to other eukaryotic transcriptional regulatory factors[4], all of which face the same precision/plasticity challenge.

Remaining Questions and Future Directions

Can the oligomeric state of GR distinguish transactivation from transrepression?

The models of transcriptional regulation by GR have consistently shown GR homodimers to be required for transactivation and monomeric GR preferred for gene repression. As described above, x-ray crystallographic, NMR, and MD analyses of the GR

DBD bound to GBSs and IR-GBSs clearly show this dimer/ monomer difference. Furthermore, work with GR dimerization deficient mutants, GRdim (GR A458T) and GRmon (A458T/I638A) solidified this concept. Ala458 makes a hydrogen bond with Ile483 on the dimer partner, and mutation of the Ala to Thr has been shown to alter GR activity in a gene-specific manner[14,27,28]. This mutation, though initially believed to be devoid of dimerization potential, still forms dimers on DNA but with diminished cooperativity[9]. To counter this issue another mutation, I638A, was made within the LBD to generate GR-mon (A458T/I638A). GRmon has been shown to be overwhelmingly monomeric in cells and maintained the ability to repress NF-κB activity[29]. This work agrees with the monomeric hypothesis by showing that not only does GR DBD crystallize as a monomer on TREs, but NMR analysis shows residues within the dimerization loop to not be perturbed when binding a TRE (**Figure 3.4**). We also show that GRdim and GRmon repressed our luciferase gene reporters slightly better than WT GR (**Figure 3.4e**).

To expand this hypothesis we aim to develop a constitutively active GR dimer with the goals of testing if this dimeric GR would preferentially activate gene transcription. The GR (both full-length and isolated domains) are monomers in solution. Therefore, to accomplish this goal we aimed to create a new surface within the GR LBD that would promote dimerization. Unlike the 3-keto SRs, the estrogen receptor (ER) is a dimer in solution (**Figure 6.2a**). The ER LBD structure shows H8, H9, H10, and loops 8-9 from each monomer interacts to form a homodimer (**Figure 6.2a,1.6a**)[30]. Therefore, we decided to create a GR LBD chimera that had elements of the dimerization interface of ER, to generate a new mutant called DIMER (**Figure 6.2b,c**). Preliminary data with this GR mutant revealed promising results (**Figure 6.2d,e**). In contrast to GRmon, which is a strong repressor of

inflammatory gene reporters, GR DIMER drove expression of the luciferase gene reporter better than WT GR. Further work with this mutant will be required to ensure that a dimeric GR is forms in cells.

Are GR selective gene modulators possible?

In the 1940s, GR's endogenous ligand, cortisol, was identified as a potent suppressor of inflammation[31]. Since then, synthetic GCs with dramatically improved affinity, potency and selectivity have become the most widely used treatment for anti-inflammatory therapies[32]. Targeting GR is not challenging, however, current GCs drive the activation of metabolic, homeostatic and growth pathways in addition to immunosuppression[33,34]. It is GR's role in transactivation that is associated with the numerous debilitating side effects associated with GC treatment. Therefore, researchers have worked to develop GCs that will allow GR to transrepress but not transactivate. These selective gene modulators are a popular idea but have thus far been unsuccessful.

The difficulty lies in the fact that GR regulates vastly different gene programs depending on the context. This is a main theme of Chapter 2. Developing improved GCs with less off target side effects will require linking ligand-induced receptor motions with selective coregulator interactions to drive GR to specific promoters[16]. We add to this work by solving the structure of GR bound to a clinically relevant ligand, TA (**Chapter 4**). As exemplified for ER, comparing ER LBD structures bound to 20 different ligands, subtle allosteric changes within LBDs could be linked to a range of ER activity[35]. Obtaining more structural and dynamic information for GR will be vital to this effort.

What techniques are on the horizon for expanding our understanding of NRs.

Full-Length Structural Studies: NRs contain a modular domain architecture consisting of 4 main domains (**Figure 1.1**): an unstructured N-terminal domain (NTD) which contains the activation function-1 (AF-1) region which interacts with coregulator proteins; a zinc finger DNA binding domain (DBD); a hinge region; and a ligand-binding domain (LBD) which contains the ligand-sensitive AF-2 surface that also enables interaction with coregulators[36,37]. Until recent advances, only the DBD and LBD were amenable to structural analysis. However, we have learned a great deal from these structures. Yet, advancements in single particle cryo-electron microscopy (EM) and x-ray crystal structures of a few intact NRs have revealed how these the intact proteins interact with DNA (**Figure 1.1c**)[38]. The first atomic resolution structure was PPAR-RXR bound to a DR1 RE[39]. Since then another two high resolution structures have been solved including HNF-4 homodimer and the LXR-RXR heterodimer bound to DR DNA REs[40,41]. These structures revealed that the DBDs and LBDs directly interact, that the "common NR architecture" is not so common, and even bound to the same DNA element the NR complexes can look different. More intact NR structures will be required to expand our understanding of allosteric control of NRs.

CRISPR-Cas: Genome editing using zinc finger nucleases or transcription activator-like effector nucleases (TALENs) provide, in principle, routes for assaying nuclear receptor response element (RE) activity *in situ.* Used in conjunction with ChIP, qPCR, RNA-seq, transcriptomics, or other *in vivo* assays, can validate RE activity and identify RE target gene. Unfortunately, the technical complexity of these methods, have left these approaches largely unused. For example, only one single glucocorticoid response element (GRE) has been validated at its endogenous locus *in vivo*. In that case, a short deletion introduced into the

first intron of the mouse circadian clock gene *Per2* fortuitously covered a GOR ~25kb downstream from the transcription start site, and produced allele-specific loss of glucocorticoid-mediated induction of *Per2* expression in mesenchymal stem cells[42]. With the emergence of CRISPR-Cas technologies[43], precise genome editing will rapidly become the gold standard for validation of NR RE activity in any chosen cell type- or physiological-context.

Concluding Remarks

The research presented in this work utilized structural biology to explore how allosteric modulators can have a drastic effect on NR function. To this end, we present an intriguing model that NR, and likely transcription factors as a whole, simply function as scaffolds to which these modulators act upon. To this end, this suggest that these modulators ultimately influence whether a gene is turned off or on. Looking forward, applying this model will be critical as we work towards understanding the highly complex process of gene regulation. NRs are an important family of proteins that are useful tools to probe these difficult scientific questions.

Figures

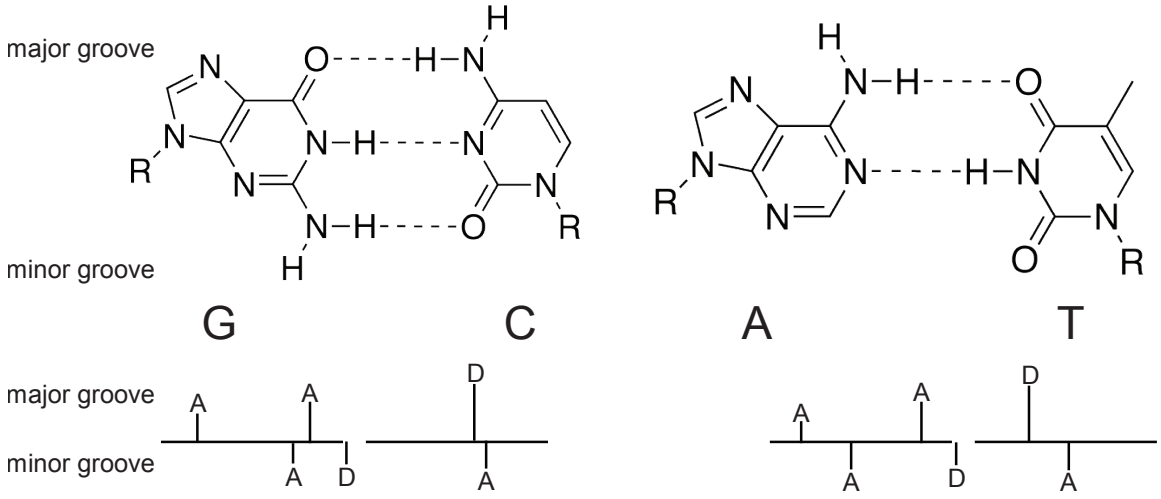


Figure 6.1: G-C and A-T base pairs.

The top edge represents the portion exposed to the major groove and the bottom would be exposed in the minor groove. Below there is a schematic indicating the relative locations of the hydrogen bond donors (D) and acceptors (A) within the major and minor grooves.

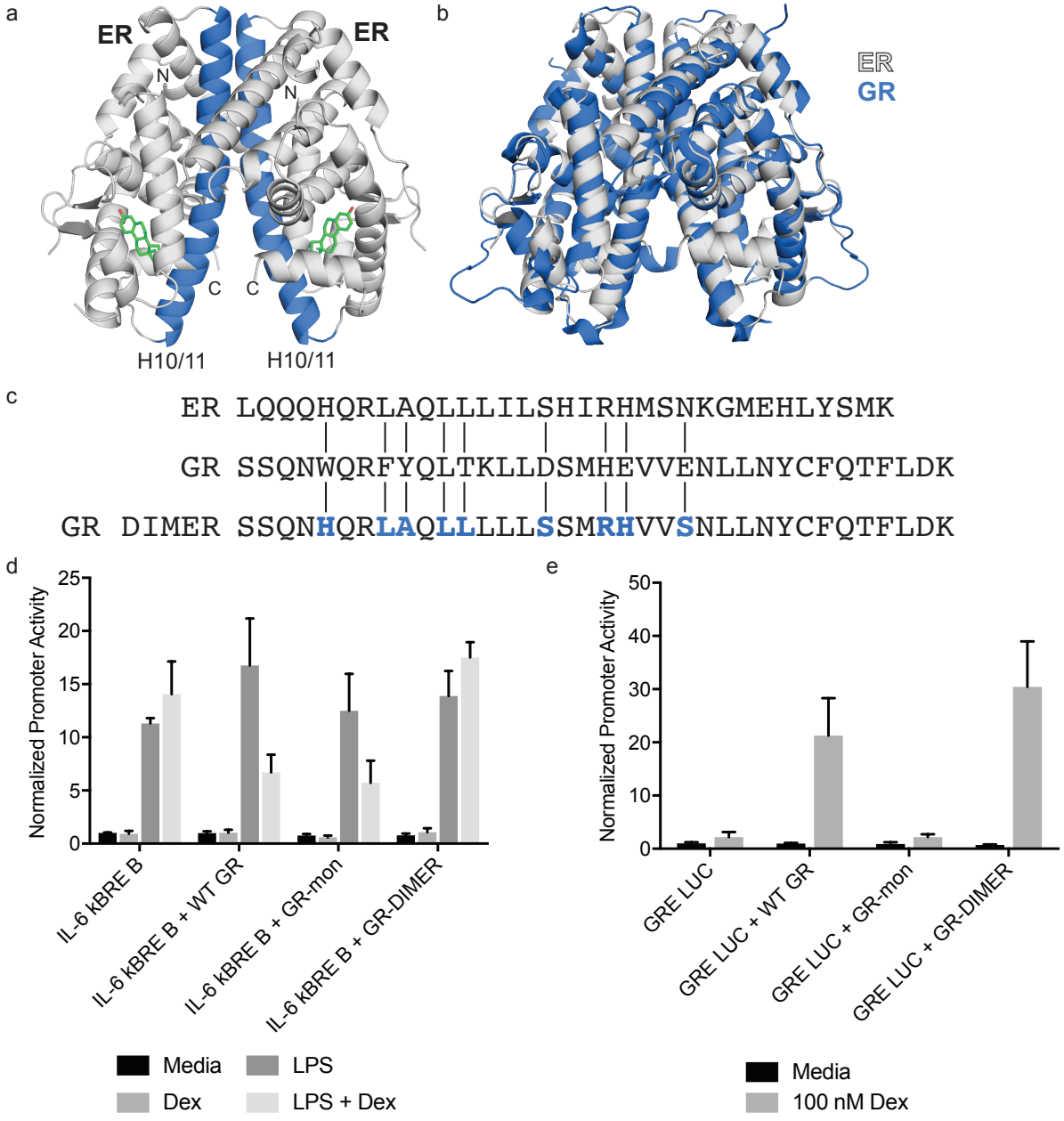


Figure 6.2: Generation of GR DIMER.

(a) Overall cartoon structure of the ER LBD homodimer. Helix 10/11 (colored blue) are the main face of the dimerization interface. Ligand shown as green sticks. (b) ER LBD

homodimer (gray) with two copies of the monomeric GR LBD (blue) overlaid. (c) The amino acid sequences from H10/11 from ER and GR were compared and mutations (bold/blue) were made to create an ER like interface. (d) Preliminary data using a luciferase reporter that contains a portion of the interleukin-6 (*IL6*) promoter, which has a NF-κB response element (κBRE). Treatment with lippopolysaccride (LPS) stimulates NF-κB to activate luciferase expression. With subsequent dexamethasone treatment, which activates GR, the luciferase reporter is repressed. This repression is only seen upon co-transfection with WT GR or the monomeric GR mutant, GRmon, but not in the presence of DIMER. (e) Preliminary data using a constitutively active luciferase reporter that contains a repeating glucocorticoid response element (GRE). Upon activation by Dex, WT GR and DIMER can drive expression of the luciferase reporter. GRmon is unable to transactivate.

References

1. Hudson, William H, and Eric A Ortlund. 2014. 'The Structure, Function and Evolution of Proteins That Bind DNA and RNA', *Nature Reviews Molecular Cell Biology*, 15.11 (Nature Publishing Group): 749–60.
2. Meijsing, Sebastiaan H, Miles A Pufall, Alex Y So, Darren L Bates, Lin Chen, and others. 2009. 'DNA Binding Site Sequence Directs Glucocorticoid Receptor Structure and Activity', *Science*, 324.5925: 407–10.
3. Watson, Lisa C, Kristopher M Kuchenbecker, Benjamin J Schiller, John D Gross, Miles A Pufall, and others. 2013. 'The Glucocorticoid Receptor Dimer Interface Allosterically Transmits Sequence-Specific DNA Signals', *Nature Structural & Molecular Biology*, 20.7: 876–83.
4. Lefstin, Jeffrey A, and Keith R Yamamoto. 1998. 'Allosteric Effects of DNA on Transcriptional Regulators.', *Nature*, 392.6679: 885–88.
5. Weikum, E. R., M. T. Knuesel, E. A. Ortlund and K. R. Yamamoto (2017). "Glucocorticoid receptor control of transcription: precision and plasticity via allostery." Nat Rev Mol Cell Biol.
6. Luisi, B. F., W. X. Xu, Z. Otwinowski, L. P. Freedman, K. R. Yamamoto and P. B. Sigler (1991). "Crystallographic analysis of the interaction of the glucocorticoid receptor with DNA." Nature **352**(6335): 497-505.
7. Meijsing, S. H., M. A. Pufall, A. Y. So, D. L. Bates, L. Chen and K. R. Yamamoto (2009). "DNA binding site sequence directs glucocorticoid receptor structure and activity." Science **324**(5925): 407-410.

8. Hudson, W. H., B. R. Kossmann, I. M. de Vera, S. W. Chuo, E. R. Weikum, G. N. Eick, J. W. Thornton, I. N. Ivanov, D. J. Kojetin and E. A. Ortlund (2016). "Distal substitutions drive divergent DNA specificity among paralogous transcription factors through subdivision of conformational space." Proc Natl Acad Sci U S A **113**(2): 326-331.
9. Meijsing, Sebastiaan H, Miles A Pufall, Alex Y So, Darren L Bates, Lin Chen, and others. 2009. 'DNA Binding Site Sequence Directs Glucocorticoid Receptor Structure and Activity', *Science*, 324.5925: 407–10.
10. Surjit, M., K. P. Ganti, A. Mukherji, T. Ye, G. Hua, D. Metzger, M. Li and P. Chambon (2011). "Widespread negative response elements mediate direct repression by agonist-liganded glucocorticoid receptor." Cell **145**(2): 224-241.
11. Hudson, W. H., C. Youn and E. A. Ortlund (2013). "The structural basis of direct glucocorticoid-mediated transrepression." Nat Struct Mol Biol **20**(1): 53-58.
12. De Bosscher, K., W. Vanden Berghe and G. Haegeman (2003). "The interplay between the glucocorticoid receptor and nuclear factor-kappaB or activator protein-1: molecular mechanisms for gene repression." Endocr Rev **24**(4): 488-522.
13. Herrlich, P. (2001). "Cross-talk between glucocorticoid receptor and AP-1." Oncogene **20**(19): 2465-2475.
14. Reichardt, H. M., K. H. Kaestner, J. Tuckermann, O. Kretz, O. Wessely, R. Bock, P. Gass, W. Schmid, P. Herrlich, P. Angel and G. Schutz (1998). "DNA binding of the glucocorticoid receptor is not essential for survival." Cell **93**(4): 531-541.
15. Karin, M. (1998). "New twists in gene regulation by glucocorticoid receptor: is DNA binding dispensable?" Cell **93**(4): 487-490.

16. Weikum, E. R., M. L. Tuntland, M. N. Murphy and E. A. Ortlund (2016). "A Structural Investigation into Oct4 Regulation by Orphan Nuclear Receptors, Germ Cell Nuclear Factor (GCNF), and Liver Receptor Homolog-1 (LRH-1)." J Mol Biol **428**(24 Pt B): 4981-4992.
17. Zhang, Jun, Michael J Chalmers, Keith R Stayrook, Lorri L Burris, Yongjun Wang, and others. 2011. 'DNA Binding Alters Coactivator Interaction Surfaces of the Intact VDR–RXR Complex', *Nature Structural & Molecular Biology*, 18.5: 556–63.
18. Moras, D. and H. Gronemeyer (1998). "The nuclear receptor ligand-binding domain: structure and function." Curr Opin Cell Biol **10**(3): 384-391.
19. Weatherman, R. V., R. J. Fletterick and T. S. Scanlan (1999). "Nuclear-receptor ligands and ligand-binding domains." Annu Rev Biochem **68**: 559-581.
20. Xu, E. H. and M. H. Lambert (2003). "Structural insights into regulation of nuclear receptors by ligands." Nucl Recept Signal **1**: e004.
21. Bledsoe, Randy K., Valerie G. Montana, Thomas B. Stanley, Chris J. Delves, Christopher J. Apolito, and others. 2002. 'Crystal Structure of the Glucocorticoid Receptor Ligand Binding Domain Reveals a Novel Mode of Receptor Dimerization and Coactivator Recognition', *Cell*, 110: 93–105.
22. He, Yuanzheng, Wei Yi, Kelly Suino-Powell, X Edward Zhou, W David Tolbert, and others. 2014. 'Structures and Mechanism for the Design of Highly Potent Glucocorticoids.', *Cell Research*, 24.6 (Nature Publishing Group): 713–26.
23. Kauppi, Björn, Clarissa Jakob, Mathias Färnegårdh, Jie Yang, Harri Ahola, and others. 2003. 'The Three-Dimensional Structures of Antagonistic and Agonistic Forms of the Glucocorticoid Receptor Ligand-Binding Domain: RU-486 Induces a

Transconformation That Leads to Active Antagonism.', *The Journal of Biological Chemistry*, 278.25: 22748–54.

24. Ricketson, D., U. Hostick, L. Fang, K.R. Yamamoto, and B.D. Darimont. 2007. 'A Conformational Switch in the Ligand-Binding Domain Regulates the Dependence of the Glucocorticoid Receptor on Hsp90', *Journal of Molecular Biology*, 368.3: 729–41.
25. Chen, Sheng-Hong, Kiriko Masuno, Samantha B Cooper, and Keith R Yamamoto. 2013. 'Incoherent Feed-Forward Regulatory Logic Underpinning Glucocorticoid Receptor Action', *Proceedings of the National Academy of Sciences*, 110.5: 1964–69.
26. Wang, Jen-Chywan, Nilesh Shah, Carlos Pantoja, Sebastiaan H. Meijsing, Joseph D. Ho, and others. 2006. 'Novel Arylpyrazole Compounds Selectively Modulate Glucocorticoid Receptor Regulatory Activity.', *Genes & Development*, 20.6: 689–99.
27. Heck, S, M Kullmann, A Gast, H Ponta, H J Rahmsdorf, and others. 1994. 'A Distinct Modulating Domain in Glucocorticoid Receptor Monomers in the Repression of Activity of the Transcription Factor AP-1', *The EMBO Journal*, 13.17: 4087–95.
28. Schiller, Benjamin J, Rajas Chodankar, Lisa C Watson, Michael R Stallcup, and Keith R Yamamoto. 2014. 'Glucocorticoid Receptor Binds Half Sites as a Monomer and Regulates Specific Target Genes', *Genome Biology*, 15.8: 418.
29. Presman, D. M., M. F. Ogara, M. Stortz, L. D. Alvarez, J. R. Pooley, R. L. Schiltz, L. Grontved, T. A. Johnson, P. R. Mittelstadt, J. D. Ashwell, S. Ganesan, G. Burton, V. Levi, G. L. Hager and A. Pecci (2014). "Live cell imaging unveils multiple domain requirements for in vivo dimerization of the glucocorticoid receptor." PLoS Biol **12**(3): e1001813.

30. Brzozowski, A. M., A. C. Pike, Z. Dauter, R. E. Hubbard, T. Bonn, O. Engstrom, L. Ohman, G. L. Greene, J. A. Gustafsson and M. Carlquist (1997). "Molecular basis of agonism and antagonism in the oestrogen receptor." Nature **389**(6652): 753-758.

31. Kendall, E. C. (1951). "The development of cortisone as a therapeutic agent." Antibiot Chemother (Northfield) **1**(1): 7-15.

32. Clark, A. R. and M. G. Belvisi (2012). "Maps and legends: the quest for dissociated ligands of the glucocorticoid receptor." Pharmacol Ther **134**(1): 54-67.

33. Cato, A. C. and E. Wade (1996). "Molecular mechanisms of anti-inflammatory action of glucocorticoids." Bioessays **18**(5): 371-378.

34. De Bosscher, K., W. Vanden Berghe and G. Haegeman (2000). "Mechanisms of anti-inflammatory action and of immunosuppression by glucocorticoids: negative interference of activated glucocorticoid receptor with transcription factors." J Neuroimmunol **109**(1): 16-22.

35. Nwachukwu, J. C., S. Srinivasan, N. E. Bruno, J. Nowak, N. J. Wright, F. Minutolo, E. S. Rangarajan, T. Izard, X. Q. Yao, B. J. Grant, D. J. Kojetin, O. Elemento, J. A. Katzenellenbogen and K. W. Nettles (2017). "Systems Structural Biology Analysis of Ligand Effects on ERalpha Predicts Cellular Response to Environmental Estrogens and Anti-hormone Therapies." Cell Chem Biol **24**(1): 35-45.

36. Carson-Jurica, M. A., W. T. Schrader and B. W. O'Malley (1990). "Steroid receptor family: structure and functions." Endocr Rev **11**(2): 201-220.

37. Kumar, R. and E. B. Thompson (1999). "The structure of the nuclear hormone receptors." Steroids **64**(5): 310-319.

38. Rastinejad, F., V. Ollendorff and I. Polikarpov (2015). "Nuclear receptor full-length architectures: confronting myth and illusion with high resolution." Trends Biochem Sci **40**(1): 16-24.PMID 19043829

39. Lou, X., G. Toresson, C. Benod, J. H. Suh, K. J. Philips, P. Webb and J. A. Gustafsson (2014). "Structure of the retinoid X receptor alpha-liver X receptor beta (RXRalpha-LXRbeta) heterodimer on DNA." Nat Struct Mol Biol **21**(3): 277-281.

40. Chandra, V., P. Huang, N. Potluri, D. Wu, Y. Kim and F. Rastinejad (2013). "Multidomain integration in the structure of the HNF-4alpha nuclear receptor complex." Nature **495**(7441): 394-398.

41. So, A. Y.-L., Teresita U Bernal, Marlisa L Pillsbury, Keith R Yamamoto, and Brian J Feldman. 2009. 'Glucocorticoid Regulation of the Circadian Clock Modulates Glucose Homeostasis', *Proceedings of the National Academy of Sciences*, 106.41: 17582–87.

42. Ran, F Ann, Patrick D Hsu, Jason Wright, Vineeta Agarwala, David A Scott, and others. 2013. 'Genome Engineering Using the CRISPR-Cas9 System', *Nature Protocols*, 8.11: 2281–2308