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:

_____ _____

Jeffrey A. Kohn Date

Novel Allosteric Communication in Nuclear Receptor Activation

By

Jeffrey A. Kohn
Doctor of Philosophy

Graduate Division of Biological and Biomedical Sciences
Molecular and Systems Pharmacology

_____
Eric A. Ortlund, Ph.D.
Advisor

_____
Christine Dunham, Ph.D.
Committee Member

_____
T.J. Murphy, Ph.D.
Committee Member

_____
Roy Sutliff, Ph.D.
Committee Member

Accepted:

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

_____
Date

Novel Allosteric Communication in Nuclear Receptor Activation

By

Jeffrey A. Kohn
Bachelor of Science, Binghamton University, The State University of New York, 2009

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 Sciences
Molecular and Systems Pharmacology
2016

Abstract

Novel Allosteric Communication in Nuclear Receptor Activation

By

Jeffrey A. Kohn

Nuclear receptors (NRs) are highly important targets for the pharmacological management of many human conditions, including cancer, inflammation, autoimmunity, metabolic syndrome, thyroid dysfunction, and reproduction. Unfortunately, NR-targeting drugs often cause significant adverse reactions. NRs are a highly interrelated family of transcription factors that are evolutionarily descended from a common ancestor, from which they inherited a conserved structural fold and mode of activity. NR activation involves the recruitment of coactivator and corepressor proteins, which promote the transcription or repression of the target gene. Thus, side-effects arise from the off-target action of a drug on close relatives of its target, and the inability of the drug to selectively control the transactivation or transrepression of the therapeutically relevant genes. To address these problems, the present work considers the structural biology of the NR family from the perspective of their molecular evolution in order to elucidate the structural mechanisms that drive ligand-regulated receptor activity in two receptors, the corticosteroid receptors (consisting of the glucocorticoid and mineralocorticoid receptors, GR and MR) and liver receptor homolog-1 (LRH-1). In all of these receptors, ligand binding affected, and was affected, by interaction of residues in a flexible region at the bottom of the receptor. In the corticosteroid receptors, mutations at this region toggled the activity of synthetic glucocorticoids between agonist and antagonist, without affecting the activity of endogenous hormones. In LRH-1, which is modulated by phospholipids (PLs), the length of the PL tails differentially stabilized this region, allowing for the selective recruitment of coactivators or corepressors. Molecular dynamics simulations demonstrated that this region was in allosteric communication with the coregulator binding surface, permitting the binding of varying ligands to promote the selective recruitment of coactivators or corepressors. Thus, this region constitutes a novel, alternate activation function in the NR ligand binding domain that may be exploitable by next generation drugs in order to improve their therapeutic profile.

Novel Allosteric Communication in Nuclear Receptor Activation

By

Jeffrey A. Kohn
Bachelor of Science, Binghamton University, The State University of New York, 2009

Advisor: Eric A. Ortlund, Ph.D.

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 Sciences
Molecular and Systems Pharmacology
2016

Table of Contents

List of Figures

List of Tables

List of abbreviations

11-βHSD: 11-β-hydroxysteroid desaturase

15-HETE: 15-Hydroxyeicosatetraenoic acid

15-KETE: 15-Ketoeicosatetraenoic acid

β-CD: β-Cyclodextrin

AA: Arachadonic acid

A-AF: Alternate activation function

AF-1: Activation function-1

AF-2: Activation function-2

AF-H: Activation function helix (helix 12, H12)

AGR: Ancestral gene resurrection

AncCR: Ancestral corticosteroid receptor

AncGR1: Ancestral glucocorticoid receptor-1

AncGR2: Ancestral glucocorticoid receptor-2

Apo: Unliganded, *i.e.* empty ligand binding pocket

AncSR1: Ancestral steroid receptor-1

AncSR2: Ancestral steroid receptor-2

AncSR3: Ancestral steroid receptor-3

AP-1: Activator protein-1

APS: Advanced photon source

AR: Androgen receptor

ATF2: Activating transcription factor 2

ANOVA: Analysis of Variance

CASTp: Computed Atlas of Surface Topography of proteins

CBP: CREB binding protein

CCP4: Collaborative Computational Project Number 4

CD1: Cluster of differentiation 1

CHAPS: 3-[(3-Cholamidopropyl)dimethylammonio]-1-propanesulfonate

CHARMM: Chemistry at Harvard Macromolecular Mechanics

CHO-K1: Chinese hamster ovary cells, strain K1

CMV: Cytomegalovirus

COOT: Crystallographic Object-Oriented Toolkit

CoRNR: Nuclear receptor corepressor motif

CTE: C-terminal extension

CYP: Cytochrome P450

DAG: Diacylglycerol

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

DBD: DNA-binding domain

DCIA: 7-diethylamino-3-((4'-(iodoacetyl)amino)phenyl)-4-methylcoumarin

Dex: Dexamethasone

DF: Dexamethasone-fluorescein

DHA: Docosahexaenoic acid

DLPC: Dilauroyl phosphatidylcholine

DLPE: Dilauroyl phosphatidylethanolamine

DMEM: Dulbecco's Modified Eagle Medium

DMSO: Dimethyl sulfoxide

DPPC: Dipalmitoyl phosphatidylcholine

DPPG: Dipalmitoyl phosphatidylglycerol

DTT: Dithiothreitol

DUPC: Diundecanoyl phosphatidylcholine

EDTA: Ethylenediaminetetraacetic acid

ENaC: Epithelial sodium channel

ER: Estrogen receptor

ESI: Electrospray ionization

FFA: Free fatty acids

FRET: Förster resonance energy transfer

FKBP52: FK506-binding protein 4

Ftz-F1: Fushi tarazu transcription factor 1

GAL4: Yeast transcription factor GAL4

GLUT4: Glucose transporter type 4

GPS2: G-protein pathway suppressor 2

GR: Glucocorticoid receptor

GRE: Glucocorticoid response element

GSK: GlaxoSmithKline

H#: Helix #

HAT: Histone acetyltransferase

HDAC: Histone deacetylace

HDX: Hydrogen-Deuterium exchange

HEK: Human embryonic kidney cells

HEPES: 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid

HIV: Human Immunodeficiency Virus

HNF-4α: Hepatocyte nuclear factor 4α

Holo: Liganded, *i.e.* occupied ligand binding pocket

Hop: HSP organizing protein

HPA: Hypothalamic pituitary adrenal axis

HRE: Hormone response element

Tukey HSD: Tukey's honest significant difference test

HSP: Heat shock protein

ID: Interaction domain (proteins); insertion device (crystallography)

IP_3: Inositol trisphosphate

IPMK: Inositol polyphosphate multikinase

IPTG: Isopropyl β-D-1-thiogalactopyranoside

Krpm: 1000x revolutions per minute

LBD: Ligand-binding domain

LBP: Ligand-binding pocket

LIC: Ligation independent coning

LXR: Liver X receptor

m/z: mass to charge ratio

MBP: Maltose binding protein

MD: Molecular dynamics

ML: Maximum likelihood

MOF: Mometasone furoate

MR: Mineralocorticoid receptor

MS: Mass spectrometry

NAMD: Nanoscale Molecular Dynamics

NBD: 7-nitro-2,1,3-benzoxadiazole

NF-κB: nuclear factor kappa-light-chain-enhancer of activated B cells

NTD: N-terminal domain

NR: Nuclear receptor

NCoR: Nuclear receptor corepressor-1

OCT-4: octamer-binding transcription factor 4

PAPC: Palmitoyl-arachidonyl phosphatidylcholine

PDK-1: Phosphoinositide-dependent kinase 1

PI4K: Phosphatidylinositol 4-kinase

PIP_n: Phosphatidylinositol phosphate n (n = 2-3)

PKC: Protein kinase C

PLA2: Phospholipase A2

POPC: Palmitoyl-oleoyl phosphatidylcholine

PTEN: Phosphatase and tensin homolog

LPA: Lysophosphatidic acid

LRH-1: Liver receptor homolog-1

PC: Phosphatidylcholine

PC1: Principal component vector 1

PC2: Principal component vector 2

PCA: Principal component analysis

PCTP: Phosphatidylcholine transfer protein

PITP: Phosphatidylinositol transfer protein

PDB: Protein databank

PE: Phosphatidylethanolamine

PG: Phosphatidylglycerol

PI: Phosphatidylinositol

PL: Phospholipid

PLK3: Polo-like kinase 3

PMSF: phenylmethanesulfonyl fluoride

PPAR: Peroxisome proliferator-activated receptor

PPRE: PPAR response element

PR: Progesterone receptor

ProSMART: Procrustes Structural Matching Alignment and Restraints Tool

PS: Phosphatidylserine

PTM: Post-translational modification

RAAS: Renin angiotensin aldosterone system

RAR: Retinoic acid receptor

RESPA: Reversible Reference System Propagator Algorithm

RJW: Richard J. Whitby

r.m.s.d.: Root mean square deviation

r.m.s.i.p.: Root mean square inner product

rpm: revolutions per minute

RXR: Retinoid X Receptor

S1P: Sphingosine-1-phosphate

SAPC: Stearoyl-arachidonyl phosphatidylcholine

S.E.M.: Standard error of the mean

SER-CAT: Southeast Regional Collaborative Access Team

SF-1: Steroidogenic factor-1

SHP: Small heterodimeric partner

SIRT1: Sirtuin 1

SM: Sphingomyelin

SMRT: Silencing mediator for retinoid or thyroid hormone receptors (nuclear receptor corepressor-2)

SNuRM: Selective nuclear receptor modulator

SR: Steroid receptor

SRA: Steroid receptor activator 1

SRC-1: Steroid receptor coactivator-1

SRC-3: Steroid receptor coactivator-3

SRE: Steroid response element

SREBP: Sterol regulatory element binding protein

StAR: Steroidogenic acute regulatory protein

START: StAR related lipid transfer domain

SUMO: Small Ubiquitin-like Modifier

SUV: Small unilamellar vesicles

TBL1X: Transducin β-Like 1X-Linked

TEV: Tobacco etch virus

TF: Transcription factor

TFIID: Transcription factor II D

TFIIF: Transcription factor II F

TIF-2: Transcriptional mediators/intermediary factor-2 (steroid receptor coactivator-2)

TMAO: Trimethylamine-N-oxide

TR: Thyroid hormone receptor

Tris: Tris(hydroxymethyl)aminomethane

USP: Ultraspiracle protein

VDR: Vitamin D receptor

VMD: Visual Molecular Dynamics

Chapter 1: Introduction

Regulation of large gene programs by nuclear receptors

The nuclear receptor superfamily (NRs) is an expansive and diverse class of transcription factors (TFs) that serve as master regulators of embryonic development, cell cycle and differentiation, metabolism, and cell death (1). Dysfunction of NR signaling contributes to many human diseases, including metabolic diseases such as obesity and diabetes, atherosclerosis, infertility, and many forms of cancer (1; 2). There are 48 different NRs in humans, 24 of which have known ligands and 24 of which are orphans, including 9 that have been adopted after identification of a ligand (Table 1.1) (3). The identified NR ligands include xenobiotics, steroids, cholesterol derivatives, retinoids, thyroid hormones, vitamin D, fatty acids, phospholipids, and heme. NR ligands tend to share the common feature of being small, lipophilic molecules that can freely diffuse across the cell membrane and traverse the blood plasma and cytosol on their own or in complex with chaperone proteins, ultimately binding to their cognate NR in order to modulate its activity (3). With very few exceptions, NRs are defined by the presence of a central DNA-binding domain that binds a short sequence of DNA, termed a NR response element, either directly upstream of the NR target gene, or on a distal enhancer element. Each member of the NR family can regulate a large number of genes (for example, ChIP studies have identified thousands of genomic DNA binding sites for ERα alone (4)), and each target gene may be regulated by a number of different NRs. Thus, the role of NRs is to detect and integrate multiple chemical and biological signals in the form of DNA and ligand recognition, post-translational modifications (PTMs), and coregulator availability, and to respond by controlling the expression of their many target genes, thereby controlling with fine precision the myriad of physiological processes that work to sustain the life of the host animal.

Nuclear receptor structure and function

Despite the overwhelming number and diversity of the cellular processes that are regulated by the NR superfamily, the overall structure and basic mechanism of operation by NRs is quite conserved. The general NR architecture consists of several modular domains (Figure 1.1).

Table 1.1. Mammalian Nuclear Receptors: Nomenclature and Ligands

Common Name	*Common Abbreviation*	*Unified Nomenclature*	*Ligands*
Androgen receptor	AR	NR3C4	androgens
Constitutive androstane receptor	CAR	NR1I3	xenobiotics
Chicken ovalbumin upstream promoter-transcription factor α	COUP-TFα	NR2F1	
Chicken ovalbumin upstream promoter-transcription factor β	COUP-TFβ	NR2F2	
Chicken ovalbumin upstream promoter-transcription factor γ	COUP-TFγ	NR2F6	
Dosage-sensitive sex reversal-adrenal hypoplasia congenital critical region on the X chromosome, gene 1	DAX-1	NR0B1	
Estrogen receptor α	ERα	NR3A1	estrogens
Estrogen receptor β	ERβ	NR3A2	estrogens
Estrogen related receptor α	ERRα	NR3B1	
Estrogen related receptor β	ERRβ	NR3B2	
Estrogen related receptor γ	ERRγ	NR3B3	
Farnesoid X receptor α	FXRα	NR1H4	bile acids
Farnesoid X receptor βa	FXRβ	NR1H5	
Germ cell nuclear factor	GCNF	NR6A1	
Glucocorticoid receptor	GR	NR3CI	glucocorticoids
Hepatocyte nuclear factor 4 α	HNF4α	NR2A1	fatty acids
Hepatocyte nuclear factor 4 γ	HNF4γ	NR2A2	fatty acids
Liver receptor homolog-1	LRH-1	NR5A2	phospholipids
Liver X receptor α	LXRα	NR1H3	oxysterols
Liver X receptor β	LXRβ	NR1H2	oxysterols
Mineralocorticoid receptor	MR	NR3C2	mineralocorticoids and glucocorticoids
Nerve-growth-factor-induced gene B	NGF1-B	NR4A1	
Neuron-derived orphan receptor 1	NOR-1	NR4A3	
Nur-related factor 1	NURR1	NR4A2	
Photoreceptor-cell-specific nuclear receptor	PNR	NR2E3	
Peroxisome proliferator-activated receptor α	PPARα	NR1C1	fatty acids
Peroxisome proliferator-activated receptor β/δ	PPARβ/δ	NR1C2	fatty acids
Peroxisome proliferator-activated receptor γ	PPARγ	NR1C3	fatty acids
Progesterone receptor	PR	NR3C3	progesterone
Pregnane X receptor	PXR	NR1I2	endobiotics and xenobiotics
Retinoic acid receptor α	RARα	NR1B1	retinoic acids
Retinoic acid receptor β	RARβ	NR1B2	retinoic acids
Retinoic acid receptor γ	RARγ	NR1B3	retinoic acids
Reverse-Erb α	REV-ERBα	NR1D1	heme
Reverse-Erb β	REV-ERBβ	NR1D2	heme

RAR-related orphan receptor α	RORα	NR1F1	sterols
RAR-related orphan receptor β	RORβ	NR1F2	sterols
RAR-related orphan receptor γ	RORγ	NR1F3	sterols
Retinoid X receptor α	RXRα	NR2B1	9-cis retinoic acid and docosahexaenoic acid
Retinoid X receptor β	RXRβ	NR2B2	9-cis retinoic acid and docosahexaenoic acid
Retinoid X receptor γ	RXRγ	NR2B3	9-cis retinoic acid and docosahexaenoic acid
Steroidogenic factor 1	SF-1	NR5A1	phospholipids
Short heterodimeric partner	SHP	NR0B2	
Tailless homolog orphan receptor	TLX	NR2E1	
Testicular orphan receptor 2	TR2	NR2C1	
Testicular orphan receptor 4	TR4	NR2C2	
Thyroid hormone receptor α	TRα	NR1A1	thyroid hormones
Thyroid hormone receptor β	TRβ	NR1A2	thyroid hormones
Vitamin D receptor	VDR	NR1I1	1α,25-dihydroxyvitamin D3 and lithocholic acid

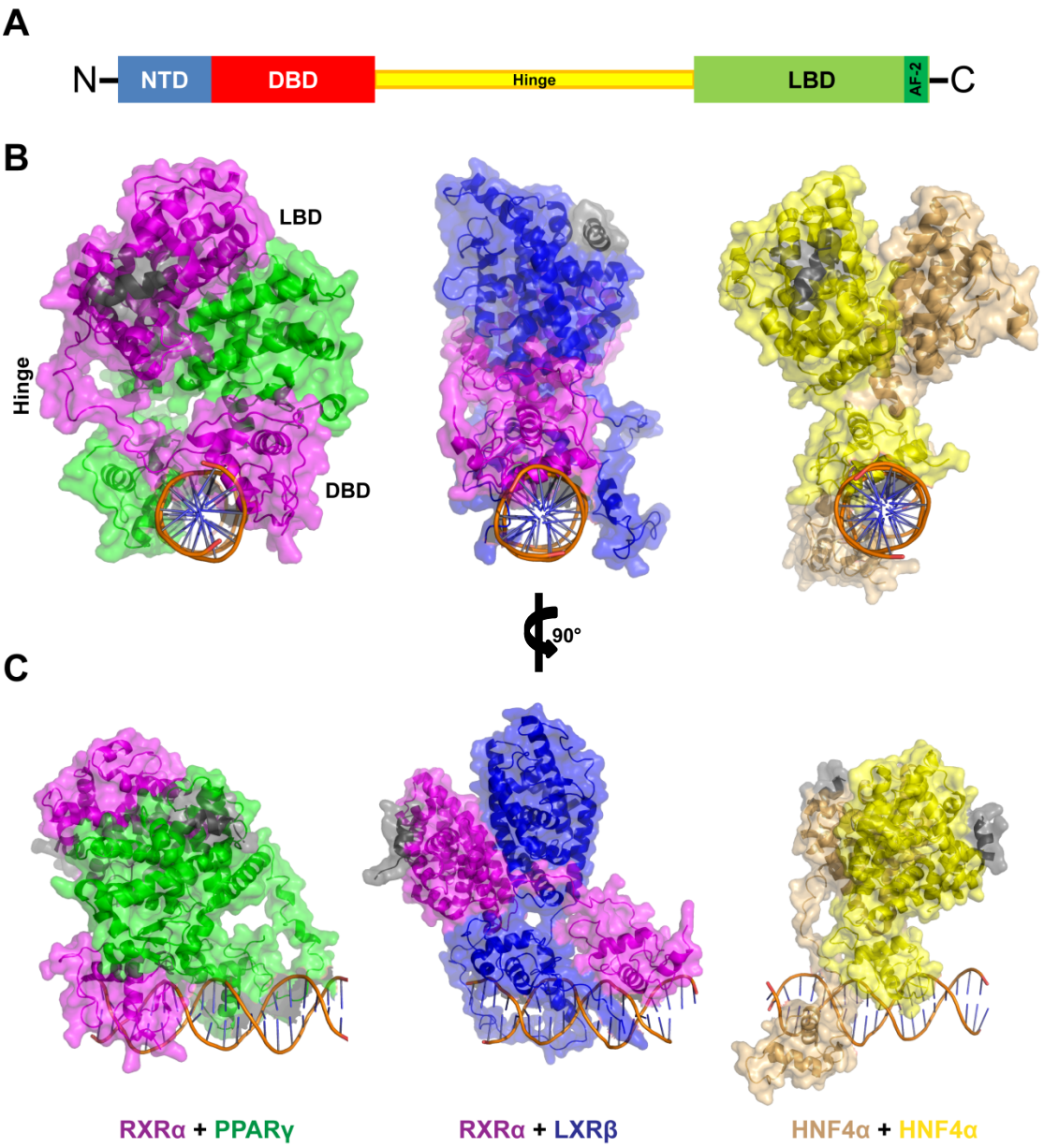

Figure 1.1. General structure of nuclear receptors.

(A) Diagram of the modular structure of the full-length NR. (B) and (C) Views of the NR DBD, hinge, and LBD, bound to DNA and coregulatory peptides. Coregulator peptides are shown in grey. NR domains indicated in panel B. Figures drawn from published X-ray crystal structures, PDBs: 3DZY, 4NQA, and 4IQR (left to right).

The N-terminal domain (NTD) is a loosely structured regulatory region that serves as a site for coregulator binding and PTMs that modulate the activity of the receptor. Following the NTD domain is the structurally-conserved zinc-finger containing DNA binding domain (DBD), which recognizes and binds to DNA sequences containing the NR hormone response element (HRE), a short DNA segment of variable sequence. The hinge region is comprised of a disordered stretch of residues of variable length that not only connects the C-terminus of the DBD to the ligand binding domain (LBD), but also contains additional sites for PTMs. The LBD is a structurally conserved domain, manifesting as an α-helical bundle that surrounds a hydrophobic ligand binding pocket (LBP) to which various small lipophilic molecules can bind. The LBD culminates in helix 12, also known as the activation function helix (H12, AF-H), a short α-helix that docks against helices 3 and 4 of the active receptor to establish the activation function 2 site (AF-2), a hydrophobic cleft that recognizes and binds to coregulator proteins, recruiting them to the transcriptional or repressive complex. Recruitment of coactivators results in transactivation of the downstream gene; recruitment of corepressors results in transrepression. Thus, NRs are able to finely control the transcription of their target genes in response to multiple cellular signals, including hormone binding, PTMs, and the cellular levels of coregulators and other NR binding partners.

NRs can be subdivided into several mechanistic types, I-IV, depending upon the type of HRE they bind and distinguishing features of their mode of activation (Figure 1.2) (5). Type I NRs are the steroid receptors (SRs). These receptors are located in the cytoplasm and complexed with heat shock proteins (HSPs) prior to ligand binding, after which they are shuttled into the nucleus, where they bind as homodimers to a HRE consisting of two inverted repeats. Type II NRs, unlike type I NRs, reside permanently in the nucleus irrespective of ligand binding and bind DNA as heterodimers with retinoid X receptors (RXRs). The binding of an agonist to type II NRs triggers their release from a corepressor complex with SMRT/NCoR and histone deacetylases (HDACs) and the recruitment of coactivators and transcriptional machinery. Type III NRs, like

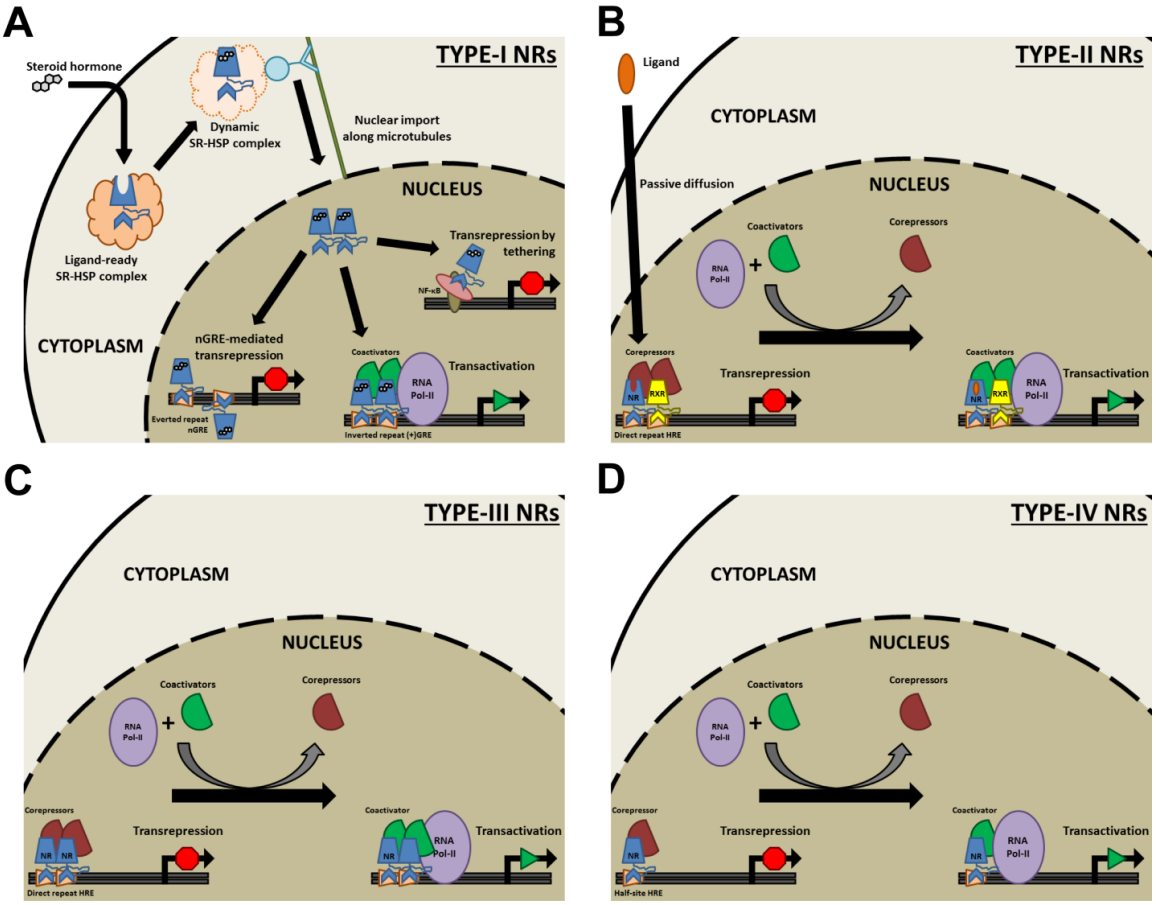

Figure 1.2. Nuclear receptor subtypes and modes of activation.

(A) Activation of type I NRs. (B) Activation of type II NRs. (C) Activation of type III NRs. (D) Activation of Type IV NRs.

type I, homodimerize upon their HREs, but recognize direct repeat sequences instead of the inverted repeat that is recognized by type I NRs. Type IV NRs are distinguished by their monomeric binding to DNA, occupying only a single half-site within their HRE.

Structure and function of the DNA-binding domain

The NR DBD is a small (typically 15-20 kDa), globular domain with a highly conserved tertiary structure, minimally composed of two helices and two distinct zinc finger modules, that recognizes and binds short sequences of DNA upstream of the target gene of the receptor (Figure 1.3 A) (6). The N-terminal helix rests within the major groove of the HRE and makes specific base contacts (Figure 1.3 C), providing the means for recognition of the HRE sequence (7). The second helix lies perpendicular to the first, forming the core of the domain, makes nonspecific contacts with the DNA backbone, and, along with the loop between these two helices, forms the DBD dimerization interface (7). The two zinc fingers are conformationally and functionally distinct from one another, with each coordinating a single Zn^{2+} ion via four cysteine residues. One Zn^{2+} ion is coordinated between the N-terminal loop of the DBD and the major groove helix, while the other Zn^{2+} ion is coordinated at the dimerization interface between the interhelical loop and the helix that forms the core of the domain (6; 7).

In addition to the major groove helix, some NR DBDs contain a C-terminal extension (CTE) immediately preceding the hinge, which can expand the dimerization interface between DBDs or create additional receptor-DNA contacts within the minor groove of the DNA sequence, increasing the specificity of the receptor for certain HREs (Figure 1.3 B, D) (6). In the sex hormone receptors, AR and PR, the CTE is necessary for recognition of specific androgen and progesterone HREs (AREs and PREs) (8; 9), but not for the recognition of general SREs, which can also be bound by GR and MR (10; 11). In these receptors, the CTE enables the high-affinity binding of direct repeat sequences, with SR DBDs lacking the CTE only able to recognize the inverted repeat sequence that defines the classical SREs (8; 12). The CTE is similarly required for

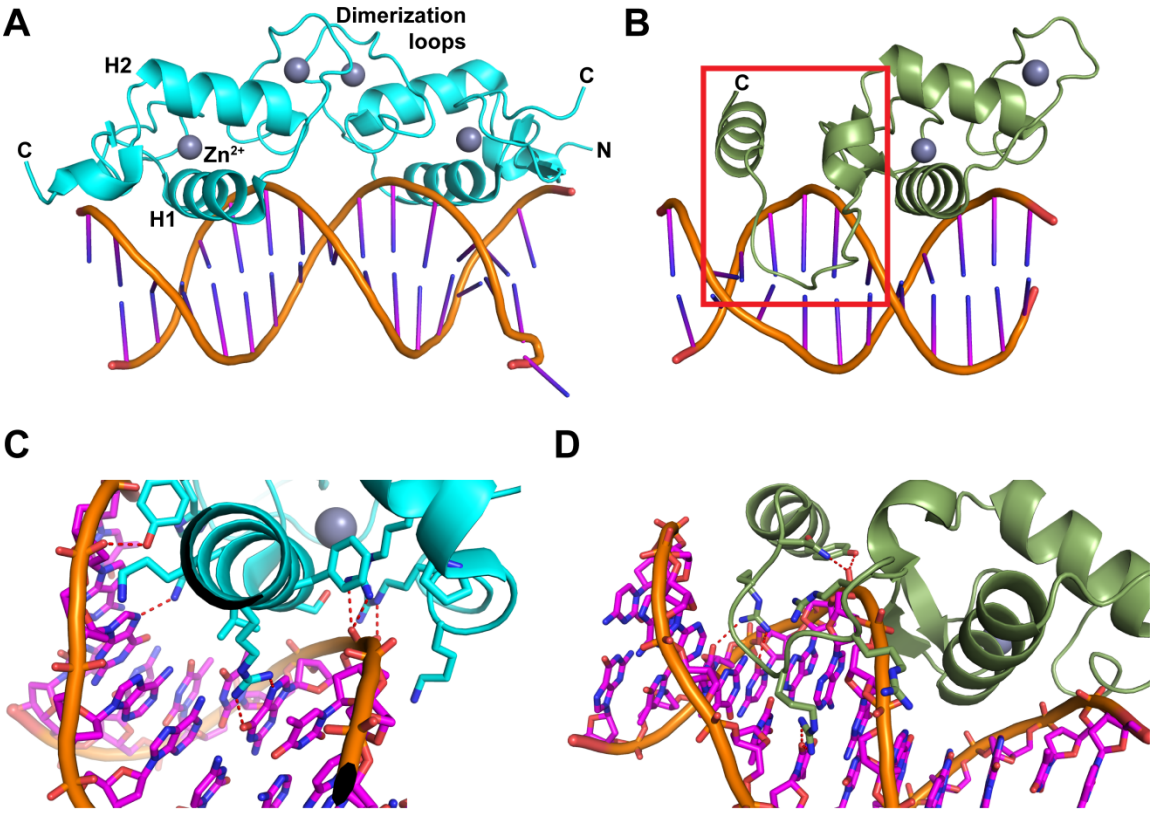

Figure 1.3. Structure of the NR DBD on DNA

(A) Dimeric structure of the GR DBD on a GRE, promoting transactivation. Zn^{2+} ions are shown as grey spheres. PDB: 3G6U. (B) Monomeric structure of the LRH-1 DBD on its HRE, with the CTE boxed in red. PDB: 2A66. (C) Specific molecular contacts made by the GR DBD. PDB: 3G6U. D) Specific molecular contacts made by the LRH-1 CTE. PDB: 2A66.

the high affinity and selective binding of ERα and ERβ to their respective EREs (13). The CTE also creates sequence-specific DNA contacts in NRs that bind DNA as monomers, including the closely related receptors ERR, SF-1, and LRH-1, allowing these receptors to bind as monomers to their HREs with high affinity (14-16). In type II NRs, the CTE can either form specific DNA contacts or contribute to receptor heterodimerization by providing a buffer space between the individual monomers (17-19).

Structure and function of the ligand-binding domain

The NR LBD is often described as a three-layer α-helical "sandwich" (Figure 1.4 A-B). The "upper" half of the LBD (with helix 9 considered by convention to be the top of the domain, see Figure 1.4 A) consists of three layers of helices and is relatively rigid compared to the bottom half (20; 21), which lacks the middle helix and has in its place a large hydrophobic cavity – the ligand binding pocket (LBP) – that can accommodate small lipophilic molecules of highly variable size and shape. The "bottom" of the receptor, defined as the bottom of helices 3, 4, and 10, plus the region comprising helix 5, the β-sheet, helix 6, and helix 7, forms the "mouth" of the LBP. This is a flexible area that allows the unliganded LBD to "breathe", subtly opening and closing in order to facilitate high affinity ligand binding by increasing the accessibility of the LBP (20; 22). Because of its partial instability, the unliganded LBD is often described as a "molten globule" (22). Ligand binding stabilizes the LBD (20), decreasing its sensitivity to proteolysis (23; 24) and backbone proton exchange (25), increasing its thermal melting temperature (26), stabilizing the outer layer of helices relative to the rest of the LBD (27), stabilizing the mouth of the LBP (25), and decreasing its hydrodynamic radius (20). The stabilized LBD then binds to coactivators or corepressors, which serve as scaffolds for the formation of the transcriptional or repressor complex, ultimately resulting in the activation or repression of the target gene.

Coregulator recruitment occurs at the AF-2 surface, a hydrophobic cleft that is formed by the docking of helix 12 against helices 4, 5, and 10. Binding of coregulator to the AF-2 occurs via the recognition of leucine-rich helical motifs, termed interaction domains (IDs), and is secured

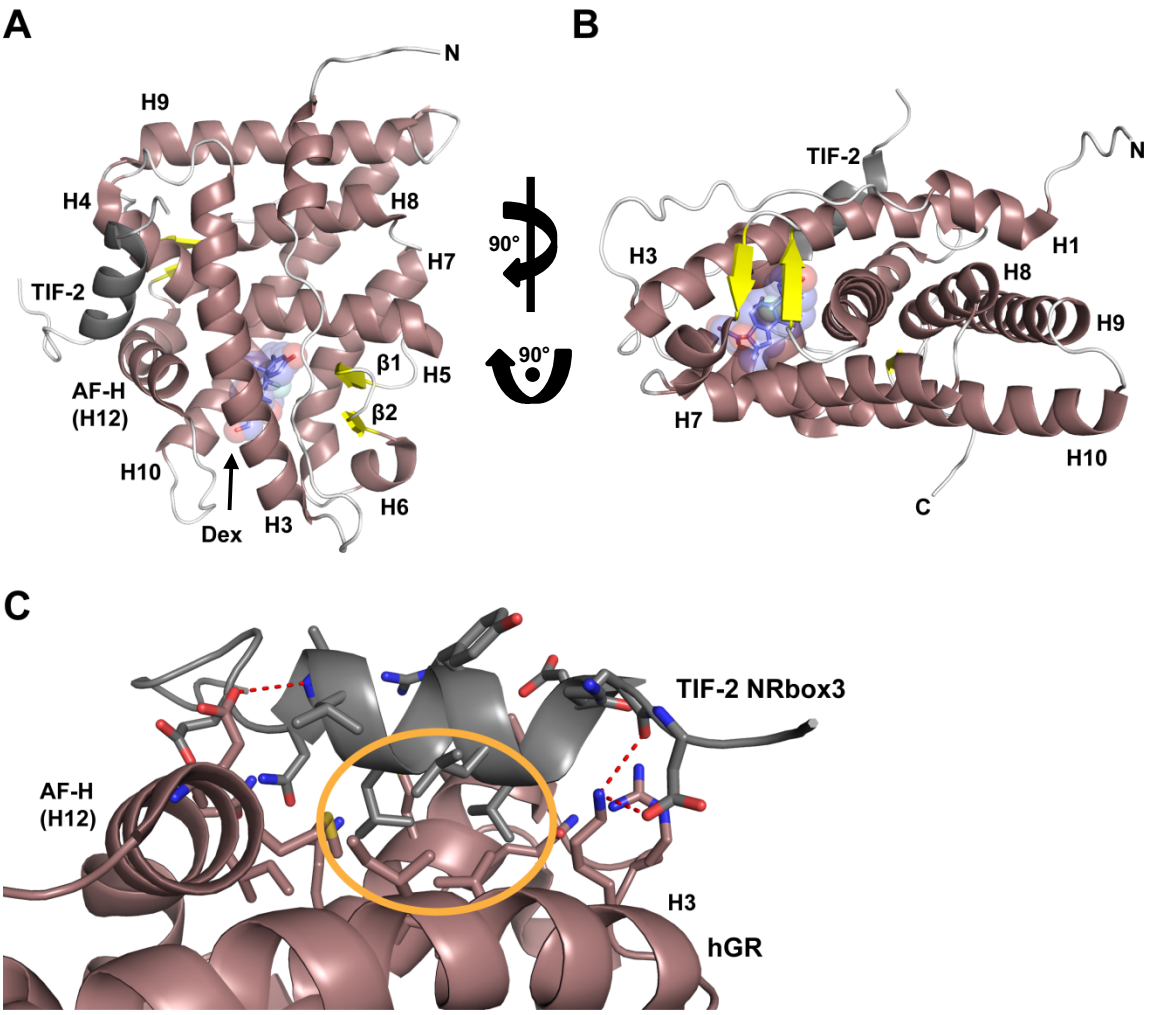

Figure 1.4. Structure of the NR LBD

(A) Front view of the GR LBD in its canonical orientation, with structural features labeled. Bound ligand, dexamethasone (dex), shown in purple. Coregulator peptide is shown in grey. PDB: 1M2Z. (B) Side view of the GR LBD, demonstrating its sandwich-like structure. (C) Close-up view of the AF-2 coregulator binding surface, showing recognition of the LxxLL motif of TIF-2 (circled in orange) and the conserved charge clamp that anchors it to the NR LBD.

via a conserved charge clamp (Figure 1.4 C). In coactivators, the ID motif is LxxLL (where "x" represents any amino acid) (28; 29), and in corepressors, this motif is extended to [I/L]xxIIxxxL (30-34). The different sizes of the coactivator and corepressor ID motifs require H12 to undergo a repositioning in order to best accommodate their binding, resulting in similar but distinct conformations of the activated and repressed NR LBD. The presence or absence of bound ligand, and the nature of the ligand itself, affect the conformation of both the LBD as a whole and the AF-2, and drive the differential recruitment of coactivators or corepressors. Early models of this conformational change, developed based upon the comparison of crystal structures of apo and holo NRs, indicated a dramatic repositioning of H12, with the H12 of apo RXRα and PPARα LBDs being fully displaced from the AF-2 in order to accommodate the extended CoRNR motif (Figure 1.5) (35; 36). This is in comparison to the classical model of the activated conformation of the NR LBD, wherein H12 is securely docked against the other helices of the AF-2, with the LxxLL motif of a coactivator bound in the cleft between H12, H3, and H5 (Figure 1.4 C) (37). The hypothesized repositioning of H12 was described as a "mouse trap" that snaps shut upon the binding of an agonist, contracting the surface area of the AF-2 in order to toggle between corepressor and coactivator recruitment (37).

It is now believed that the mouse-trap model is incorrect, and that the repositioning of H12 between active and repressed conformations involves much more subtle motions within the AF-2, rather than a full dissociation of H12 from the surface of the LBD (38). This revised model is supported by computational and experimental data. For example, molecular dynamics simulations of the binding and release of retinoic acid to RARγ implicated H12 in ligand dissociation, but failed to demonstrate an extended conformation of H12 (39-41). Similar simulations on TR (42-44), ER (45), and VDR (46) likewise suggest that a dramatic repositioning of H12 is not involved in ligand binding. In the case of TR, HDX experiments show that H12 protects regions of the LBD in the presence or absence of ligand, indicating H12 is bound in both

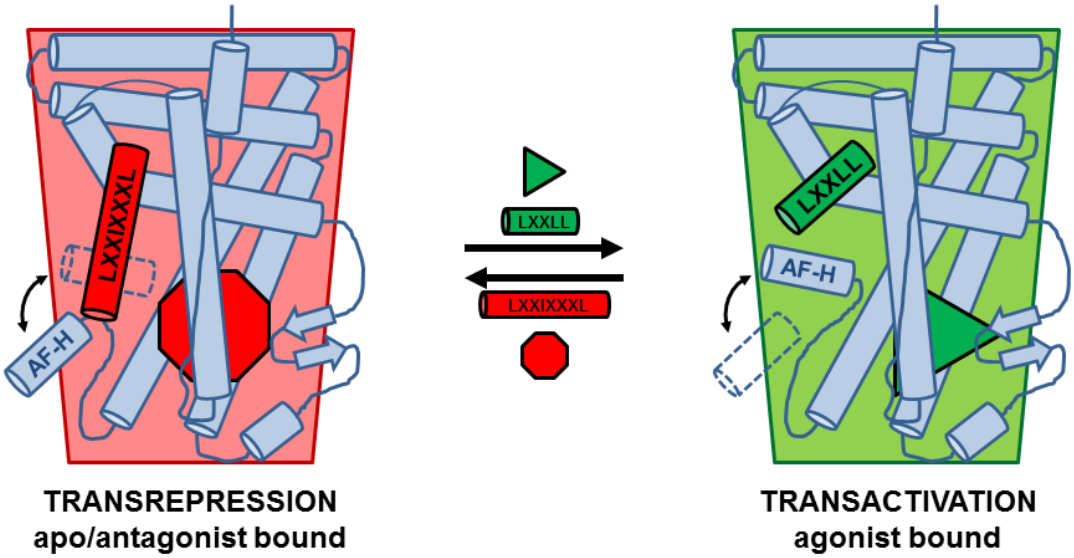

Figure 1.5. Mousetrap model of NR activation

The mousetrap model is based upon a dramatic repositioning of the AF-H in response to the binding of an agonist (green triangle) or antagonist/apo (red octagon). In the active state, the AF-H is securely docked against the side of the LBD, completing the AF-2 surface and enabling the binding of LxxLL-containing coactivators. In the repressed state, the AF-H is separated from the LBD, and corepressors bind via recognition of an extended leucine-rich motif.

cases (47). Furthermore, fluorescence anisotropy experiments of PPARγ LBD fluorescently labeled at its C-terminus demonstrate that H12 remains folded against the LBD in both its apo and holo forms (38; 48). Thus, it is now understood that transactivation and transrepression result from the differential stabilization of the conformationally dynamic LBD in one of two subtly distinct conformations, driven by allosteric communication between the ligand and coregulator (Figure 1.6).

The N-terminal domain and hinge

The NTD assists in recruiting coregulators and other proteins to the activated or repressed NR complex, thereby regulating the activity of the NR as a whole. The NTD exhibits very low sequence homology among the NR family and is highly variable in length, ranging from complete nonexistence to over 500 residues in the steroid receptor family (49). Contained within the NTD of many NRs is the activation function 1 (AF-1), a protein interaction surface composed of one or more stretches of residues that are intrinsically disordered under normal circumstances, but can adopt an induced α-helical secondary structure upon the addition of organic compounds (kosmotropes) such as trifluoroethanol and TMAO (50; 51), or interaction with binding partners, including the coactivators SRC-1 and TIF-2 (47; 52; 53), the corepressors SMRT and NCoR-1 (54), the transcription factors TFIID (55) and TFIIF (56), TATA binding protein (53), and the long noncoding RNA SRA (57). Ordered helical structure can also be directly induced by PTMs (particularly phosphorylation) (58; 59), and indirectly by the binding of the NR DBD to DNA (mediated by JBP2) (60-63) or ligand to the LBD (64). Induced structure within the NTD is strongly and positively correlated with NR activation, and is thought to be required for ligand-independent transactivation (49; 64).

The NR hinge, like the NTD, is a highly unstructured region of variable length that is poorly conserved among NRs (59), Once thought to be merely a flexible linker between the DBD and the LBD, it is now understood that the hinge contributes markedly to receptor activity in several ways. The beginning of the hinge region can make base-specific contacts with DNA,

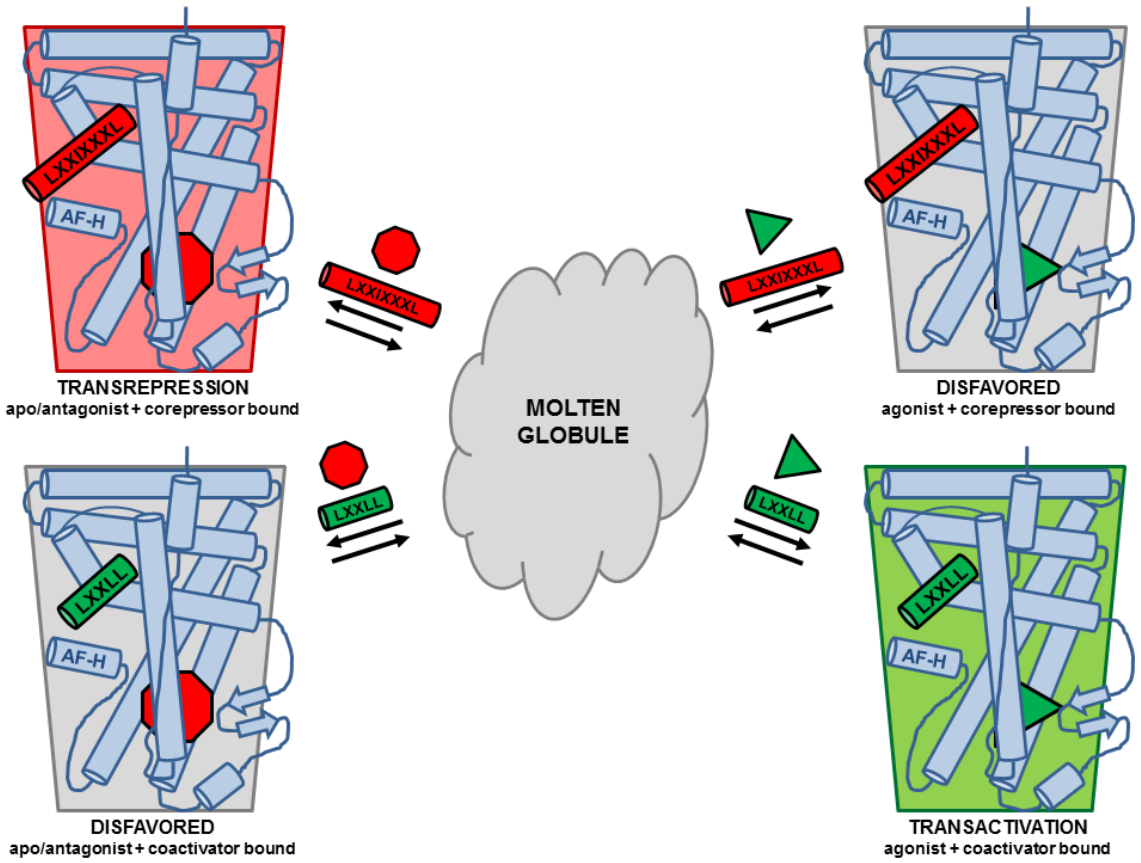

Figure 1.6. Revised model of NR activation

The fully unliganded NR LBD exists in a conformationally unstable molten globule, which becomes stabilized upon the binding of ligand and coregulator. When the bound ligand and coregulator both promote transactivation or transrepression pathways, the NR LBD complex is in ligand/coregulator "agreement", becoming stable and permitting transactivation or transrepression activity. When ligand and coregulator promote opposing pathways, the NR LBD complex is in ligand/coregulator "disagreement", a less stable conformation that does not promote its downstream activity.

enhancing the high affinity binding of NRs to specific HREs (discussed in this text as the C-terminal extension of the DBD). The hinge is enriched in basic residues that serve as a nuclear localization signal (NLS), permitting transit of the NR into the nucleus. The hinge also serves as the site of PTMs, including phosphorylation, acetylation, methylation, SUMOylation, and ubiquitination (59; 65; 66).

Nuclear receptor coregulators

The existence of NR coregulators was first inferred by the observation of squelching. In this phenomenon, the NR-dependent transactivation of a reporter gene was diminished by the transfection of a plasmid expressing a second NR known to transactivate the gene (67). It was hypothesized that transactivation was dependent upon the recruitment of a limited cellular pool of protein cofactors that are recruited by both NRs, and that the overexpression of the second NR competed with the first for this pool, preventing transactivation of the reporter gene by the first (67). In 1995, the first coregulators – the coactivator SRC-1, and the corepressors NCoR-1 and SMRT – were cloned (68-70). Since then, over 300 NR coregulators have been identified, indicative of their pervasive role in transcriptional regulation and enabling the complex pathways that allow the selective regulation of individual genes (71). These proteins are divided into two classes: coactivators, which promote the transcription of a gene, and corepressors, which prevent gene transcription.

Coactivators are a broadly defined class of long, intrinsically disordered proteins that serve as scaffolds for the assembly of the active transcriptional complex (Figure 1.7). This involves the recruitment of transcription factors (including NRs), histone acetyltransferases and other chromatin modifying enzymes, and co-coregulators such as p300/CRB (72). Similarly to the NR NTD, the assembly of this complex induces a helical structure in the coactivator (72). The transactivation complex ultimately recruits RNA polymerase II, resulting in transcription of the target gene (73). NR LBDs recruit coactivators to DNA in response to the binding of an agonist via the recognition of the NR box motif, a hydrophobic helix containing the amino acid sequence

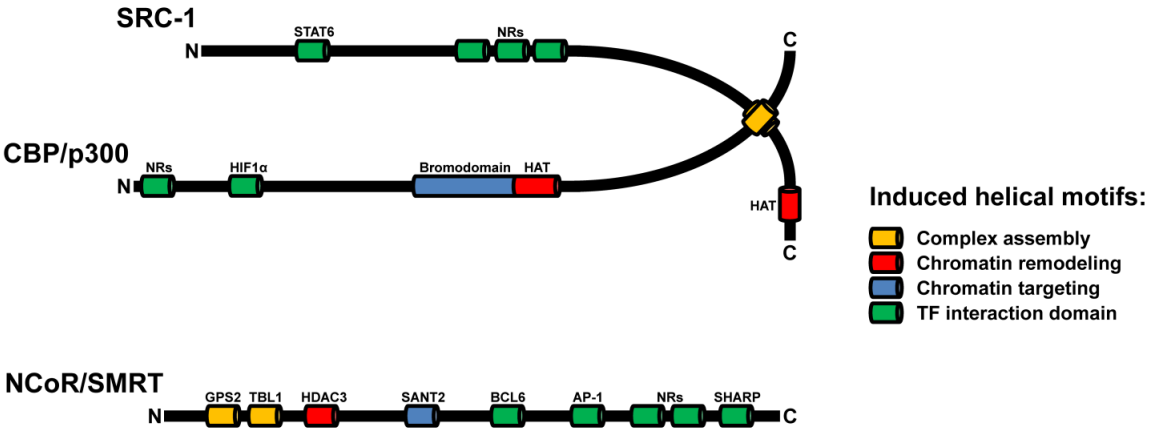

Figure 1.7. Modular structure of NR coregulators.

Schematic showing the scaffold-like organization of the representative coactivator SRC-1, co-coactivator CBP/p300, and corepressor NCoR/SMRT. Induced helical interaction domains are indicated as cylinders and color coded according to the functions of the proteins they recruit (see legend). Representative examples of binding partners are indicated above each helix.

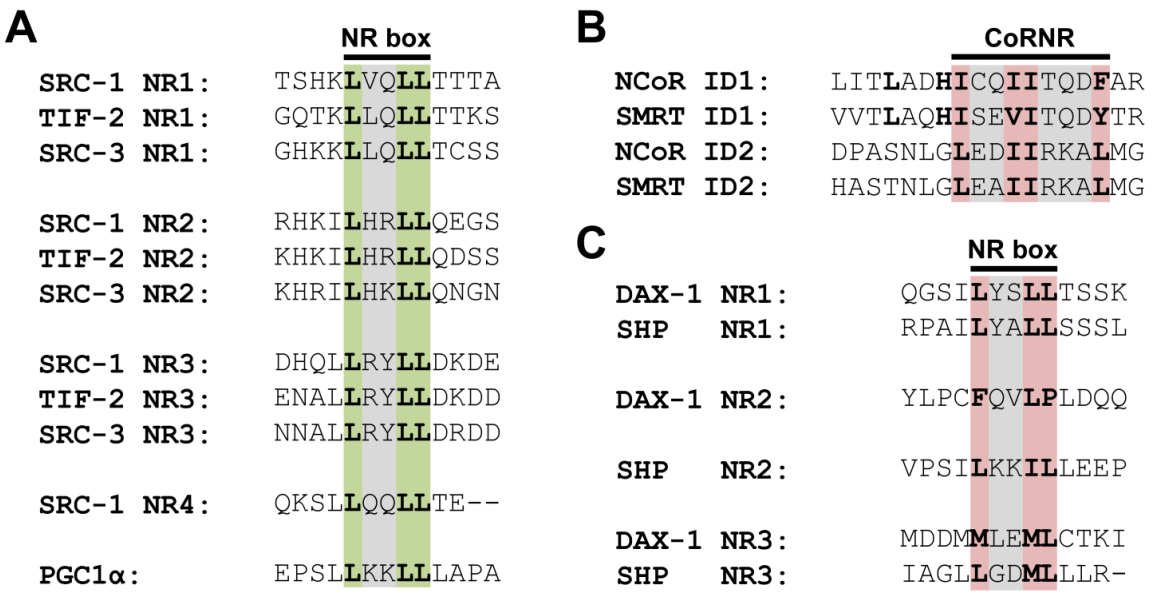

Figure 1.8. NRs recruit coregulators via recognition of conserved hydrophobic motifs.

Sequence alignment of conserved leucine-rich motifs of (A) coactivators, (B) classical corepressors, and (C) atypical NR0B corepressors. NR recognition sequences are highlighted with the interacting hydrophobic residues highlighted in color.

LxxLL, where "x" represents any amino acid and "L" can sometimes be substituted by other hydrophobic residues (Figure 1.8 A). Common coactivators include SRC-1, TIF2 (SRC-2), SRC-3, and PGC1α.

Corepressors, like coactivators, are scaffolds for the assembly of a multimeric heterocomplex (Figure 1.7). The two classical corepressors, NCoR1 and SMRT (NCoR2), form the stable core of a repression complex that also includes GPS2, TBL1X, and HDAC3 (74). The core complex is recruited to DNA by its interaction with transcription factors, including NRs, and silences genes primarily through chromatin remodeling via histone deacetylation. NRs interact with NCoR1/SMRT via the recognition of the CoRNR motif, a hydrophobic sequence analogous to the coactivator NR box motif, typically [I/L]xxIIxxxL (Figure 1.8 B) (72).

In addition to the classical corepressors, NRs also interact with the atypical corepressors DAX-1 (NR0B1) and SHP (NR0B2), two orphan members of the NR superfamily. These NRs lack DBDs, instead consisting of only the LBD subunit and, in the case of DAX-1, a disordered N-terminal regulatory region (Figure 1.9 A, B) (75). Unlike NCoR/SMRT, SHP and DAX-1 interact with the AF-2 surface using an LxxLL motif (Figure 1.8 C), blocking coactivator recruitment, and are therefore able to repress an active NR (76-79). DAX-1 and SHP are key regulators of steroidogenesis and cholesterol metabolism, respectively, and are expressed in steroidogenic tissues (DAX-1) (80) and the liver (SHP) (81; 82). The NR0B receptors both regulate and are regulated by the NR5A subfamily (SF-1 and LRH-1), with which they are coexpressed (82-84). In steroidogenic tissues, SF-1 activation causes the upregulation of DAX-1, which then represses SF-1 activity as part of a negative feedback loop (Figure 1.9 C); similarly, LRH-1 and SHP regulate each other in the liver (75; 81).

Conservation of nuclear receptor sequence and structure.

As mentioned earlier, the members of the NR superfamily display a high amount of structural homology. This is particularly true within their well-ordered DBDs and LBDs, each of which shares a common fold across the different NRs and across different species. Furthermore,

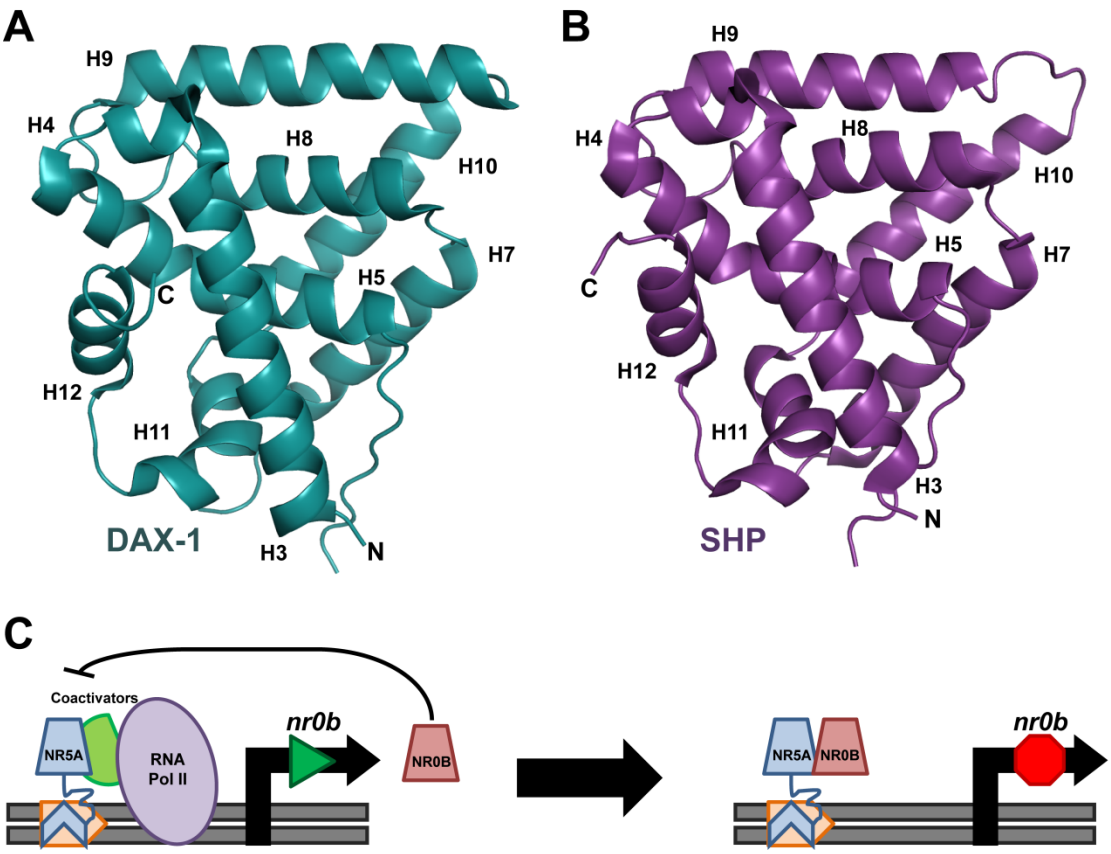

Figure 1.9. Structure and mode of action of atypical NR0B corepressors.

NR LBD-like structure of (A) DAX-1 (PDB: 3F5C) and (B) SHP (PDB: 4NUF). (C) Negative feedback of NR5A transactivation occurs via upregulation of NR0B NR corepressors.

while the different NRs show some variations in their mechanisms of activity, their general mode of activity is the same: activate or repress the expression of a target gene in response to some signal by recruiting a transactivation or transrepression complex to DNA. While their similarity may seem surprising in light of the diverse physiological roles the different NRs play in the body, the reason for their conservation is actually quite simple. Every single member of the NR superfamily is evolutionary descended from a single common ancestor, from which the modular structure, conserved folds, and general mechanism of activity was inherited (Figure 1.10) (85). This ancestral gene was duplicated, and over evolutionary time its copies diverged, duplicated, and diverged again into the extant NRs seen today. A gene family is thus comprised not only of its modern members, but also includes the ancestral genes from which they evolved. The study of this process is called molecular evolution.

Evolution of protein families and the resurrection of ancient genes

Molecular Evolution

That life on Earth arose due to evolution by natural selection is an undisputed concept that is central to our understanding of biology in the modern age. While evolution is typically considered from the perspective of whole organisms, which over generations undergo genetic mutations and selection events in order to produce new descendant species, so too can the mutation and selection of a genome produce novel genes, proteins, and biological systems. Molecular evolution is the study by which biological macromolecules can change over time.

Mutation of a DNA sequence during the course of evolution can manifest at the protein level in two ways: change in protein fold, and change of protein function. In fold evolution, mutation of one or more protein residues causes the descendant protein to adopt a 3-dimensional conformation that is different from its ancestral state, which may or may not be associated with functional changes (86). In functional evolution, sequence changes cause the descendant protein to perform a different task or role than that of its ancestor, regardless of the presence or absence

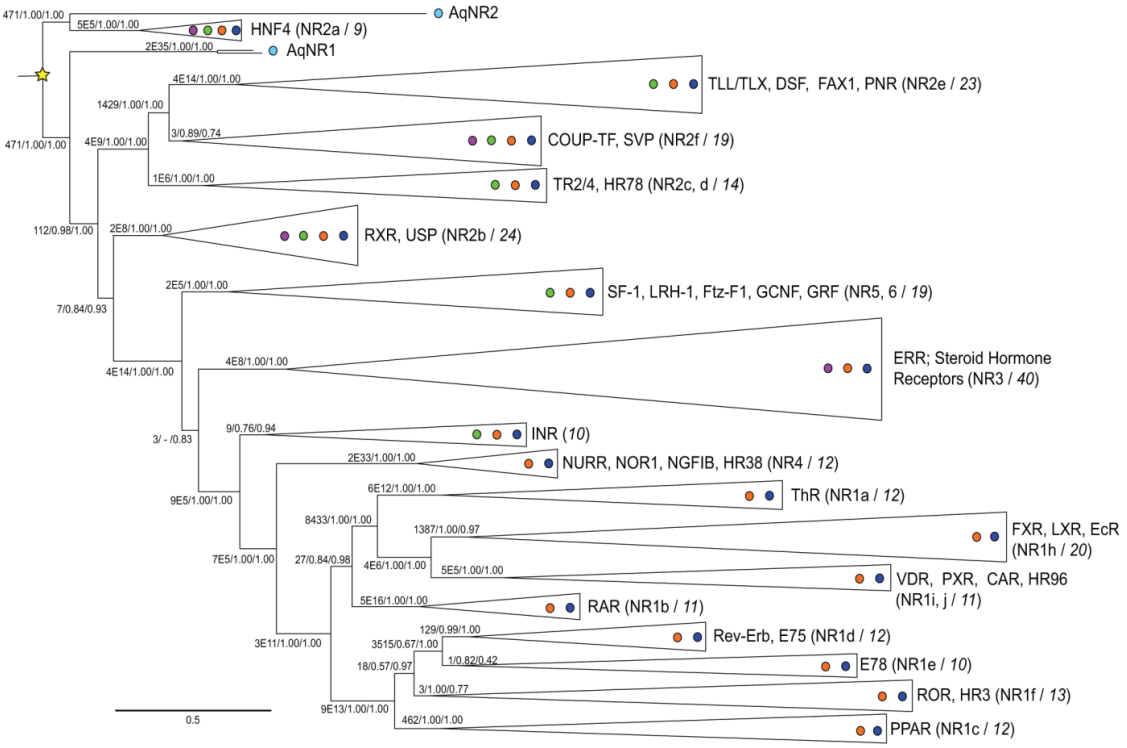

Figure 1.10. Evolution of the NR superfamily

Reduced representation of NR phylogeny inferred by maximum likelihood. Yellow star represents the common ancestor to all NRs. Colored circles indicate the presence of each clade in major metazoan taxa: sponges (light blue), placozoans (purple), Cnidaria (green), protostomes (orange), and deuterostomes (dark blue). Clades are labeled with their common protein names. Branch labels show support measured as approximate likelihood ratios (the ratio of the likelihood of the best tree with that node to the best tree without it), Bayesian posterior probabilities, and chi-square confidence estimates (the probability of a likelihood ratio at least as great as the observed ratio if the node is not resolved on the true tree). Scale bar represents the probability of substitutions per site. INRs: invertebrate-only nuclear receptors with no standard nomenclature. Figure and caption adapted from open-access published research (91).

of any changes to its fold (87). New genes evolve when a gene duplication event occurs and becomes fixed in a population (88). Because redundant genes rarely confer a selective advantage to the organism (with a few notable exceptions, *e.g.* rRNA copy number (89)), the duplicate gene is free to mutate and diverge (90). As the duplicate gene mutates, it may lose its expression or function to become a pseudogene, or evolve a new biological function. New functions can arise from neofunctionalization, whereby one copy of the duplicate gene retains its original function and the other evolves an entirely new function, or subfunctionalization, whereby the duplicate genes each diverge from their ancestral function, with each gene retaining a different but complementary function of the ancestral gene. The daughter genes are known as paralogs when they diverge within a single species, or orthologs when they diverge in the context of a speciation event (92). Over evolutionary time, this process happens repeatedly, giving rise to gene families. A gene family thus includes the ancestral genes in addition to the extant descendants. While the ancestral genes may have been extinct millions of years ago, it is now possible to physically resurrect them, enabling their study in the laboratory.

Ancestral Gene Resurrection

Ancestral Gene Resurrection (AGR) was first conceptualized in 1963 by Drs. Linus Pauling and Emile Zuckerkandl, who proposed that, with future technology, it could one day be possible to determine the sequence of and synthesize ancestral genes (93). Advances in computing power and DNA synthesis techniques over the next few decades made their vision feasible; in the modern age, it is now possible to reconstruct the DNA sequence of an ancestral gene, verify its accuracy using robust statistical methods, physically resurrect the gene by synthesizing the reconstructed DNA sequence, and use the resurrected gene in routine laboratory experiments (Figure 1.11) (92).

The accurate reconstruction of the ancestral gene sequence is of paramount importance to AGR, and several methods have been developed to infer the correct sequence. It was initially presumed that the ancestral state of a gene could be correctly derived from the consensus

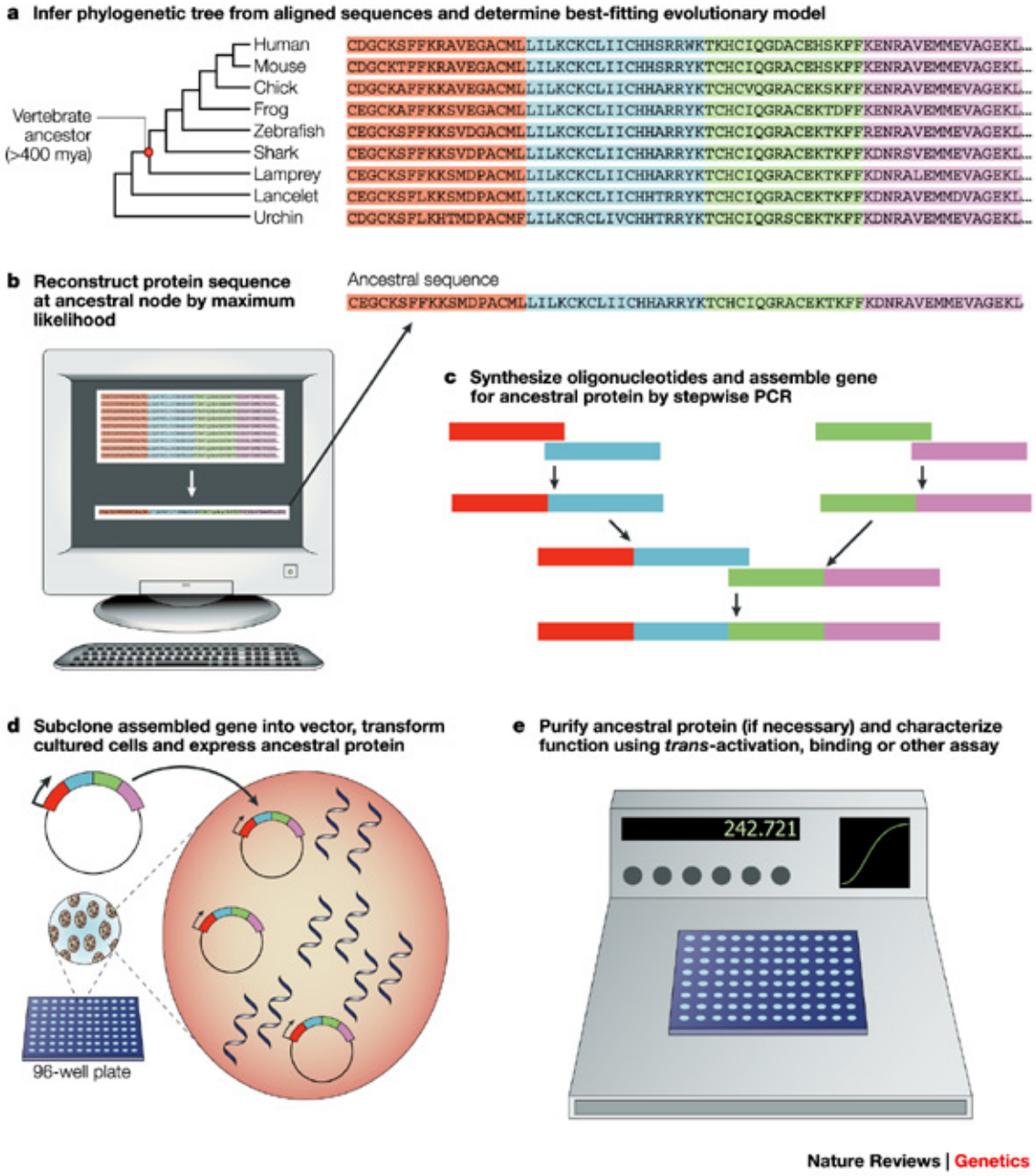

Figure 1.11. Summary of ancestral gene resurrection.

Schematic showing the overall strategy of AGR. Figure adapted from published work and used with permission of Nature Publishing Group (London, UK) (92).

sequence of its extant descendants, without regard to their phylogenetic relationship (92). However, this method was flawed, in that the ancestral sequence that was inferred was highly influenced by the input genes. Thus, an ancestral sequence reconstructed from a selection of extant genes that overrepresented a particular phylogenetic clade was artificially biased towards the consensus sequence of that clade (92).

It was for this reason that the maximum parsimony method was developed in the 1970s, and first applied in the 1980s. The defining feature of maximum parsimony is its consideration of the phylogenetic relationships among extant sequences when reconstructing the ancestral state of a recent common ancestor (92). The reconstructions produced by the maximum parsimony method are highly accurate when reconstructing the recent ancestor of several closely related genes; while there are very few organisms that preserved the ancestral sequence of their genes, a study that compared reconstructed sequences to these preserved sequences found the reconstruction to be 98.6% accurate (94). Despite its improved accuracy, however, maximum parsimony has several limitations that prevent its utility for resurrecting very ancient sequences. Maximum parsimony becomes less accurate as terminal sequences become less similar over evolutionary time. When a particular site in a sequences changes multiple times during the course of evolution, maximum parsimony often assumes that each of these changes are equally likely, with no way of determining which of these possible states is the correct one (95), Furthermore, maximum parsimony utilizes the Fitch algorithm in its reconstructions, which assumes that every possible evolutionary change is equally probable (92). In truth, the evolutionary process introduces a selective bias towards silent and advantageous mutations over deleterious ones.

The maximum likelihood method (ML) was developed in 1995 to address these pitfalls (96). The ML algorithm uses Bayesian statistics to evaluate the posterior probability – the likelihood that an initial event occurred given the observation of subsequent events – of the potential sequences of internal nodes on a phylogenetic tree. The sequence calculated to have the highest likelihood is inferred to be the ancestral sequence (92). Because it applies robust

statistical methods to the resurrection, ML has several advantages over maximum parsimony. ML incorporates empirical evidence about the molecular evolution of the gene family, such as selective bias and the length of the phylogenetic branch, into the sequence reconstruction. Thus ML can resolve ancestral states that appear ambiguous when evaluated by the parsimony algorithm, especially when the input sequences are highly divergent. Furthermore, the phylogenetic models used in the ML analysis are not assigned by the user, but rather are inferred from a statistical evaluation of the sequence data. Finally, by calculating the Bayesian posterior probability of each possible ancestral state, ML allows the user to determine the most likely sequence based upon robust statistics, conferring high confidence to the accuracy of the sequence reconstruction (92; 95).

Once the nucleotide sequence of an ancestral gene is reconstructed computationally, the gene can be resurrected physically via DNA synthesis. The advent of modern DNA synthesis techniques has made the physical resurrection of an ancestral gene a simple task; genes of up to 15 kb can now be synthesized at prices as low as US $0.23/bp, with techniques such as overlap extension PCR making the synthesis of even longer genes possible. The synthesized gene can then be used in cell-based assays or recombinantly expressed and purified for use in biochemical assays in order to test its function. A functional gene product lends credence to the accuracy of the sequence reconstruction (92).

NRs studied in this work

The corticosteroid receptors

The steroid receptors (SRs) serve as key regulators of cell growth, sexual development, reproduction, cardiovascular homeostasis, and the immune response (20). In humans, the SR family consists of five paralogs: the estrogen receptor (ER), the progesterone receptor (PR), the androgen receptor (AR), the mineralocorticoid receptor (MR), and the glucocorticoid receptor (GR).

The SRs can be subcategorized based upon both the chemical nature of their ligand and the physiological roles they play. ER responds to 3-hydroxysteroids (*i.e.* estrogens), defined by the phenolic hydroxyl group on the 3'-carbon of the steroid A ring, while AR, PR, MR, and GR respond to 3-ketosteroids, which have a non-aromatic A ring with a ketone on the 3'-carbon. The modern human SRs are evolutionarily derived from a single common ancestor, the ancestral steroid receptor-1 (AncSR1). Resurrection of AncSR1 demonstrated it to be an ER-like receptor, responding to 3-hydroxysteroids. A gene duplication event gave rise to the ancestral SR-2 (AncSR2), which evolved specificity for 3-ketosteroids recognized by its descendants AR, PR, MR, and GR (97). Further duplications produced the ancestral SR-3 (AncSR3) and the ancestral corticoid receptor (AncCR), each of which subsequently duplicated and diverged into the modern PR and AR, and the modern MR and GR.

As type I NRs (Figure 1.2 A), SRs recognize and bind short inverted repeat DNA sequences in response to the binding of small, lipophilic ligands derived ultimately from cholesterol. Prior to activation, apo SRs are located in the cytoplasm in a stable, ATP-dependent heterocomplex with the heat shock proteins (HSPs) HSP90, HSP70, HSP40, HSP70-HSP90 organizing protein (Hop), and p23 (98). This complex not only serves to stabilize the receptor against degradation, but also holds the LBP in a semi-open state that is more freely accessible to the ligand, increasing the binding affinity of steroid to receptor by two orders of magnitude (98-100). Upon steroid binding, the SR-HSP complex becomes dynamic; the SR LBP collapses around the steroid (22; 98), Hop and HSP70 dissociate from the complex and are replaced by the immunophilin FKBP52, and the steroid-bound SR freely associates and dissociates with the HSP complex (98; 101). FKBP52 mediates the association of the dynamic SR-HSP complex with dynein, and the entire complex is shuttled into the nucleus via the microtubule network (101-104). Inside the nucleus, the SR dissociates from the HSP heterocomplex and binds as a homodimer to its recognition sequence, an inverted repeat DNA upstream of their target genes (5), and regulates expression by the mutually exclusive recruitment of coactivators and

corepressors to the AF-2 surface via the recognition of LxxLL/FxxFF or Lxxx[I/H]Ixxx[I/L] motifs, respectively (28-33).

Together, GR and MR are known as the corticosteroid receptors, so named for their recognition of the steroid hormones produced by the adrenal cortex, cortisol and aldosterone. The two receptors diverged from a single ancestor, AncCR, approximately 470 million years ago (105); the extant human proteins show 94% amino acid sequence identity in their DBDs and 57% homology in their LBDs (106). When resurrected, AncCR displayed MR-like behavior, hallmarked by high-affinity (low to sub nM) activation by both aldosterone and cortisol (105; 107). In contrast, the human GR is activated exclusively by cortisol, but at a much lower affinity (high nM) (107).

The human GR responds selectively to cortisol, a steroid hormone produced by the zona fasciculata of the adrenal cortex, the secretion of which is tightly controlled by the hypothalamic-pituitary-adrenal (HPA) axis as part of the normal circadian rhythm and as part of the stress response (108). GR is ubiquitously expressed throughout the body, and regulates the expression of many genes related to development, metabolism, bone homeostasis, stress, mood regulation, hematopoiesis, inflammation, and the immune response (108). Like most NRs, GR utilizes both transactivation and transrepression pathways. GR transactivation occurs via the canonical type I NR mechanism, with the agonist-bound homodimeric GR recruiting coactivators and the transcriptional complex to its target genes. GR transrepression occurs via several mechanisms, which may occur in a DNA-dependent or independent manner. DNA-dependent GR transrepression may occur via the competitive binding of GR to the response element for another TF (109), by the crosstalk of GR with other TFs on composite response elements (109), or by the binding of GR to negative GREs (nGREs) (110; 111). DNA-independent repression occurs primarily through the tethering of GR to other DNA-bound TFs, such as AP-1 (112), Stat3 (113), and NF-κB (114).

Though it regulates a plethora of physiological processes, GR is most commonly targeted by pharmaceuticals for its anti-inflammatory and immunosuppressive functions (108). Hydrocortisone (synthetic cortisol), triamcinolone acetonide, mometasone furoate, and fluticasone propionate are widely prescribed for the routine management of allergies, asthma, and allergic dermatitis. Due to its ubiquitous expression and diverse range of functions, prolonged systemic administration of synthetic glucocorticoids invariably cause undesirable side effects. Thus, whenever possible, glucocorticoids are applied locally. When administered systemically, glucocorticoid drugs such as dexamethasone and prednisone are effective treatments for autoimmune disorders such as systemic lupus erythematosus, for preventing organ rejection after transplant, or for certain cancers due to their antiangiogenic and antiproliferative effects (108).

The mineralocorticoid receptor, unlike GR, is more limited in its physiological functions. While GR responds selectively but with low affinity to cortisol, MR responds with high affinity and equipotency to both cortisol and aldosterone (105; 107). The synthesis and release of aldosterone is regulated by the renin-angiotensin-aldosterone system (RAAS), a feedback loop analogous to the HPA axis that controls the release of cortisol. The typical plasma concentration of cortisol is 100- to 1000-fold higher than that of aldosterone (115). In order to prevent MR overactivation by glucocorticoids, MR is often coexpressed with 11-β-hydroxysteroid dehydrogenase (11βHSD), which catalyzes the dehydrogenation of cortisol into cortisone, which does not activate MR (116). MR expression is limited to the kidneys, cardiovascular system, brain, colon, inflammatory cells, and adipose tissue (115; 117). In the colon and kidneys, MR is expressed in polarized epithelia, where it helps to maintain sodium and water balance by regulating the expression of the epithelial sodium channel (ENaC) (118) and the basolateral Na^+/K^+-ATPase pump (119). In nonepithelial tissue, MR acts as a high-affinity glucocorticoid receptor, though the extent of its functions in this context is not fully elucidated (115; 116). In healthy heart, MR regulates electrophysiology and cardiomyocyte growth, with overstimulation contributing to cardiac hypertrophy and heart failure (120; 121). MR has been shown to induce

adipogenesis (122; 123). In monocytes and macrophages, MR promotes the inflammatory response, and contributes to the development of cardiac fibrosis (124; 125). In the brain, MR regulates the RAAS and the stress response via the HPA axis (126-128).

For decades, manipulation of MR signaling has been a mainstay in the management of hypertension and other cardiac diseases. The first MR antagonist, spironolactone, was introduced in 1959, and is still widely prescribed today. MR is also indirectly targeted by inhibiting the RAAS, thereby diminishing the release of aldosterone; angiotensin converting enzyme (ACE) inhibitors, angiotensin II receptor antagonists, and novel renin inhibitors such as aliskiren resolve hypertension in this way. Thus MR, like GR, is a highly valuable pharmacological target for the treatment of human disease.

Liver receptor homolog-1

An introduction to LRH-1 may be found in Chapter 2.

Current state of nuclear receptor pharmacology

Nuclear receptors as pharmaceutical targets

The fact that NRs control many physiological processes related to human disease, combined with their ability to do so in response to the binding of small molecules, has made NRs very attractive drug targets. Indeed, NRs are a heavily targeted and extraordinarily profitable subset of receptors: 13% of all FDA approved drugs target NRs (second only to GPCRs, at 26%) (129), with a total market share of over 30 billion USD (130). NR-targeting drugs include many household names, such as cortisone and its derivatives (anti-inflammatory drugs that target GR), hormonal birth control (ER and PR), anticancer agents (*e.g.* dexamethasone and tamoxifen, targeting GR and ER, respectively) and thiazolidinediones (antidiabetic PPARγ agonists) (1).

The benefit of NR drugs does not come without a price. Most of the currently approved NR targeting compounds carry the risk of unwanted side effects, which can range in severity from those that are trivial, to those that cause discontinuation of the treatment regimen, to those that are

fatal. The causes of these side-effects are highly variable across the different receptors, but stem from a handful of traits that are inherent to the NR family. From an evolutionary standpoint, NRs are descended from a single common ancestor protein, from which they inherited a highly conserved structure (85). Drugs that target one NR may show some activity for other NRs, particularly in closely paralogous subclades such as the SR family. For example, the MR antagonist spironolactone also causes gynecomastia, impotence, and menstrual irregularities due to its actions on AR and PR (131; 132). Furthermore, NRs show bimodal activity, performing both transactivation and transrepression, frequently in response to the same ligand. Thus, therapeutic effects resulting from pharmacological transrepression may be compromised by side effects caused from the transactivation activity of the same receptor, and vice-versa. This is perfectly exemplified by agonists of the glucocorticoid receptor, which are widely used for the anti-inflammatory and immunosuppressant effects that result from GR transrepression, but whose transactivation pathways cause the development of Cushing's syndrome, a condition characterized by obesity, glucose intolerance, thinning skin, osteoporosis, and hypertension (133). Moreover, it is not a guarantee that all of the genes controlled by a NR's transactivation or transrepression pathway are desirable to achieve the therapeutic goal; for instance, activation of LRH-1 causes the upregulation of CYP7A1, CYP8B1, and GLUT4 (134; 135), but also causes the upregulation of aromatase, GREB1, and Oct4, all of which are oncogenic and can contribute to breast, prostate, and pancreatic tumor formation (136-138). The ubiquitous expression throughout the body of many NRs may limit the ability to target a specific organ or system. The putative ideal NR drug, therefore, is one that is exquisitely selective for its target receptor, and can selectively modulate the genes that promote the therapeutic effect without affecting those that contribute to unwanted side effects.

Questions and hypotheses addressed in this work

The design of effective and selective drugs requires a thorough understanding of the structural mechanisms that drive both the binding of a small molecule ligand to its target receptor,

and the conformational changes induced in the receptor that produce the desired therapeutic effect. Because a protein's fold is driven by evolution, an approach that combines the insights offered by molecular evolution and structural biology may be of greater assistance to the rational design of robust therapeutics than either of these approaches alone. Thus, the broad goal of the research presented in this dissertation is to use this combined approach to explain the mechanism of selectivity of NR-targeting ligands for their cognate receptors, with a particular focus on the corticosteroid receptors and LRH-1.

LRH-1 (and its paralog SF-1) is unique among ligand-activated receptors in that its activity can be directly modulated by the binding of phospholipids (PLs) (see also Chapter 2) (134). Unlike other PL-responsive receptors, which bind PLs by recognizing only the PL head-group while the PL tails are embedded in a PL bilayer, LRH-1 binds to PLs outside the membrane, fully engulfing the PL tails in its LBP. In order to effectively target LRH-1 pharmacologically, a robust understanding of the mechanisms by which PL binding drives receptor activity is required. Unfortunately, our understanding thereof is limited to the knowledge that PLs bind LRH-1, and that the exogenous medium-chain phosphatidylcholine species dilauroyl phosphatidylcholine (DLPC) uniquely activates LRH-1 by prohibiting the binding of the corepressor SHP (25; 134). The work presented in this dissertation seeks to answer two important questions: 1) what are the structural mechanisms by which DLPC disfavors the binding of SHP in order to promote LRH-1 activation?; and, 2) how is LRH-1 regulated by endogenous PLs?

In contrast to LRH-1, the corticosteroid receptors are a well-studied group of proteins for which many established drugs are already in use. However, even the best of these drugs cause adverse effects, due in part to their off target activity on other members of the SR family. The root cause of SR cross pharmacology is related to the fact that SRs are a very closely related family of receptors and show high structural homology in their LBDs, even for NRs. And while their cognate ligands have had over 500 million years to evolve exquisite selectivity for their

target receptors (or evolve mechanisms to compensate for less-exquisite selectivity, such as 11-βHSD in the case of cortisol and GR/MR (116)), synthetic steroids have only existed for a few decades, and have failed to achieve such selectivity in that time. In recent years, an experimental approach that combined structural biology with molecular evolution and ancestral gene resurrection has been successful in elucidating the structural mechanisms by which steroid hormone selectivity has evolved over time (97; 107; 139-144). By measuring the activity of a steroid hormone for both the extant and ancestral SRs, one can pinpoint where along the course of SR evolution each receptor evolved structural features that enable or prohibit activity by that hormone. The points along evolutionary time at which a receptor lineage evolved to respond or select against a hormone represents a switch in selectivity. High-resolution structures of the hormone-receptor complexes on either side of this selectivity switch can then identify the structural changes that enabled or prohibited hormone recognition. In the present work, we propose that the same approach can be adapted to determine the structural mechanisms that drive synthetic steroid selectivity (or lack thereof), using the synthetic glucocorticoid mometasone furoate (MOF) as a representative ligand. Identifying the structural features that enable or prohibit the binding of a drug would then enable one to design novel compounds that takes advantage of these features, improving selectivity against receptors for which the existing drug is promiscuous, thereby reducing the risk of unwanted side-effects that result from the off-target binding to these receptors.

References

1. Gronemeyer H, Gustafsson JA, Laudet V. 2004. Principles for modulation of the nuclear receptor superfamily. *Nat Rev Drug Discov* 3:950-64
2. Burris TP, Solt LA, Wang Y, Crumbley C, Banerjee S, et al. 2013. Nuclear receptors and their selective pharmacologic modulators. *Pharmacol Rev* 65:710-78
3. Evans RM, Mangelsdorf DJ. 2014. Nuclear Receptors, RXR, and the Big Bang. *Cell* 157:255-66
4. Kininis M, Kraus WL. 2008. A global view of transcriptional regulation by nuclear receptors: gene expression, factor localization, and DNA sequence analysis. *Nucl Recept Signal* 6:e005

5. Mangelsdorf DJ, Thummel C, Beato M, Herrlich P, Schutz G, et al. 1995. The nuclear receptor superfamily: the second decade. *Cell* 83:835-9
6. Helsen C, Kerkhofs S, Clinckemalie L, Spans L, Laurent M, et al. 2012. Structural basis for nuclear hormone receptor DNA binding. *Mol Cell Endocrinol* 348:411-7
7. Luisi BF, Xu WX, Otwinowski Z, Freedman LP, Yamamoto KR, Sigler PB. 1991. Crystallographic analysis of the interaction of the glucocorticoid receptor with DNA. *Nature* 352:497-505
8. Schoenmakers E, Alen P, Verrijdt G, Peeters B, Verhoeven G, et al. 1999. Differential DNA binding by the androgen and glucocorticoid receptors involves the second Zn-finger and a C-terminal extension of the DNA-binding domains. *Biochem J* 341 (Pt 3):515-21
9. Roemer SC, Donham DC, Sherman L, Pon VH, Edwards DP, Churchill ME. 2006. Structure of the progesterone receptor-deoxyribonucleic acid complex: novel interactions required for binding to half-site response elements. *Mol Endocrinol* 20:3042-52
10. Ham J, Thomson A, Needham M, Webb P, Parker M. 1988. Characterization of response elements for androgens, glucocorticoids and progestins in mouse mammary tumour virus. *Nucleic Acids Res* 16:5263-76
11. Denison SH, Sands A, Tindall DJ. 1989. A tyrosine aminotransferase glucocorticoid response element also mediates androgen enhancement of gene expression. *Endocrinology* 124:1091-3
12. Haelens A, Verrijdt G, Callewaert L, Christiaens V, Schauwaers K, et al. 2003. DNA recognition by the androgen receptor: evidence for an alternative DNA-dependent dimerization, and an active role of sequences flanking the response element on transactivation. *Biochem J* 369:141-51
13. Melvin VS, Harrell C, Adelman JS, Kraus WL, Churchill M, Edwards DP. 2004. The role of the C-terminal extension (CTE) of the estrogen receptor alpha and beta DNA binding domain in DNA binding and interaction with HMGB. *J Biol Chem* 279:14763-71
14. Gearhart MD, Holmbeck SM, Evans RM, Dyson HJ, Wright PE. 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
15. Little TH, Zhang Y, Matulis CK, Weck J, Zhang Z, et al. 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. Solomon IH, Hager JM, Safi R, McDonnell DP, Redinbo MR, Ortlund EA. 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
17. Lee MS, Sem DS, Kliewer SA, Provencal J, Evans RM, Wright PE. 1994. NMR assignments and secondary structure of the retinoid X receptor alpha DNA-binding domain. Evidence for the novel C-terminal helix. *Eur J Biochem* 224:639-50
18. Rastinejad F, Perlmann T, Evans RM, Sigler PB. 1995. Structural determinants of nuclear receptor assembly on DNA direct repeats. *Nature* 375:203-11
19. Chandra V, Huang P, Hamuro Y, Raghuram S, Wang Y, et al. 2008. Structure of the intact PPAR-gamma-RXR- nuclear receptor complex on DNA. *Nature* 456:350-6
20. Nagy L, Schwabe JW. 2004. Mechanism of the nuclear receptor molecular switch. *Trends Biochem Sci* 29:317-24
21. Nolte RT, Wisely GB, Westin S, Cobb JE, Lambert MH, et al. 1998. Ligand binding and co-activator assembly of the peroxisome proliferator-activated receptor-gamma. *Nature* 395:137-43

22. Gee AC, Katzenellenbogen JA. 2001. Probing conformational changes in the estrogen receptor: evidence for a partially unfolded intermediate facilitating ligand binding and release. *Mol Endocrinol* 15:421-8
23. Leng X, Tsai SY, O'Malley BW, Tsai MJ. 1993. Ligand-dependent conformational changes in thyroid hormone and retinoic acid receptors are potentially enhanced by heterodimerization with retinoic X receptor. *J Steroid Biochem Mol Biol* 46:643-61
24. Keidel S, LeMotte P, Apfel C. 1994. Different agonist- and antagonist-induced conformational changes in retinoic acid receptors analyzed by protease mapping. *Mol Cell Biol* 14:287-98
25. Musille PM, Pathak MC, Lauer JL, Hudson WH, Griffin PR, Ortlund EA. 2012. Antidiabetic phospholipid-nuclear receptor complex reveals the mechanism for phospholipid-driven gene regulation. *Nat Struct Mol Biol* 19:532-7, S1-2
26. Watkins RE, Davis-Searles PR, Lambert MH, Redinbo MR. 2003. Coactivator binding promotes the specific interaction between ligand and the pregnane X receptor. *J Mol Biol* 331:815-28
27. Pissios P, Tzameli I, Kushner P, Moore DD. 2000. Dynamic stabilization of nuclear receptor ligand binding domains by hormone or corepressor binding. *Mol Cell* 6:245-53
28. Heery DM, Kalkhoven E, Hoare S, Parker MG. 1997. A signature motif in transcriptional co-activators mediates binding to nuclear receptors. *Nature* 387:733-6
29. Darimont BD, Wagner RL, Apriletti JW, Stallcup MR, Kushner PJ, et al. 1998. Structure and specificity of nuclear receptor-coactivator interactions. *Genes Dev* 12:3343-56
30. Hu X, Lazar MA. 1999. The CoRNR motif controls the recruitment of corepressors by nuclear hormone receptors. *Nature* 402:93-6
31. Webb P, Anderson CM, Valentine C, Nguyen P, Marimuthu A, et al. 2000. The nuclear receptor corepressor (N-CoR) contains three isoleucine motifs (I/LXXII) that serve as receptor interaction domains (IDs). *Mol Endocrinol* 14:1976-85
32. Perissi V, Staszewski LM, McInerney EM, Kurokawa R, Krones A, et al. 1999. Molecular determinants of nuclear receptor-corepressor interaction. *Genes Dev* 13:3198-208
33. Nagy L, Kao HY, Love JD, Li C, Banayo E, et al. 1999. Mechanism of corepressor binding and release from nuclear hormone receptors. *Genes Dev* 13:3209-16
34. Cohen RN, Brzostek S, Kim B, Chorev M, Wondisford FE, Hollenberg AN. 2001. The specificity of interactions between nuclear hormone receptors and corepressors is mediated by distinct amino acid sequences within the interacting domains. *Mol Endocrinol* 15:1049-61
35. Bourguet W, Ruff M, Chambon P, Gronemeyer H, Moras D. 1995. Crystal structure of the ligand-binding domain of the human nuclear receptor RXR-alpha. *Nature* 375:377-82
36. Xu HE, Stanley TB, Montana VG, Lambert MH, Shearer BG, et al. 2002. Structural basis for antagonist-mediated recruitment of nuclear co-repressors by PPARalpha. *Nature* 415:813-7
37. Renaud JP, Rochel N, Ruff M, Vivat V, Chambon P, et al. 1995. Crystal structure of the RAR-gamma ligand-binding domain bound to all-trans retinoic acid. *Nature* 378:681-9
38. Batista MR, Martinez L. 2013. Dynamics of nuclear receptor Helix-12 switch of transcription activation by modeling time-resolved fluorescence anisotropy decays. *Biophys J* 105:1670-80
39. Blondel A, Renaud JP, Fischer S, Moras D, Karplus M. 1999. Retinoic acid receptor: a simulation analysis of retinoic acid binding and the resulting conformational changes. *J Mol Biol* 291:101-15
40. Kosztin D, Izrailev S, Schulten K. 1999. Unbinding of retinoic acid from its receptor studied by steered molecular dynamics. *Biophys J* 76:188-97

41. Carlsson P, Burendahl S, Nilsson L. 2006. Unbinding of retinoic acid from the retinoic acid receptor by random expulsion molecular dynamics. *Biophys J* 91:3151-61
42. Martinez L, Polikarpov I, Skaf MS. 2008. Only subtle protein conformational adaptations are required for ligand binding to thyroid hormone receptors: simulations using a novel multipoint steered molecular dynamics approach. *The journal of physical chemistry. B* 112:10741-51
43. Martinez L, Sonoda MT, Webb P, Baxter JD, Skaf MS, Polikarpov I. 2005. Molecular dynamics simulations reveal multiple pathways of ligand dissociation from thyroid hormone receptors. *Biophys J* 89:2011-23
44. Martinez L, Webb P, Polikarpov I, Skaf MS. 2006. Molecular dynamics simulations of ligand dissociation from thyroid hormone receptors: evidence of the likeliest escape pathway and its implications for the design of novel ligands. *J Med Chem* 49:23-6
45. Shen J, Li W, Liu G, Tang Y, Jiang H. 2009. Computational insights into the mechanism of ligand unbinding and selectivity of estrogen receptors. *The journal of physical chemistry. B* 113:10436-44
46. Perakyla M. 2009. Ligand unbinding pathways from the vitamin D receptor studied by molecular dynamics simulations. *Eur Biophys J* 38:185-98
47. Figueira AC, Saidemberg DM, Souza PC, Martinez L, Scanlan TS, et al. 2011. Analysis of agonist and antagonist effects on thyroid hormone receptor conformation by hydrogen/deuterium exchange. *Mol Endocrinol* 25:15-31
48. Kallenberger BC, Love JD, Chatterjee VK, Schwabe JW. 2003. A dynamic mechanism of nuclear receptor activation and its perturbation in a human disease. *Nat Struct Biol* 10:136-40
49. Simons SS, Jr., Edwards DP, Kumar R. 2014. Minireview: dynamic structures of nuclear hormone receptors: new promises and challenges. *Mol Endocrinol* 28:173-82
50. Dahlman-Wright K, Baumann H, McEwan IJ, Almlof T, Wright AP, et al. 1995. Structural characterization of a minimal functional transactivation domain from the human glucocorticoid receptor. *Proc Natl Acad Sci U S A* 92:1699-703
51. Reid J, Kelly SM, Watt K, Price NC, McEwan IJ. 2002. Conformational analysis of the androgen receptor amino-terminal domain involved in transactivation. Influence of structure-stabilizing solutes and protein-protein interactions. *J Biol Chem* 277:20079-86
52. Khan SH, Awasthi S, Guo C, Goswami D, Ling J, et al. 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. *J Biol Chem* 287:44546-60
53. Kumar R, Moure CM, Khan SH, Callaway C, Grimm SL, et al. 2013. Regulation of the structurally dynamic N-terminal domain of progesterone receptor by protein-induced folding. *J Biol Chem* 288:30285-99
54. Hodgson MC, Shen HC, Hollenberg AN, Balk SP. 2008. Structural basis for nuclear receptor corepressor recruitment by antagonist-liganded androgen receptor. *Mol Cancer Ther* 7:3187-94
55. Ford J, McEwan IJ, Wright AP, Gustafsson JA. 1997. Involvement of the transcription factor IID protein complex in gene activation by the N-terminal transactivation domain of the glucocorticoid receptor in vitro. *Mol Endocrinol* 11:1467-75
56. Kumar R, Betney R, Li J, Thompson EB, McEwan IJ. 2004. Induced alpha-helix structure in AF1 of the androgen receptor upon binding transcription factor TFIIF. *Biochemistry* 43:3008-13
57. Lanz RB, McKenna NJ, Onate SA, Albrecht U, Wong J, et al. 1999. A steroid receptor coactivator, SRA, functions as an RNA and is present in an SRC-1 complex. *Cell* 97:17-27

58. Garza AM, Khan SH, Kumar R. 2010. Site-specific phosphorylation induces functionally active conformation in the intrinsically disordered N-terminal activation function (AF1) domain of the glucocorticoid receptor. *Mol Cell Biol* 30:220-30
59. Hill KK, Roemer SC, Churchill ME, Edwards DP. 2012. Structural and functional analysis of domains of the progesterone receptor. *Mol Cell Endocrinol* 348:418-29
60. Hall JM, McDonnell DP, Korach KS. 2002. Allosteric regulation of estrogen receptor structure, function, and coactivator recruitment by different estrogen response elements. *Mol Endocrinol* 16:469-86
61. Wardell SE, Kwok SC, Sherman L, Hodges RS, Edwards DP. 2005. Regulation of the amino-terminal transcription activation domain of progesterone receptor by a cofactor-induced protein folding mechanism. *Mol Cell Biol* 25:8792-808
62. Hill KK, Roemer SC, Jones DN, Churchill ME, Edwards DP. 2009. A progesterone receptor co-activator (JDP2) mediates activity through interaction with residues in the carboxyl-terminal extension of the DNA binding domain. *J Biol Chem* 284:24415-24
63. Garza AS, Khan SH, Moure CM, Edwards DP, Kumar R. 2011. Binding-folding induced regulation of AF1 transactivation domain of the glucocorticoid receptor by a cofactor that binds to its DNA binding domain. *PLoS One* 6:e25875
64. Kumar R, Thompson EB. 2003. Transactivation functions of the N-terminal domains of nuclear hormone receptors: protein folding and coactivator interactions. *Mol Endocrinol* 17:1-10
65. Chalkiadaki A, Talianidis I. 2005. SUMO-dependent compartmentalization in promyelocytic leukemia protein nuclear bodies prevents the access of LRH-1 to chromatin. *Mol Cell Biol* 25:5095-105
66. Clinckemalie L, Vanderschueren D, Boonen S, Claessens F. 2012. The hinge region in androgen receptor control. *Mol Cell Endocrinol* 358:1-8
67. McKenna NJ, Lanz RB, O'Malley BW. 1999. Nuclear receptor coregulators: cellular and molecular biology. *Endocr Rev* 20:321-44
68. Chen JD, Evans RM. 1995. A transcriptional co-repressor that interacts with nuclear hormone receptors. *Nature* 377:454-7
69. Horlein AJ, Naar AM, Heinzel T, Torchia J, Gloss B, et al. 1995. Ligand-independent repression by the thyroid hormone receptor mediated by a nuclear receptor co-repressor. *Nature* 377:397-404
70. Kurokawa R, Soderstrom M, Horlein A, Halachmi S, Brown M, et al. 1995. Polarity-specific activities of retinoic acid receptors determined by a co-repressor. *Nature* 377:451-4
71. Lonard DM, O'Malley B W. 2007. Nuclear receptor coregulators: judges, juries, and executioners of cellular regulation. *Mol Cell* 27:691-700
72. Millard CJ, Watson PJ, Fairall L, Schwabe JW. 2013. An evolving understanding of nuclear receptor coregulator proteins. *J Mol Endocrinol* 51:T23-36
73. Bulynko YA, O'Malley BW. 2011. Nuclear receptor coactivators: structural and functional biochemistry. *Biochemistry* 50:313-28
74. Millard JC, Schwabe RJW. 2015. Assembly and Regulation of Nuclear Receptor Corepressor Complexes. In *Nuclear Receptors: From Structure to the Clinic*, ed. JI McEwan, R Kumar:155-75. Cham: Springer International Publishing. Number of 155-75 pp.
75. Ehrlund A, Treuter E. 2012. Ligand-independent actions of the orphan receptors/corepressors DAX-1 and SHP in metabolism, reproduction and disease. *J Steroid Biochem Mol Biol* 130:169-79
76. Johansson L, Bavner A, Thomsen JS, Farnegardh M, Gustafsson JA, Treuter E. 2000. The orphan nuclear receptor SHP utilizes conserved LXXLL-related motifs for interactions with ligand-activated estrogen receptors. *Mol Cell Biol* 20:1124-33

77. Johansson L, Thomsen JS, Damdimopoulos AE, Spyrou G, Gustafsson JA, Treuter E. 1999. The orphan nuclear receptor SHP inhibits agonist-dependent transcriptional activity of estrogen receptors ERalpha and ERbeta. *J Biol Chem* 274:345-53
78. Suzuki T, Kasahara M, Yoshioka H, Morohashi K, Umesono K. 2003. LXXLL-related motifs in Dax-1 have target specificity for the orphan nuclear receptors Ad4BP/SF-1 and LRH-1. *Mol Cell Biol* 23:238-49
79. Zhang H, Thomsen JS, Johansson L, Gustafsson JA, Treuter E. 2000. DAX-1 functions as an LXXLL-containing corepressor for activated estrogen receptors. *J Biol Chem* 275:39855-9
80. Guo W, Burris TP, McCabe ER. 1995. Expression of DAX-1, the gene responsible for X-linked adrenal hypoplasia congenita and hypogonadotropic hypogonadism, in the hypothalamic-pituitary-adrenal/gonadal axis. *Biochem Mol Med* 56:8-13
81. Goodwin B, Jones Sa, Price RR, Watson Ma, McKee DD, et al. 2000. A regulatory cascade of the nuclear receptors FXR, SHP-1, and LRH-1 represses bile acid biosynthesis. *Molecular cell* 6:517-26
82. Lu TT, Makishima M, Repa JJ, Schoonjans K, Kerr TA, et al. 2000. Molecular basis for feedback regulation of bile acid synthesis by nuclear receptors. *Molecular cell* 6:507-15
83. Ikeda Y, Swain A, Weber TJ, Hentges KE, Zanaria E, et al. 1996. Steroidogenic factor 1 and Dax-1 colocalize in multiple cell lineages: potential links in endocrine development. *Mol Endocrinol* 10:1261-72
84. Ikeda Y, Takeda Y, Shikayama T, Mukai T, Hisano S, Morohashi KI. 2001. Comparative localization of Dax-1 and Ad4BP/SF-1 during development of the hypothalamic-pituitary-gonadal axis suggests their closely related and distinct functions. *Dev Dyn* 220:363-76
85. Amero SA, Kretsinger RH, Moncrief ND, Yamamoto KR, Pearson WR. 1992. The origin of nuclear receptor proteins: a single precursor distinct from other transcription factors. *Mol Endocrinol* 6:3-7
86. Grishin NV. 2001. Fold change in evolution of protein structures. *J Struct Biol* 134:167-85
87. Todd AE, Orengo CA, Thornton JM. 2001. Evolution of function in protein superfamilies, from a structural perspective. *J Mol Biol* 307:1113-43
88. Ohno S, ed. 1970. *Evolution by Gene Duplication*: Springer Berlin Heidelberg.
89. Klappenbach JA, Dunbar JM, Schmidt TM. 2000. rRNA operon copy number reflects ecological strategies of bacteria. *Appl Environ Microbiol* 66:1328-33
90. Zhang J. 2003. Evolution by gene duplication: an update. *Trends in Ecology & Evolution* 18:292-8
91. Bridgham JT, Eick GN, Larroux C, Deshpande K, Harms MJ, et al. 2010. Protein evolution by molecular tinkering: diversification of the nuclear receptor superfamily from a ligand-dependent ancestor. *PLoS Biol* 8
92. Thornton JW. 2004. Resurrecting ancient genes: experimental analysis of extinct molecules. *Nat Rev Genet* 5:366-75
93. Pauling L, Zuckerkandl E. 1963. Chemical paleogenetics. Molecular "Restoration Studies" of Extinct Forms of Life. *Acta chem. scand* 17:9-16
94. Hillis DM, Bull JJ, White ME, Badgett MR, Molineux IJ. 1992. Experimental phylogenetics: generation of a known phylogeny. *Science* 255:589-92
95. Zhang J, Nei M. 1997. Accuracies of ancestral amino acid sequences inferred by the parsimony, likelihood, and distance methods. *J Mol Evol* 44 Suppl 1:S139-46
96. Yang Z, Kumar S, Nei M. 1995. A new method of inference of ancestral nucleotide and amino acid sequences. *Genetics* 141:1641-50

97. Thornton JW. 2001. Evolution of vertebrate steroid receptors from an ancestral estrogen receptor by ligand exploitation and serial genome expansions. *Proc Natl Acad Sci U S A* 98:5671-6
98. Pratt WB, Morishima Y, Osawa Y. 2008. The Hsp90 chaperone machinery regulates signaling by modulating ligand binding clefts. *J Biol Chem* 283:22885-9
99. Bresnick EH, Dalman FC, Sanchez ER, Pratt WB. 1989. Evidence that the 90-kDa heat shock protein is necessary for the steroid binding conformation of the L cell glucocorticoid receptor. *J Biol Chem* 264:4992-7
100. Kaul S, Murphy PJ, Chen J, Brown L, Pratt WB, Simons SS, Jr. 2002. Mutations at positions 547-553 of rat glucocorticoid receptors reveal that hsp90 binding requires the presence, but not defined composition, of a seven-amino acid sequence at the amino terminus of the ligand binding domain. *J Biol Chem* 277:36223-32
101. Pratt WB, Galigniana MD, Harrell JM, DeFranco DB. 2004. Role of hsp90 and the hsp90-binding immunophilins in signalling protein movement. *Cell Signal* 16:857-72
102. Harrell JM, Murphy PJ, Morishima Y, Chen H, Mansfield JF, et al. 2004. Evidence for glucocorticoid receptor transport on microtubules by dynein. *J Biol Chem* 279:54647-54
103. Thadani-Mulero M, Nanus DM, Giannakakou P. 2012. Androgen receptor on the move: boarding the microtubule expressway to the nucleus. *Cancer Res* 72:4611-5
104. Galigniana MD, Erlejman AG, Monte M, Gomez-Sanchez C, Piwien-Pilipuk G. 2010. The hsp90-FKBP52 complex links the mineralocorticoid receptor to motor proteins and persists bound to the receptor in early nuclear events. *Mol Cell Biol* 30:1285-98
105. Bridgham JT, Carroll SM, Thornton JW. 2006. Evolution of hormone-receptor complexity by molecular exploitation. *Science* 312:97-101
106. Arriza JL, Weinberger C, Cerelli G, Glaser TM, Handelin BL, et al. 1987. Cloning of human mineralocorticoid receptor complementary DNA: structural and functional kinship with the glucocorticoid receptor. *Science* 237:268-75
107. Ortlund EA, Bridgham JT, Redinbo MR, Thornton JW. 2007. Crystal structure of an ancient protein: evolution by conformational epistasis. *Science* 317:1544-8
108. Kadmiel M, Cidlowski JA. 2013. Glucocorticoid receptor signaling in health and disease. *Trends Pharmacol Sci* 34:518-30
109. Kassel O, Herrlich P. 2007. Crosstalk between the glucocorticoid receptor and other transcription factors: molecular aspects. *Mol Cell Endocrinol* 275:13-29
110. Surjit M, Ganti KP, Mukherji A, Ye T, Hua G, et al. 2011. Widespread negative response elements mediate direct repression by agonist-liganded glucocorticoid receptor. *Cell* 145:224-41
111. Hudson WH, Youn C, Ortlund EA. 2013. The structural basis of direct glucocorticoid-mediated transrepression. *Nat Struct Mol Biol* 20:53-8
112. Tuckermann JP, Reichardt HM, Arribas R, Richter KH, Schutz G, Angel P. 1999. The DNA binding-independent function of the glucocorticoid receptor mediates repression of AP-1-dependent genes in skin. *J Cell Biol* 147:1365-70
113. Langlais D, Couture C, Balsalobre A, Drouin J. 2012. The Stat3/GR interaction code: predictive value of direct/indirect DNA recruitment for transcription outcome. *Mol Cell* 47:38-49
114. Auphan N, DiDonato JA, Rosette C, Helmberg A, Karin M. 1995. Immunosuppression by glucocorticoids: inhibition of NF-kappa B activity through induction of I kappa B synthesis. *Science* 270:286-90
115. Bauersachs J, Jaisser F, Toto R. 2015. Mineralocorticoid receptor activation and mineralocorticoid receptor antagonist treatment in cardiac and renal diseases. *Hypertension* 65:257-63
116. Funder JW. 2005. Mineralocorticoid receptors: distribution and activation. *Heart Fail Rev* 10:15-22

117. Yang J, Young MJ. 2009. The mineralocorticoid receptor and its coregulators. *J Mol Endocrinol* 43:53-64
118. Rossier BC, Pradervand S, Schild L, Hummler E. 2002. Epithelial sodium channel and the control of sodium balance: interaction between genetic and environmental factors. *Annu Rev Physiol* 64:877-97
119. Horisberger JD, Lemas V, Kraehenbuhl JP, Rossier BC. 1991. Structure-function relationship of Na,K-ATPase. *Annu Rev Physiol* 53:565-84
120. Kuster GM, Kotlyar E, Rude MK, Siwik DA, Liao R, et al. 2005. Mineralocorticoid receptor inhibition ameliorates the transition to myocardial failure and decreases oxidative stress and inflammation in mice with chronic pressure overload. *Circulation* 111:420-7
121. Qin W, Rudolph AE, Bond BR, Rocha R, Blomme EA, et al. 2003. Transgenic model of aldosterone-driven cardiac hypertrophy and heart failure. *Circ Res* 93:69-76
122. Armani A, Marzolla V, Fabbri A, Caprio M. 2015. Cellular mechanisms of MR regulation of adipose tissue physiology and pathophysiology. *J Mol Endocrinol* 55:R1-10
123. Caprio M, Feve B, Claes A, Viengchareun S, Lombes M, Zennaro MC. 2007. Pivotal role of the mineralocorticoid receptor in corticosteroid-induced adipogenesis. *FASEB J* 21:2185-94
124. Marzolla V, Armani A, Feraco A, De Martino MU, Fabbri A, et al. 2014. Mineralocorticoid receptor in adipocytes and macrophages: a promising target to fight metabolic syndrome. *Steroids* 91:46-53
125. Rickard AJ, Morgan J, Tesch G, Funder JW, Fuller PJ, Young MJ. 2009. Deletion of mineralocorticoid receptors from macrophages protects against deoxycorticosterone/salt-induced cardiac fibrosis and increased blood pressure. *Hypertension* 54:537-43
126. Joels M, Karst H, DeRijk R, de Kloet ER. 2008. The coming out of the brain mineralocorticoid receptor. *Trends Neurosci* 31:1-7
127. Pietranera L, Brocca ME, Cymeryng C, Gomez-Sanchez E, Gomez-Sanchez CE, et al. 2012. Increased expression of the mineralocorticoid receptor in the brain of spontaneously hypertensive rats. *J Neuroendocrinol* 24:1249-58
128. ter Heegde F, De Rijk RH, Vinkers CH. 2015. The brain mineralocorticoid receptor and stress resilience. *Psychoneuroendocrinology* 52:92-110
129. Overington JP, Al-Lazikani B, Hopkins AL. 2006. How many drug targets are there? *Nat Rev Drug Discov* 5:993-6
130. Moore JT, Collins JL, Pearce KH. 2006. The nuclear receptor superfamily and drug discovery. *ChemMedChem* 1:504-23
131. Corvol P, Michaud A, Menard J, Freifeld M, Mahoudeau J. 1975. Antiandrogenic effect of spirolactones: mechanism of action. *Endocrinology* 97:52-8
132. Kolkhof P, Borden SA. 2012. Molecular pharmacology of the mineralocorticoid receptor: prospects for novel therapeutics. *Mol Cell Endocrinol* 350:310-7
133. Newell-Price J, Bertagna X, Grossman AB, Nieman LK. 2006. Cushing's syndrome. *Lancet* 367:1605-17
134. Lee JM, Lee YK, Mamrosh JL, Busby Sa, Griffin PR, et al. 2011. A nuclear-receptor-dependent phosphatidylcholine pathway with antidiabetic effects. *Nature*
135. Bolado-Carrancio A, Riancho JA, Sainz J, Rodriguez-Rey JC. 2014. Activation of nuclear receptor NR5A2 increases Glut4 expression and glucose metabolism in muscle cells. *Biochemical and biophysical research communications* 446:614-9
136. Clyne CD, Speed CJ, Zhou J, Simpson ER. 2002. Liver receptor homologue-1 (LRH-1) regulates expression of aromatase in preadipocytes. *The Journal of biological chemistry* 277:20591-7

137. Benod C, Vinogradova MV, Jouravel N, Kim GE, Fletterick RJ, Sablin EP. 2011. Nuclear receptor liver receptor homologue 1 (LRH-1) regulates pancreatic cancer cell growth and proliferation. *Proc Natl Acad Sci U S A* 108:16927-31
138. Chand AL, Wijayakumara DD, Knower KC, Herridge KA, Howard TL, et al. 2012. The orphan nuclear receptor LRH-1 and ERalpha activate GREB1 expression to induce breast cancer cell proliferation. *PLoS One* 7:e31593
139. Thornton JW, Need E, Crews D. 2003. Resurrecting the ancestral steroid receptor: ancient origin of estrogen signaling. *Science* 301:1714-7
140. Carroll SM, Bridgham JT, Thornton JW. 2008. Evolution of hormone signaling in elasmobranchs by exploitation of promiscuous receptors. *Mol Biol Evol* 25:2643-52
141. Bridgham JT, Brown JE, Rodriguez-Mari A, Catchen JM, Thornton JW. 2008. Evolution of a new function by degenerative mutation in cephalochordate steroid receptors. *PLoS Genet* 4:e1000191
142. Bridgham JT, Ortlund EA, Thornton JW. 2009. An epistatic ratchet constrains the direction of glucocorticoid receptor evolution. *Nature* 461:515-9
143. Carroll SM, Ortlund EA, Thornton JW. 2011. Mechanisms for the evolution of a derived function in the ancestral glucocorticoid receptor. *PLoS Genet* 7:e1002117
144. Eick GN, Colucci JK, Harms MJ, Ortlund EA, Thornton JW. 2012. Evolution of minimal specificity and promiscuity in steroid hormone receptors. *PLoS Genet* 8:e1003072

Chapter 2: Phospholipid-driven gene regulation

Summary

Phospholipids, well known for their fundamental role in cellular structure, play critical signaling roles via their derivatives and cleavage products acting as second messengers in signaling cascades. Recent work has shown that intact PLs act as signaling molecules in their own right by modulating the activity of nuclear hormone transcription factors responsible for tuning genes involved in metabolism, lipid flux, steroid synthesis and inflammation. As such, PLs have been classified as novel hormones. This review highlights recent work in PL-driven gene regulation with a focus on the unique structural features of phospholipid-sensing transcription factors and what sets them apart from well-known soluble phospholipid transporters.

This chapter has been slightly modified from the published manuscript:

Musille PM,* Kohn JA,* Ortlund EA. 2013. Phospholipid – Driven gene regulation. *FEBS Letters* 587:1238-46

*These authors contributed equally to the writing of this manuscript

Introduction

Phospholipids

Phospholipids (PLs) are ubiquitous to all forms of life serving as the major constituent of the membranes that isolate and protect cells from their external environment, and segregate organelles from the greater cellular milieu. PLs are composed of two hydrophobic tails, donated by a diacylglycerol (DAG), and a hydrophilic head group containing a phosphate, which is frequently conjugated to an additional hydrophilic metabolite (Figure 2.1). This amphipathic, bipartite structure drives their spontaneous assembly into bilayers, which compartmentalize the cell and harbor an assortment of proteins, glycans, and other lipids that play critical roles in cell structure, function, metabolism, and signaling.

PLs as signaling molecules

Though best known for their role in membrane construction, PLs play integral roles in a number of cellular signaling cascades at and within the membrane bilayer (1). Arguably the most familiar of these are the IP_3/DAG and Akt cascades. In the former, membrane-bound PI-bisphosphate (PIP_2) is cleaved by PLC to yield inositol trisphosphate (IP_3) and DAG; IP_3 is released into the cytoplasm and triggers the release of Ca^{2+} from the endoplasmic reticulum, while DAG remains in the plasma membrane and activates PKC (2). PI-trisphosphate (PIP_3) is instrumental in recruiting Akt to the plasma membrane, where it is activated by PDK-1 (3). In more recent years, additional PL derivatives have been implicated in cell signaling. Lysophospholipids, single-chain PLs that include sphingosine-1-phosphate (S1P) and lysophosphatidic acid (LPA), were found to bind and activate G protein-coupled receptors (GPCRs) upstream of Ras homolog gene family, member A (RhoA) activation, affecting numerous signaling responses (4). Furthermore, a family of tail-oxidized PLs are now known to play central roles in the regulation of the plasma membrane and the innate immune system (5). PLs have therefore emerged as key players in the signal cascades that control many vital biological processes.

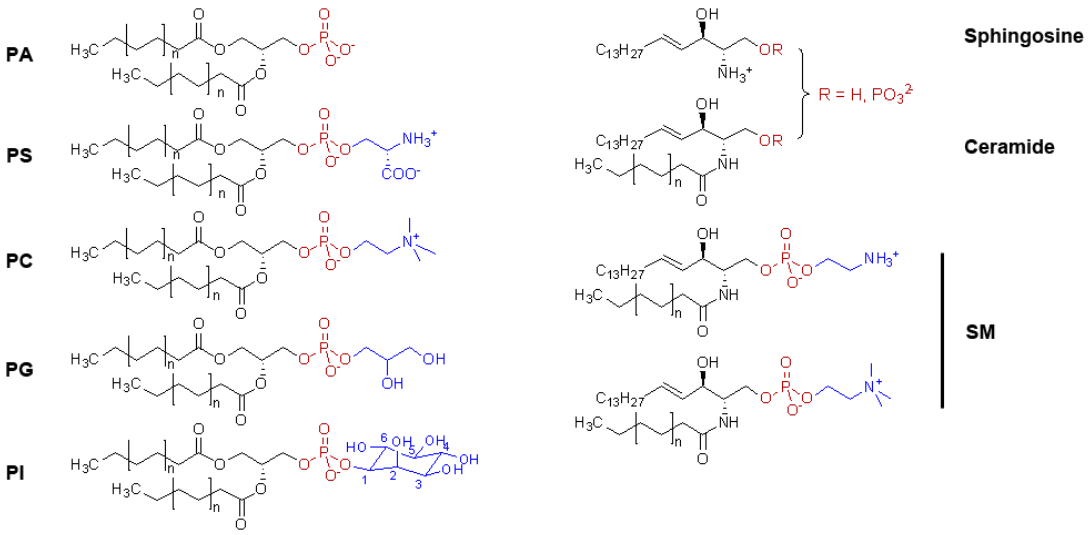

Figure 2.1. Structures of major phospholipid species.

PLs consist of a hydrophobic diacyl tail (black), a phosphate (red), and a polar head group (blue). PA: phosphatidic acid; PS: phosphatidylserine; PC: phosphatidylcholine; PG: phosphatidyl glycerol; PI: phosphatidylinositol; SM: sphingomyelin.

PLs outside the membrane

A significant fraction of the cellular PL pool resides outside of the membrane, particularly inside the nucleus. While some of this subpopulation may have structural roles as part of chromatin or the nuclear lamin (6), it is now evident that there is a PL signaling system distinct from that which occurs within the membrane bilayer (7). PIs again are at the core of the known nuclear lipid signaling pathways (8), and while the nature of nuclear PLs remains enigmatic, it is now understood that PI and PIPs have important functions in the regulation of protein-chromatin interactions (9). The close association of PLs with DNA (10) suggests that, in addition to their roles in cell structure and signal transduction, PLs play a role in driving gene expression and regulation.

PLs are a new class of hormone

Ernest Starling coined the term "hormone" in 1905, long before the isolation of the first nuclear receptor (NR) in 1958, to describe a substance that is able to travel throughout an organism serving as a chemical messenger to alter cell behavior. PLs have long been thought of as synthesis material for some hormones, but new evidence suggests they are transmitting their own unique signals to alter transcriptional patterns. The vast majority of evidence for direct PL-mediated transcription is among the NR family of transcription factors.

Nuclear Receptors: lipid regulated transcription factors

Nuclear receptor structure and function

NRs are a family of ligand regulated transcription factors that are activated by a diverse group of lipophilic ligands including fatty acids, cholesterol derivatives, steroid hormones, vitamins, dietary components, and xenobiotics (11-14). These ligands, primarily derived from lipids, act as messengers by transmitting chemical information that reflects the body's nutritional and endocrine states (15). This allows for the coordination of growth, reproduction, and

homeostasis, and allows the body to appropriately respond to events, such as eating a meal, exercise, or stress.

NRs share a highly conserved multi-domain architecture including a variable N-terminal domain, often referred to as the activation function 1 (AF-1), a DNA binding domain (DBD), a flexible linker region, and a ligand binding domain (LBD) that contains a ligand sensitive transcriptional switch, the AF-2 (12; 13). Ligand dependent NR activation is centered on the LBD, a helical bundle containing a lipophilic cavity that can accommodate ligands. The hydrophobic pockets within NRs typically vary in size and shape to match their cognate hormone (13; 14). A mobile ligand sensing helix, termed the activation function helix (AF-H), responds to a bound ligand by rotating and packing against the LBD. This repositioning completes the AF-2 surface, enabling interaction with coactivator proteins contained in chromatin modifying complexes that promote gene transcription (12). In the absence of ligand, NRs preferentially interact with corepressor complexes which displace the "active AF-H" from the body of the protein resulting in transcriptional repression (12). Similarly, NR antagonists alter AF-H positioning to either prevent coactivator binding or promote binding of corepressor proteins to inhibit transcription.

NRs ligands are invariably hydrophobic and freely diffuse across membranes to allow for long-range signal transmission. In this way, hormones affect diverse groups of gene programs involved in pathophysiology ranging from diabetes to cancer, making NRs ideal targets for pharmacological intervention. As such, NR-targeting drugs have a myriad of uses ranging from cancer treatments, and contraceptives, to treating allergic reactions and metabolic disorders and represent a major industrial and academic investment in basic research and drug development (14; 16; 17).

PL-driven NR activation

To date, four NRs have been identified as PL-binding proteins: liver receptor homolog 1 (LRH-1) and steroidogenic factor 1 (SF-1), members of the NR5a class of steroidogenic factor-

like NRs; peroxisome proliferator-activated receptor alpha (PPARα), a member of the NR1 thyroid hormone receptor-like family of receptors; and ultraspiracle (USP), the insect homolog of the retinoid X receptor. This review will focus on the compelling evidence for PLs role in regulating these receptors, as well as a family of PL transporters that stimulate NR transactivation.

Case Studies

LRH-1

LRH-1 is a member of the NR5, or Ftz-f1, subfamily of NR's, and regulates the expression of genes involved in development, lipid and glucose homeostasis, steroidogenesis, and cell proliferation (18; 19). During the early stages of development, LRH-1 is responsible for maintaining levels of OCT-4, considered to be a master regulator of pluripotency (20). Disruption of the LRH-1 gene in mice leads to the loss of Oct4 expression in the epiblast, causing lethality at embryonic day 6.5 (21). Over expression of LRH-1 is sufficient to reprogram murine somatic cells to pluripotent cells without simultaneous overexpression of OCT-4. This makes LRH-1 the only known transcription factor that can replace OCT-4 in the cellular reprogramming identifying it as a new stem cell factor (22). It is unknown what role LRH-1 plays in OCT4 regulation beyond development, however, the receptor was recently shown to regulate OCT4 expression in human cancer stem cells (23).

In adults, LRH-1 is expressed in liver, pancreas, intestine, brain and sex glands such as the ovaries and placenta (18; 24). In the liver, LRH-1 is a master regulator of lipid homeostasis (19), regulating bile acid and cholesterol flux through regulation of CYP7A1, which catalyzes the rate-limiting step in bile acid synthesis (18). LRH-1 also regulates the transcription a number of other lipid, bile, and cholesterol synthesis enzymes and transporters required in the processes of lipid transport to the liver and elimination (25-32). Recently, LRH-1 has been identified as a direct transcriptional regulator of glucokinase, responsible for glucose capture in the liver (33).

Disruption of the LRH-1 gene in healthy livers not only disrupted lipogenesis but also resulted in reduced glycogen synthesis and glycolysis in response to acute and prolonged glucose exposure. Taken together, these studies demonstrate LRH-1's influences on metabolic homeostasis by linking PL levels to glucose and lipid metabolism.

LRH-1 is also expressed in preadipocytes and adipocytes surrounding estrogen receptor positive breast cancer cells (24). Here, in conjunction with GATA and protein kinase A, LRH-1 drives the expression of *CYP19* (aromatase), increasing the local estrogen concentration to fuel tumor growth (24; 34). Additionally, LRH-1 appears to take part in a positive feedback loop with active estrogen receptor further enhancing these effects (35).

In the colon, LRH-1 plays a markedly different role in cancer development and progression. Here, LRH-1 has been shown to synergize with the beta-catenin/TCF transcriptional complex to enhance the expression of cell proliferation, growth and survival genes such as cyclin's D1 and E1 (21). Additionally, LRH-1 has also been found to be overexpressed in gastric cancer (36).

Bound E. coli PLs offer the first clue that LRH-1 may be PL regulated.

In 2003, the crystal structure of mouse LRH-1 was reported, showing the receptor held in an active conformation in the absence of a ligand or co-regulatory peptide (37). This structure suggested that LRH-1 might act in a ligand-independent manner, discouraging efforts to pursue LRH-1 as a drug target despite its therapeutic potential. In 2005, however, subsequent crystal structures of human LRH-1 all revealed a large >1,400 $Å^3$ ligand binding pocket (LBP) occupied by a diverse array of PLs including PG, PE, and a rare phosphatidylglycerol-phosphoglycerol (38-40). Mutations designed to reduce PL binding showed decreased transcriptional activity in reporter gene assays and a decrease in the ability to recruit coregulators and coregulator fragments both *in vitro* and in cells (39; 41). These exciting new findings showed for the first time that LRH-1 might be regulated by PLs.

LRH-1 – PIP interactions

To identify plausible mammalian PL ligands, Krylova et al. assessed binding of LRH-1 to immobilized PLs which revealed that LRH-1 bound to a range of PLs, but bound most strongly to PIP2 and PIP3 species (40). Lipid binding was confirmed through non-denaturing mass spectrometry (40). LBP mutations designed to prevent lipid binding decreased the ability of LRH-1 to bind these immobilized lipids (40). Notably, this assay did not show PC binding for either LRH-1 or SF-1 (40), both of which were later shown to be activated by PC in cells and bind PC *in vitro* (41; 42).

DLPC

Recently, Lee et al. showed that both human and mouse LRH-1 are specifically activated by the exogenous medium chain phosphatidylcholine isoforms – diundecanoyl (DUPC, PC 11:0/11:0) and dilauroyl (DLPC, PC 12:0/12:0) phosphatidylcholine (43). These medium chain PC agonists selectively activate the receptor in luciferase assays, increase the ability of LRH-1 to interact with the coactivators and increase the production of LRH-1 target genes (43). Moreover, DLPC lowers serum lipid levels and reduces blood glucose levels in diabetic mice in a LRH-1 dependent manner (43). The X-ray crystal structure of the LRH-1–DLPC complex in combination with hydrogen-deuterium exchange assays confirmed that DLPC interacts directly with LRH-1 and revealed the mechanism dictating DLPC-driven transcriptional activation (41). Unlike other NRs that rely on intra-protein interactions to coordinate activation, LRH-1 relies on intramolecular contacts between distal residues in the LBP and the PL to sense and transmit ligand status to the AF-H (41). Additionally, generation and characterization of apo LRH-1, showed that ligand free LRH-1 LBD has a highly destabilized structure that is profoundly stabilized by lipids (41). DLPC simultaneously enhanced co-activator peptide recruitment while disfavoring repressor peptide interaction (41). These recent results show for the first time that LRH-1 is able to dynamically respond to a PL ligand.

SF-1

SF-1, another member of the Ftz-F1 NR5A subfamily, is a key regulator of steroidogenesis and the development of steroidogenic organs, such as the adrenal cortex and gonads (44). It is expressed primarily in these tissues, and in tissues along the steroid hormone regulatory axes, including the hypothalamus and pituitary gland (45; 46). Genes involved in nearly all stages of steroid biosynthesis are regulated by SF-1, including those that encode HMG-CoA synthase (47), cholesterol transporters (48-50), 3β steroid dehydrogenase, and many of the cytochrome P450 enzymes that catalyze the conversion of cholesterol into steroid hormones (51).

Dysfunction of SF-1 has been linked to a number of human disorders (52; 53). Mutations in SF-1 have been detected in patients with disorders in sexual development (54-57), ovarian insufficiency (55), and adrenal failure ,(56), while SF-1 dysregulation has been linked to endometriosis (58) and adrenocortical carcinoma (59). Like LRH-1, SF-1 makes an alluring drug target, yet a robust understanding of its ligand-binding properties is only now emerging.

However, some headway has been made in identifying synthetic compounds that act upon SF-1. In 2008, a number of inverse agonists for SF-1 were identified (60-62). Not only could these compounds inhibit SF-1-dependent gene transcription in luciferase assays, they also inhibited StAR expression in human adrenocortical cells (60), suggesting a possible therapeutic value in the treatment of adrenocortical cancers. Isoquinolone-derived inverse agonists were subsequently shown to inhibit the expression of CYP21 and CYP17 mRNA *in vitro*, with a concurrent reduction in the secretion of aldosterone, cortisol, and DHEA-S, and inhibition of adrenocortical carcinoma cell proliferation (5; 63). These results indicate that pharmacological modulation of SF-1 may be a viable strategy in treating adrenocortical carcinomas, and possibly other human diseases. However, more research is needed to understand the intricacies of ligand-driven SF-1 activity, before its full potential as a drug target can be realized.

E. coli PL binding from early structural studies

The first crystal structures of SF-1 were reported in 2005, showing the LBD in complex with copurified *E. coli* medium chain PG and PE species (38; 40; 64). The binding of SF-1 to immobilized eukaryotic PLs was tested along with LRH-1, and it was found that SF-1 could bind to an array of PL species, including PA, PI, PIP_2, and PIP_3, with a preference for PIPs phosphorylated at the 3- and 5-carbons (40). Coactivator recruitment was enhanced by Pes (38; 64) and PCs (64), identifying diverse PLs as activating ligands *in vitro*.

PA versus sphingosine

The discovery that SF-1 could bind exogenous PLs intensified the search for its endogenous ligands. By 2007, mass spectrometry experiments had identified sphingosine, lysoSM, PA, PE, and PI bound to SF-1 that had been immunoprecipitated from human adrenocarcinoma cells (65; 66). Further analysis showed that sphingosine acts as a SF-1 antagonist, blocking cAMP-stimulated CYP17 reporter gene activity and coactivator recruitment, which could be negated by inhibiting the acid ceramidases that produce sphingosine from ceramide, or by introducing mutations into the LBP that abrogated sphingosine binding (65). Subsequently, it was found that PA activated SF-1-dependent CYP17 expression and transcriptional activity, SF-1 heterocomplex assembly, and steroidogenesis. These effects could be inhibited by sphingosine or by LBP mutations (66).

These data suggest a model, wherein SF-1 is maintained in an inactive conformation by sphingosine under basal conditions (65; 67) and is activated by the binding of PA, which is generated subsequent to ACTH/cAMP signaling (66). The two different lipid species have opposing effects on the activity of SF-1, suggesting a regulatory mechanism in which the levels of these two lipids control the expression of genes linked to SF-1.

PIP2 versus PIP3

While no structures of a SF-1–PI or SF-1–PIP complex have been reported, modeling studies showed that phosphorylated PIs may be stabilized by several histidine residues around the

mouth of the SF-1 LBP (68). Mutations to these residues greatly impaired exchange of bacterial PG with PIP2 and PIP3 and diminished SF-1 transcriptional activity, suggesting that the binding of PIPs to SF-1 is a biologically relevant interaction (68). Indeed, IPMK phosphorylates PIP2 only when bound to SF-1, increasing downstream gene transcription; likewise, PTEN cleaves PIP3 only when complexed with SF-1, attenuating downstream activity (69). Thus, the PIP–SF-1 interaction appears to introduce a regulatory mechanism not previously seen in NRs, in which the phosphorylation status of a bound ligand dictates the activity of its receptor.

PPARs

The peroxisome proliferator-activated receptors (PPARs α, β/δ, and γ) are members of the NR1C subfamily of NRs and play integral roles in the regulation of lipid metabolism and inflammation (70-72). PPARs form heterodimers with the retinoid X receptor (RXR) (73), and recognize an array of ligands, including fatty acids, eicosinoids, and oxidized lipid products (72).

PPARα and PC 16:0/18:1

PPARα is expressed in the heart, liver, kidney, muscle, and brown adipose tissue (74). As a fatty acid binding protein, PPARα regulates the expression of many proteins involved in cellular fatty acid homeostasis (75-77) and systemic lipid balance (78). It has been implicated in atherosclerosis and dyslipidemia, and prolonged activation has been linked to oxidative damage and liver cancer (79). As such, PPARα is an important pharmacological target. Fibrates, a class of drugs used to treat dyslipidemia, are pharmacological agonists of PPARα, and exert their therapeutic effects by lowering triglyceride levels (80).

PPARα is known to bind to many natural free fatty acids (FFAs) and while these are likely physiological ligands, proving that these are *bona fide* endogenous activators is technically challenging. Like PLs, FFAs are typically insoluble, partitioning into droplets, membranes and soluble lipid binding proteins making direct correlations between binding affinity and activation difficult. It is clear, however, that μM levels of exogenous FFAs (1 – 50 μM) activate PPARs *in vivo* and in animals (81). This is on par with PL-dependent transactivation among NR5A

receptors, which display EC50 values ranging from 30 – 100 μM for activating PC and PE isoforms (38; 42). This affinity for FFAs and PLs among nuclear receptors is likely a result of their "generous" lipid binding pockets which allow binding to an array of lipid metabolites.

In 2009, mass spectrometry experiments identified PC 16:0/18:1 as one of several lipids bound to PPARα isolated from murine liver tissue, and the only one whose presence was dependent on fatty acid synthase (FAS) (81). Binding of this PC species was selective for PPARα over PPARδ and PPARγ, and could be enhanced *in vivo* by FAS induction, and inhibited by treatment with a PPARα agonist (81). Additionally, PC 16:0/18:1 treatment stimulated PPARα-dependent gene expression and decreased fatty liver symptoms in mice, lending further credence to its suggested role as an endogenous PPARα agonist (81).

PPARγ and tail-oxidized PLs

PPARγ, which regulates glucose and fatty acid metabolism, is an important target in the treatment of type II diabetes, and is the receptor upon which the thiazolidinedione class of drugs acts (82). In addition to metabolic regulation, PPARγ is known to be an important player in anti-inflammatory pathways (83). Recently, 15-KETE- and 15-HETE PE, two oxidized PE species, were shown to activate PPARγ *in vitro*. Reporter gene assays showed a dose dependent activation in HEK293 cells cotransfected with PPARγ and a PPRE-luciferase construct, and in macrophages harvested from PPRE-EGFP transgenic mice. Furthermore, these oxidized PEs induce the PPARγ-dependent expression of CD36 in human monocytes (84). Unoxidized PE showed no PPARγ activation, suggesting that PPARγ may specifically recognize oxidized PLs. While the formation of oxidized PEs is not dependent on lipases, it remains possible that phospholipase A (PLA) isoforms may liberate oxidized fatty acids, which are also known PPAR activators. Earlier work showed that oxidized PLs bind directly to the LBP, and PPARγ protects these oxidized PLs from phospholipase A1 mediated cleavage; however, this same work showed that PLA1 treated oxidized PLs had a similar ability to stimulate PPARγ transactivation relative

to untreated oxidized PLs (85). For PPARα, however, PLA2 appears to be required for activation by oxidized PLs (86).

USP

Ultraspiracle protein (USP) was identified as the *Drosophila* homolog of mammalian RXR in 1990 (87; 88). Its major function is to serve as a binding partner for the ecdysone receptor (EcR); this heterodimer is a vital regulator of molting and metamorphosis, which is triggered by the binding of 20-hydroxyecdysone (20E) to the EcR subunit (89). However, USP itself can bind to several farnesoid insect juvenile hormones (90), and it is hypothesized to be a ligand-activated NR in its own right (91).

E. Coli PLs

Crystal structures of USP consistently show bacterially-derived PL bound in the LBP (92-95), stabilizing the receptor in an antagonist conformation (93). While most data implicate farnesoid derivatives as the endogenous USP ligand, it is conceivable that insect PLs may play a role in USP-mediated gene regulation, given the emerging role of PLs in other NR pathways. Insects have coopted PLs in the regulation of SREBP processing and nuclear translocation and may have independently evolved PL sensitive NRs. A comparison of the USP-PL crystal structures reveals a nearly identical mode of PL binding versus LRH-1 and SF-1.

PL transport and PL dependent coactivation

PPAR and PC-TP

In addition to direct NR-mediated gene expression, PLs have been shown to indirectly affect gene regulation through lipid shuttling proteins such as phosphatidylcholine transfer protein (PC-TP). PC-TP is a member of the steroidogenic acute regulatory protein (StAR)-related lipid transfer (START) domain superfamily that shares a common fold for lipid binding (96; 97). PC-TP is exquisitely selective for PCs (98), and was originally shown to catalyze both one-for-one PC exchange, and net PC transfer between membranes (99-101). PC-TP has since been identified

as an important metabolic regulator, participating in hepatobiliary cholesterol, lipoprotein, glucose and fatty acid metabolism as well as brown fat-mediated thermogenesis (102).

Consistent with PC-TP's participation in metabolic processes, it has been identified as a binding partner for multiple metabolic proteins (103). Arguably, the most interesting of these interactions is with PPAR-α (104). In addition to PPAR-α regulating the expression of PC-TP, PC-TP was shown to up regulate the transcriptional activity of both PPAR-α and HNF-4α (104). The mechanism of this effect on the transcriptional activity of NRs is not currently understood. Additionally, the context in which NRs bind to PL transporters is also unclear. There is a possibility that in addition to its role in the distribution of lipids in membranes, PC-TP may also deliver PL ligands to PL-sensitive receptors.

Structural Analysis of PL binding proteins

What does it take to bind to PLs as a ligand?

With a large aliphatic surface and significant conformational freedom for the bulk of the molecular structure, PLs certainly do not look like traditional NR ligands (Figure 2.1). Interaction with the hydrophobic tails, while energetically favorable, does not permit specificity by the usual suspects (*e.g.* H-bonds, salt bridges, cation-π interactions). Below, we discuss the distinction between soluble PL transporters and proteins that utilize the information contained in the PL headgroup to drive intermolecular signaling.

Shuttlers versus transcription factors

Structurally characterized soluble PL transport proteins such as PC-TP and PITPα, fully engulf PLs, interacting substantially with both the lipid tails and the headgroup (Figure 2.2 E-F) (98; 105). Headgroup specificity is generated via H-bonds, ionic interactions and cation-π interactions via residues located at the core of the protein. The lipid tails extend toward the protein surface but remain protected from bulk solvent. This binding mode is in stark contrast to PL-binding NRs, which bury PL tails and present the headgroup at the protein surface (Figure 2.2

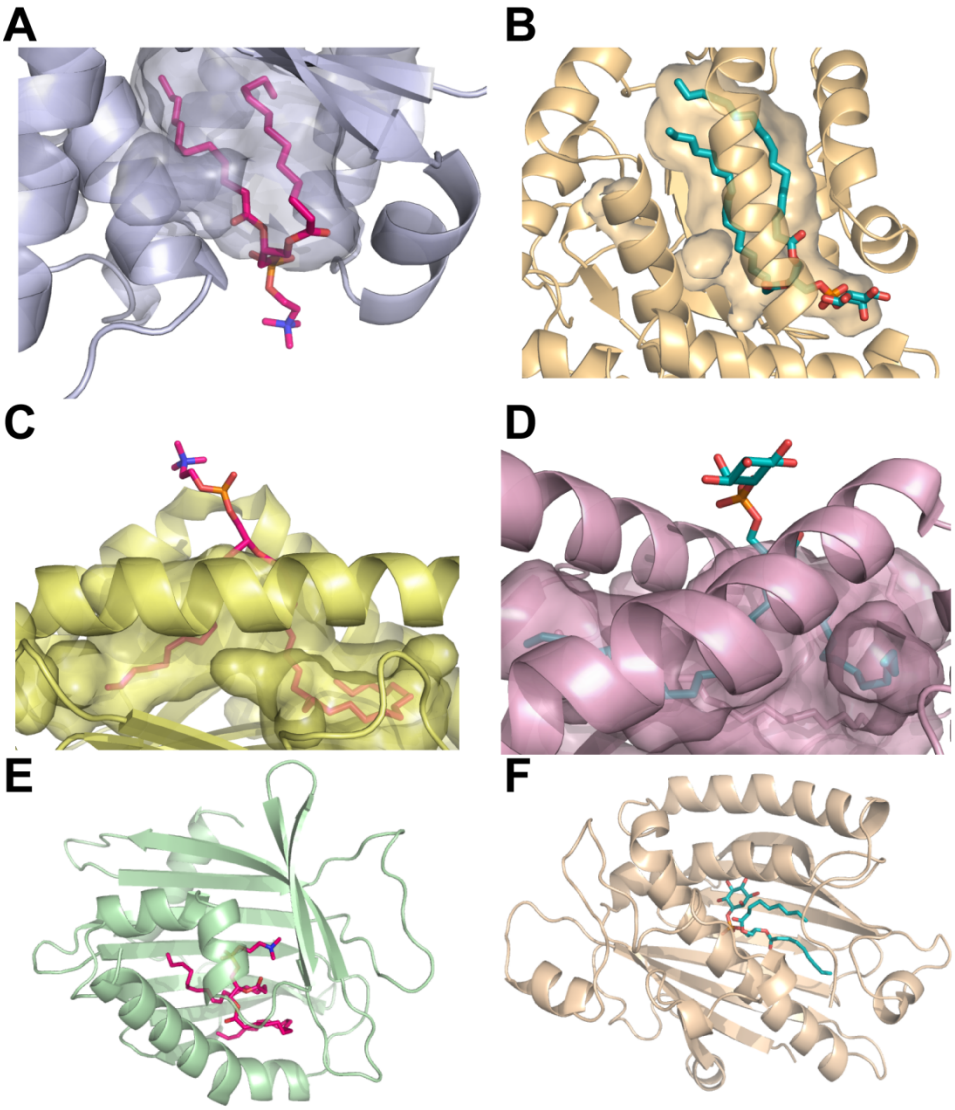

Figure 2.2. Crystal structures of soluble PL signaling proteins.

Proteins are depicted as ribbons with bound phospholipids represented as sticks (O, red; P, magenta; N, blue). Molecular surfaces are shown to highlight the ligand binding pockets. (A) LRH-1 (slate) bound to DLPC (magenta) (41), (B) SFH-1 (tan) bound to PI (cyan) (106)**,** (C) CD-1 (yellow) bound to PC (magenta) (107)**,** (D) CD-1(pink) bound to PI (cyan) (108) showing the bound ligands with lipid head-groups exposed to solvent. In contrast, the lipid shuttling proteins (E) PC-TP (light green) bound to PC (magenta) (98) and (F) PITP (almond) bound to PI (cyan) (109) completely engulf their lipid ligands.

A). The average LBP volume in PC-TP and PITPα is 2297 and 3000 $Å^3$, respectively; this is nearly twice as large as the LRH-1, SF-1 and USP LBPs. The molecular volumes of their bound lipids, however, are 874 and 552 $Å^3$, for PCTP and PITPα respectively. It is tempting to speculate the excess cavity volume and "tails out" PL conformation may be due to the requirement that transporters deliver their PL cargo to a target membrane or PL binding receptor prohibiting tight molecular interactions. Consistent with these observations, holo structures of PC-TP and PITPα show that atomic disorder increases distally from the headgroup suggesting less than optimal contacts are made with the PL tails which have vastly more potential energy to contribute to theprotein-ligand interaction.

Parallels in the immune system

Both exogenous and endogenous PLs have been implicated as lipid antigens capable of activating natural killer T cells when presented by CD1 proteins localized on human antigen presenting cells (107; 108). CD1 proteins play a critical role in presenting both pathogen derived lipids and glycoproteins to initiate cell-mediated immunity (110). Like NRs, CD1 glycocproteins bind PLs in a "tails-first" orientation with the PL headgroup exposed to the protein surface. The binding and presentation of both PC and PI by CD1b and CD1d, respectively, is remarkably similar to the presentation of PLs by NRs (Figure 2.2 A and 2.2 C-D), whereby the lipid tails are buried and the headgroup is exposed to solvent. Thus, PL headgroup presentation may be a hallmark of PL dependent signaling.

Comparison to the PL PI/PC transporter Sec14

Sec14, originally defined by its ability to promote the movement of PC and PI between membranes, is now known as an integrator of PL signaling at the membrane (111). To accomplish this, Sec14 senses both PC and PI levels to stimulate PI4-K mediated PI phosphorylation – a process critical for vesicle biogenesis. Interestingly, Sec14 requires both PC binding and PI binding for activity (106), however, a PC/PI exchange model has been proposed whereby PC binding facilitates PI loading. While a direct interaction between Sec14 and PI4-K

has not been observed, presentation of PI for decoration requires that the inositol moiety is accessible to protein surface (Figure 2.2 B). Indeed, while Sec14 completely buries the PC headgroup, the inositol ring of PI requires only the movement of few side chains to access the solvent. These observations parallel what we know for LRH-1/SF-1; they both are capable of binding PC and PI and presentation of the phosphorylated inositol headgroup is required for signaling (SF-1). Furthermore, since DLPC binding has not yet been tested *in vivo*, it is possible that the PC binding ability of LRH-1 and SF-1 may facilitate the loading of PI in a similar exchange reaction.

PL presentation as a model for PL dependent signaling.

Unlike widely prevalent PL binding domains such as PHD fingers that recognize PLs in the context of a membrane (112), NRs engulf PLs "tails first" making extensive hydrophobic contact with more than 15 residues and up to three hydrogen bonds near the surface of the receptor (113). It is clear that most of the binding energy is derived from interaction with the aliphatic tails, which in all known structures, intertwine to fill large 1300-1750 $Å^3$ binding pocket that starts at the core of the protein and terminates at the protein surface. Lipid tails occupy the very core of the receptor greatly enhancing protein stability (41). In this way, PLs act as folding nuclei much like the hormones in other NR family members (114). However, the vast diversity among PLs and the potential for lipid modifications suggests that PL dependent transcription factors may serve to integrate varying and complex signals to tune gene expression. This represents an added layer of complexity on the already complicated cistrome in which coregulators, DNA, chromatin modifying enzymes and accessory proteins orchestrate coordinated gene expression.

Closing Remarks

Evolution has generated a highly complex system to control energy homeostasis, including allosteric mechanisms within key metabolic enzymes, and the nutritional control of

gene expression via transcription factors. Lipids are a major source of energy for the cell, and it is well known that the composition and availability of these lipids plays a central role in regulating glycolysis. Direct PL sensing by nuclear hormone receptors tie PL levels not only into glucose and lipid homeostasis but to steroid synthesis, reproduction, inflammation, development and cell differentiation (Figure 2.3).

Given the molecular properties of PLs, it is no surprise that PL-driven transcription factors have been largely recalcitrant to drug design. Proteins with large hydrophobic pockets typically require large ligands and the potential for specific interactions within core of the LBP are slim. While there have been a few successes in designing specific compounds targeting these receptors, improving these compounds and predicting their binding modes remain challenging. Clearly, modulating PL-driven transcriptional pathways remains an untapped therapeutic opportunity and advances in this area of research are desperately needed.

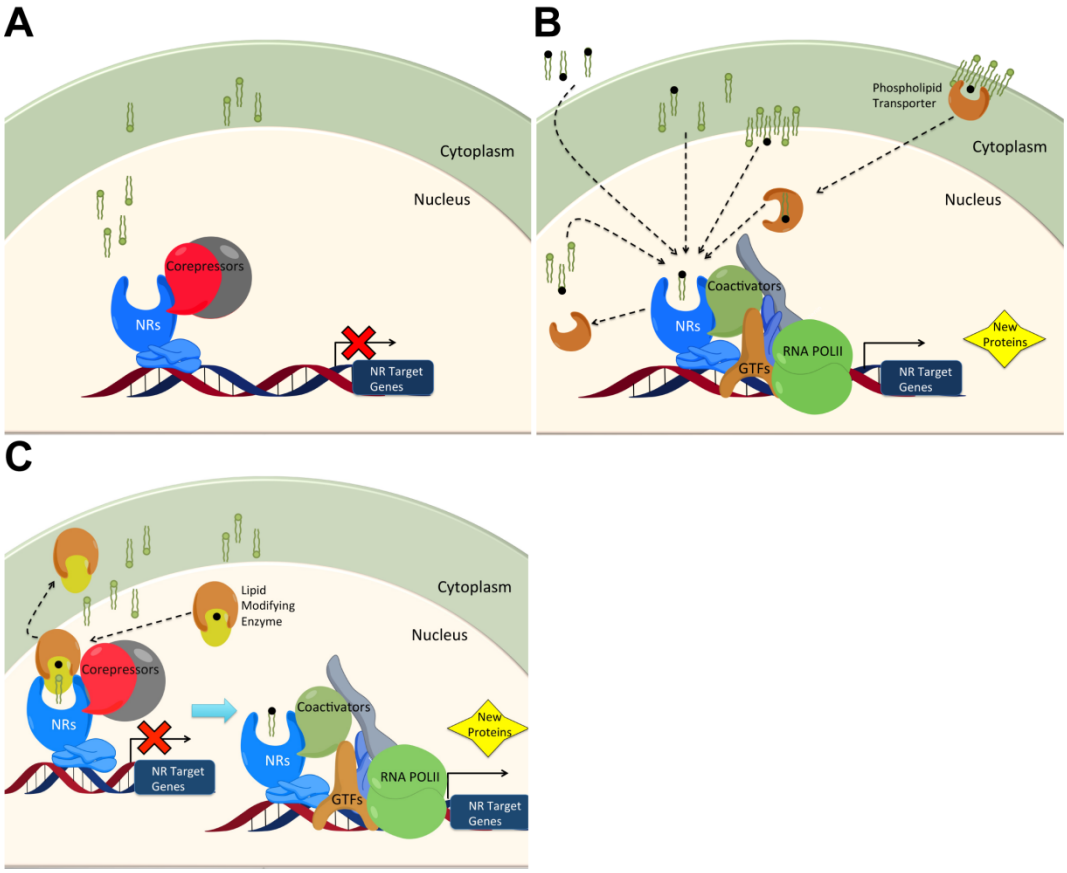

Figure 2.3. Phospholipid mediated transcription control.

(A) In the absence of a PL agonist NRs are bound to corepressor proteins and block transcription. (B) Activating PLs from exogenous, membrane bound or cytoplasmic sources bind to NRs or are potentially delivered by PL transporter proteins. Once an activating PL is bound to the NR coactivator complexes along with other general transcription factors (GTFs) and RNA polymerase initiate the transcription of genes. (C) NRs can also be bound to non-activating lipids with lipid modifying enzymes altering the lipid in place to become an activating lipid.

References

1. Michell RH. 1975. Inositol phospholipids and cell surface receptor function. *Biochim Biophys Acta* 415:81-47
2. Berridge MJ, Irvine RF. 1984. Inositol trisphosphate, a novel second messenger in cellular signal transduction. *Nature* 312:315-21
3. Song G, Ouyang G, Bao S. 2005. The activation of Akt/PKB signaling pathway and cell survival. *J Cell Mol Med* 9:59-71
4. Xiang SY, Dusaban SS, Brown JH. 2013. Lysophospholipid receptor activation of RhoA and lipid signaling pathways. *Biochim Biophys Acta* 1831:213-22
5. O'Donnell VB, Murphy RC. 2012. New families of bioactive oxidized phospholipids generated by immune cells: identification and signaling actions. *Blood* 120:1985-92
6. Irvine RF, Divecha N. 1992. Phospholipids in the nucleus--metabolism and possible functions. *Semin Cell Biol* 3:225-35
7. Irvine RF. 2003. Nuclear lipid signalling. *Nat Rev Mol Cell Biol* 4:349-60
8. Irvine RF. 2002. Nuclear lipid signaling. *Sci STKE* 2002:re13
9. Viiri K, Maki M, Lohi O. 2012. Phosphoinositides as regulators of protein-chromatin interactions. *Sci Signal* 5:pe19
10. Fraschini A, Albi E, Gahan PB, Viola-Magni MP. 1992. TEM cytochemical study of the localization of phospholipids in interphase chromatin in rat hepatocytes. *Histochemistry* 97:225-35
11. McEwan IJ. 2009. Nuclear receptors: one big family. *Methods in molecular biology* 505:3-18
12. Nagy L, Schwabe JW. 2004. Mechanism of the nuclear receptor molecular switch. *Trends Biochem Sci* 29:317-24
13. Huang P, Chandra V, Rastinejad F. 2010. Structural overview of the nuclear receptor superfamily: insights into physiology and therapeutics. *Annu Rev Physiol* 72:247-72
14. Sladek FM. 2011. What are nuclear receptor ligands? *Mol Cell Endocrinol* 334:3-13
15. Kliewer SA, Lehmann JM, Willson TM. 1999. Orphan nuclear receptors: shifting endocrinology into reverse. *Science* 284:757-60
16. Schulman IG, Heyman RA. 2004. The flip side: Identifying small molecule regulators of nuclear receptors. *Chemistry & biology* 11:639-46
17. Moore JT, Collins JL, Pearce KH. 2006. The nuclear receptor superfamily and drug discovery. *ChemMedChem* 1:504-23
18. Fernandez-Marcos PJ, Auwerx J, Schoonjans K. 2011. Emerging actions of the nuclear receptor LRH-1 in the gut. *Biochim Biophys Acta* 1812:947-55
19. Lee Y-k, Moore DD. 2008. Liver receptor homolog-1, an emerging metabolic modulator. *Frontiers in bioscience : a journal and virtual library* 13:5950-8
20. Kellner S, Kikyo N. 2010. Transcriptional regulation of the Oct4 gene, a master gene for pluripotency. *Histology and histopathology* 25:405-12
21. Botrugno Oa, Fayard E, Annicotte J-S, Haby C, Brennan T, et al. 2004. Synergy between LRH-1 and beta-catenin induces G1 cyclin-mediated cell proliferation. *Molecular cell* 15:499-509
22. Heng JC, Feng B, Han J, Jiang J, Kraus P, et al. 2010. The nuclear receptor Nr5a2 can replace Oct4 in the reprogramming of murine somatic cells to pluripotent cells. *Cell Stem Cell* 6:167-74
23. Sung B, Do HJ, Park SW, Huh SH, Oh JH, et al. 2012. Regulation of OCT4 gene expression by liver receptor homolog-1 in human embryonic carcinoma cells. *Biochemical and biophysical research communications*

24. Clyne CD, Speed CJ, Zhou J, Simpson ER. 2002. Liver receptor homologue-1 (LRH-1) regulates expression of aromatase in preadipocytes. *The Journal of biological chemistry* 277:20591-7
25. Goodwin B, Jones Sa, Price RR, Watson Ma, McKee DD, et al. 2000. A regulatory cascade of the nuclear receptors FXR, SHP-1, and LRH-1 represses bile acid biosynthesis. *Molecular cell* 6:517-26
26. Chen F, Ma L, Dawson PA, Sinal CJ, Sehayek E, et al. 2003. Liver receptor homologue-1 mediates species- and cell line-specific bile acid-dependent negative feedback regulation of the apical sodium-dependent bile acid transporter. *J Biol Chem* 278:19909-16
27. del Castillo-Olivares A, Gil G. 2000. Role of FXR and FTF in bile acid-mediated suppression of cholesterol 7alpha-hydroxylase transcription. *Nucleic Acids Res* 28:3587-93
28. Delerive P, Galardi CM, Bisi JE, Nicodeme E, Goodwin B. 2004. Identification of liver receptor homolog-1 as a novel regulator of apolipoprotein AI gene transcription. *Mol Endocrinol* 18:2378-87
29. Freeman LA, Kennedy A, Wu J, Bark S, Remaley AT, et al. 2004. The orphan nuclear receptor LRH-1 activates the ABCG5/ABCG8 intergenic promoter. *J Lipid Res* 45:1197-206
30. Inokuchi A, Hinoshita E, Iwamoto Y, Kohno K, Kuwano M, Uchiumi T. 2001. Enhanced expression of the human multidrug resistance protein 3 by bile salt in human enterocytes. A transcriptional control of a plausible bile acid transporter. *J Biol Chem* 276:46822-9
31. Lee YK, Schmidt DR, Cummins CL, Choi M, Peng L, et al. 2008. Liver receptor homolog-1 regulates bile acid homeostasis but is not essential for feedback regulation of bile acid synthesis. *Mol Endocrinol* 22:1345-56
32. Luo Y, Liang CP, Tall AR. 2001. The orphan nuclear receptor LRH-1 potentiates the sterol-mediated induction of the human CETP gene by liver X receptor. *J Biol Chem* 276:24767-73
33. Oosterveer MH, Mataki C, Yamamoto H, Harach T, Moullan N, et al. 2012. LRH-1-dependent glucose sensing determines intermediary metabolism in liver. *J Clin Invest* 122:2817-26
34. Bouchard MF, Taniguchi H, Viger RS. 2005. Protein kinase A-dependent synergism between GATA factors and the nuclear receptor, liver receptor homolog-1, regulates human aromatase (CYP19) PII promoter activity in breast cancer cells. *Endocrinology* 146:4905-16
35. Annicotte J-S, Chavey C, Servant N, Teyssier J, Bardin A, et al. 2005. The nuclear receptor liver receptor homolog-1 is an estrogen receptor target gene. *Oncogene* 24:8167-75
36. Wang S-L, Zheng D-Z, Lan F-H, Deng X-J, Zeng J, et al. 2008. Increased expression of hLRH-1 in human gastric cancer and its implication in tumorigenesis. *Molecular and cellular biochemistry* 308:93-100
37. Sablin EP, Krylova IN, Fletterick RJ, Ingraham HA. 2003. Structural basis for ligand-independent activation of the orphan nuclear receptor LRH-1. *Molecular cell* 11:1575-85
38. Wang W, Zhang C, Marimuthu A, Krupka HI, Tabrizizad M, et al. 2005. The crystal structures of human steroidogenic factor-1 and liver receptor homologue-1. *Proc Natl Acad Sci U S A* 102:7505-10
39. Ortlund EA, Lee Y, Solomon IH, Hager JM, Safi R, et al. 2005. Modulation of human nuclear receptor LRH-1 activity by phospholipids and SHP. *Nature structural & molecular biology* 12:357-63
40. Krylova IN, Sablin EP, Moore J, Xu RX, Waitt GM, et al. 2005. Structural analyses reveal phosphatidyl inositols as ligands for the NR5 orphan receptors SF-1 and LRH-1. *Cell* 120:343-55

41. Musille PM, Pathak MC, Lauer JL, Hudson WH, Griffin PR, Ortlund EA. 2012. Antidiabetic phospholipid-nuclear receptor complex reveals the mechanism for phospholipid-driven gene regulation. *Nat Struct Mol Biol* 19:532-7, S1-2
42. Lee JM, Lee YK, Mamrosh JL, Busby Sa, Griffin PR, et al. 2011. A nuclear-receptor-dependent phosphatidylcholine pathway with antidiabetic effects. *Nature*
43. Moore D. Targeting nuclear receptors to treat type 2 diabetes. *Proc. 14th International Congress of Endocrinology*, *Kyoto, Japan*, *2010*:
44. Hoivik EA, Lewis AE, Aumo L, Bakke M. 2010. Molecular aspects of steroidogenic factor 1 (SF-1). *Mol Cell Endocrinol* 315:27-39
45. Shinoda K, Lei H, Yoshii H, Nomura M, Nagano M, et al. 1995. Developmental defects of the ventromedial hypothalamic nucleus and pituitary gonadotroph in the Ftz-F1 disrupted mice. *Dev Dyn* 204:22-9
46. Ingraham HA, Lala DS, Ikeda Y, Luo X, Shen WH, et al. 1994. The nuclear receptor steroidogenic factor 1 acts at multiple levels of the reproductive axis. *Genes Dev* 8:2302-12
47. Mascaro C, Nadal A, Hegardt FG, Marrero PF, Haro D. 2000. Contribution of steroidogenic factor 1 to the regulation of cholesterol synthesis. *Biochem J* 350 Pt 3:785-90
48. Sugawara T, Holt JA, Kiriakidou M, Strauss JF, 3rd. 1996. Steroidogenic factor 1-dependent promoter activity of the human steroidogenic acute regulatory protein (StAR) gene. *Biochemistry* 35:9052-9
49. Cao G, Garcia CK, Wyne KL, Schultz RA, Parker KL, Hobbs HH. 1997. Structure and localization of the human gene encoding SR-BI/CLA-1. Evidence for transcriptional control by steroidogenic factor 1. *J Biol Chem* 272:33068-76
50. Lopez D, Shea-Eaton W, McLean MP. 2001. Characterization of a steroidogenic factor-1-binding site found in promoter of sterol carrier protein-2 gene. *Endocrine* 14:253-61
51. Parker KL, Schimmer BP. 1997. Steroidogenic factor 1: a key determinant of endocrine development and function. *Endocr Rev* 18:361-77
52. Schimmer BP, White PC. 2010. Minireview: steroidogenic factor 1: its roles in differentiation, development, and disease. *Mol Endocrinol* 24:1322-37
53. Ferraz-de-Souza B, Lin L, Achermann JC. 2011. Steroidogenic factor-1 (SF-1, NR5A1) and human disease. *Mol Cell Endocrinol* 336:198-205
54. Correa RV, Domenice S, Bingham NC, Billerbeck AE, Rainey WE, et al. 2004. A microdeletion in the ligand binding domain of human steroidogenic factor 1 causes XY sex reversal without adrenal insufficiency. *J Clin Endocrinol Metab* 89:1767-72
55. Camats N, Pandey AV, Fernandez-Cancio M, Andaluz P, Janner M, et al. 2012. Ten novel mutations in the NR5A1 gene cause disordered sex development in 46,XY and ovarian insufficiency in 46,XX individuals. *J Clin Endocrinol Metab* 97:E1294-306
56. Achermann JC, Ito M, Hindmarsh PC, Jameson JL. 1999. A mutation in the gene encoding steroidogenic factor-1 causes XY sex reversal and adrenal failure in humans. *Nat Genet* 22:125-6
57. Lin L, Achermann JC. 2008. Steroidogenic factor-1 (SF-1, Ad4BP, NR5A1) and disorders of testis development. *Sex Dev* 2:200-9
58. Bulun SE, Utsunomiya H, Lin Z, Yin P, Cheng YH, et al. 2009. Steroidogenic factor-1 and endometriosis. *Mol Cell Endocrinol* 300:104-8
59. Pianovski MA, Cavalli LR, Figueiredo BC, Santos SC, Doghman M, et al. 2006. SF-1 overexpression in childhood adrenocortical tumours. *Eur J Cancer* 42:1040-3
60. Del Tredici AL, Andersen CB, Currier EA, Ohrmund SR, Fairbain LC, et al. 2008. Identification of the first synthetic steroidogenic factor 1 inverse agonists: pharmacological modulation of steroidogenic enzymes. *Mol Pharmacol* 73:900-8

61. Madoux F, Li X, Chase P, Zastrow G, Cameron MD, et al. 2008. Potent, selective and cell penetrant inhibitors of SF-1 by functional ultra-high-throughput screening. *Mol Pharmacol* 73:1776-84
62. Roth J, Madoux F, Hodder P, Roush WR. 2008. Synthesis of small molecule inhibitors of the orphan nuclear receptor steroidogenic factor-1 (NR5A1) based on isoquinolinone scaffolds. *Bioorg Med Chem Lett* 18:2628-32
63. Doghman M, Cazareth J, Douguet D, Madoux F, Hodder P, Lalli E. 2009. Inhibition of adrenocortical carcinoma cell proliferation by steroidogenic factor-1 inverse agonists. *J Clin Endocrinol Metab* 94:2178-83
64. Li Y, Choi M, Cavey G, Daugherty J, Suino K, et al. 2005. Crystallographic identification and functional characterization of phospholipids as ligands for the orphan nuclear receptor steroidogenic factor-1. *Mol Cell* 17:491-502
65. Urs AN, Dammer E, Sewer MB. 2006. Sphingosine regulates the transcription of CYP17 by binding to steroidogenic factor-1. *Endocrinology* 147:5249-58
66. Li D, Urs AN, Allegood J, Leon A, Merrill AH, Jr., Sewer MB. 2007. Cyclic AMP-stimulated interaction between steroidogenic factor 1 and diacylglycerol kinase theta facilitates induction of CYP17. *Mol Cell Biol* 27:6669-85
67. Urs AN, Dammer E, Kelly S, Wang E, Merrill AH, Jr., Sewer MB. 2007. Steroidogenic factor-1 is a sphingolipid binding protein. *Mol Cell Endocrinol* 265-266:174-8
68. Sablin EP, Blind RD, Krylova IN, Ingraham JG, Cai F, et al. 2009. Structure of SF-1 bound by different phospholipids: evidence for regulatory ligands. *Mol Endocrinol* 23:25-34
69. Blind RD, Suzawa M, Ingraham HA. 2012. Direct modification and activation of a nuclear receptor-PIP(2) complex by the inositol lipid kinase IPMK. *Sci Signal* 5:ra44
70. Bensinger SJ, Tontonoz P. 2008. Integration of metabolism and inflammation by lipid-activated nuclear receptors. *Nature* 454:470-7
71. Desvergne B, Wahli W. 1999. Peroxisome proliferator-activated receptors: nuclear control of metabolism. *Endocr Rev* 20:649-88
72. Berger J, Moller DE. 2002. The mechanisms of action of PPARs. *Annu Rev Med* 53:409-35
73. Miyata KS, McCaw SE, Marcus SL, Rachubinski RA, Capone JP. 1994. The peroxisome proliferator-activated receptor interacts with the retinoid X receptor in vivo. *Gene* 148:327-30
74. Auboeuf D, Rieusset J, Fajas L, Vallier P, Frering V, et al. 1997. Tissue distribution and quantification of the expression of mRNAs of peroxisome proliferator-activated receptors and liver X receptor-alpha in humans: no alteration in adipose tissue of obese and NIDDM patients. *Diabetes* 46:1319-27
75. Martin G, Schoonjans K, Lefebvre AM, Staels B, Auwerx J. 1997. Coordinate regulation of the expression of the fatty acid transport protein and acyl-CoA synthetase genes by PPARalpha and PPARgamma activators. *J Biol Chem* 272:28210-7
76. Motojima K, Passilly P, Peters JM, Gonzalez FJ, Latruffe N. 1998. Expression of putative fatty acid transporter genes are regulated by peroxisome proliferator-activated receptor alpha and gamma activators in a tissue- and inducer-specific manner. *J Biol Chem* 273:16710-4
77. Dreyer C, Keller H, Mahfoudi A, Laudet V, Krey G, Wahli W. 1993. Positive regulation of the peroxisomal beta-oxidation pathway by fatty acids through activation of peroxisome proliferator-activated receptors (PPAR). *Biol Cell* 77:67-76
78. Schoonjans K, Peinado-Onsurbe J, Lefebvre AM, Heyman RA, Briggs M, et al. 1996. PPARalpha and PPARgamma activators direct a distinct tissue-specific transcriptional response via a PPRE in the lipoprotein lipase gene. *EMBO J* 15:5336-48

79. Pyper SR, Viswakarma N, Yu S, Reddy JK. 2010. PPARalpha: energy combustion, hypolipidemia, inflammation and cancer. *Nucl Recept Signal* 8:e002
80. Forman BM, Chen J, Evans RM. 1997. Hypolipidemic drugs, polyunsaturated fatty acids, and eicosanoids are ligands for peroxisome proliferator-activated receptors alpha and delta. *Proc Natl Acad Sci U S A* 94:4312-7
81. Chakravarthy MV, Lodhi IJ, Yin L, Malapaka RR, Xu HE, et al. 2009. Identification of a physiologically relevant endogenous ligand for PPARalpha in liver. *Cell* 138:476-88
82. Huang JV, Greyson CR, Schwartz GG. 2012. PPAR-gamma as a therapeutic target in cardiovascular disease: evidence and uncertainty. *J Lipid Res* 53:1738-54
83. Martin H. 2009. Role of PPAR-gamma in inflammation. Prospects for therapeutic intervention by food components. *Mutat Res* 669:1-7
84. Hammond VJ, Morgan AH, Lauder S, Thomas CP, Brown S, et al. 2012. Novel Keto-phospholipids Are Generated by Monocytes and Macrophages, Detected in Cystic Fibrosis, and Activate Peroxisome Proliferator-activated Receptor-gamma. *J Biol Chem* 287:41651-66
85. Davies SS, Pontsler AV, Marathe GK, Harrison KA, Murphy RC, et al. 2001. Oxidized alkyl phospholipids are specific, high affinity peroxisome proliferator-activated receptor gamma ligands and agonists. *J Biol Chem* 276:16015-23
86. Delerive P, Furman C, Teissier E, Fruchart J, Duriez P, Staels B. 2000. Oxidized phospholipids activate PPARalpha in a phospholipase A2-dependent manner. *FEBS Lett* 471:34-8
87. Oro AE, McKeown M, Evans RM. 1990. Relationship between the product of the Drosophila ultraspiracle locus and the vertebrate retinoid X receptor. *Nature* 347:298-301
88. Henrich VC, Sliter TJ, Lubahn DB, MacIntyre A, Gilbert LI. 1990. A steroid/thyroid hormone receptor superfamily member in Drosophila melanogaster that shares extensive sequence similarity with a mammalian homologue. *Nucleic Acids Res* 18:4143-8
89. Schwedes CC, Carney GE. 2012. Ecdysone signaling in adult Drosophila melanogaster. *J Insect Physiol* 58:293-302
90. Jones G, Sharp PA. 1997. Ultraspiracle: an invertebrate nuclear receptor for juvenile hormones. *Proc Natl Acad Sci U S A* 94:13499-503
91. Jones D, Jones G. 2007. Farnesoid secretions of dipteran ring glands: what we do know and what we can know. *Insect Biochem Mol Biol* 37:771-98
92. Clayton GM, Peak-Chew SY, Evans RM, Schwabe JW. 2001. The structure of the ultraspiracle ligand-binding domain reveals a nuclear receptor locked in an inactive conformation. *Proc Natl Acad Sci U S A* 98:1549-54
93. Billas IM, Moulinier L, Rochel N, Moras D. 2001. Crystal structure of the ligand-binding domain of the ultraspiracle protein USP, the ortholog of retinoid X receptors in insects. *J Biol Chem* 276:7465-74
94. Billas IM, Iwema T, Garnier JM, Mitschler A, Rochel N, Moras D. 2003. Structural adaptability in the ligand-binding pocket of the ecdysone hormone receptor. *Nature* 426:91-6
95. Browning C, Martin E, Loch C, Wurtz JM, Moras D, et al. 2007. Critical role of desolvation in the binding of 20-hydroxyecdysone to the ecdysone receptor. *J Biol Chem* 282:32924-34
96. Iyer LM, Koonin EV, Aravind L. 2001. Adaptations of the helix-grip fold for ligand binding and catalysis in the START domain superfamily. *Proteins* 43:134-44
97. Ponting CP, Aravind L. 1999. START: a lipid-binding domain in StAR, HD-ZIP and signalling proteins. *Trends Biochem Sci* 24:130-2
98. Roderick SL, Chan WW, Agate DS, Olsen LR, Vetting MW, et al. 2002. Structure of human phosphatidylcholine transfer protein in complex with its ligand. *Nature structural biology* 9:507-11

99. Wirtz KW, Devaux PF, Bienvenue A. 1980. Phosphatidylcholine exchange protein catalyzes the net transfer of phosphatidylcholine to model membranes. *Biochemistry* 19:3395-9
100. Kamp HH, Wirtz WA, Baer PR, Slotboom AJ, Rosenthal AF, et al. 1977. Specificity of the phosphatidylcholine exchange protein from bovine liver. *Biochemistry* 16:1310-6
101. Johnson LW, Zilversmit DB. 1975. Catalytic properties of phospholipid exchange protein from bovine heart. *Biochim Biophys Acta* 375:165-75
102. Kang HW, Wei J, Cohen DE. 2010. PC-TP/StARD2: Of membranes and metabolism. *Trends in endocrinology and metabolism: TEM* 21:449-56
103. Kanno K, Wu MK, Agate DS, Fanelli BJ, Wagle N, et al. 2007. Interacting proteins dictate function of the minimal START domain phosphatidylcholine transfer protein/StarD2. *J Biol Chem* 282:30728-36
104. Kang HW, Kanno K, Scapa EF, Cohen DE. 2010. Regulatory role for phosphatidylcholine transfer protein/StarD2 in the metabolic response to peroxisome proliferator activated receptor alpha (PPARalpha). *Biochim Biophys Acta* 1801:496-502
105. Zheng J, Singh VK, Jia Z. 2005. Identification of an ITPase/XTPase in Escherichia coli by structural and biochemical analysis. *Structure* 13:1511-20
106. Schaaf G, Ortlund EA, Tyeryar KR, Mousley CJ, Ile KE, et al. 2008. Functional anatomy of phospholipid binding and regulation of phosphoinositide homeostasis by proteins of the sec14 superfamily. *Mol Cell* 29:191-206
107. Giabbai B, Sidobre S, Crispin MD, Sanchez-Ruiz Y, Bachi A, et al. 2005. Crystal structure of mouse CD1d bound to the self ligand phosphatidylcholine: a molecular basis for NKT cell activation. *Journal of immunology* 175:977-84
108. Gadola SD, Zaccai NR, Harlos K, Shepherd D, Castro-Palomino JC, et al. 2002. Structure of human CD1b with bound ligands at 2.3 A, a maze for alkyl chains. *Nature immunology* 3:721-6
109. Tilley SJ, Skippen A, Murray-Rust J, Swigart PM, Stewart A, et al. 2004. Structure-function analysis of human [corrected] phosphatidylinositol transfer protein alpha bound to phosphatidylinositol. *Structure* 12:317-26
110. Jullien D, Stenger S, Ernst WA, Modlin RL. 1997. CD1 presentation of microbial nonpeptide antigens to T cells. *J Clin Invest* 99:2071-4
111. Bankaitis VA, Mousley CJ, Schaaf G. 2010. The Sec14 superfamily and mechanisms for crosstalk between lipid metabolism and lipid signaling. *Trends Biochem Sci* 35:150-60
112. Gozani O, Karuman P, Jones DR, Ivanov D, Cha J, et al. 2003. The PHD finger of the chromatin-associated protein ING2 functions as a nuclear phosphoinositide receptor. *Cell* 114:99-111
113. Ingraham Ha, Redinbo MR. 2005. Orphan nuclear receptors adopted by crystallography. *Current opinion in structural biology* 15:708-15
114. Gee AC, Katzenellenbogen JA. 2001. Probing conformational changes in the estrogen receptor: evidence for a partially unfolded intermediate facilitating ligand binding and release. *Mol Endocrinol* 15:421-8

Chapter 3: Unexpected allosteric network contributes to LRH-1 co-regulator selectivity

Summary

Phospholipids (PLs) are unusual signaling hormones sensed by the nuclear receptor liver receptor homolog-1 (LRH-1), which has evolved a novel allosteric pathway to support appropriate interaction with coregulators depending on ligand status. LRH-1 plays an important role in controlling lipid and cholesterol homeostasis and is a potential target for the treatment of metabolic and neoplastic diseases. While the prospect of modulating LRH-1 via small molecules is exciting, the molecular mechanism linking PL structure to transcriptional coregulator preference is unknown. Previous studies showed that binding to an activating PL-ligand, such as dilauroylphosphatidylcholine (DLPC), favors LRH-1's interaction with transcriptional coactivators to upregulate gene expression. Both crystallographic and solution-based structural studies showed that DLPC binding drives unanticipated structural fluctuations outside of the canonical activation surface in an alternate activation function (AF) region, encompassing the β-sheet–H6 region of the protein. However, the mechanism by which dynamics in the alternate AF influences coregulator selectivity remains elusive. Here we pair x-ray crystallography with molecular modeling to identify an unexpected allosteric network that traverses the protein ligand binding pocket and links these two elements to dictate selectivity. We show that communication between the alternate AF region and classical AF2 is correlated with the strength of the coregulator interaction. This work offers the first glimpse into the conformational dynamics that drive this unusual PL-mediated nuclear hormone receptor activation.

This chapter has been slightly modified from the published manuscript:

Musille PM,* Kossmann BR,* Kohn JA* Ivanov I Ortlund EA. 2016. Unexpected allosteric network contributes to LRH-1 coregulator selectivity. *J Biol Chem* 291:1411-26

*These authors contributed equally to the writing of this manuscript

Introduction

Phospholipids (PLs) are best known for their structural role in membranes and as synthesis material for potent signaling molecules, such as dicacylglycerol, leukotrienes, and inositol phosphates. Recent evidence, however, suggests intact PLs are able to directly modulate the activity of transcription factors involved in lipid homeostasis, such as sterol regulatory element-binding protein 1 (SREBP-1), and some members of the nuclear receptor (NR) family of ligand-regulated transcription factors, including peroxisome proliferator activated receptor α (PPARα; NR1C1), steroidogenic factor 1 (SF-1; NR5A1) and human liver receptor homologue-1 (LRH-1; NR5A2) (1-4). LRH-1 regulates the expression of genes central to embryonic development, cell cycle progression, steroid synthesis, lipid and glucose homeostasis, and local immune function (5-12). Thus, LRH-1 is an enticing pharmaceutical target for the treatment of metabolic and neoplastic diseases (6).

Although the endogenous ligand for hLRH-1 is currently unknown, oral treatment with the exogenous PL agonist dilauroylphosphatidylcholine (PC 12:0-12:0; DLPC) lowers serum lipid levels, reduces liver fat accumulation, and improves glucose tolerance in a LRH-1 dependent manner in a diabetic mouse model (13). Activation of LRH-1 by DLPC drives increased glucose uptake by muscle and increases the rate of both glycolysis and glycogen synthesis with a concomitant reduction in fatty acid metabolism (14). These observations suggest LRH-1 agonists may resolve glucose homeostasis related-diseases. New evidence suggests that LRH-1 may also be targeted to relieve chronic ER stress. Activation of LRH-1 by synthetic DLPC or the small molecule RJW100 induces *Plk3*, which is required for the activation of ATF2 and the induction of its target genes, which play a key role in resolving ER stress (15). Given its potential therapeutic value, LRH-1 has been the subject of multiple attempts to identify small molecule modulators (16-19). These attempts have been met with mixed success due in part to our limited understanding of LRH-1's mechanism of activation.

We have shown that DLPC is able to bind directly to the LRH-1 ligand binding domain (LBD) and activate the receptor by affecting receptor dynamics in an alternate activation function (AF) region, encompassing the β-sheet–H6 region of the protein, to alter co-regulator binding preference (20). Importantly, it seems that DLPC may promote activation by relieving LRH-1 from repression by the non-canonical co-repressor NR SHP, which mimics a co-activator using the canonical Leu-X-X-Leu-Leu (where X is any amino acid) nuclear coactivator interaction motif (21; 22). In the absence of ligand, the alternate AF is highly dynamic and mutations that restrict motion in this region ablate transactivation (20). SHP is a robust corepressor of LRH-1-mediated transactivation in the liver can recognize both apo LRH-1 and LRH-1 when bound to a non-ideal ligand such as bacterial PLs *in vitro* (21; 23; 24). It is unclear how LRH-1 discriminates between SHP and coactivators such as TIF2 that bind using a similar LxxLL motif to recognize the active NR orientation. Further, how does human LRH-1 recognize coactivators in the absence of ligand? How do PLs varying only in their acyl tail composition show differing abilities to drive transactivation? Which ligand/coregulator states are appropriate for *in silico* ligand design?

This incomplete understanding of what dictates LRH-1's PL and coregulator selectivity limits our ability to guide the design of robust small molecule modulators for this intriguing pharmacological target. To address these questions, we have generated a novel crystal structure of the LRH-1–TIF2 complex in an apo state, as well as a higher resolution structure of LRH-1 bound to *E. coli* PLs. These crystal structures, in combination with novel lipid binding assays, molecular dynamics simulations and principle component analysis (PCA) have allowed us to identify an unexpected allosteric network that may contribute to PL-mediated NR signaling and coregulator selectivity.

Experimental Procedures

Reagents

Chemicals were purchased from Sigma, Fisher or Avanti PLs. pMALCH10T and the vector for His tagged TEV were a gift from John Tesmer (UT Austin). pLIC_MBP and pLIC_HIS were gifts from John Sondek (UNC, Chapel Hill). Peptides were synthesized by RS Synthesis (Louisville, KY). DNA oligonucleotide primers were synthesized by IDT (Coralville, IA USA).

Protein expression and purification

The human LRH-1 LBD (residues 291–541) was purified as described previously (25). Purified protein was dialyzed against 60 mM NaCl, 100 mM ammonium acetate (pH 7.4), 1 mM DTT, 1 mM EDTA and 2 mM CHAPS and concentrated using centrifugal filters with a 10-kDa cutoff to 5–7 mg ml^{-1}. For apo LRH-1 crystallization, purified LRH-1 LBD was incubated with 1,2- ditetracosanoyl-sn-glycero-3-phosphocholine (PC 24:0–24:0) (Avanti Polar Lipids) and GSK8470, a weak and labile agonist, at a final PC24:ligand:protein ratio of 20:3:1 (17). The receptor was purified away from unbound PC 24:0–24:0 and the weakly bound agonist by size exclusion chromatography, dialyzed against 60 mM NaCl, 100 mM ammonium acetate, pH 7.4, 1 mM DTT, 1 mM EDTA and 2 mM CHAPS and concentrated to 5–7 mg ml^{-1}.

Structure determination

Both the apo LRH-1 LBD–TIF2 complex and the LRH-1 LBD–*E. coli* PL–TIF2 complex crystals were generated by hanging-drop vapor diffusion at 20 °C from solutions containing 1 μl of protein at 6.5 mg ml^{-1} in complex with a peptide derived human TIF2 NR box 3 (+H3N-KENALLRYLLDKDD-CO2−) at a 1:4 molar ratio and 1 μl of the following crystal mixture: 0.7-1 M di-Sodium Malonate, 0.1 M HEPES pH 7.4, 0.5% Jeffamine ED-2001. Crystals were cryoprotected in crystallant containing 20% (v/v) glycerol and flash-frozen in liquid N2. Data for the apo LRH-1 LBD–TIF2 NRBox3 complex were collected to 1.75 Å resolution at 100 K using a wavelength of 0.9999 at 22-BM at the Southeast Regional Collaborative Access Team (SER-

CAT) at the Advanced Photon Source and were processed and scaled with HKL2000 (26). Data for the LRH-1 LBD–*E. coli* PL–TIF2 complex were collected to 1.75 Å resolution at 100 K using a wavelength of 0.9999 Å at 22-ID at the Southeast Regional Collaborative Access Team (SER-CAT) at the Advanced Photon Source and were processed and scaled with HKL2000 (26). Initial phases for both structures were determined using LRH-1 PDB 1YOK as a molecular replacement search model. The structures were refined using the PHENIX suite of programs, and model building was carried out in COOT (27; 28). The final model for the LRH-1–TIF2 complex contains LRH-1 residues 300–538 and TIF2 residues 742–752; it shows good geometry, with 98.4% and 1.6% of the residues in the favored and allowed regions of the Ramachandran plot, respectively. The final model for the LRH-1–*E. coli* PL–TIF2 NRbox3 complex contains LRH-1 residues 298–538 and TIF2 residues 743–750; it shows good geometry, with 98.7% and 1.3% of the residues in the favored and allowed regions of the Ramachandran plot, respectively. Data collection and refinement statistics are listed in Table 1. Coordinates and structure factors have been deposited with the Protein Data Bank under accession codes 4PLD and 4PLE.

Local Conformational Analysis

ProSMART is an alignment tool that provides a conformation-independent structural comparison of two proteins based upon the alignment of corresponding overlapping fragments of the protein chains (29). We performed ProSMART analyses among five LRH-1 structures with different bound ligands and coregulator peptides, representing different activation states: apo-SHP (fully repressed; PDB: 4DOR), apo-TIF2, *E. coli* PL-SHP (PDB: 1YUC), *E. coli* PL-TIF2, and DLPC-TIF2 (fully activated; PDB: 4DOS). This allowed for a detailed analysis of the local structural dissimilarities between two proteins independently of their global conformations. The local backbone conformation of available LRH-1 crystal structures were compared to generate the Procrustes score, which 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

Synthesis of NBD-DLPE

DLPE (dilauroylphosphatidylethanolamine; 50 mg, 90 μmol), NBD-Cl (4-chloro-7-nitrobenzofuran; 50 mg, 250 μmol), and triethylamine (17.5 uL) were dissolved in 5 mL 1:1 $CHCl_3$:MeOH and stirred for 2 h at room temperature. The reaction mixture was dried, reconstituted in a minimal volume of $CHCl_3$, and purified by TLC on silica in 9:1 $CHCl_3$:MeOH (R_f = 0.36). The product was extracted with $CHCl_3$, filtered, and evaporated to yield 37 mg (50 μmol, 56% yield) NBD-DLPE. Product identity and purity was verified by mass spectrometry, with a single peak corresponding to NBD-DLPE at m/z 741.38671.

Phospholipid binding assays

To characterize PL-binding, we developed an equilibrium based FRET assay using DCIA-labeled LRH-1 LBD as the donor and NBD-DLPE as the acceptor. Recombinant LRH-1 from *E. coli* was fluorescently labeled with DCIA (7-diethylamino-3-((4'-(iodoacetyl)amino)phenyl)-4-methylcoumarin; Molecular Probes, Inc.; Eugene, Oregon USA) according to manufacturer instructions, and further purified by gel filtration chromatography to remove excess dye. All experiments were performed in assay buffer containing 150 mM NaCl, 10 mM Tris HCl (pH 7.4), 5% glycerol, and 0.1% N-octyl-β-D-glucopyranoside. All PL stocks were prepared as small unilamellar vesicles via sonication from evaporated chloroform stocks reconstituted in assay buffer. The binding affinity of NBD-DLPE to LRH-1 was measured using a constant concentration of 150 nM unlabeled or DCIA-LRH-1, and 0 – 100 μM NBD-DLPE. Competition assays were performed with constant concentrations of 150 nM DCIA-LRH-1 and 5 μM NBD-DLPE, with 0 – 100 μM competing PL. Fluorescence intensity was measured on a Synergy 4 plate reader (Biotek; Winooski, VT USA) equipped with 380/20 nm excitation and 460/40 nm emission filters. All assays were performed in triplicate on black 384-well plates in a total volume of 50 μL. Data for unlabeled LRH-1 were subtracted from corresponding DCIA-LRH-1 data to remove background fluorescence, and all background-corrected data were

expressed as percent fluorescence intensity of fully unbound DCIA-LRH-1 (*i.e.* 0 M NBD-DLPE). Data were processed with GraphPad Prism 5 (GraphPad Software, Inc.).

Reporter gene assays

Transactivation of wild type and mutant LRH-1 was measured via luciferase-based reporter gene assay. HEK 293T cells were seeded into 24-well plates and incubated at 37°C in complete media (DMEM supplemented with 10% charcoal/dextran-stripped FBS and 1% penicillin-streptomycin) until approximately 90% confluent. Each well was then transiently transfected in OptiMem using Lipofectamine 3000 with plasmids encoding firefly luciferase under control of the *shp* promoter (SHP-luc; 500 ng/well), renilla luciferase under constitutive activation via the CMV reporter (pRLCMV; 10 ng/well), and wild-type or mutant LRH-1 in the pCI mammalian expression vector (100 ng/well). Transfection was ended after 4h incubation at 37°C via the replacement of transfection mixture with complete media, and cells were incubated overnight. Luciferase activity was measured using the Dual-Glo luciferase assay system (Promega; Madison, WI USA). Statistical analyses were performed in GraphPad Prism 5 (GraphPad Software, Inc.; La Jolla, CA USA), via one-factor ANOVA followed by Dunnett's multiple comparison test using wild-type LRH-1 as a control. Data are the results of five independent experiments. All mutations were introduced into the wildtype LRH-1/pCI construct using the QuikChange II Lightning Multi site-directed mutagenesis kit (Agilent Technologies, Inc.; Santa Clara, CA USA).

Model Construction for Molecular Dynamics

Five models were constructed to examine the structural and allosteric impacts of ligand/co-regulator agreement: 1) apoLRH-1 – TIF2 NRBox3, 2) LRH-1 – DLPC – TIF2 NRBox3, 3) LRH-1 – E. coli PLs – TIF2 NRBox3, 4) apoLRH-1 – SHP NRBox2, 5) LRH-1 – E. coli PLs – SHP NRBox2. Agreement is defined here by simultaneous binding of an activating lipid and coactivator or by the binding of a corepressor in the absence of ligand. In every case, residues 297-540 from the LRH-1 LBD form the core of the complex, with additions of 2-3 residues at

either terminus as necessary to maintain consistent sequences between models, using the program xLeap, part of the AmberTools11 suite (30). All five systems were solvated with TIP3P water in a rectangular box with equilibrated dimensions of 67 Å X 70 Å X 72 Å and neutralized with sodium and chloride ions to a salt concentration of 0.15 M.

Briefly, the first model containing the LRH-1 LBD, in the apo state, bound to a TIF2 co-activator peptide was modeled directly from the novel apoLRH-1 – TIF2 NRBox3 crystal structure. The second model, containing DLPC in the binding pocket, bound to a TIF2 co-activator peptide was modeled directly from PDB ID 4DOS (20). The third system, comprised of the LRH-1 LBD with the *E. coli* PL in the binding pocket, bound to a TIF2 peptide was modeled from the novel LRH-1 – *E. coli* PLs – TIF2 NRBox3 crystal structure. While electron density in the crystal structure is insufficient to identify the head group of the bound lipid, mass spectrometry results suggest phosphatidylglycerol and phosphatidylethanolamine to be the predominant PL isoforms (20). Thus, we modeled a bacterial phosphatidylethanolamine with 16 and 18 carbons on the sn1 and sn2 position, respectively, derived from PDB ligand EPH, which is herein referred to as *E. coli* PL. The fourth model consists of the LRH-1 LBD, in the apo state, bound to a SHP co-repressor peptide, constructed from the LRH-1 LBD (derived from the apoLRH-1 LBD – TIF2 NRBox3 structure), with the SHP peptide (PDB ID: 4DOR) (20) modeled in place of TIF2 via superposition of LRH-1 LBD residues 340-382 and 533-538. The charge clamp specific contacts between LRH-1 residues Arg361 and Glu534 and the SHP peptide were enforced with harmonic restraints during the equilibration phase of the molecular dynamics simulation and released before the production runs. The final model, LRH-1 containing DLPC in the binding pocket and bound to a SHP co-repressor peptide was constructed from PDB: 4DOS (20) with the SHP co-repressor modeled in place of TIF2 as described in the previous model.

Molecular Dynamics

The CHARMM27 (31) force field for lipids and proteins was employed for all simulations. All systems were subjected to 10,000 steps of steepest-descent minimization, heated

to 300 K under the canonical ensemble for 100 ps. Finally, positional restraints were incrementally released first on the protein sidechains, followed by the backbone, under the isobaric-isothermal ensemble. Production runs were performed under constant pressure and temperature, totaling 212 ns of unrestrained molecular dynamics for each system, with 12 ns discarded as equilibration, resulting in 200 ns of production simulation time per system. All simulations were performed with NAMD 2.9 (32), employing the r-RESPA (33) multiple time step method, with bonded and short-range interactions evaluated every 2 fs and long-range electrostatics evaluated every 4 fs with the smooth Particle Mesh Ewald method (34). The short-range non-bonded interactions were calculated used a cutoff of 10 Å with a switching function at 8.5 Å. The integration time step was 2 fs and the SHAKE algorithm was applied to fix the bonds between the hydrogens and the heavy atoms. Parameters and topology for the *E. coli* PL ligand were obtained from the general lipid parameters available in CHARMM27.

Analysis methodology

For all analyses, 10,000 evenly spaced frames were taken from the 200 ns production runs to allow for sufficient statistical sampling. Covariance matrices were constructed using the program Carma (35) over all alpha-carbons to produce per-residue statistics. The NetworkView plugin (36) in VMD (37), along with the programs subopt, included in the NetworkView package, and Carma were employed to produce dynamical networks for each system, along with suboptimal path analyses. The ptraj module of AmberTools11 was used for structural averaging as well as Cartesian principal component analysis over protein backbone atoms and, over the same 10,000-frame trajectories used for the covariance analyses. Principal components were projected onto the molecular dynamics trajectories, with snapshots binned according to their displacements along the components. Temporal correlations between modes are lost in this approach but heavily sampled regions of the conformational subspace are more easily identified.

Results

Structure of the apo LRH-1 LBD – TIF complex:

LRH-1 is able to bind to both coactivators and corepressor proteins in the absence of a ligand. To visualize the structural perturbations necessary to bind to coactivators in its apo form, we crystallized apo LRH-1 LBD bound to a fragment of the coactivator TIF2 and determined its structure to 1.75 Å (Figure 3.1 A). There is no visible electron density to support modeling a bound ligand. The opening to the ligand binding pocket is constricted by 2 Å, which reduces the volume of the ligand binding pocket from 1554 $Å^3$ in the LRH-1 – TIF2 – DLPC complex to roughly 940 $Å^3$. This is in stark contrast to the apo LRH-1 LBD – SHP NRBox2 structure reported previously, which lacks electron density for the entirety of the alternate AF (Figure 3.1 A vs C). Unlike the ligand binding pocket of rodent LRH-1, the ligand binding pocket constriction is not stabilized by any intramolecular interactions (38). However, it is possible that the alternate AF, which comprises nearly one third of the binding pocket, may be visible due to fortuitous interactions with a crystallographic symmetry mate. Regardless, this shows remarkable flexibility of the ligand binding pocket.

The structure contains a single CHAPS detergent molecule that docks on H10 and H12 via hydrophobic interactions and two hydrogen bonds between the CHAPS 7-OH and Glu-514, and the CHAPS phosphate and Tyr-518. CHAPS also makes extensive contact with a crystallographic symmetry mate (Figure 3.1 C). Thus, two molecules within the crystal create a cleft for CHAPS binding.

Improved structure of the LRH-1 LBD – E. coli PL – TIF2 complex

To generate a more accurate model for molecular dynamics studies, and as a control in our crystallization experiments, we crystallized the LRH-1 LBD – *E. coli* PL – TIF2 complex and determined its structure to 1.75 Å (Figure 3.2 A). This represents an improved resolution over the existing LRH-1 LBD – *E. coli* PL – TIF2 structures, which were both solved to 2.5 Å (39; 40). The structure is highly similar to the previous structures with an r.m.s.d. of 0.6 Å over main chain

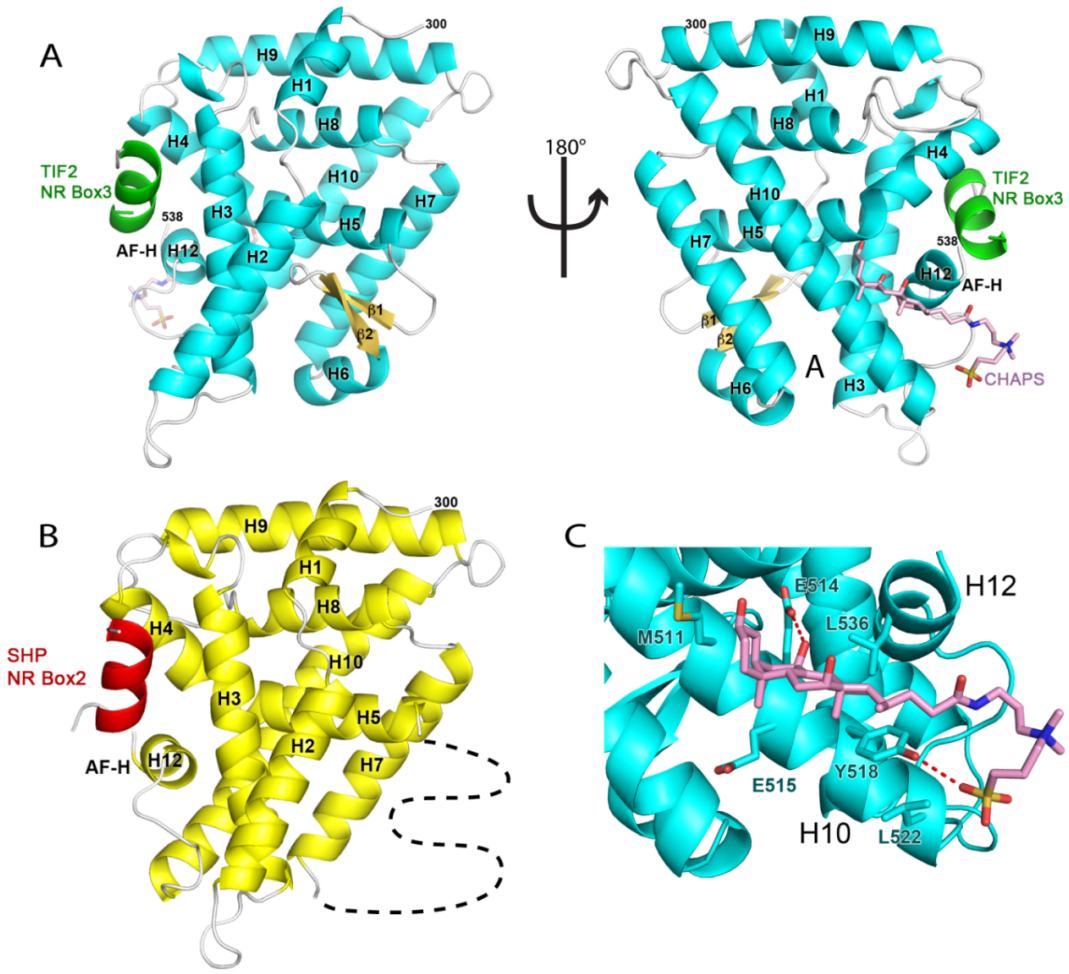

Figure 3.1. Structure of the apo LRH-1 LBD—TIF complex.

(A) Ribbon diagram of apo LRH-LBD (α-helices, *teal*; β-strands, *yellow*) with the TIF2 NR box 3 peptide (*orange*). The surface bound CHAPS is depicted as sticks (C, *pink*; O, *red*; S, *yellow*; N, *blue*). The AF-2 surface is defined by H3, H4 and H12. (B) Ribbon diagram of apo LRH-SHP NRBox2 complex (PDB ID 4DOR) with the unobserved alternate AF region (defined by β1-2 and H6) represented by a dashed line. (C) Close up view of the bound CHAPS molecules included in the crystallization buffer.

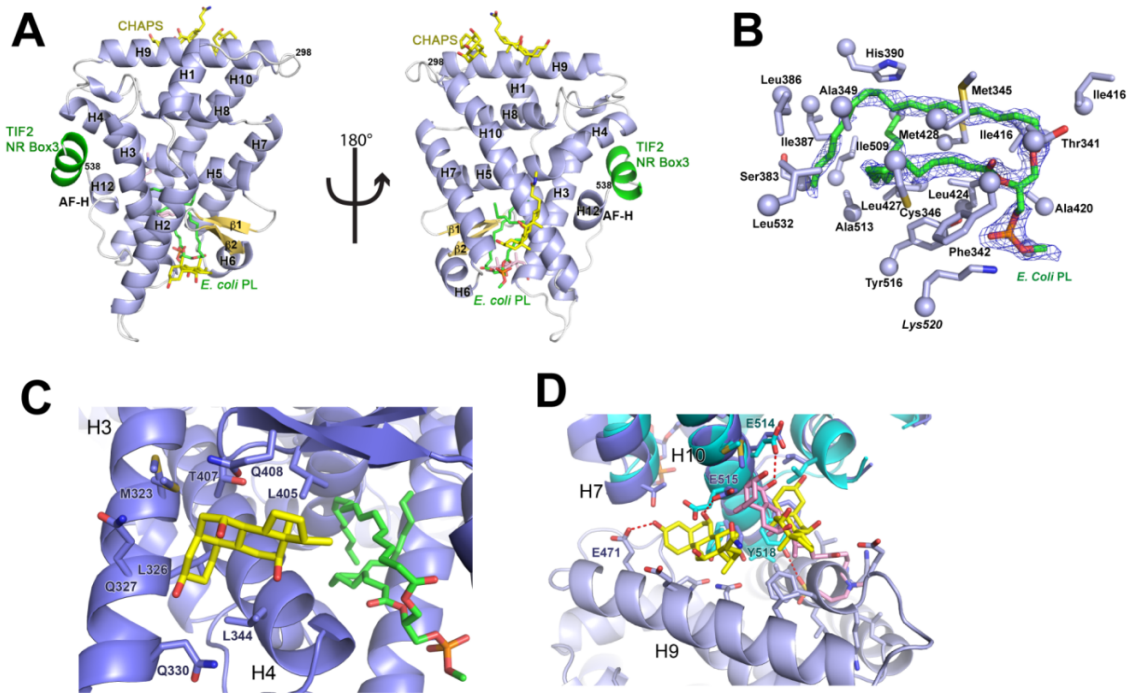

Figure 3.2. Structure of the LRH-1 LBD—*E. coli* PL—TIF2 complex.

(A) Ribbon diagram of *E.* coli PL bound LRH-LBD (α-helices, *teal*; β-strands, *yellow*) with the TIF2 NR box 3 peptide (*green*). The bound *E.* coli PL is depicted as sticks (C, green; O, red; P, magenta) The surface bound CHAPS is depicted as sticks (C, *yellow*; O, *red*; S, *yellow*; N, *blue*). (B) 2Fo – Fc electron density (contoured at 1 σ) for the bound *E. coli* PL observed in this structure, along with side chains lining the ligand-binding pocket of hLRH-1 that contact this ligand. (C), Close up view of the bound CHAPS molecules included in the crystallization buffer along H3 and H4 in close proximity to the bound PL. Residues within 4.2 Å are depicted as sticks. (D) Close up view of the bound CHAPS along H9 which interact with a crystallographic symmetry mate and in a postion overalapping the CHAPS site in the apo LRH-1 – TIF2 complex. Residues within 4.2 Å of CHAPS are depicted as sticks.

Table 3.1. Data collection and refinement statistics for novel LRH-1 complexes

	LRH-1 – TIF2 NRbox3	LRH-1 – *E. coli* PL – TIF2 NRbox3
Data collection		
Space group	$P2_12_12_1$	$P2_1$
Cell dimensions		
a, *b*, *c* (Å)	45.8, 65.7, 83.5	65.9, 76.9, 100.8
α, β, γ (°)	90.0, 90.0, 90.0	90.0, 95.5, 90.0
Resolution (Å)	1.75 (1.81 – 1.75)*	1.75 (1.81 – 1.75)*
R_{merge}	6.6 (30.6)	6.6 (30.9)
$I / \sigma I$	18.99 (2.8)	12.8 (3.2)
Completeness (%)	99.4 (96.22)	92.6 (63.8)
Redundancy	3.9 (3.3)	3.6 (3.2)
Refinement		
Resolution (Å)	1.75	1.75
No. reflections	25933	6751
R_{work} / R_{free}	18.7 / 22.4	20.67 / 23.4
No. atoms		
Protein	2026	8117
Ligand/ion	42	493
Water	137	378
B-factors		
Protein	23.9	27.0
Ligand/ion	29.2	37.5
Water	29.4	32.7
R.m.s. deviations		
Bond lengths (Å)	0.008	0.006
Bond angles (°)	1.41	1.03
PDB	4PLD	4PLE

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

atoms and maintains *E. coli* PLs in the binding pocket (Figure 3.2 B). The lipid acyl tails show a decrease in electron density near their termini, which is similar to previous observations for the bound *E. coli* PLs and the LRH-1 – DLPC complex (20) (Figure 3.2 B). This observation further supports the hypothesis that LRH-1 specifically recognizes its PL ligands near the glycerolphosphate backbone and the exact position of the acyl tails is less important than the amount of space they occupy in the deeper portion of the ligand binding pocket.

The structure contains three CHAPS detergent molecules that dock onto the surface of the protein and make interactions with crystallographic symmetry mates. One CHAPS molecule is secured in the cleft between H3 and the β-sheet via hydrophobic interactions. A second CHAPS molecule mediates contact between two copies of the LRH-1 monomer and is secured by hydrophobic interactions along H10 and a hydrogen bond with Glu-515 of one monomer, and hydrophobic interactions along H9 and a hydrogen bond with Glu-471 of the second monomer. The third CHAPS is adjacent to the second, and also mediates contact between two LRH-1 monomers via hydrophobic interactions with H10 of the first monomer and H9 of the second, but does not make any hydrogen bonds with either monomer. The CHAPS molecules contacting two LRH-1 monomers are unique relative to the apo LRH-1 LBD – TIF complex, while the CHAPS occupying the site near H10 shows a partial overlap with the well-ordered CHAPS in the apo structure (Figure 3.2 D). In contrast to the excellent electron density for the CHAPS bound in the apo LRH-1 LBD – TIF complex, the CHAPS bound at this site in the E. coli lipid bound complex shows electron density for only the sterane ring. This is likely due to greater thermal motion or reduced CHAPS occupancy at this site in the crystal. Interestingly, CHAPS is docked at regions within LRH-1 that show most exchange in HDX studies and the most conformational fluctuations in crystal structures. It is possible that these are sites for protein-protein or protein-lipid interaction in the cell.

LRH-1 can bind to several PLs (2; 25; 40; 41), yet only PCs have been shown to drive transactivation (13; 20; 40; 41). It is unclear why LRH-1 responds only to PCs in cells; this may

be intrinsic to the receptor or due to uncharacterized PL transporters capable of delivering PC ligands. In order to elucidate the mechanisms by which PLs differentially activate LRH-1, it is critical to determine the effects of head group and tail variation on binding. To characterize differential PL binding, we developed a FRET-based PL-binding assay monitoring the ability of NBD-labeled DLPE to bind to DCIA-labeled LRH-1 (Figure 3.3 A). This binding event quenches the DCIA fluorescence, which can be recovered upon the competitive binding of unlabeled lipids (Figure 3.3 B, E).

Prior to engulfing PLs, LRH-1 must extract PL from the lipid membrane – a step typically conducted by PL transport proteins that contain amphipathic structural elements to facilitate partitioning in membranes (42-44). In the absence of lipid chaperones, we find that LRH-1 extracts and binds PC, PG, and PI with micromolar affinity, but cannot extract PE, PS, PA, SM, ceramide, or sphingosine (Figure 3.3 C). Thus, LRH-1's ability to bind PLs from vesicles is sensitive to the nature of the head group. However, addition of 5 mM β-cyclodextrin, a small molecule chaperone widely used for the delivery of hydrophobic small molecules, enables the binding of PC, PE, PS, PG, PA, with low micromolar affinity (Figure 3.3 C). LRH-1 is unable to bind sphingosine and ceramide despite the presence of β-cyclodextrin suggesting that extraction from vesicles is not a limiting factor; rather, these lipids do not fit well within the ligand binding pocket. These extracts contain a range of PL isoforms and the PC mixture showed the highest maximum displacement of bound NBD-DLPE (Figure 3.3 E and data not shown).

We then investigated LRH-1's intrinsic selectivity for PL tail composition by testing a range of saturated PCs. Surprisingly, only PCs with mid-length chains of 8-16 carbons bind to LRH-1, with DLPC showing the strongest affinity. We observed no change in binding with the inclusion of 5 mM β-cyclodextrin (data not shown). These findings mirror previously published activation data, which demonstrate that LRH-1 is most strongly activated by the 11- and 12-carbon saturated PCs, DUPC and DLPC (13). Thus, PL selectivity is driven by the length of the fatty acid tails *in vitro* suggesting that the amount of space filled by the acyl tails is a critical

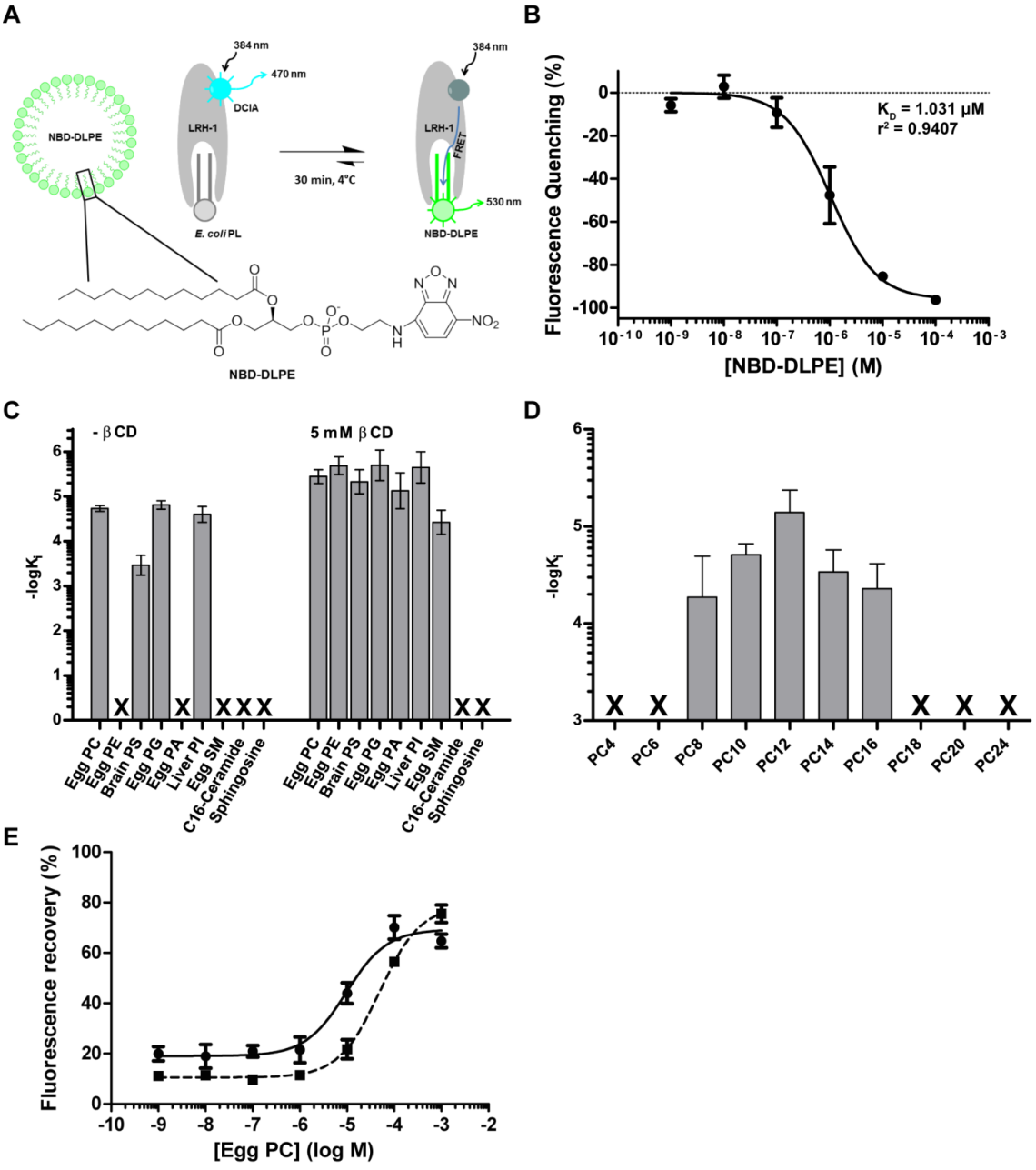
A
384 nm
470 nm
DCIA
LRH-1
NBD-DLPE
E. coli PL
30 min, 4°C
384 nm
LRH-1
FRET
530 nm
NBD-DLPE
NBD-DLPE
B
Fluorescence Quenching (%)
K_D = 1.031 µM
r² = 0.9407
[NBD-DLPE] (M)
C
- β CD
5 mM β CD
-logK_i
Egg PC
Egg PE
Brain PS
Egg PG
Egg PA
Liver PI
Egg SM
C16-Ceramide
Sphingosine
D
-logK_i
PC4
PC6
PC8
PC10
PC12
PC14
PC16
PC18
PC20
PC24
E
Fluorescence recovery (%)
[Egg PC] (log M)

Figure 3.3. LRH-1 *in vitro* lipid binding profile.

Binding affinities of LRH-1 to PLs of differing head group and tail compositions. (A) PL binding was measured relative to probe ligand NBD-DLPE via FRET quenching of DCIA-labeled LRH-1. (B) Binding affinity of LRH-1 to NBD-DLPE probe. (C) Relative binding affinities of competing PLs of differing head groups; 5 mM β-cyclodextrin added as indicated. (D) Relative binding affinities of competing saturated PCs of differing tail lengths. Data are reported are the means + S.E.M. of three independent experiments. The presence of an X instead of a bar indicates that no binding was observed. (E) Example of an individual competitive binding curves for NBD-DLPE displacement. Solid line represents the inclusion of 5 mM β-cyclodextrin while the dashed line is without 5 mM β-cyclodextrin as described in the methods.

determinant of binding.

Co-regulator binding interactions are altered by ligand status

The canonical model of NR activation revolves primarily around a mobile ligand-sensing helix (H12), termed the AF-H. When a receptor is bound to an agonist the AF-H packs against helices 3 and 4 of the LBD forming a surface, termed activation function 2 (AF-2), which enables interaction with coactivator proteins containing a LxxLL motif (45). This helical peptide inserts its leucines into a groove on the AF-2 surface and is further stabilized by a charge clamp interactions with Arg 361 on H3 and Glu 534 on the AF-H. An equivalent charge clamp is conserved across NRs and represents a general mechanism for activation (46). LRH-1, like some other orphan NRs, is able to form a productive AF-2 in the absence and presence of ligands in available crystal structures. This makes inferences regarding ligand potency based on backbone positioning within the AF-2 alone challenging. Nevertheless, we compared coregulator binding at the AF-2 across all available crystals structures and observed that regardless of the ligand state, Arg 361 on H3 forms the expected charge clamp interaction. In contrast, we were surprised to find that Glu 534, on the AF-H, does not make the expected charge clamp interaction with coregulator peptide under all circumstances (Figure 3.4). This does not appear to be an artifact of crystal packing. Instead, the conformation of Glu 534 correlates with an agreement between the ligand and the bound coregulator peptide. Agreement is defined here by simultaneous binding of an activating lipid and coactivator or by the binding of a corepressor in the absence of ligand. When apo, or bound to a poorly activating ligand, Glu 534 is rotated out of hydrogen bond distance with the coactivator TIF2 peptide (Figure 3.4 A-C). In contrast, when LRH-1 is bound to a strong agonist such as DLPC, Glu 534 makes the expected hydrogen bond with a backbone amide of the TIF2 peptide (Figure 3.4 D). This charge clamp interaction is also observed in apo or *E. coli* PL bound LRH-1 when complexed to a peptide derived from the corepressor SHP (Figure 3.4 E-F). These observations suggest that LRH-1 has an extensive allosteric network that

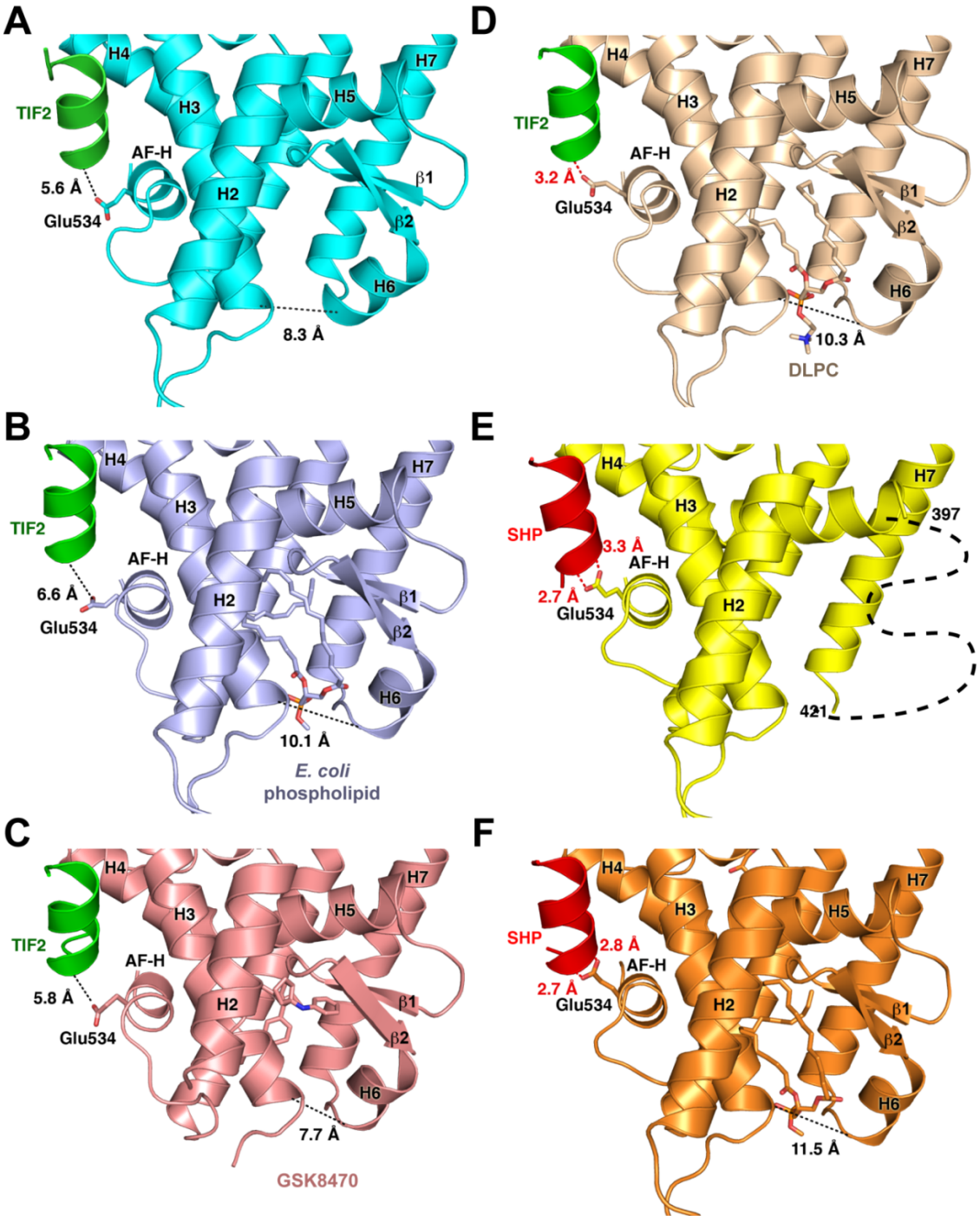
A
H4
H7
TIF2
H3
H5
AF-H
5.6 Å
H2
β1
Glu534
β2
H6
8.3 Å
D
H4
H7
H3
H5
TIF2
AF-H
3.2 Å
H2
β1
Glu534
β2
H6
10.3 Å
DLPC
B
H4
H7
H3
H5
TIF2
AF-H
6.6 Å
H2
β1
Glu534
β2
H6
10.1 Å
E. coli
phospholipid
E
H4
H7
H3
397
SHP
3.3 Å
AF-H
2.7 Å
H2
Glu534
421
C
H4
H7
H3
H5
TIF2
AF-H
5.8 Å
H2
β1
Glu534
β2
H6
7.7 Å
GSK8470
F
H4
H7
H3
H5
SHP
2.8 Å
AF-H
2.7 Å
H2
β1
Glu534
β2
H6
11.5 Å

Figure 3.4. AF-2 charge clamp engagement is dictated by ligand-coregulator combination. Ligand binding pocket entrance measurements and analysis of Glu 534 – peptide charge clamp engagement for the (A) apo LRH-1–TIF2 complex, (B) LRH-1–*E. coli* PL–TIF2 complex, (C) LRH-1–GSK8470–TIF2 complex (PDB ID: 3PLZ), (D) LRH-1–DLPC–TIF2 complex (PDB ID: 4DOS), (E) apo LRH-1–SHP complex (PDB ID: 4DOR), (F) LRH-1–*E. coli* PL–SHP complex (PDB ID: 1YUC).

appropriately tunes the receptors ability to stabilize very similar LxxLL motifs present in coactivators and corepressors.

Ligand and coregulator drive differential effects on local residue environment

Supposition of multiple LRH-1 – ligand structures revealed only subtle differences in the coregulator binding surface. We therefore used ProSMART to compare the local residue environment to identify how differential ligand and coregulator peptide binding affects local structure (29). Caution of course must be taken with the interpretation of these results since the crystal structures included in this analysis are derived from multiple crystal forms. LRH-1 shows the greatest conformational similarity between structures where both ligand and coregulator status are in agreement within the structural complex (Figure 3.5). Greater conformational dissimilarity is seen when one or both complexes are not in ligand-coregulator agreement, indicating that such agreement is crucial in maintaining a stable complex, regardless of whether that complex is activated or repressed. In all coregulator states, the addition of a ligand stabilizes the alternate AF region compared to apo, as demonstrated by the high structural dissimilarity seen in this region compared to the apo-TIF2 structure (Figure 3.5 E, F, H), and the fact that this area could not be modeled in the apo-SHP complex. As expected, the highest structural dissimilarity is seen in the AF-2 and alternate AF (β-sheet/H6), the respective interaction sites for the coregulator peptide and PL head group. SHP poorly discriminates between apo and bacterial PL – bound receptor, and shows high structural similarity throughout the ligand binding domain (Figure 3.5 A). In contrast, the LRH-1 TIF2 complexes show strong differences with LRH-1 SHP complexes regardless of ligand status, even in cases where the ligand is the same or nonexistent (Figure 3.5 B-E, G, I). Thus, unlike SHP, TIF2-bound conformations are sensitive to the nature of the bound ligand. All LRH-1 – TIF2 complexes exhibit moderate or high structural dissimilarity in both the AF-H and the preceding loop, and the alternate AF region (Figure 3.5 F, H, J). The greatest agreement among the LRH-1 – TIF2 complexes is seen between the *E. coli* PL and DLPC bound

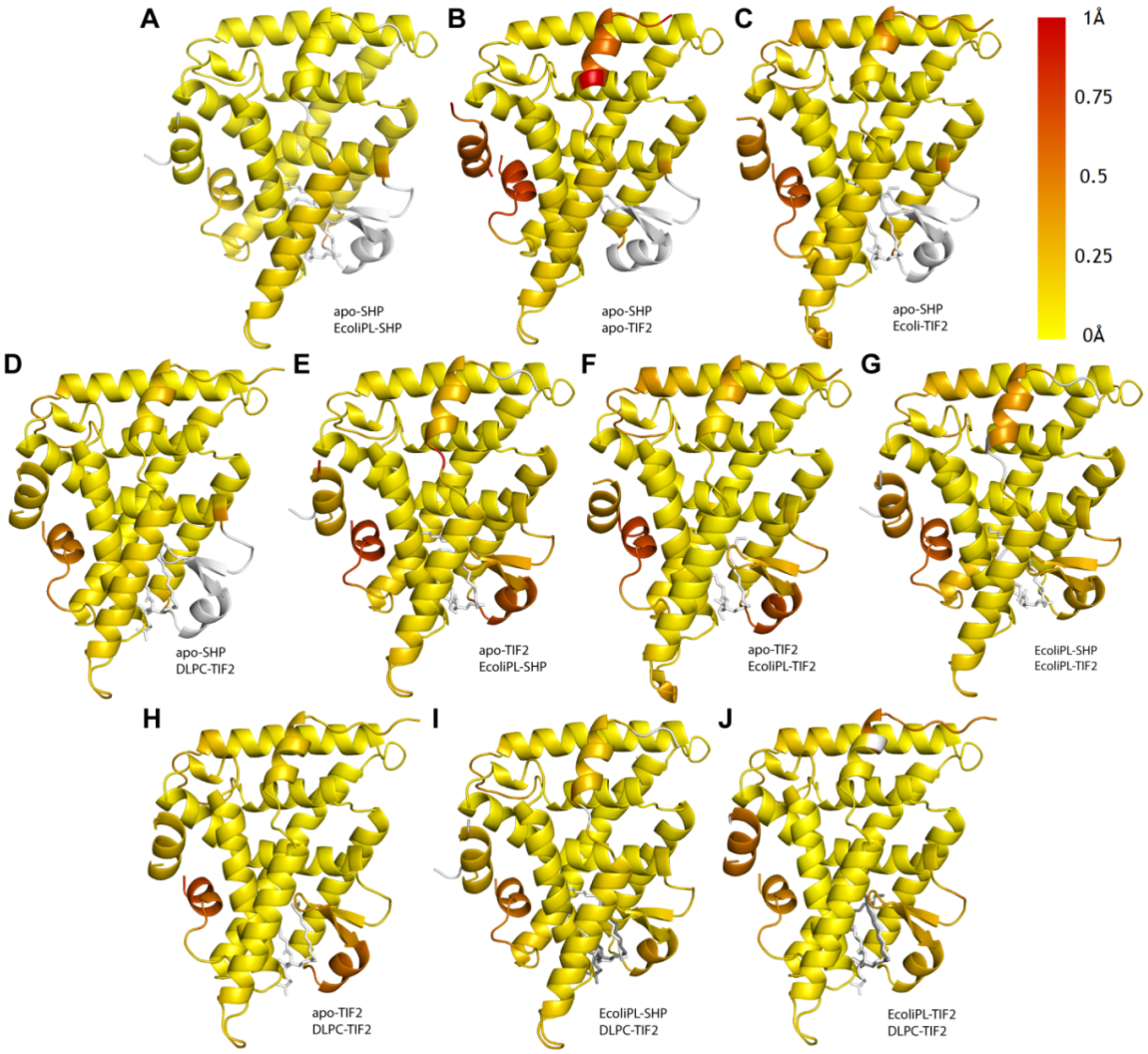

Figure 3.5. ProSMART central residue analysis of LRH-1 complexes.

ProSMART analysis of LRH-1 with differentially bound ligands and coregulator peptides. Models were colored by the Procrustes score of the central residue of an aligned fragment pair according to the legend at top right. Areas colored white were omitted from the analysis. The following pairwise comparisons were made: (A) apo-SHP (PDB ID: 4DOR) vs *E. coli* PL-SHP (PDB ID: 1YUC); (B) apo-SHP vs apo-TIF2; (*C*) apo-SHP vs *E. coli* PL-TIF2; (*D*) apo-SHP vs DLPC-TIF2 (PDB ID: 4DOS); (E) apo-TIF2 vs *E. coli* PL-SHP; (F) apo-TIF2 vs *E. coli* PL-TIF2; (*G*) *E. coli* PL-SHP vs *E. coli* PL-TIF2; (H) apo-TIF2 vs DLPC-TIF2; (I) *E. coli* PL-SHP vs DLPC-TIF2; (J) *E. coli* PL-TIF2 vs DLPC-TIF2.

structures (Figure 3.5 J), indicating that while TIF2 is sensitive to the presence or absence of a ligand, it does not strongly discriminate between ligands so long as one is present. This is consistent with previous coregulator recruitment studies, which show only a 3-fold difference in binding affinities between TIF2 and *E. coli* PL or DLPC bound LRH-1 (20.1 μM and 6.5 μM, respectively) (20). Taken together, the ProSMART analyses suggest that ligand-coregulator agreement promotes the stabilization of LRH-1 into either an active or repressed conformation, with detectable but subtle structural differences between these conformations. These conformational variations are also in line with the prior HDX data showing conformational variation between the same structural elements (*i.e.* the alternate AF, the AF2 and in H8 and 9) (1).

The Activated LRH-1 Complex Exhibits Coordinated Motions

To analyze the dynamic coupling of structural elements in LRH-1, we computed cross-correlation (normalized covariance) matrices for the C-α atoms in each of the five systems with the program Carma (35). A covariance matrix contains a great deal of information regarding the dynamics within a system, in this case describing the correlation of motions $\mathbf{r}_i$ and $\mathbf{r}_j$ for residues i and j, taken from their respective α Carbons. Element (i, j) of the covariance matrix is calculated as $\left\langle \left(r_i - \langle r_j \rangle\right)\left(r_j - \langle r_i \rangle\right) \right\rangle$. In essence, this type of covariance matrix provides a way of visualizing whether the motions of two residues within a complex are correlated, anti-correlated or non-correlated. A cross-correlation matrix is simply a covariance matrix that is normalized to vary between -1 (perfectly anti-correlated) and 1 (perfectly correlated) (Figure 3.6).

The motions in residues within helices 4 through 9 of the LRH-1 LBD become correlated upon lipid binding in the presence of a co-activator (Figure 3.6 C). The correlation of these motions in the lipid and co-activator bound systems is muted in the apo states as well as the DLPC-bound LBD in complex with the SHP peptide (Figure 3.6 A, D). This suggests that both lipid and co-regulator binding impact an allosteric network through the LRH-1 core, requiring the

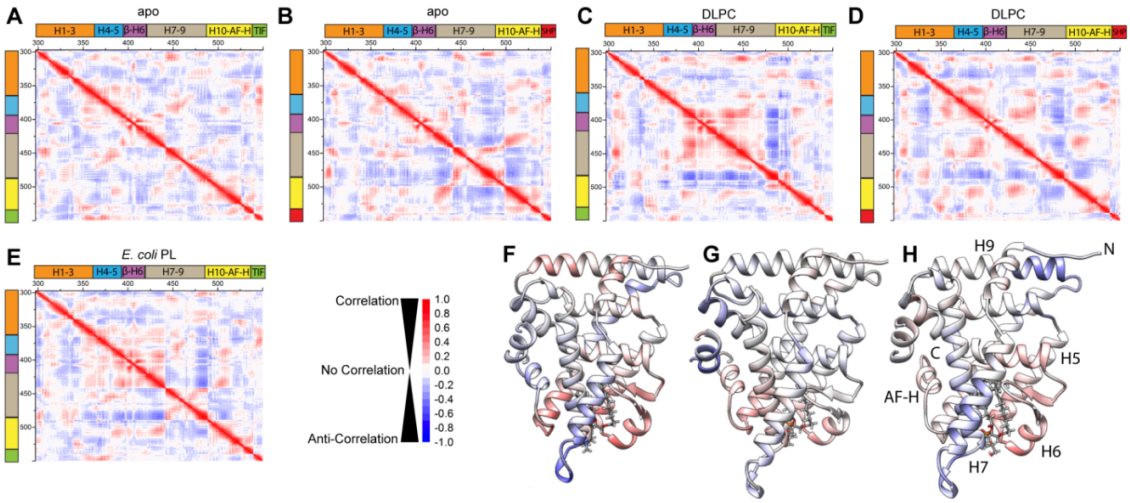

Figure 3.6. Correlated motion in LRH-1—PL—coregulator systems.

Cross-correlation matrices showing correlated and anti-correlated motion over the 200 ns MD simulation for (A) apo LRH-1–TIF2 complex, (B) apo LRH-1–SHP complex, (C) LRH-1–DLPC–TIF2 complex, (D) LRH-1–DLPC–SHP complex, (E) LRH-1–*E. coli* PL–TIF2 complex. Cross-correlation between protein residues and the lipid head group phosphorus atom are mapped to the protein structure in (F) LRH-1–DLPC–TIF2 complex, (G) LRH-1–DLPC–SHP complex, (H) LRH-1–E. coli PL–TIF2 complex.

lipid pocket and AF-H elements to be in agreement to yield an active complex. Lipid may therefore allow correlated motions in LRH-1 to favor TIF2 binding while in the apo state these motions are eliminated, thereby favoring SHP binding. We have mapped cross-correlation between the lipid head group phosphorus atom and all protein residues, for each lipid-bound system studied, onto the LRH-1 structure (Figure 3.6 F-H). We find that the lipid displays some positive covariance with the β-H6 region of the complex, and some negative covariance with H9 and H2. The DLPC – LRH-1 – SHP system shows similar behavior, but with smaller magnitudes, likely due to its disagreement status.

MD Simulations Demonstrate Communication between β-sheet–H6 and the AF-H through Helices 3, 4, and 5

We have previously discovered that LRH-1 contains an alternate activation function region that encompasses the β-sheet–H6 portion the ligand-binding pocket. Our data suggested that the dynamics of this region are coupled to ligand binding and receptor activation (20). To identify the relevant communication pathways contributing to these observations, we constructed dynamical networks to identify the most prevalent communication pathways between the β-sheet–H6 region and the bound co-regulator (Figure 3.6). Dynamical networks, as defined in the field of network theory, describe the communication pathways between components of a system. In a dynamical network, every component is taken to be a "node" and a communication between two nodes defines an "edge." In the methodology employed here, each protein residue's α carbon is a node and any two nodes must be within a distance cutoff of 4.5 Å for 75% of the MD trajectory and the strength of communication between two nodes, or the "edge weight," is determined from the covariance between the two nodes. A communication path between two distant nodes is then a chain of edges that connect them and the optimal path transmits communication between two nodes through the fewest number of edges possible and is likely to carry more communication than any other single path. The optimal path and a relatively small set of slightly longer suboptimal paths are expected to carry the majority of communications between

two edges. Monitoring the strength and number of suboptimal paths between two distant nodes can yield detailed insight into the strength of communication, or in macromolecular systems, allostery.

These pathways show much stronger communication when the lipid pocket and AF-H domain states are in agreement than otherwise (Figure 3.7 A-E). The number of communication pathways increases greatly upon lipid pocket – AF-H state agreement, especially expanding outward from the β-sheet–H6 region and into the co-regulator itself. This strongly supports our previous hypothesis that the β-sheet–H6 and AF-H regions communicate to control LRH-1 activation. Furthermore, the vast majority of communication paths proceed through helices 3, 4, and 5. These same helices showed the most protection from deuterium exchange in prior HDX studies suggesting that their rigidity may facilitate the flow of information through the receptor (20). Therefore, the allosteric pathway between the β-sheet–H6 region and AF-H like traverse through helices 3, 4, and 5 (Figure 3.6 B, D). These helices present an optimal tether between the allosteric switches. It is interesting to note that many of the mutations that affect LRH-1 PL binding, coregulator sensitivity, and overall activation lie directly on or immediately adjacent to this pathway (Figure 3.7 F) (20; 25; 47).

Structural and Dynamical Rationale for Lipid and Co-regulator Agreement

To identify and functionally significant collective motions of the residues forming the allosteric network within LRH-1, we employed principal component analysis (PCA) (48). In PCA, the C-α covariance matrix is diagonalized to yield eigenvectors, denoted as principal modes, and eigenvalues, representing the mean square fluctuation along each principal mode. Projections of the MD trajectory onto the principal modes are called the principal components. By reducing the dimensionality of the data, PCA recapitulates the most important dynamical features from the MD trajectories. Thus, the first few principal modes, known as the essential dynamics, are likely to describe the collective, global motions of LRH-1 involved in the allosteric response to ligand and coregulator binding.

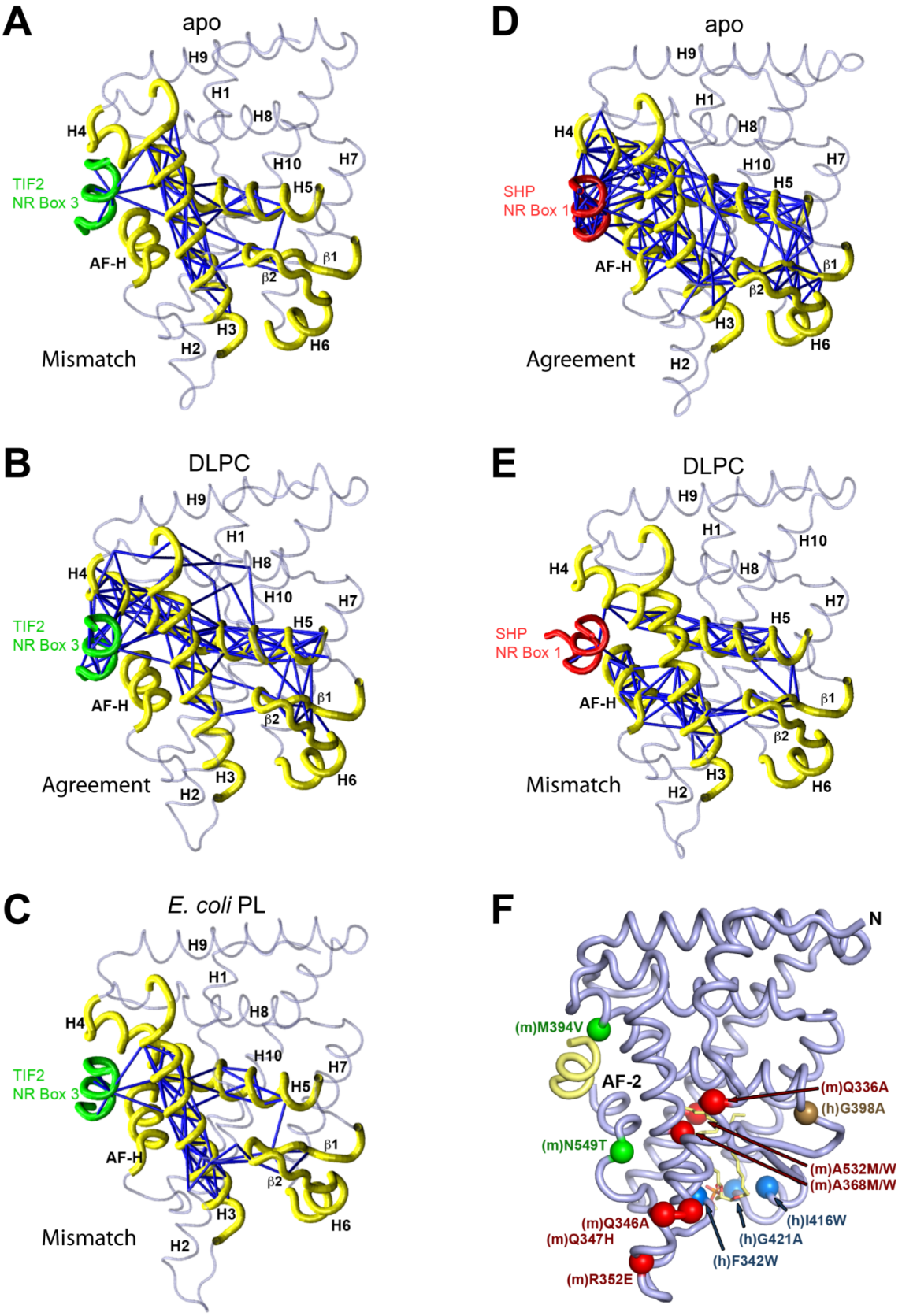
A
apo
H9
H1
H8
H4
H10
H7
TIF2
NR Box 3
H5
AF-H
β1
β2
H3
H6
H2
Mismatch
D
apo
H9
H1
H8
H4
H10
H7
SHP
NR Box 1
H5
AF-H
β1
β2
H3
H6
H2
Agreement
B
DLPC
H9
H1
H8
H4
H10
H7
TIF2
NR Box 3
H5
AF-H
β1
β2
H3
H6
H2
Agreement
E
DLPC
H9
H1
H10
H4
H8
H7
SHP
NR Box 1
H5
AF-H
β1
β2
H3
H6
H2
Mismatch
C
E. coli PL
H9
H1
H8
H4
H10
H7
TIF2
NR Box 3
H5
AF-H
β1
β2
H3
H6
H2
Mismatch
F
N
(m)M394V
AF-2
(m)Q336A
(h)G398A
(m)N549T
(m)A532M/W
(m)A368M/W
(h)I416W
(m)Q346A
(m)Q347H
(h)G421A
(h)F342W
(m)R352E

Figure 3.7. Allosteric paths from binding pocket to coregulator.

Allosteric communication pathways between the β-sheet–H6 and co-regulator binding regions of the LRH-1 LBD in the (A) apo LRH-1–TIF2, (B) LRH-1–DLPC–TIF2, (C) LRH-1–*E. coli* PL–TIF2, (D) apo LRH-1–SHP and (E) LRH-1–DLPC–SHP complexes. Cartoon loop view of LRH-1 showing thick loops (*yellow*, LRH-1; *green*, TIF2; *red*, SHP) for regions of the protein identified along the allosteric path. (F) LRH-1 mutations that alter PL binding or coregulator recruitment lie on or adjacent to the allosteric pathway. Known mutations of mouse (m) or human (h) LRH-1 LBD are shown as C-α spheres on the LRH-1 protein backbone. Mutations shown in green enhance the degree of LRH-1 activation in response to coactivator binding; mutations shown in red selectively decrease LRH-1 sensitivity to SHP without affecting overall activation; mutations shown in brown decrease overall LRH-1 activity without affecting PL binding; mutations shown in blue decrease PL binding and overall activity.

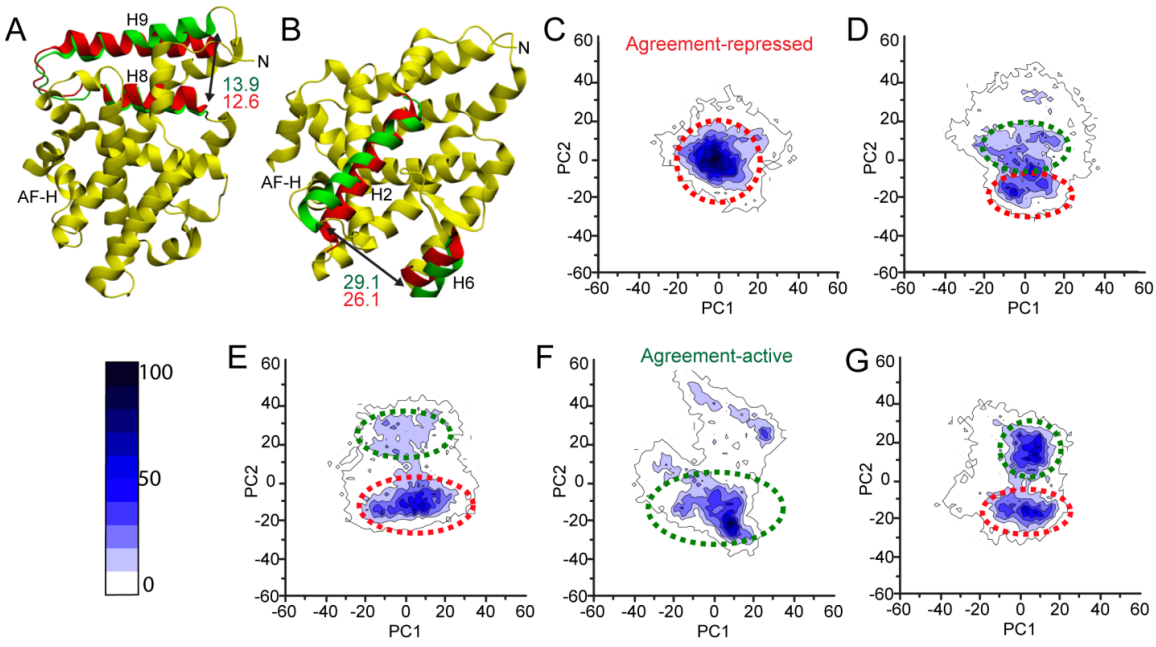

Figure 3.8. Biologically relevant principal modes identified from the projections of the MD trajectories on PC1 vs. PC2.

(A) An outward swing of helix 9 contributes to PC1, while (B) opening motions at the mouth of the lipid binding pocket result in translation along PC2. (C-G) Projections of snapshots taken from MD onto PC1 and PC2 in (C) apo LRH-1–SHP, (D) LRH-1–DLPC–SHP, (E) apo LRH-1–TIF2, (F) LRH-1–DLPC–TIF2, and (G) LRH-1–*E. coli* PL–TIF2 complexes. Higher densities indicate more populated regions of the conformational subspace. Scale bar indicates how many snapshots (out of 10,000) were collected within a contour. Green and red rings indicate activating and repressing regions of the subspace, respectively.

We have identified two modes that are indicative of the lipid-binding pocket's state and the bound coregulator, named PC1 and PC2 and have projected snapshots from the molecular dynamics trajectories onto these modes in Figure 3.8. To ensure comparability and uniformity of the modes studied, we optimized the total root mean square inner product (r.m.s.i.p.) across all systems' essential dynamics. The r.m.s.i.p. method for optimizing subspace overlap does not guarantee that the same mode number will be selected from each system, as some variation in the ordering of principal modes is expected, even for highly similar systems (49). A table of the modes chosen for PC1 and PC2 and the dot products between these modes are included in Table 3.2.

In the projections (Figure 3.8 C-G), areas of high density indicate regions of high conformational probability. Snapshots from the most densely populated regions of each system's conformational subspace were collected and averaged to obtain representative structures for comparison (Figure 3.8 A,B). PC1 is characterized by an outward motion of helix 9 relative to helix 8 and the core of the LBD, with the distance from N332 to T422 measuring 29.1 Å in the DLPC – LRH-1 –TIF2 model and 26.1 Å in the apo-LRH-1 – SHP model (Figure 3.8 A). PC2 consists primarily of an opening motion near the mouth of the lipid-binding pocket, with the distance from Q444 to N487 measuring 13.9 Å in the most prevalent conformation in DLPC – LRH-1 – TIF2 and just 12.6Å for the dominant apo-LRH-1-SHP conformation (Figure 3.8 B).

Projections of the MD trajectories onto these principal modes (Figure 3.8 C-G) illustrates that DLPC binding promotes conformations with high values of PC2, while apo- and bacterial long-tail lipid bound states tend to exhibit conformations of lower PC2 magnitude. Coregulator binding influences the dominant conformation's placement along PC1, with all TIF2-bound complexes exhibiting primary centroids near +10 and SHP-bound complexes exhibiting centroids near -10. It is worth noting that the long-tail E. coli lipid and apo-TIF2 complexes both share two common clusterings, with the former maintaining nearly equal populations near each center and the latter undergoing a population shift toward the repression-promoting region of the subspace.

Table 3.2. Modes chosen for PC1 and PC2 and dot products between modes.

PC1	**apo SHP : 2**	**dlpc SHP : 3**	**apo TIF : 3**	**dlpc TIF : 3**	**pl TIF : 3**	**Average**	**+/-**
apo SHP : 2	1.0000	0.0756	0.0626	0.4221	0.4954	0.4111	0.3834
dlpc SHP : 3	0.0756	1.0000	0.1587	0.4538	0.2609	0.3898	0.3692
apo TIF : 3	0.0626	0.1587	1.0000	0.2138	0.2337	0.3338	0.3783
dlpc TIF : 3	0.4221	0.4538	0.2138	1.0000	0.6050	0.5389	0.2931
pl TIF : 3	0.4954	0.2609	0.2337	0.6050	1.0000	0.5190	0.3112

PC2	**apo SHP : 4**	**dlpc SHP : 1**	**apo TIF : 1**	**dlpc TIF : 1**	**pl TIF : 1**	**Average**	**+/-**
apo SHP : 4	1.0000	0.3522	0.0943	0.5088	0.0987	0.4108	0.3734
dlpc SHP : 1	0.3522	1.0000	0.1388	0.5782	0.2195	0.4578	0.3457
apo TIF : 1	0.0943	0.1388	1.0000	0.1446	0.0122	0.2780	0.4071
dlpc TIF : 1	0.5088	0.5782	0.1446	1.0000	0.3446	0.5153	0.3184
pl TIF : 1	0.0987	0.2195	0.0122	0.3446	1.0000	0.3350	0.3923

*Modes and dot products selected for PC1 and PC2 via RMSIP. Each system label is formatted as "binding-pocket-state coregulator : mode". Averages of dot products and standard deviations for each system with respect to each other system are included. In each PC, the LRH-1—DLPC—TIF2 system (highlighted in green) was the most central eigenvector, having the highest average inner product with the other systems.

These results show that lipid binding and coregulator binding both impact motions in the LRH-1 LBD and that combinations of those motions result in only two distinct and stable conformations: repressing and activating. We also observed that "disagreement" complexes exist in mixed populations between the two states.

R.m.s.d. alignment of the resultant repressing and activating structures (Figure 3.8 A-B), respectively, reveals an upward shift in helices 2 and 3 in the activated structure perturbs the AF-H backbone by an r.m.s.d. of 1.2 Å. This alters the binding position of the coregulator and provides a mechanism by which binding pocket status directly impacts coregulator choice through PC2. Similarly, overlaying the average repressing structure from both the apo-LRH-1 – TIF and apo-LRH-1 – SHP, the differing binding position of the coregulator presses outward on helix 4, causing a slight rearrangement in helices 8 and 9, leading to the motion observed in PC1. Interestingly, the large motions identified in PC1 and PC2 encompass the same regions showing the highest conformational movement in previous HDX studies (see supplementary Figure 3 from published work (20)). In these prior studies, apo LRH-1 shows rapid exchange in helices 2, 3 and 6 (PC1) and helices 8 and 9 (PC2) with complete exchange of these elements occurring in 60 seconds. These same elements are the most sensitive to ligand status showing the strongest projection in the LRH-1 – phospholipid complex.

Modest disruption of interhelical interactions along the allosteric pathway reduces, but does not eliminate, LRH-1 activity

In order to verify that the allosteric pathway we identified plays a role in LRH-1 transcriptional activity, we generated mutant forms of LRH-1 designed to perturb the communication network between phospholipid and coregulator. We took care to avoid residues that make direct contact with the ligand, the AF-2 (coregulator binding) surface or the β-sheet–H6 (alternate AF) region since these would all be expected to reduce LRH-1 transactivation. We instead sought to disrupt LRH-1's allostery one shell of residues away from these surfaces. Helix 5 was identified as a central feature of the pathway (Figure 3.7 B, D), as its junction with helix 10

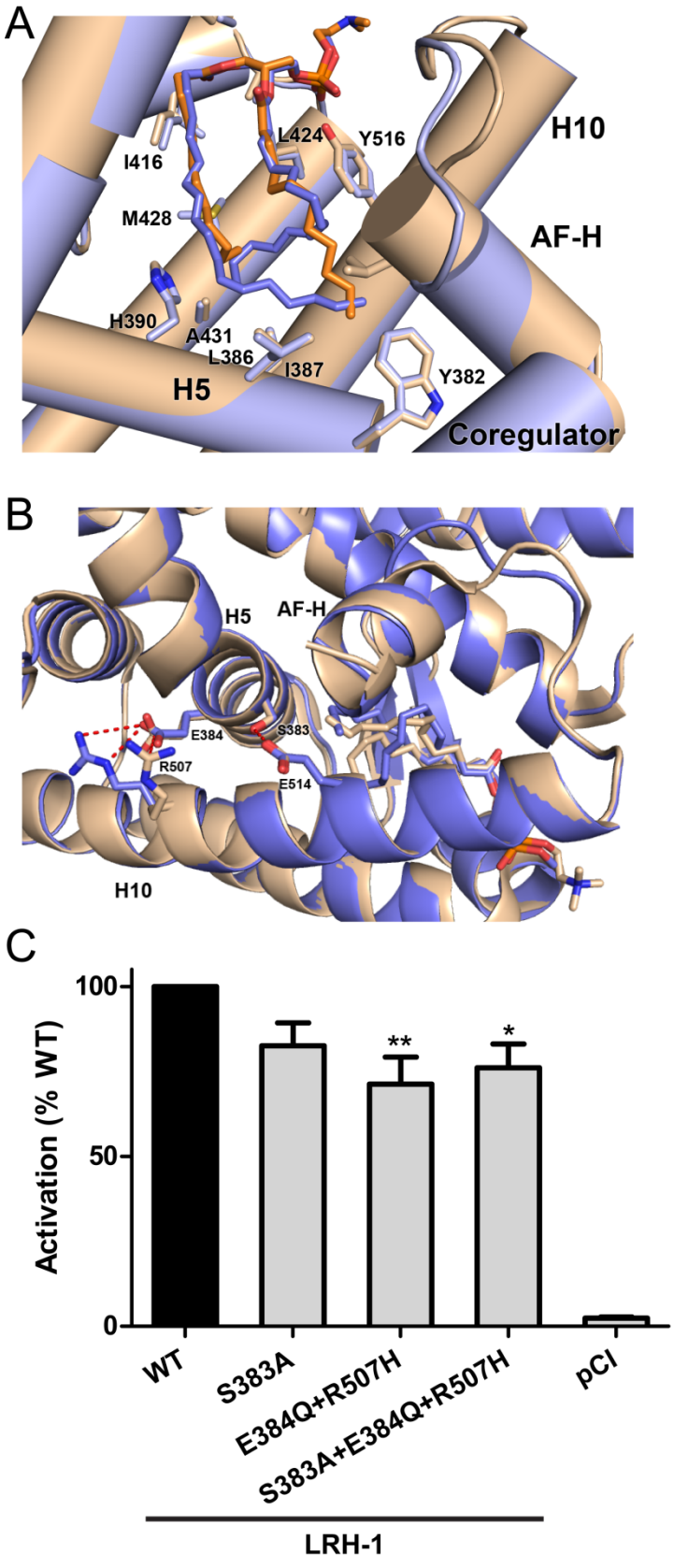
A
L424
Y516
H10
I416
M428
AF-H
H390
A431
L386
I387
H5
Y382
Coregulator
B
H5
AF-H
E384
S383
R507
E514
H10
C
Activation (% WT)
100
50
0
**
*
WT
S383A
E384Q+R507H
S383A+E384Q+R507H
pCI
LRH-1

Figure 3.9. Subtle disruption of residues on or near the allosteric pathway reduces LRH-1 activation.

(A) Close up view of the PL binding pocket of DLPC- (beige/orange) and *E. coli* PL-bound LRH-1 (blue). Helices are shown as cylinders and helix 3 has been hidden. Residues within 4.2 Å of the phospholipid are depicted as sticks. (B) The junction of helices 5 and 10 is displays hydrogen bonding (red dashes) between S383 and E514, and electrostatic interactions between E384 and R507. In the active conformation, helix 12 docks against this junction to support the AF-2 coregulator binding surface, driving gene transactivation and transrepression. (C) Abolition of the electrostatic interaction between helices 5 and 10 via at E384Q/R507H mutation causes a subtle but significant reduction in LRH-1 transcriptional activity. LRH-1 activity was measured via luciferase reporter gene assay in transiently transfected HEK 293T cells. Data are the combined results of 5 independent experiments. Statistical significance is represented as *: $p < 0.05$; **: $p < 0.01$.

creates the cleft against which helix 12 docks to establish the AF-2 coregulator binding surface (Figure 3.9 A). Moreover, the acyl tails of long-chain PLs dock against these helices. We hypothesized that differences in tail length or unsaturation may exert variable amounts of pressure against these helices, which is then transmitted along the allosteric network to the AF-2 site, affecting coregulator binding. The junction between helices 5 and 10 displays hydrogen bonding between residues S383 and E514, and electrostatic interactions between residues E384 and R507 (Figure 3.9 B). To disrupt these interactions, we generated mutant forms of LRH-1 (S383A, E384Q + R507H, and S383A + E384Q + R507H) and measured their transcriptional activity in HEK 293T cells via luciferase reporter gene assay.

These mutant forms of LRH-1 all showed a slight decrease in basal activity, which achieved statistical significance specifically upon disruption of the electrostatic interaction between E384 and R507 (Figure 3.9 C). Importantly, none of these mutations fully abrogated LRH-1 activity, indicating that these mutations did not fatally inactivate the receptor.

Discussion

Robust signaling pathways must not only respond to activating ligands but must discriminate against the wrong ones to reduce noise (50). For LRH-1, this challenge is amplified since its ligands include highly abundant intact PLs that comprise a large fraction of cell membranes. It is possible that LRH-1 displays an intrinsic set of selection criteria for PL isoforms, that PL delivery to the receptor is facilitated by a soluble lipid transport proteins, or a combination of the two. Our results show that LRH-1 is able to bind a wide range of PLs *in vitro*, but can extract only PCs, PGs, and PIs from a membrane/ vesicle without assistance from a molecular chaperone. Inclusion of a non-specific lipid chaperone, β-cyclodextrin, permits the binding of all glycerophospholpids tested. This is in line with structural studies since the majority of recognition occurs via contacts with the lipid tails and phosphoglycerol backbone. Thus, LRH-1 lipid preference is driven more so by the composition of the PL tails than by the head group,

which protrudes from the receptor surface. Remarkably, while LRH-1 can readily accommodate a range of medium-chain saturated PLs, affinity is highest for the 11- and 12-carbon PCs shown to selectively drive receptor activation in cells (13).

Lipid mediated allosteric control of a protein-protein binding interface

Intact PLs are unusual ligands and LRH-1 has evolved to respond to them, via a novel allosteric pathway to support appropriate interaction with coregulators depending on the ligand status. The idea that ligand binding can drive the selective recruitment of different coregulators has been hypothesized before; previous MD studies have indicated that the SHP – LRH-1 interaction is weakened upon the binding of phosphatidylserine (PS) to the apo receptor, while binding of DAX-1 and PROX1 is strengthened (51), suggesting that an avenue exists for communication between the LBP and the AF-2 cleft. While no studies have demonstrated a role for PS in the regulation of LRH-1's target genes, recent HDX studies that compared LRH-1 bound to *E. coli* PLs and DLPC demonstrated increased flexibility in both the mouth of the LBP and the AF-2 region in DLPC-bound LRH-1 (20). Furthermore, stabilizing the mouth of the LBP in apo hLRH-1 by replacing residues 419-424 with the corresponding mouse LRH-1 sequence enhances binding of the coactivators TIF-2 and PGC1α (52). In the absence of PLs, the receptor accesses a greater amount of conformational space and readily interacts with corepressors. Medium-chain PLs appear to promote productive motions that favor coactivator interaction and disfavor SHP interaction, perhaps by suppressing non-activating (non-productive) motions to drive selective interaction with coregulators. LRH-1's allosteric network connecting the β-sheet–H6 region may be an evolutionary adaptation that allowed LRH-1 to sense these unusually large ligands and discriminate against fatty acids and cholesterol-derived ligands which would also fit in the receptor's large hydrophobic pocket.

Ideally, structure-function work should be performed and interpreted in the context of the full-length protein. Obtaining a structure of the intact receptor has been challenging, likely due to the large amount of disorder in the linker region connecting the DNA and ligand binding

domains. Thus, we modeled systems for which there was empirical structural and biochemical data. In addition, LRH-1 transactivation has been shown to be affected by posttranslational modifications located on the hinge (*i.e.* phosphorylation, acetylation and SUMOylation) (12). Phosphorylation of the serine residues S238 and S243 in the hinge region of the human LRH-1 by the mitogen-activated protein kinase ERK1/2 enhances its activity (53). LRH-1 also been shown to be acetylated in the basal state and is bound by the small heterodimer partner (SHP)-sirtuin 1 (SIRT1) transrepressive complex. Surprisingly SIRT1 does not modulate LRH-1 directly, thus what is driving the acetylation and deacetylation of LRH-1 is not established (54). LRH-1 transactivation is also controlled by SUMO conjugation to lysine 289 (55). SUMOylation was shown to drive LRH-1 localization in nuclear bodies, whereby SUMO-conjugated LRH-1 is preferentially sequestered in these bodies preventing it from binding to DNA (55). Recently, Dr. Kristina Schoonjans's lab showed that SUMOylated LRH-1 interacts with PROX-1, a corepressor, to control 25% of LRH-1 gene targets in the liver. Mutation of lysine 289 to an arginine specifically ablates PROX-1 interaction, without affecting other canonical coregulator interactions.

Emerging evidence suggests that NR activation does not occur via the classically described "mouse trap" model, whereby the AF-H swings from an inactive to active state upon agonist binding. Both experimental and modeling studies are inconsistent with radical repositioning of H12 away from the AF-2 in apoNRs (56-59). Rather, subtle local conformational adaptations are observed in H12 as well as other regions within the LBD such as the H11-H12 loop, H3 and H5 (59). These subtle conformational differences between structures may be functionally important, representing a shift between conformational ensembles, but are difficult to identify via inspection of superimposed crystal structures. Previous work with both steroid receptors and fatty acid sensing NRs, have also revealed remarkable flexibility in this region comprising bottom half of the ligand binding pocket including H3, H6-H7 and H11 (60; 61). In the absence of ligand, NRs are partially unfolded. Recent NMR studies focused on PPARγ show

that in the apo state only half of the expected peaks appear on the intermediate exchange timescale (milliseconds-to-microseconds). NMR supports a model whereby NRs sample a range of conformations in the apo state. Full-agonists drive this equilibrium towards a more classically active conformation by protecting residues comprising the ligand binding pocket and AF-2 from intermediate exchange, while partial agonists only partially stabilize the regions of the receptor (62). The β-sheet region may also play an important role in mediating PPARγ's response to ligands (56). While the dynamics in this region are important for mediating ligand action, activation by partial agonists is mediated by the ability of a solvent inaccessible serine residue in this region to be phosphorylated (56).

Given LRH-1's limited selectivity criteria *in vitro*, it is possible that access to endogenous ligands are controlled both temporally and spatially by phospholipid transfer proteins. For example, phospholipid transfer proteins such as phosphatidylinositol transfer protein α and phosphatidylcholine transfer protein are both capable of transporting intact PLs into the nucleus (63; 64). The effect of tail unsaturation has also not yet been studied, but it is likely that the bends introduced by cis unsaturation would allow the LRH-1 ligand binding pocket to accommodate longer-chain acyl tails promoting potent receptor activation. Given the diverse composition of PL tails *in vivo*, these studies are best guided by lipodimics-based identification of endogenous PL ligands. Current limitations in the ability to isolate LRH-1 from mammalian tissue have limited the field's ability to identify endogenous ligands, though these studies are underway.

References

1. Musille PM, Kohn JA, Ortlund EA. 2013. Phospholipid – Driven gene regulation. *FEBS Letters*
2. Blind RD, Sablin EP, Kuchenbecker KM, Chiu HJ, Deacon AM, et al. 2014. The signaling phospholipid PIP3 creates a new interaction surface on the nuclear receptor SF-1. *Proc Natl Acad Sci U S A* 111:15054-9
3. Liu S, Brown JD, Stanya KJ, Homan E, Leidl M, et al. 2013. A diurnal serum lipid integrates hepatic lipogenesis and peripheral fatty acid use. *Nature* 502:550-4

4. Walker AK, Jacobs RL, Watts JL, Rottiers V, Jiang K, et al. 2011. A conserved SREBP-1/phosphatidylcholine feedback circuit regulates lipogenesis in metazoans. *Cell* 147:840-52
5. Fernandez-Marcos PJ, Auwerx J, Schoonjans K. 2010. Emerging actions of the nuclear receptor LRH-1 in the gut. *Biochim Biophys Acta*
6. Moore D. Targeting nuclear receptors to treat type 2 diabetes. *Proc. 14th International Congress of Endocrinology*, *Kyoto, Japan*, *2010*:
7. Oosterveer MH, Schoonjans K. 2013. Hepatic glucose sensing and integrative pathways in the liver. *Cellular and molecular life sciences : CMLS*
8. Zhang C, Large MJ, Duggavathi R, DeMayo FJ, Lydon JP, et al. 2013. Liver receptor homolog-1 is essential for pregnancy. *Nature Medicine* 19:1061-6
9. Gerrits H, Parade MC, Koonen-Reemst AM, Bakker NE, Timmer-Hellings L, et al. 2014. Reversible infertility in a liver receptor homologue-1 (LRH-1)-knockdown mouse model. *Reproduction, fertility, and development* 26:293-306
10. Kelly VR, Xu B, Kuick R, Koenig RJ, Hammer GD. 2010. Dax1 Up-Regulates Oct4 Expression in Mouse Embryonic Stem Cells via LRH-1 and SRA. *Molecular endocrinology (Baltimore, Md.)* 24:1-11
11. Venteclef N, Jakobsson T, Steffensen KR, Treuter E. 2011. Metabolic nuclear receptor signaling and the inflammatory acute phase response. *Trends in endocrinology and metabolism: TEM*:1-11
12. Stein S, Schoonjans K. 2014. Molecular basis for the regulation of the nuclear receptor LRH-1. *Current opinion in cell biology* 33C:26-34
13. Lee JM, Lee YK, Mamrosh JL, Busby Sa, Griffin PR, et al. 2011. A nuclear-receptor-dependent phosphatidylcholine pathway with antidiabetic effects. *Nature*
14. Bolado-Carrancio A, Riancho JA, Sainz J, Rodriguez-Rey JC. 2014. Activation of nuclear receptor NR5A2 increases Glut4 expression and glucose metabolism in muscle cells. *Biochemical and biophysical research communications* 446:614-9
15. Mamrosh JL, Lee JM, Wagner M, Stambrook PJ, Whitby RJ, et al. 2014. Nuclear receptor LRH-1/NR5A2 is required and targetable for liver endoplasmic reticulum stress resolution. *eLife* 3:e01694-e
16. Whitby RJ, Dixon S, Maloney PR, Delerive P, Goodwin BJ, et al. 2006. Identification of small molecule agonists of the orphan nuclear receptors liver receptor homolog-1 and steroidogenic factor-1. *Journal of medicinal chemistry* 49:6652-5
17. Whitby RJ, Stec J, Blind RD, Dixon S, Leesnitzer LM, et al. 2011. Small Molecule Agonists of the Orphan Nuclear Receptors Steroidogenic Factor-1 (SF-1, NR5A1) and Liver Receptor Homologue-1 (LRH-1, NR5A2). *Journal of medicinal chemistry* 1
18. Busby S, Nuhant P, Cameron M, Mercer BA, Hodder P, et al. 2010. Discovery of Inverse Agonists for the Liver Receptor Homologue-1 (LRH1; NR5A2). In *Probe Reports from the NIH Molecular Libraries Program*. Bethesda (MD). Number of.
19. Benod C, Carlsson J, Uthayaruban R, Hwang P, Irwin JJ, et al. 2013. Structure-based discovery of antagonists of nuclear receptor LRH-1. *J Biol Chem* 288:19830-44
20. Musille PM, Pathak MC, Lauer JL, Hudson WH, Griffin PR, Ortlund EA. 2012. Antidiabetic phospholipid-nuclear receptor complex reveals the mechanism for phospholipid-driven gene regulation. *Nat Struct Mol Biol* 19:532-7, S1-2
21. Goodwin B, Jones Sa, Price RR, Watson Ma, McKee DD, et al. 2000. A regulatory cascade of the nuclear receptors FXR, SHP-1, and LRH-1 represses bile acid biosynthesis. *Molecular cell* 6:517-26
22. Zhi X, Zhou XE, He Y, Zechner C, Suino-Powell KM, et al. 2014. Structural insights into gene repression by the orphan nuclear receptor SHP. *Proc Natl Acad Sci U S A* 111:839-44

23. Goodwin B, Watson MA, Kim H, Miao J, Kemper JK, Kliewer SA. 2003. Differential regulation of rat and human CYP7A1 by the nuclear oxysterol receptor liver X receptor-alpha. *Mol Endocrinol* 17:386-94
24. Lu TT, Makishima M, Repa JJ, Schoonjans K, Kerr TA, et al. 2000. Molecular basis for feedback regulation of bile acid synthesis by nuclear receptors. *Molecular cell* 6:507-15
25. Ortlund EA, Lee Y, Solomon IH, Hager JM, Safi R, et al. 2005. Modulation of human nuclear receptor LRH-1 activity by phospholipids and SHP. *Nat Struct Mol Biol* 12:357-63
26. Otwinowski Z, Minor W. 1997. [20] Processing of X-ray diffraction data collected in oscillation mode. 276:307-26
27. Murshudov GN, Skubak P, Lebedev AA, Pannu NS, Steiner RA, et al. 2011. REFMAC5 for the refinement of macromolecular crystal structures. *Acta crystallographica. Section D, Biological crystallography* 67:355-67
28. Echols N, Grosse-Kunstleve RW, Afonine PV, Bunkoczi G, Chen VB, et al. 2012. Graphical tools for macromolecular crystallography in PHENIX. *Journal of applied crystallography* 45:581-6
29. Nicholls RA, Fischer M, McNicholas S, Murshudov GN. 2014. Conformation-independent structural comparison of macromolecules with ProSMART. *Acta crystallographica. Section D, Biological crystallography* 70:2487-99
30. Wang J, Wang W, Kollman PA, Case DA. 2006. Automatic atom type and bond type perception in molecular mechanical calculations. *Journal of molecular graphics & modelling* 25:247-60
31. B. R. Brooks REB, B. D. Olafson, D. J. States, S. Swaminathan, and M. Karplus. 1983. CHARMM: A Program for Macromolecular Energy, Minimization, and Dynamics Calculations. *Journal of Computational Chemistry* 4:187-217
32. Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, et al. 2005. Scalable molecular dynamics with NAMD. *Journal of Computational Chemistry* 26:1781-802
33. Tuckerman M, Berne BJ, Martyna GJ. 1992. Reversible Multiple Time Scale Molecular-Dynamics. *J Chem Phys* 97:1990-2001
34. Essmann U, Perera L, Berkowitz ML, Darden T, Lee H, Pedersen LG. 1995. A Smooth Particle Mesh Ewald Method. *J Chem Phys* 103:8577-93
35. Glykos NM. 2006. Software news and updates. Carma: a molecular dynamics analysis program. *Journal of Computational Chemistry* 27:1765-8
36. Sethi A, Eargle J, Black AA, Luthey-Schulten Z. 2009. Dynamical networks in tRNA:protein complexes. *Proc Natl Acad Sci U S A* 106:6620-5
37. Humphrey W, Dalke A, Schulten K. 1996. VMD: visual molecular dynamics. *Journal of molecular graphics* 14:33-8, 27-8
38. Sablin EP, Krylova IN, Fletterick RJ, Ingraham HA. 2003. Structural basis for ligand-independent activation of the orphan nuclear receptor LRH-1. *Molecular cell* 11:1575-85
39. Wang W, Zhang C, Marimuthu A, Krupka HI, Tabrizizad M, et al. 2005. The crystal structures of human steroidogenic factor-1 and liver receptor homologue-1. *Proceedings of the National Academy of Sciences of the United States of America* 102:7505-10
40. Krylova IN, Sablin EP, Moore J, Xu RX, Waitt GM, et al. 2005. Structural analyses reveal phosphatidyl inositols as ligands for the NR5 orphan receptors SF-1 and LRH-1. *Cell* 120:343-55
41. Blind RD, Suzawa M, Ingraham HA. 2012. Direct Modification and Activation of a Nuclear Receptor-PIP2 Complex by the Inositol Lipid Kinase IPMK. *Sci Signal* 5
42. Schaaf G, Ortlund EA, Tyeryar KR, Mousley CJ, Ile KE, et al. 2008. Functional anatomy of phospholipid binding and regulation of phosphoinositide homeostasis by proteins of the sec14 superfamily. *Mol Cell* 29:191-206

43. Kanno K, Wu MK, Scapa EF, Roderick SL, Cohen DE. 2007. Structure and function of phosphatidylcholine transfer protein (PC-TP)/StarD2. *Biochim Biophys Acta* 1771:654-62
44. Schouten A, Agianian B, Westerman J, Kroon J, Wirtz KW, Gros P. 2002. Structure of apo-phosphatidylinositol transfer protein alpha provides insight into membrane association. *EMBO J* 21:2117-21
45. Nagy L, Schwabe JW. 2004. Mechanism of the nuclear receptor molecular switch. *Trends Biochem Sci* 29:317-24
46. Li Y, Lambert MH, Xu HE. 2003. Activation of nuclear receptors: a perspective from structural genomics. *Structure (London, England : 1993)* 11:741-6
47. Sablin EP, Krylova IN, Fletterick RJ, Ingraham HA. 2003. Structural basis for ligand-independent activation of the orphan nuclear receptor LRH-1. *Mol Cell* 11:1575-85
48. Levy RM, Srinivasan AR, Olson WK, McCammon JA. 1984. Quasi-harmonic method for studying very low frequency modes in proteins. *Biopolymers* 23:1099-112
49. Skjaerven L, Martinez A, Reuter N. 2011. Principal component and normal mode analysis of proteins; a quantitative comparison using the GroEL subunit. *Proteins* 79:232-43
50. Atkins WM. 2014. Biological messiness vs. biological genius: Mechanistic aspects and roles of protein promiscuity. *J Steroid Biochem Mol Biol*
51. Burendahl S, Treuter E, Nilsson L. 2008. Molecular dynamics simulations of human LRH-1: the impact of ligand binding in a constitutively active nuclear receptor. *Biochemistry* 47:5205-15
52. Musille PM, Pathak M, Lauer JL, Griffin PR, Ortlund EA. 2013. Divergent sequence tunes ligand sensitivity in phospholipid-regulated hormone receptors. *J Biol Chem* 288:20702-12
53. Lee YK, Choi YH, Chua S, Park YJ, Moore DD. 2006. Phosphorylation of the hinge domain of the nuclear hormone receptor LRH-1 stimulates transactivation. *J Biol Chem* 281:7850-5
54. Chanda D, Xie YB, Choi HS. 2010. Transcriptional corepressor SHP recruits SIRT1 histone deacetylase to inhibit LRH-1 transactivation. *Nucleic Acids Res* 38:4607-19
55. Chalkiadaki A, Talianidis I. 2005. SUMO-dependent compartmentalization in promyelocytic leukemia protein nuclear bodies prevents the access of LRH-1 to chromatin. *Mol Cell Biol* 25:5095-105
56. Hughes TS, Chalmers MJ, Novick S, Kuruvilla DS, Chang MR, et al. 2012. Ligand and receptor dynamics contribute to the mechanism of graded PPARgamma agonism. *Structure* 20:139-50
57. Martinez L, Polikarpov I, Skaf MS. 2008. Only subtle protein conformational adaptations are required for ligand binding to thyroid hormone receptors: simulations using a novel multipoint steered molecular dynamics approach. *The journal of physical chemistry. B* 112:10741-51
58. Batista MR, Martinez L. 2013. Dynamics of nuclear receptor Helix-12 switch of transcription activation by modeling time-resolved fluorescence anisotropy decays. *Biophys J* 105:1670-80
59. Mackinnon JA, Gallastegui N, Osguthorpe DJ, Hagler AT, Estebanez-Perpina E. 2014. Allosteric mechanisms of nuclear receptors: insights from computational simulations. *Mol Cell Endocrinol* 393:75-82
60. Bledsoe RK, Montana VG, Stanley TB, Delves CJ, Apolito CJ, et al. 2002. Crystal structure of the glucocorticoid receptor ligand binding domain reveals a novel mode of receptor dimerization and coactivator recognition. *Cell* 110:93-105
61. Gee AC, Katzenellenbogen JA. 2001. Probing conformational changes in the estrogen receptor: evidence for a partially unfolded intermediate facilitating ligand binding and release. *Mol Endocrinol* 15:421-8

62. Kojetin DJ, Burris TP. 2013. Small molecule modulation of nuclear receptor conformational dynamics: implications for function and drug discovery. *Mol Pharmacol* 83:1-8
63. Nile AH, Bankaitis VA, Grabon A. 2010. Mammalian diseases of phosphatidylinositol transfer proteins and their homologs. *Clin Lipidol* 5:867-97
64. Kang HW, Kanno K, Scapa EF, Cohen DE. 2010. Regulatory role for phosphatidylcholine transfer protein/StarD2 in the metabolic response to peroxisome proliferator activated receptor alpha (PPARalpha). *Biochim Biophys Acta* 1801:496-502

Chapter 4: Regulation of LRH-1 by endogenous lipids – preliminary findings

Introduction

Due to its role as a master regulator of many genes, tight control of LRH-1 activity is crucial for development and normal function in digestive and reproductive organs. Regulation of LRH-1 occurs via several means, including up- or downregulation of the *nr5a2* gene, and PTMs such as SUMOylation (1-3). In recent years, phospholipids (PLs) have emerged as regulatory ligands for LRH-1 and its close paralog SF-1 (4-11). While the exogenous medium-chain PC species DLPC has been shown to activate LRH-1 by preventing the binding of corepressors (5; 6), to date no endogenously relevant PLs have been identified as bona-fide regulatory ligands for LRH-1. The ability of PLs to regulate gene transcription as NR ligands is a novel and interesting concept (12). PLs are unusual NR ligands, due to their typical residence in the membrane, requiring their extraction in order to bind to the NR LBD. Furthermore, PLs have a modular structure that integrates information from multiple cellular pathways into a single molecule via variation of the fatty acyl tails and the polar headgroups. The sheer diversity of PLs may enable exquisite control over the activity of LRH-1 (or other NRs), since distinct PLs may promote the recruitment of different coregulator proteins. However, the diversity of the PL family also makes it difficult to predict and assay which species are relevant for NR regulation. Thus, the following questions remain: which PLs act as endogenous regulatory ligands for LRH-1? By what structural mechanism does PL binding drive LRH-1 coregulator recruitment and receptor activity?

To answer these questions, we conducted pulldown and mass spectrometry experiments in order to identify the endogenous PLs that bound to LRH-1. We then determined the binding affinity of the PLs identified, and capacity of these PLs to activate or repress LRH-1. Furthermore, in order to elucidate the mechanism of PL delivery to LRH-1, we analyzed the role of the lipid transfer protein PCTP in LRH-1 activity, and its capacity to directly transfer PC to LRH-1.

Experimental Procedures

Phospholipid pulldown

Recombinant LRH-1 was expressed in *E. coli* and purified as described in Chapter 4 and used without removal of *E. coli* lipids. LRH-1 (1 mg/mL) was incubated overnight at 4°C with SUVs (100 μM DLPC or 1/10 dilution from chloroform extract of indicated mouse tissue) or HepG2 total lysate in an incubation buffer consisting of 150 mM NaCl, 20 mM Tris-HCl pH 7.4, 5% glycerol, 0.1% N-octyl-β-glucopyrranoside, and 5 mM β-cyclodextrin. The protein was loaded onto His-Trap spin columns, washed 3x with incubation buffer containing 25 mM imidazole, and eluted with incubation buffer containing 250 mM imidazole. Bound lipids were extracted with 65:34:1 $CHCl_3$:MeOH:H_2O and analyzed by electrospray ionization mass spectrometry on a QTrap 6500 (ABSciex) , with precursor ion scanning for phosphatidylcholine species (fragment m/z 184.10).

LRH-1 Phospholipid binding assay

The affinity of LRH-1 for the indicated phospholipids was measured using the assay described in Chapter 3 (page 73).

Cell culture

HEK 293T cells were cultured in complete media consisting of DMEM supplemented with 10% FBS and penicillin/streptomycin. Cells were passaged twice per week. Transfections were performed in OptiMEM using the Lipofectamine 3000 transfection reagent.

LRH-1 activation assay.

HEK 293T cells were seeded into 24 well plates and incubated at 37°C for 48 h. Each well was transiently cotransfected with 500 ng plasmid encoding firefly luciferase under control of the SHP promoter (SHP-luc), 10 ng plasmid encoding renilla luciferase under control of the constitutively active CMV promoter (pRLCMV), and 100 ng plasmid encoding the full-length human LRH-1 in the pCI vector backbone. After 4 h incubation at 37°C, the transfections were ended via replacement of the transfection mixture with 100 μM phospholipid or 10 μM A1 as

indicated; each phospholipid mixture was prepared by diluting phospholipid from a 10 mM ethanol stock into complete media and sonicating for 15 min in a bath sonicator (Avanti). The cells were incubated in the phospholipid mixture for 24 h at 37°C, after which luminescence was measured on a Biotek Synergy 4 plate reader using the Dual-Glo Luciferase Assay System (Promega). The activity of firefly luciferase was normalized to renilla luciferase. Data are presented in triplicate as mean and S.E.M., and processed in Graphpad Prism 6 (GraphPad, Inc).

Expression of PCTP

Human PCTP was recombinantly expressed in *E. coli* strain BL21 DE3 pLysS and grown in TB supplemented with 0.1% soy lecithin to provide a source of PC ligand. Briefly, TB cultures were grown at 37°C to OD_{600} 0.4, then at 22°C to OD_{600} 0.7, at which point they were induced by addition of 1 mM IPTG and grown overnight at 22°C. Cell debris was pelleted by centrifugation at 4000×g at 4°C. After one freeze/thaw cycle at -80°C, the cells were lysed by sonication in 4x (volume) NiA buffer containing 150 mM NaCl, 20 mM Tris pH 7.4, 5% glycerol, and 25 mM imidazole, supplemented with lysozyme, DNase, and 100 μM PMSF. The lysate was cleared by centrifugation at 17Krpm in a JA-20 rotor at 4°C for 30 min, and loaded onto a 5 mL nickel affinity column. The column was washed with 5 CV NiA and 5 CV NiA + 5% NiB (NiA + 250 mM imidazole), and eluted with 50% NiB in NiA. Protein was concentrated and further purified for use in assays by size exclusion chromatography into buffer containing 150 mM NaCl, 20 mM Tris pH 7.4, 100 μM DTT, and 100 μM EDTA.

Loading of PCTP with PC-NBD probe

Purified PCTP was loaded with the fluorescent probe PC 14:0,6:0-NBD (PC-NBD; 1-myristoyl-2-{6-[(7-nitro-2-1,3-benzoxadiazol-4-yl)amino]hexanoyl}-sn-glycero-3-phosphocholine; Avanti Polar Lipids) as follows. PCTP (100 μM) was added to 300 μM PC-NBD donor vesicles (300 μM PC-NBD + 15 μM DPPG (dipalmitoleyl phosphatidylglycerol)) and incubated overnight at room temperature in the dark with gentle rocking in the size exclusion buffer described above. PCTP was repurified in the dark by size exclusion into the same buffer.

Concentration of PCTP loaded with PC-NBD was measured by the absorbance of NBD at 460 nM.

PCTP—LRH-1 direct phospholipid transfer assay

Direct transfer of PC-NBD to LRH-1 was measured using an adaptation of the FRET binding assay described in Chapter 3. Briefly, the quenching of DCIA—LRH-1 (150 nM) in response to increasing concentrations of PC-NBD vesicles or PCTP loaded with PC-NBD (1 nM – 100 μM) was measured in PCTP size exclusion buffer (described above) at 37°C at the time of addition and every 30 s for 72 min.

PCTP catalyzed LRH-1 phospholipid competition assay

The capacity of PCTP to catalyze the transfer of POPC to LRH-1 was measured as follows. The fluorescence recovery of DCIA—LRH-1 (150 nM) pre-quenched by the addition of PC-NBD (5 μM) was measured in response to incubation with competitor vesicles (100 μM POPC + 5 μM DPPG) and varying amounts of PCTP (1 nM – 10 μM), in PCTP size exclusion buffer for 2 h at 37°C.

Results

To gain insight into the role of endogenous PLs as regulatory ligands, we sought to identify the PLs that bind LRH-1 in an endogenous cellular environment. We were unable to harvest a useful quantity of LRH-1 from mammalian cells, so we instead incubated recombinantly expressed LRH-1 with lipid extracts from mouse liver and heart, and HepG2 cell lysate, and identified by mass spectrometry the lipids that bound. Phosphatidylcholine species were selected via precursor ion scanning and identified by the total carbon content and degree of unsaturation of their combined acyl tails.

LRH-1 binds to PCs that contained long-chain mono- and polyunsaturated acyl tails (Figures 4.1-3, Table 4.1). Included were several abundant structural PCs (typically found in a

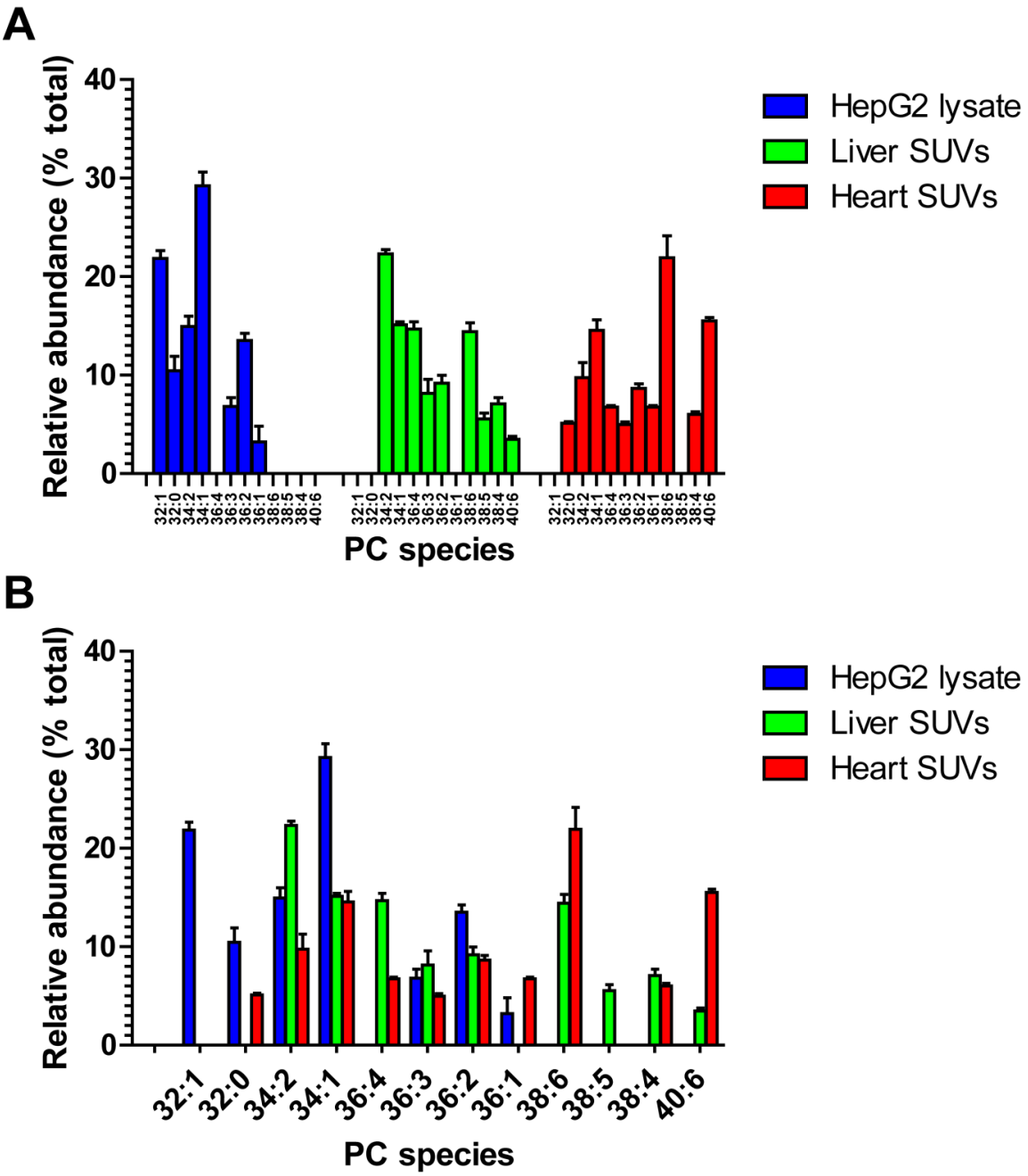

Figure 4.1. Identification of LRH-1—binding PCs by mass spectrometry.

(A) LRH-1 binding PCs organized by source tissue. (B) LRH-1 binding PCs ordered by tail composition. Carbon content is described as #carbon:#unsaturation.

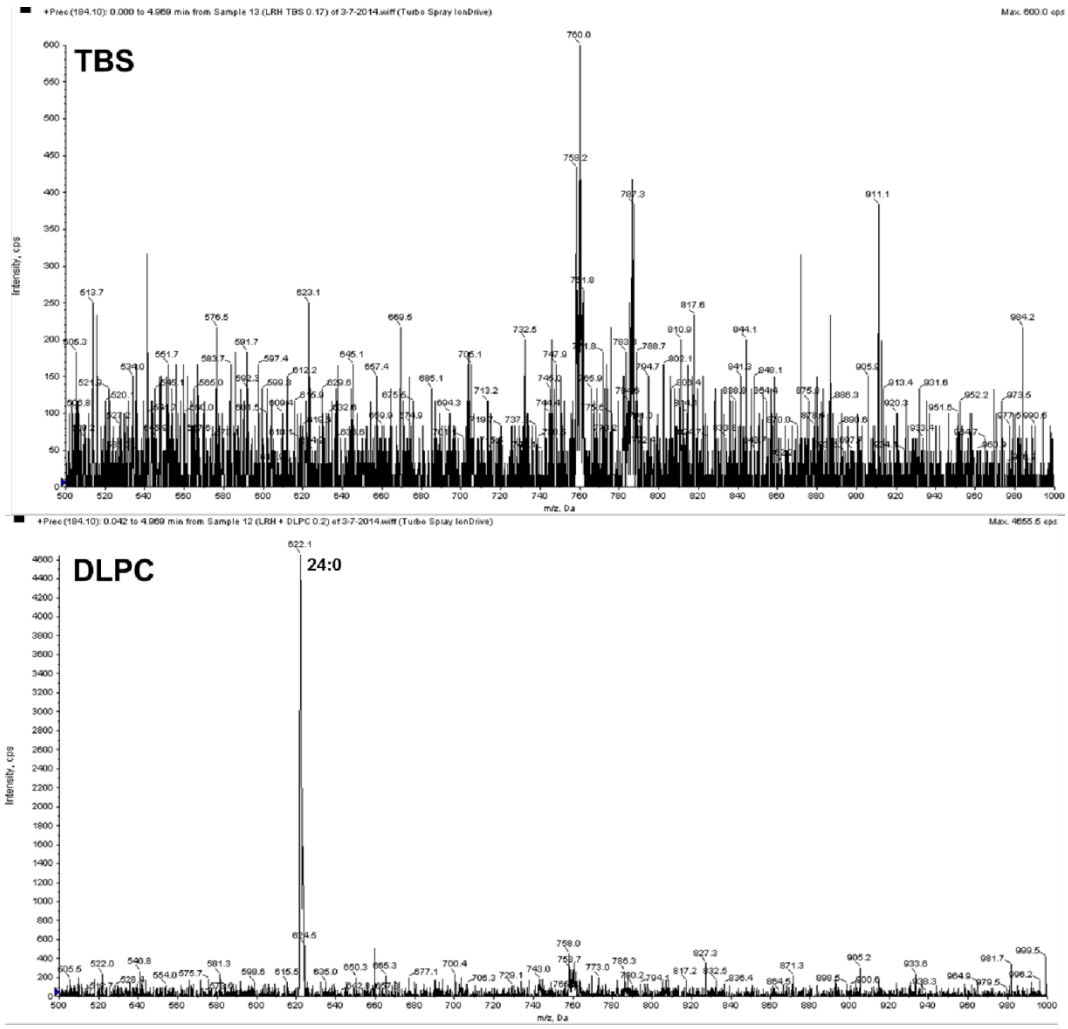

Figure 4.2. Spectra of PCs bound to LRH-1 in control incubations.

To verify PC exchange, LRH-1 was incubated with buffer (TBS; top) or 100 μM DLPC vesicles (bottom) as indicated. PCs were identified by their total carbon content and unsaturation based upon the m/z of the observed peaks.

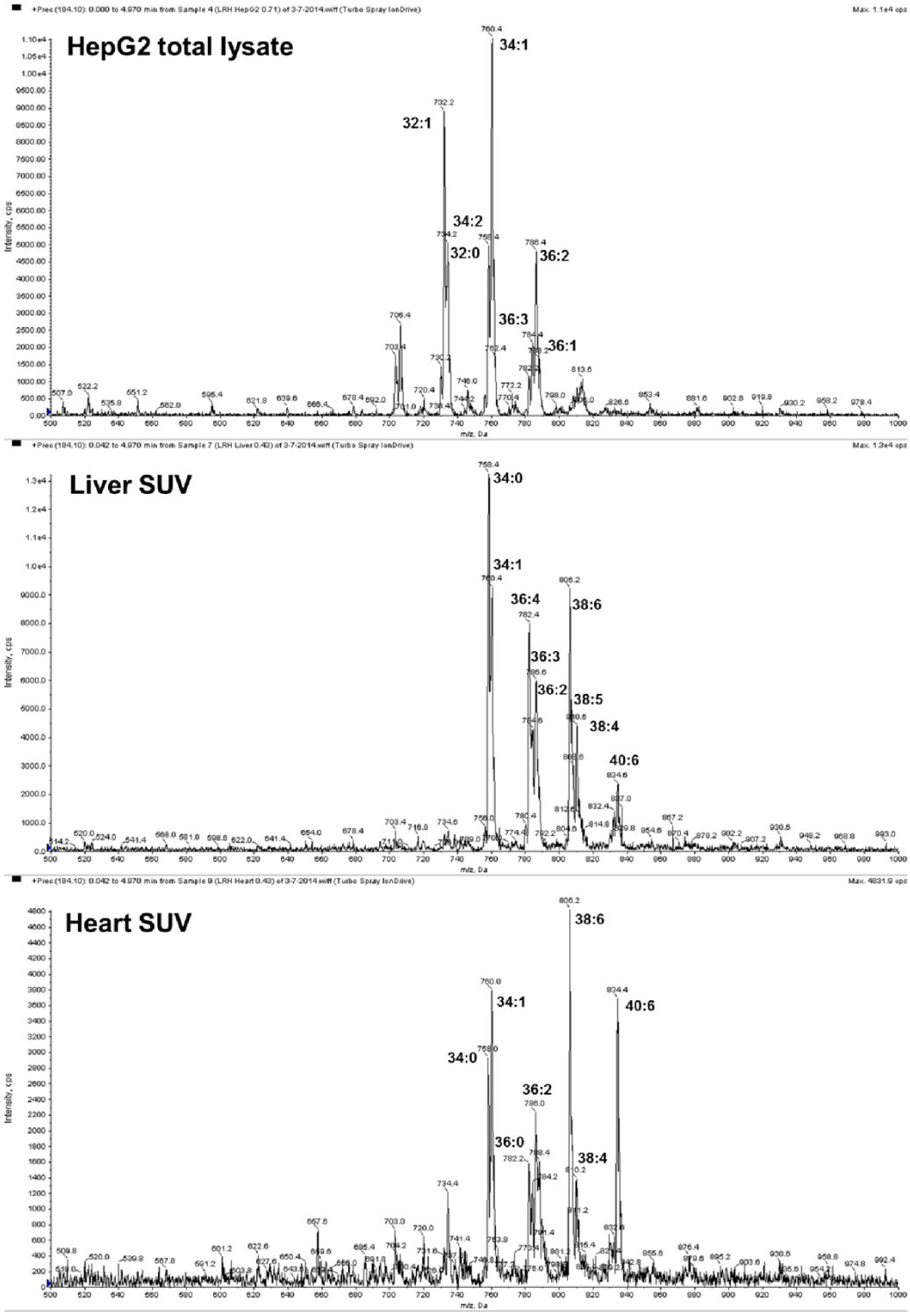

Figure 4.3. Spectra of PCs bound to LRH-1 from natural lipid extracts.

LRH-1 was incubated with lipid extracts as indicated. PCs were identified by their total carbon content and unsaturation based upon the m/z of the observed peaks.

Table 4.1. Relative abundance of PC species

PC total carbon content	HepG2 lysate (%)	Liver SUV (%)	Heart SUV (%)	Probable PC species
32:1	22	0	0	16:0,16:1
32:0	10	0	5	16:0,16:0
34:2	15	22	10	16:0,18:2
34:1	29	15	14	16:0,18:1
36:4	0	15	7	16:0,20:4
36:3	7	8	5	18:0,18:3
36:2	14	9	9	18:0,18:2
36:1	3	0	7	18:0,18:1
38:6	0	14	22	16:0,22:6
38:5	0	6	0	18:0,20:5
38:4	0	7	6	18:0,20:4
40:6	0	3	15	18:0,22:6

plasma membrane) with tails of 16-18 carbons and 0-2 unsaturations, and PCs that contain arachidonic acid (20:4) and docosahexaenoic acid (22:6). These PCs showed comparable binding affinity to LRH-1 (Figure 4.4 A), but, unlike DLPC, did not activate or repress the SHP promoter relative to vehicle (Figure 4.4 B).

We next sought to evaluate the role of PCTP in LRH-1 regulation. Inhibition of PCTP with a small molecule antagonist decreased the transcriptional activity of LRH-1 on the SHP promoter by about 50% (Figure 4.5 A), indicating that the intracellular transfer of PCs was important to LRH-1 activity. To test whether or not PCTP delivered PCs directly to the LRH-1 LBD, we loaded recombinantly expressed human PCTP and with PC-NBD (PC 14:0,6:0-NBD), a tail-labeled fluorescent PC probe that quenches the fluorescence of DCIA-labeled LRH-1 as described in Chapter 3, and measured the ability of the probe-loaded PCTP to transfer PC-NBD to LRH-1. Surprisingly, LRH-1 uptake of PC-NBD from PCTP was slightly less efficient than uptake of PC-NBD directly from vesicles (Figure 4.5 B-D). Furthermore, PCTP did not catalyze the transfer of POPC (PC 16:0,18:1) from vesicles to LRH-1 (Figure 4.5 E, F) *in vitro*. Thus, while PCTP may be critical in regulating LRH-1 activity, these preliminary results failed to detect transfer of a labeled PC ligands directly to LRH-1.

Discussion

The *in vitro* binding and pulldown data presented in this work indicate that LRH-1 readily binds PCs with long chain unsaturated acyl tails with an affinity comparable to that of the unnatural PL DLPC, and ten-fold higher than that of the natural PL DPPC (see Figures 3.3, 4.4). The distribution of fatty acyl tails of the PCs bound to LRH-1 roughly approximates the overall fatty acid distribution seen in their source tissues (13; 14); this is congruent with the similar binding affinities seen among the different PC species (Figure 4.4 A), and suggests that ligand regulation of LRH-1 is likely not driven by the selection of a high-affinity PC. Instead, regulation of LRH-1 may be achieved by their ability to drive differential coregulator recruitment, or by the

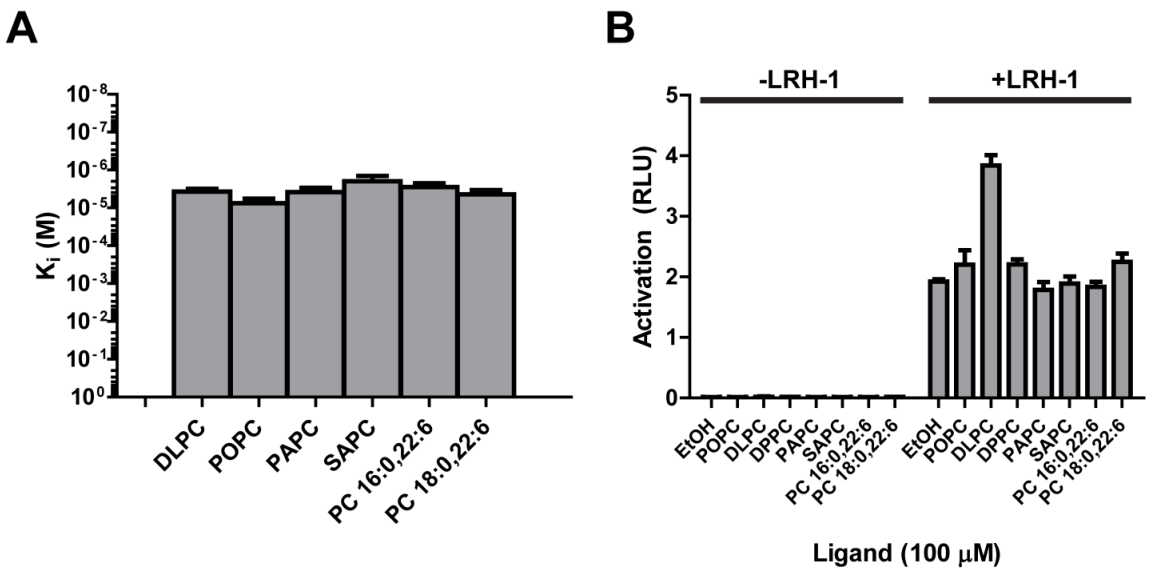

Figure 4.4. LRH-1 binding and activation of endogenous PCs.

(A) Relative binding affinities of endogenous PCs to LRH-1. Binding was relative to 5 μM NBD-DLPE probe. Data are reported as mean ± S.E.M. of triplicate experiments. (B) Transctivation of the SHP promoter by LRH-1 in HEK 293T cells in response to treatment with PCs. Luminescence of firefly luciferase under control of the SHP promoter has been normalized to luminescence of constitutively expressed renilla luciferase. Data are presented as the mean ± S.E.M. of triplicate experiments. PL abbreviations are as follows: DLPC = dilauroyl phosphatidylcholine (PC 12:0,12:0); POPC = 1-palmitoyl-2-oleoyl phosphatidylcholine (PC 16:0,18:1); PAPC = 1-palmitoyl-2-arachidonyl phosphatidylcholine (PC 16:0,20:4); SAPC = 1-oleoyl-2-arachidonyl phosphatidylcholine (PC 18:0,20:4).

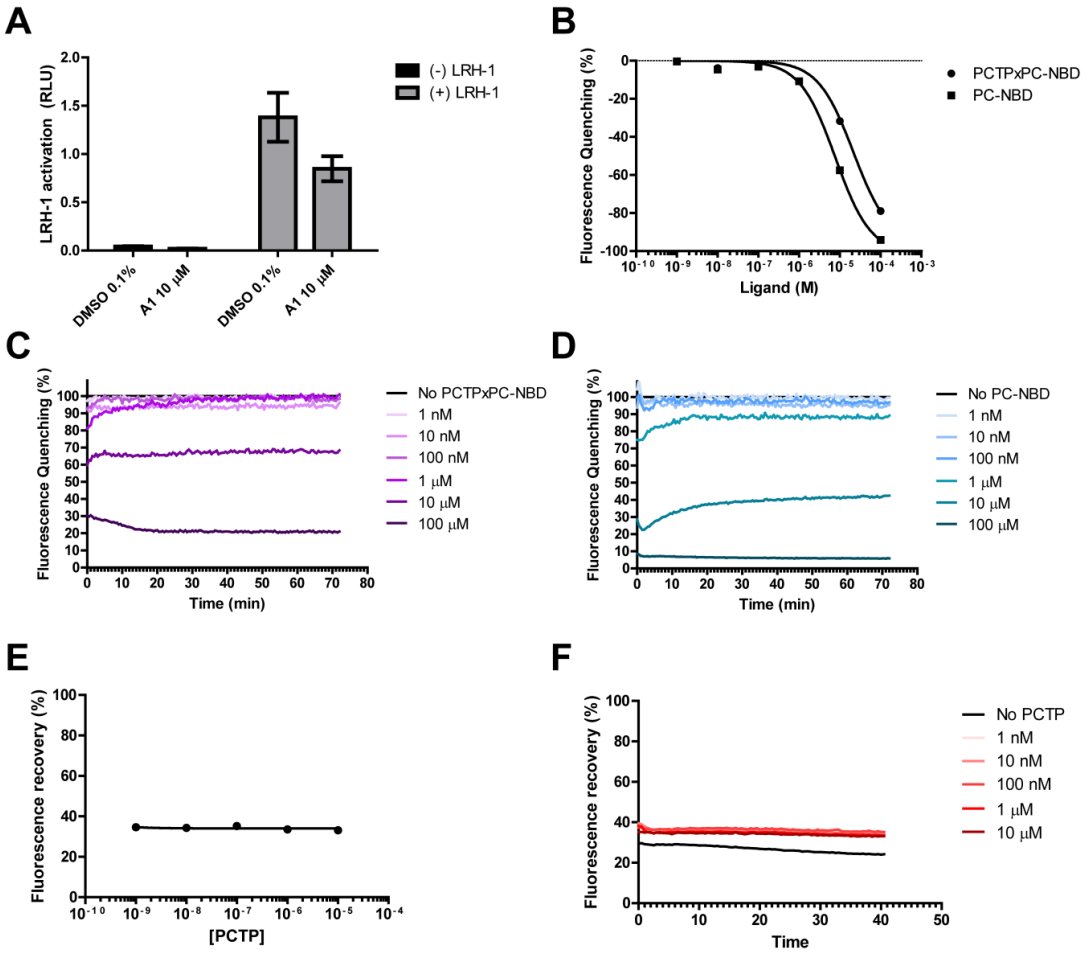

Figure 4.5. Role of PCTP in LRH-1 ligand acquisition.

(A) Chemical inhibition of PCTP reduces LRH-1 transactivation on the SHP promoter by 50% in HEK 293T reporter gene assays. (B) PCTP loaded with PC-NBD is slightly less effective than PC-NBD vesicles at delivering PC to LRH-1 (72 h incubation at 37°C). (C) Kinetics of delivery of PC-NBD from PCTP to LRH-1. (D) Kinetics of extraction of PC-NBD directly from vesicles by LRH-1. (E) PCTP fails to facilitate the transfer of POPC from donor vesicles (100 µM) to LRH-1 even at micromolar concentrations (40 min incubation at 37°C). (F) Kinetics of PCTP-mediated transfer of POPC from donor vesicles to LRH-1.

capacity of chaperone proteins to differentially deliver PCs to LRH-1. Because there were no statistically significant differences in the binding affinities among these lipids, and because none of them appeared to induce robust activation of LRH-1 on the SHP promoter, their role as regulatory ligands remains unclear.

Interestingly, but not unexpectedly, introducing cis double bonds permitted the binding of PLs with long-chain (>18C) fatty acids that would not bind when saturated (see Figure 3.3). It is noteworthy that both acyl tails of a PL lie along helices 3 and 5 when bound to LRH-1, and that each tail must bend sharply in order for the PL to fit in the LBP (see Chapter 3, Figure 3.9 A, B). Both of these helices are part of the allosteric network that links the mouth of the LBP to the AF-2. Cis unsaturated fatty acids are naturally bent by their double bond, with the extent of their bending proportional to the amount of unsaturation, but saturated fatty acids in solution tend to adopt a straight-chain conformation; thus, while the energy required to maintain an unsaturated fatty acid comes from the double bond, the energy required to bend the carbon chain to fit in the LRH-1 LBP must therefore come from the protein itself – specifically via the contact of the acyl chain with helices 3 and 5. Their connection to the allosteric network may enable these helices to act as unsaturation sensors that communicate with the AF-2, permitting the differential recruitment of coregulators. Additionally, while this work focused on LRH-1 binding to PC species, it is possible that further control of LRH-1 activity may be achieved via the variation of the PL head group. Interestingly, in addition to PCs containing the more common palmitic, stearic, and oleic acids, LRH-1 co-purifed with and readily bound to PCs containing arachidonic acid (AA) and docosahexaenoic acid (DHA). AA and DHA are both precursors to pro- and anti-inflammatory signals, and are stored in the cell as components of PLs. LRH-1 has been identified as a negative regulator of inflammation via several mechanisms, including the repression of acute phase proteins (15), the activation of extra-adrenal production of glucocorticoids (16), and the upregulation of genes controlling the production of 15-hydroxyeicosatetraenoic acid (15-HETE), a potent PPARγ agonist that is required for the IL-13 induced alternative activation of

macrophages (17). Poignantly, 15-HETE is produced from AA in this cascade, and while the nature of PL regulation of LRH-1 remains elusive, it is conceivable that recognition of AA-containing PLs could promote the activity of LRH-1 in this context. Similarly, DHA-containing PLs may modulate LRH-1 activity in other inflammatory pathways, and PLs containing the energy-rich palmitic, stearic, and oleic acids may be involved in metabolic feedback via LRH-1.

PLs are unusual ligands in that they are insoluble, and instead form membranes and other supramolecular structures. In order for LRH-1 to bind PLs as observed in crystal structures, the PL must be extracted from the membrane. This extraction can be performed by a lipid transfer protein, such as PCTP, or by LRH-1 itself. While the normal function of PCTP is to transfer PCs between membranes, we hypothesized that it may also be responsible for extracting PCs from the plasma membrane and delivering them to LRH-1. Surprisingly, this did not seem to be the case (Figure 4.5 B-D), despite the demonstrated importance of PCTP function for LRH-1 activity (Figure 4.5 A). We have shown that LRH-1 is able to bind PLs of all head groups with comparable affinity, but only in the presence of β-cyclodextrin, which increases the fraction of soluble, monomeric PL. LRH-1 cannot extract PS, PA, PE, or SM at physiological concentrations in the absence of β-cyclodextrin, but PG, PI, and PC are readily extracted. Thus, the role of PCTP in LRH-1 regulation may be instead to deliver activating PCs to a pool of PLs that is accessible to LRH-1, from which LRH-1 directly extracts the PC. This may enable a means to control the access of LRH-1 to different PCs, via their differential affinity for PCTP. While LRH-1 seems to bind to most biologically relevant PCs indiscriminately, PCTP shows a preference for PCs with polyunsaturated tails (18). PCTP may therefore be able to promote the delivery of these PCs to LRH-1 *in vivo*, increasing the likelihood that LRH-1 encounters polyunsaturated over monounsaturated PCs, despite the lack of a significant difference in the affinity of these lipids for LRH-1 itself. Intriguingly, PCTP shows highest affinity for PCs containing AA (18). Taken together, the effect of PCTP inhibition on LRH-1 activity, the observation that LRH-1 bound AA-containing PCs with highest affinity (albeit by a modest margin, see figure 4.4 A), and the

knowledge that LRH-1 regulates the genes that convert AA to 15-HETE may point towards AA as a bona-fide LRH-1 regulatory ligand in the context of PCs. Future research should compare the coregulator binding profile of LRH-1 loaded with the PCs identified in this chapter in order to determine the effect that the fatty acyl tails of endogenous PCs have on driving differential coregulator recruitment. Furthermore, the identification of endogenous ligands bound to LRH-1 harvested from mammalian tissue remains an elusive but critically important goal.

References

1. Stein S, Schoonjans K. 2014. Molecular basis for the regulation of the nuclear receptor LRH-1. *Current opinion in cell biology* 33C:26-34
2. Chalkiadaki A, Talianidis I. 2005. SUMO-dependent compartmentalization in promyelocytic leukemia protein nuclear bodies prevents the access of LRH-1 to chromatin. *Mol Cell Biol* 25:5095-105
3. Stein S, Oosterveer MH, Mataki C, Xu P, Lemos V, et al. 2014. SUMOylation-dependent LRH-1/PROX1 interaction promotes atherosclerosis by decreasing hepatic reverse cholesterol transport. *Cell Metab* 20:603-13
4. Ortlund EA, Lee Y, Solomon IH, Hager JM, Safi R, et al. 2005. Modulation of human nuclear receptor LRH-1 activity by phospholipids and SHP. *Nature structural & molecular biology* 12:357-63
5. Lee JM, Lee YK, Mamrosh JL, Busby Sa, Griffin PR, et al. 2011. A nuclear-receptor-dependent phosphatidylcholine pathway with antidiabetic effects. *Nature*
6. Musille PM, Pathak MC, Lauer JL, Hudson WH, Griffin PR, Ortlund EA. 2012. Antidiabetic phospholipid-nuclear receptor complex reveals the mechanism for phospholipid-driven gene regulation. *Nat Struct Mol Biol* 19:532-7, S1-2
7. Krylova IN, Sablin EP, Moore J, Xu RX, Waitt GM, et al. 2005. Structural analyses reveal phosphatidyl inositols as ligands for the NR5 orphan receptors SF-1 and LRH-1. *Cell* 120:343-55
8. Sablin EP, Blind RD, Krylova IN, Ingraham JG, Cai F, et al. 2009. Structure of SF-1 bound by different phospholipids: evidence for regulatory ligands. *Mol Endocrinol* 23:25-34
9. Blind RD, Suzawa M, Ingraham HA. 2012. Direct modification and activation of a nuclear receptor-PIP(2) complex by the inositol lipid kinase IPMK. *Sci Signal* 5:ra44
10. Blind RD, Sablin EP, Kuchenbecker KM, Chiu HJ, Deacon AM, et al. 2014. The signaling phospholipid PIP3 creates a new interaction surface on the nuclear receptor SF-1. *Proc Natl Acad Sci U S A* 111:15054-9
11. Sablin EP, Blind RD, Uthayaruban R, Chiu HJ, Deacon AM, et al. 2015. Structure of Liver Receptor Homolog-1 (NR5A2) with PIP3 hormone bound in the ligand binding pocket. *J Struct Biol* 192:342-8
12. Musille PM, Kohn JA, Ortlund EA. 2013. Phospholipid – Driven gene regulation. *FEBS Letters*
13. Valencak TG, Ruf T. 2011. Feeding into old age: long-term effects of dietary fatty acid supplementation on tissue composition and life span in mice. *J Comp Physiol B* 181:289-98

14. Kelley DS, Bartolini GL, Newman JW, Vemuri M, Mackey BE. 2006. Fatty acid composition of liver, adipose tissue, spleen, and heart of mice fed diets containing t10, c12-, and c9, t11-conjugated linoleic acid. *Prostaglandins Leukot Essent Fatty Acids* 74:331-8
15. Venteclef N, Smith JC, Goodwin B, Delerive P. 2006. Liver receptor homolog 1 is a negative regulator of the hepatic acute-phase response. *Mol Cell Biol* 26:6799-807
16. Coste A, Dubuquoy L, Barnouin R, Annicotte JS, Magnier B, et al. 2007. LRH-1-mediated glucocorticoid synthesis in enterocytes protects against inflammatory bowel disease. *Proc Natl Acad Sci U S A* 104:13098-103
17. Lefevre L, Authier H, Stein S, Majorel C, Couderc B, et al. 2015. LRH-1 mediates anti-inflammatory and antifungal phenotype of IL-13-activated macrophages through the PPARgamma ligand synthesis. *Nat Commun* 6:6801
18. de Brouwer AP, Bouma B, van Tiel CM, Heerma W, Brouwers JF, et al. 2001. The binding of phosphatidylcholine to the phosphatidylcholine transfer protein: affinity and role in folding. *Chem Phys Lipids* 112:109-19

Chapter 5: Deciphering modern glucocorticoid cross-pharmacology using ancestral corticosteroid receptors

Summary

Steroid receptors (SRs) are the largest family of metazoan transcription factors, and control genes involved in development, endocrine signaling, reproduction, immunity, and cancer. The entire hormone-receptor system is driven by a molecular switch triggered by the binding of small lipophilic ligands (1; 2). This makes the SRs ideal pharmaceutical targets, yet even the best clinically approved synthetic steroidal agonists are prone to cross-reactivity and off-target pharmacology. The mechanism underlying this promiscuity is derived from the fact that SRs share common structural features derived from their evolutionary relationship. More often than not, rational attempts to probe SR drug selectivity via mutagenesis fail even when high quality structural and functional data is available, due to the fact that important mutations often result in nonfunctional receptor. This highlights the fact that steroid receptors suffer from instability preventing in depth mutational analysis and hampering crystallization of key receptor-ligand complexes. We have taken a unique approach to address this problem by using a resurrected ancestral protein to determine the structure of a previously intractable complex, identifying the structural mechanisms that confer activation and selectivity for a widely used glucocorticoid, mometasone furoate (MOF). Moreover, we identify a single residue located outside of the ligand binding pocket that controls MOF antagonism vs. agonism in the human mineralocorticoid receptor.

This chapter has been slightly modified from the published manuscript:

Kohn JA, Deshpande K, Ortlund EA. 2012. Deciphering modern glucocorticoid cross-pharmacology using ancestral corticosteroid receptors. *J Biol Chem* 287:16267-75

Introduction

Complex life depends on intra- and intercellular communication, whereby secreted messengers are detected by specific receptors in order to regulate metabolism, reproduction, cell cycles, and more. This coordination tightly controls cellular activity within the higher organism. Poor coordination of these processes can result in many health concerns, including metabolic disorders, reproductive diseases, and cancer. Over time, a vast repertoire of receptors has evolved to respond to small chemical stimuli, making them attractive pharmacological targets. However, since most receptors belong to large classes of evolutionary-related proteins that show high structural similarity, targeting a single receptor subtype is a major challenge. Poor selectivity can cause serious off-target side effects, as seen in the treatments of major depression (3), heart disease (4; 5), asthma (4; 5), and allergies (6).

To fully understand the mechanisms supporting receptor-ligand recognition, selectivity, and activation, robust structure-function relationships must be built from extensive mutational analysis and ligand design. This analysis is hindered for several reasons. First, amino acid residues conferring protein function and ligand specificity between homologous receptors can be difficult to identify among the vastly more prevalent neutral mutations that accumulate over time (7). Second, restrictive mutations that are not directly related to the protein-ligand interaction can accumulate in extant proteins, preventing the tolerance of function-shifting mutations (8). Third, many mutations are destabilizing and result in loss of protein function, complicating the distinction between an effect that is specific to the protein-ligand interaction *vs.* an effect that is globally inactivating to the protein (7). While most conclusions are currently drawn from function-*killing* mutations, the insight needed to understand ligand selectivity among a class of homologous proteins would be better drawn from function-*shifting* mutations that preserve receptor activation.

These problems have hindered the design of selective drugs that target human steroid receptors (SRs). SRs are a family of ligand regulated transcription factors that control genes

involved in development, endocrine signaling, reproduction, immunity, and cancer.(9) This makes them attractive pharmaceutical targets. While SRs show exquisite selectivity for their endogenous hormones, SR-targeting drugs tend to be promiscuous and cause many off-target side-effects (10; 11). This is because SRs – consisting of the estrogen receptor (ER), progesterone receptor (PR), androgen receptor (AR), mineralocorticoid receptor (MR), and glucocorticoid receptor (GR) – descended from a common ancestor >500 million years ago (Figure 5.1 A), and show high structural similarity (9; 12). In the absence of ligand, SRs remain partially unfolded and associate with heat shock proteins (13; 14). This instability is necessary to permit the conformational changes that drive receptor activation upon ligand binding (7; 14-16). but it has limited our ability to probe receptor-ligand interactions via mutagenesis, as many mutations of interest disable the protein entirely.

With the advancement of whole-gene synthesis and pioneering efforts made in computational and evolutionary biology, it is now possible to predict and "resurrect" ancestral genes. Ancestral gene reconstruction (AGR) is used to study the molecular evolution of a biological system (12; 17-19), but shows promising applications to the process of drug design. By comparing two ancestral proteins from nodes on an evolutionary tree, we are provided with a smaller subset of possible amino acid replacements to dissect between related proteins that have different ligand specificities. Our efforts can be focused on fewer residues when probing structure-function relationships than when looking only at extant proteins. This approach therefore allows us to avoid interference from neutral and restrictive mutations that have accumulated over time. Furthermore, unlike many extant proteins, ancestral proteins show remarkable tolerance towards changes in function-shifting residues making them more stable under laboratory conditions.

We hypothesized that one could exploit these ancestral proteins to understand cross-pharmacology in human steroid receptors. Ancestral SRs are more tolerant to mutation than their extant descendants (20), and their molecular and structural evolution have already been

characterized (8; 12; 20-23). AncSRs therefore make an effective model to study the structural mechanisms of SR pharmacology. To achieve this goal, we determined the structure of the ancestral glucocorticoid receptor 2 (AncGR2) ligand binding domain (LBD) in complex with a fragment of human transcription intermediary factor 2 (TIF-2) and mometasone furoate (MOF). We draw upon functionally important historical amino acid substitutions to elucidate the mechanisms driving GR activation for this widely used glucocorticoid. Furthermore, we use a combination of structural analysis and functional assays to explain the selectivity of this drug against MR and AR and strong cross-reactivity with PR.

Experimental Procedures

Protein expression and purification

AncGR2 (Genbank accession: EF631976.1) in a pMALCH10T vector was transformed into *Escherichia coli* strain BL21 (DE3) and expressed as an MBP-His fusion. Cultures (1.3 L in TB) were grown to an OD_{600} of 0.6-0.7 and induced with 400 μM IPTG and 50 μM mometasone furoate at 30°C for 4 h. Cell mass was collected by centrifugation at 4 krpm for 15 minutes, lysed, and purified by nickel affinity chromatography. The MBP-His tag was cleaved by TEV protease at 4°C overnight with simultaneous dialysis into a buffer containing 300 mM NaCl/20 mM tris pH 7.4/5% glycerol, and purified to homogeneity by nickel affinity followed by gel filtration chromatography.

Crystallization, data collection, structural refinement

Pure AncGR2 was concentrated to 3-5 mg mL^{-1} in a buffer containing 300 mM NaCl/20 mM tris pH 7.4/5% glycerol/50 μM CHAPS/50 μM MOF. Crystals were grown via hanging drop vapor diffusion at 4°C from solutions containing 0.75 μL AncGR2-TIF2-MOF solution, 0.75 μL crystallant (1.5-3 M ammonium formate), and a dodecapeptide derived from the glucocorticoid receptor coactivator human TIF2 ($^{+}H_3N$-ENALLRYLLDKD-CO_2^{-}, Synbiosci). Crystals were cryoprotected by immersion in crystallant containing 25% glycerol and flash frozen in liquid

nitrogen. Data to a resolution of 2.5 Å were collected at the South East Regional Collaborative Access Team at the Advanced Photon Source (Argonne, IL). The structure of the AncGR2/MOF/TIF2 complex was solved by molecular replacement using PHASER in the CCP4 software suite. Model building and refinement were performed using Refmac and COOT. Cavity volumes were calculated using CASTp and figures were generated in PyMOL. The refined AncGR2-MOF structure has been deposited into the PDB (accession number 4E2J).

Mutagenesis

Wild-type AncGR1.1 and AncGR2 were subcloned into a pMCSG7-MBP-His expression vector, and the following mutations were created from these constructs: AncGR1.1-S106P, AncGR1.1-S106P+L111Q (AncGR1.1-PQ), AncGR1.1-S106P+L111A (AncGR1.1-PA), AncGR2-P106S, AncGR2-P106S+Q111L (AncGR2-SL), and AncGR2-P106S+Q111A (AncGR2-SA). All mutagenesis was performed using Quikchange II XL (Stratagene).

Ligand binding assays

Wild-type or mutant AncGR1 or AncGR2 were expressed as above, and assayed prior to TEV cleavage as purified MBP fusion proteins. All FP experiments were performed in a buffer containing 150 mM NaCl, 10 mM HEPES pH 7.4, 5 mM DTT 3 mM EDTA and 0.005% Tween-20. Binding affinity to dexamethasone-fluorescein (DM) was measured with a constant concentration of 12 nM DM and variable protein concentration of 10^{-10} – 10^{-5} M. Competition assays were performed at a protein concentration 1.2 times its binding affinity to DM, in the presence of 12 nM DM and 10^{-10} – 10^{-5} M competing ligand. Data was processed in Prism 5 (La Jolla, CA). Statistical significance was determined by two-factor ANOVA, and individual comparisons were made with Tukey HSD post-hoc tests.

In-cell activation assays

All human, ancestral, and mutant LBDs were cloned into a pSG5 expression vector immediately following a GAL4-DBD and a GR hinge sequence. CHO-K1 cells were grown and maintained in phenol red free complete αMEM (Gibco) supplemented with 10% charcoal-dextran

stripped FBS (Gibco) and penicillin/streptomycin. Cells grown in 96-well assay plates were transfected at 70-90% confluence with 1 ng receptor, 100 ng UAS-driven firefly luciferase reporter (pFRluc), and 0.1 ng constitutive *Renilla* luciferase reporter (phRLtk), for 4 h using Lipofectamine 2000 in OPTIMEM (Invitrogen). Transfections were ended by replacement with complete αMEM, and cells were allowed to recover overnight. After recovery, cells were treated in triplicate with 10^{-12} – 10^{-6} M ligand or vehicle (DMSO) in complete αMEM for 24 h (final working DMSO 1%), then assayed with Dual-Glo luciferase substrate (Promega). Firefly activity was normalized to *Renilla* activity, and fold increase in activation was calculated relative to vehicle control. Dose-response curves were generated in Graphpad Prism 5 (La Jolla, CA). Statistical significance was determined by two-factor ANOVA, and individual comparisons were made with Tukey HSD post-hoc tests.

Results

AncGR2-TIF2-MOF crystal structure

MOF is a powerful topical anti-inflammatory drug for the skin and airways and is the active ingredient of Nasonex, Asmanex, and Elocon (24). While MOF has been in clinical use for over 24 years, it suffers from severe cross-pharmacology resulting in unwanted side effects, limiting its use to topical applications. MOF strongly activates GR, cross reacts with PR, and is selective against AR and MR. The ternary AncGR2-TIF2-MOF crystal structure reveals the structural basis for MOF binding to vertebrate GR's (Figure 5.1 B-E; Table 5.1). The hydrogen bond network that is required for activation of corticoid receptors (25) is intact, stabilized by a dipole-dipole interaction between MOF C21-Cl and AncGR2 N33. MOF binding requires a rearrangement of the H6-H7 region of the receptor to accommodate the large 17α furoate moiety, inducing a 200 $Å^3$ (1.3x) increase in the volume of the ligand-binding pocket relative to dexamethasone; this highlights the ability of SRs to expand their ligand binding pockets to accommodate exogenous ligands (26; 27). The AncGR2-TIF2-MOF structure also reveals that a

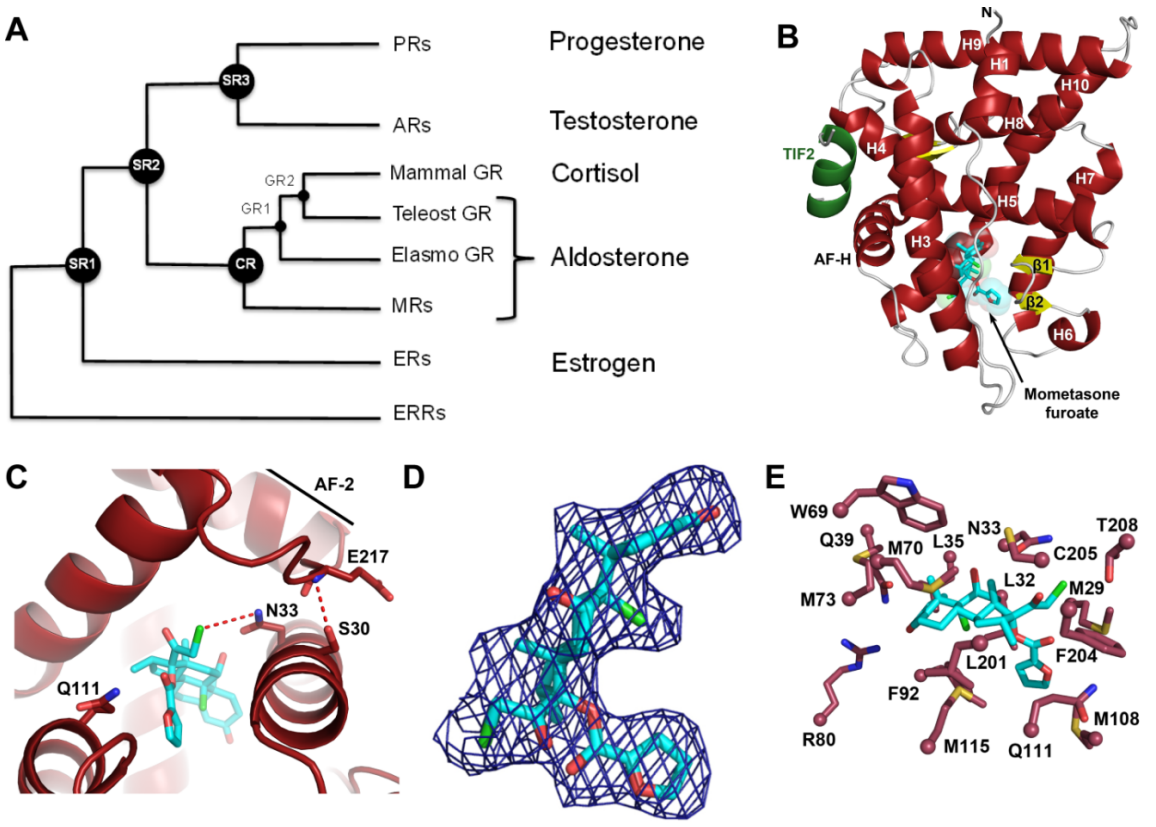

Figure 5.1. Evolutionary history of corticosteroid receptors and structure of AncGR2—MOF—TIF2 complex.

(A) Simplified phylogenetic tree depicting the evolution of corticosteroid receptors; activating hormones listed at right. (B) Structure of the Ancestral Glucocorticoid Receptor 2 (red) in complex with human TIF2 (green) and mometasone fuorate (MOF; cyan). (C) AncGR2 ligand binding pocket residues (red) with MOF shown (cyan). (D) 2Fo-Fc omit map of MOF in the AncGR2 LBP. (E) AncGR2 LBP residues within 4.2 Å of MOF.

Table 5.1. Data collection and refinement statistics for the novel AncGR2—MOF—TIF2 complex

Resolution (highest shell), Å	2.50 Å (2.59 – 2.50)
Space Group	$P6_1$
Unit Cell Dimensions	
a, b, c (Å)	a=104.4, b=104.4, c=143.9
α, β, γ (°)	α=β=90.0°, γ= 120.0°
No. of Reflections	30710
[a]Rsym (highest shell)	7.7% (42.2%)
Completeness (highest shell)	99.90% (98.96%)
Ave. Redundancy (highest shell)	8.0 (7.9)
I/σ	29.3 (5.3)
Monomers per asymmetric unit (AU)	2
No. of protein atoms/AU	4195
No. of ligand atoms/AU (+GOL+FMT)	85
No. of waters/AU	151
[b]$R_{working}$ ([c]R_{free})	20.5 (25.5)
Ave. B-factors, Å^2	
Protein	45.0
Ligand	53.5
Water	45.5
R.m.s. deviations	
Bond lengths, Å	0.005
Bond angles, °	1.078
PDB ID	4E2J

[a] $R_{sym} = \Sigma|I- \langle I \rangle|/ \Sigma|I|$, where I is the observed intensity and $\langle I \rangle$ is the average intensity of several symmetry-related observations.

[b] $R_{working} = \Sigma||F_o|-|F_c||/ \Sigma|F_o$, where F_o and F_c are the observed and calculated structure factors, respectively.

[c] $R_{free} = \Sigma||F_o|-|F_c||/ \Sigma|F_o$ for 7% of the data not used at any stage of the structural refinement.

strong H-bond is not possible between MOF and Q111 of GR (Figure 5.1 C) – an interaction that plays a critical role in the specific recognition of 17α-OH substituted ligands and is absolutely required for cortisol activation (8). Instead, hydrophobic interactions replace this interaction in a fashion analogous to the structure of the GR fluticasone furoate complex (28).

Structural and evolutionary basis for PR cross-reactivity

The AncGR2-TIF2-MOF structure allows for the direct structural comparison of GR-MOF and PR-MOF complexes, and reveals how additional space in the H6-H7 loop region is created to accommodate the furoate moiety. PR residues 791-ESSF-794 on H7 appear to play a key role in allowing strong MOF binding by expanding the pocket via a conserved E791-S793 H-bond between the H6-H7 loop and H7 (Figure 5.2 A). Steric bulk provided by PR residue F794 between helices 7 and 3 maintains space for the 17α furoate moiety and contributes a hydrophobic interaction via the aromatized side chain (26). This motif is strictly conserved among PRs but is not present in AncSR2 (the common ancestor of all 3-keto SRs) or AncSR3 (the common ancestor of PR and AR) (Figure 5.1 A, 5.2 B). Therefore, PR response to MOF was probably a late evolutionary derivation resulting in this cross-reactivity.

Structural and evolutionary basis for selectivity against MR

Our structure also suggests a mechanism for the selectivity of MOF against MR and AR. Helices 6 and 7, which border the 17α binding area, are partially unwound and stabilized by P637/106 in GR/AncGR2, accommodating the furoate moiety (Figure 5.3 A, left); MR, AR, and AncGR1 have a serine at this site that caps H7, positioning the helix within 2.5 Å of where the furoate would rest, creating a steric incompatibility (Figure 5.3 A, right). We have shown in previous work that, during the evolution of GR, S106P and L111Q substitutions were critical in both reshaping the H6-H7 region of the receptor and in generating a new H-bond with the 17-OH moiety of cortisol, the endogenous glucocorticoid. To test the effect of reversing these critical substitutions with respect to MOF binding affinity we generated two AncGR2 mutants: P106S

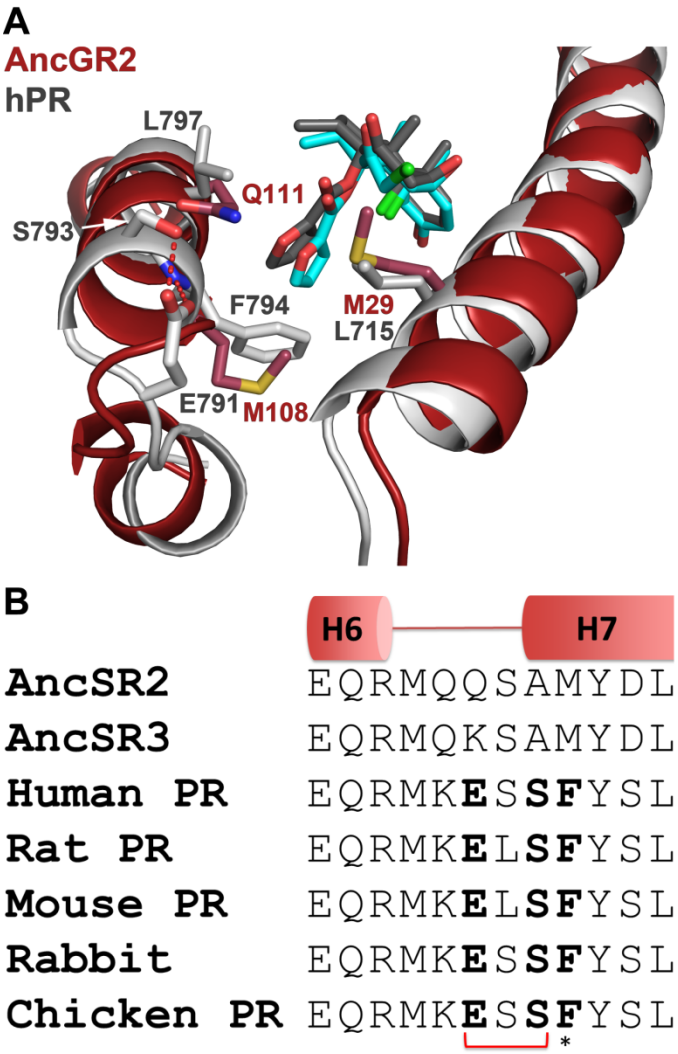

Figure 5.2. Structural basis for off-target activation of PR.

(A) The human PR-MOF complex (1SR7; white) superimposed on the AncGR2-MOF complex (red). PR residue F794 maintains space for the 17α furoate moiety and contributes a hydrophobic interaction via the aromatized side chain (15). PR residues 791-ESSF-794 on H7 appear to play a key role in allowing strong MOF binding by positioning the H6-H7 loop and H7 via a conserved E791-S793 H-bond. (B) This motif is strictly conserved among extant PRs but is not present in AncSR2.

and P106S/Q111L (AncGR2-SL), and measured their binding affinities for cortisol, dexamethasone, and MOF using fluorescence polarization competition assays against dexamethasone-fluorescein (DF). The P106S reversal reduces the affinity of AncGR2 for all three ligands by only one order of magnitude (Figure 5.3 B). This result was surprising since the P106S mutation wasidentified as the driving force behind the H6-H7 rearrangement required to open space in the receptor for specific recognition of hormones with C17 substituents (13; 22). Since MOF binding requires this structural rearrangement (Figure 5.3 A), P106 likely plays a role in stabilizing the H6-H7 loop in a productive binding mode but is not absolutely required to induce this structural change. AncGR2-SL, which is known to be inactive to endogenous ligands (8), did not bind DF (Figure 5.5 F). This prevented competition assays on this mutant, but suggested that H7 indeed repositions to place L111 in contact with the C17 position of the steroid. This generates a polar incompatibility with C17-OH containing steroids such as cortisol and dexamethasone and introduces a steric clash with the bulky furoate substituent of MOF. To test this hypothesis we generated a P106S/L111A mutant (AncGR2-SA), designed to alleviate this steric clash in the P106 background, which restored binding to MOF and dexamethasone (Figure 5.3 B). As expected, cortisol binding was only marginally restored since it does not contain the additional bulky hydrophobic group present on MOF to stabilize the core of the receptor in the absence of the critical Q111 – 17 OH H-bond. Interestingly dexamethasone binding was more fully restored than cortisol, presumably due to additional interactions on its modified backbone. Thus, the H6-H7 region of the receptor can adopt an expanded conformation in the absence of P106, suggesting that the H6-H7 region of GRs is inherently flexible, allowing it to adapt to ligand-induced perturbation. This reshapes our understanding of the role of the H6-H7 region within the LBD in the recognition of synthetic glucocorticoids.

We have shown previously that AncGR1, which preceded the evolution of AncGR2, is a low sensitivity MR-like receptor with activation by both mineralocorticoids and glucocorticoids (22). Since MOF is selective against MR we reasoned that MOF would display similar selectivity

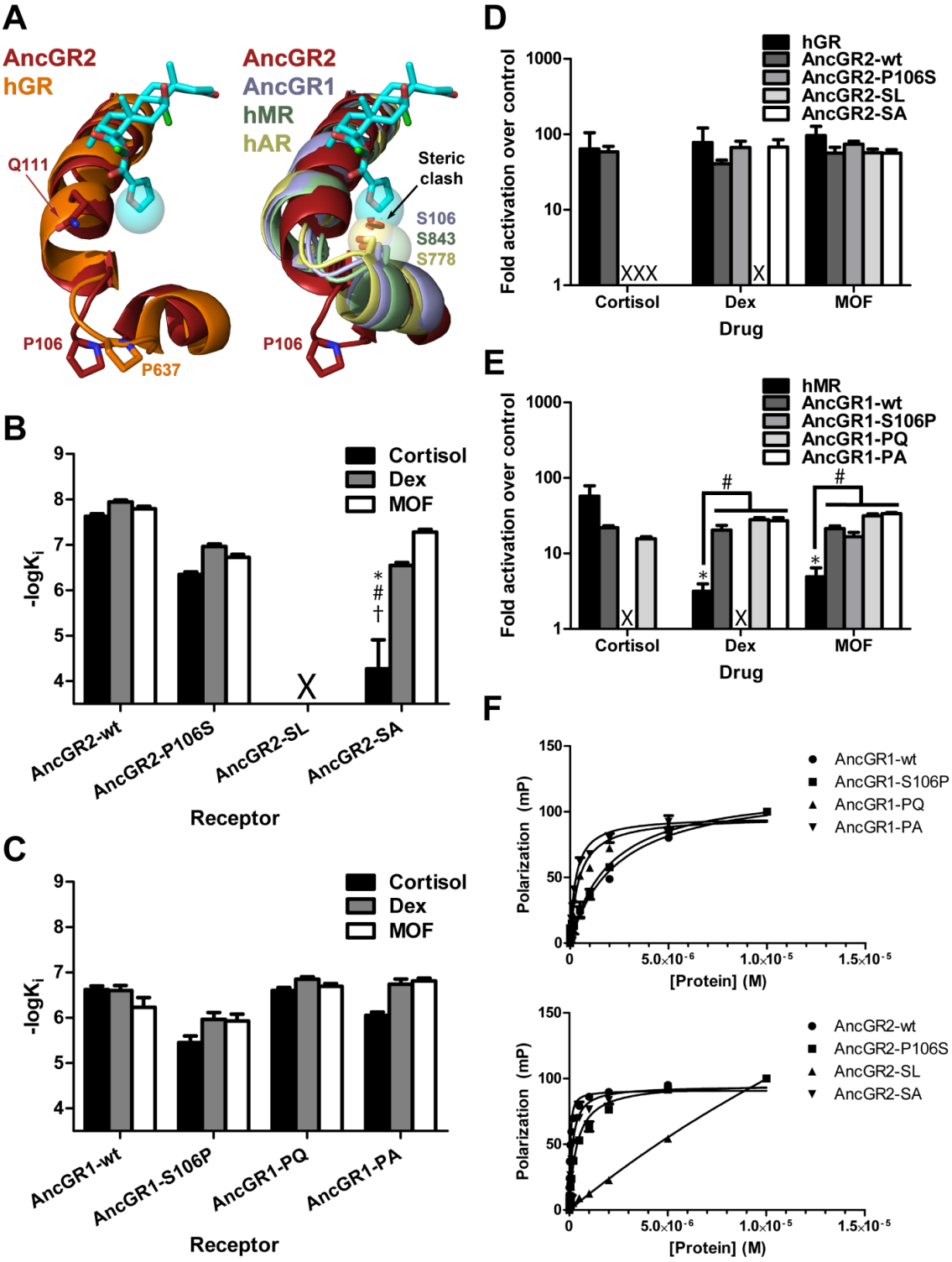
A
AncGR2
hGR
Q111
P106
P637
AncGR2
AncGR1
hMR
hAR
Steric clash
S106
S843
S778
P106
B
Cortisol
Dex
MOF
-logKi
AncGR2-wt
AncGR2-P106S
AncGR2-SL
AncGR2-SA
Receptor
C
Cortisol
Dex
MOF
-logKi
AncGR1-wt
AncGR1-S106P
AncGR1-PQ
AncGR1-PA
Receptor
D
hGR
AncGR2-wt
AncGR2-P106S
AncGR2-SL
AncGR2-SA
Fold activation over control
XXX
X
Cortisol
Dex
MOF
Drug
E
hMR
AncGR1-wt
AncGR1-S106P
AncGR1-PQ
AncGR1-PA
Fold activation over control
Cortisol
Dex
MOF
Drug
F
AncGR1-wt
AncGR1-S106P
AncGR1-PQ
AncGR1-PA
Polarization (mP)
[Protein] (M)
AncGR2-wt
AncGR2-P106S
AncGR2-SL
AncGR2-SA
Polarization (mP)
[Protein] (M)

Figure 5.3. Binding of MOF by modern and ancestral SRs requires expansion of the ligand binding pocket.

(A) Structure of the 17α binding pocket. Like PR and AncGR2, GR has an extended H6-7 loop conformation (left); MR, AR, and AncGR1 have a tightened H6-7 that would create a steric incompatibility with the furoate (right). (B, C) Binding affinities of AncGR2 (B) and AncGR1 (C) mutants to the indicated ligand were measured by fluorescence polarization competition with dexamethasone-fluorescein. AncGR2-SL did not bind dexamethasone-fluorescein, and competition experiments could not be performed for this receptor. Statistical analyses were performed using two-factor ANOVA with Tukey HSD post-hoc tests used for individual comparisons; comparisons found to be statistically significant to $p < 0.05$ are marked (*, compared to same ligand binding for wild-type receptor; #, compared to dexamethasone binding for same mutant; †, compared to MOF binding for same mutant). (D, E) Receptor activation for GR-like (D) and MR-like (E) receptors was measured by dual luciferase reporter gene activation in transiently transfected CHO-K1 cell cultures. Mean ± S.E.M. shown, n = 3. Statistical analyses were performed using two-factor ANOVA with Tukey HSD post-hoc tests used for individual comparisons; comparisons found to be statistically significant to $p < 0.05$ are marked (*, compared to activation of the same receptor by cortisol; #, comparisons made as marked on figure). X – No binding or activation observed. (F) Binding of dexamethasone-fluorescein probe to AncGR1 and AncGR2 wt and mutants.

against AncGR1. Surprisingly, MOF bound AncGR1 with an affinity comparable to dexamethasone and cortisol (Figure 5.3 C), indicating that the AncGR1 H7 had already acquired the plasticity needed to accommodate the bulky furoate moiety. The forward mutations AncGR1-S106P and S106P/L111Q (AncGR1-PQ) had no effects specific to a particular ligand, but the S106P/L111A mutations (AncGR1-PA) selectively reduced cortisol binding, while leaving dexamethasone and MOF binding unaffected. This is presumably due to the removal of a 17α interaction. These data show that receptor-ligand interactions at the 17α site are important for effective ligand binding, though poor interactions here can be surmounted by stronger interactions elsewhere along the ligand scaffold.

A single residue controls MOF selectivity and transcriptional activity

To determine the structural differences between GR and MR that govern MOF recognition, we characterized the ability of MOF to drive luciferase reporter gene activation across the entire ancestral corticosteroid phylogeny. While MOF only very weakly activates MR (Figure 5.4 A), it strongly activates AncGR1 and AncCR with a sub-nanomolar potency, comparable to the strong activation seen in AncGR2 and GR (Figure 5.4 A, B). This provides further evidence that the corticoid receptors from AncCR to AncGR2 are able to unwind H6-H7 to accommodate the MOF 17α furoate moiety without requiring the S106P substitution.

Furthermore, we show that MOF exhibits its selectivity for GR over MR not via a difference in potency, but rather in efficacy: while MOF binds MR, with a ~100 nM potency, it is unable to stabilize an active receptor conformation (25). Thus, MR must have accumulated epistatic changes that prohibit activation from this drug. We therefore examined the importance of residues that changed on the lineage leading to MR with respect to MOF activation. Mutation of residues in or near the ligand binding pocket had no significant impact on MOF activation without affecting receptor activation towards cortisol and dexamethasone, consistent with our FP competition assays (Figure 5.3 D, E, and data not shown). We therefore looked for changes outside of the ligand binding pocket and activation function surface. Residues R116 and Q120 in

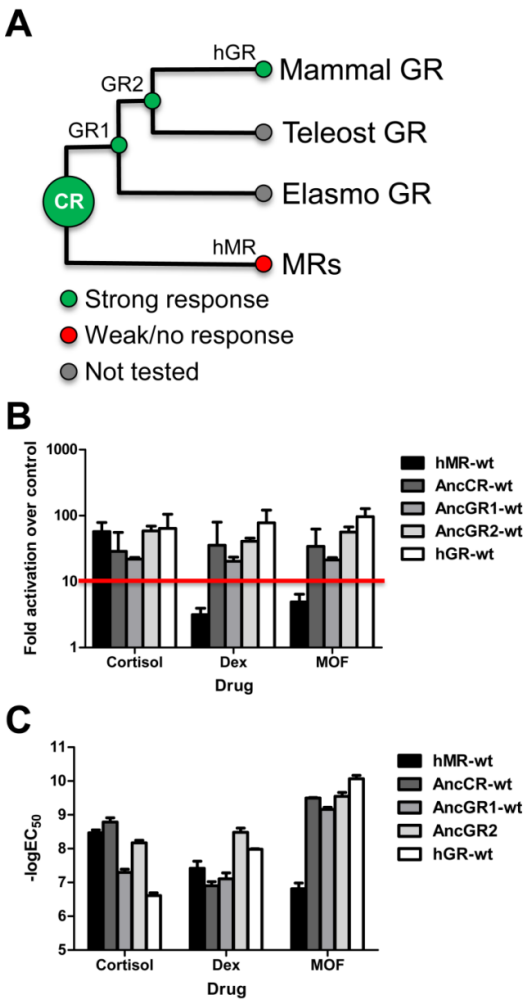

Figure 5.4. Activation of modern and ancestral corticosteroid receptors by synthetic glucocorticoids.

(A) Corticoid receptor phylogeny with response to synthetic glucocorticoids (dex/MOF) shown. Full agonism is shown in green, weak or no agonism shown in red. Human GR and MR were used to represent extant mammalian GR and MR. (B) Fold activation and (C) potency of corticosteroid receptor LBDs was measured via dual luciferase reporter gene activation in transiently transfected CHO-K1 cells. Mean ± S.E.M. shown, n = 3. For the purpose of this research, activation below 10 fold over control (B, red line) was considered weak agonism/antagonism, while activation above this threshold was considered full agonism.

AncCR, corresponding respectively to H853 and L857 in MR, are located on the solvent exposed face of H7 and interact with the main chain of the loop between H5 and β1, at AncCR residues G87 and M89 (MR residues S824 and F826) (Figure 5.5 A). These residues are ~14-18 Å from the furoate moiety of MOF and closest to the B ring of the steroid (10-16Å), yet, reversal of all four residues in MR to their ancestral states (MR-GMRQ) completely restored MOF activation (Figure 5.5 B). We further narrowed down the cause of this effect to the single residue at MR site 853. Reversal of this site via the substitution H853R confers full MOF activation (Figure 5.5 B). In wild-type MR, H853 interacts with the main chain atoms in the H5-β1 loop and appears to stabilize the MR-like configuration of H7, which must unwind to support activation by ligands with bulky C17α substituents. The stronger interaction provided by an arginine substitution at this site stabilizes MR-H853R to enable MOF activation (Figure 5.5 A, 5.6). Importantly, these changes are neutral with respect to activation by cortisol: neither the EC_{50} nor activation of cortisol is affected by the MOF selectivity mutations (Figure 5.5 B, C), indicating that the structural determinants of MOF activation are unique from those that support the endogenous ligand recognition. Introducing the equivalent forward substitutions in AncCR (AncCR-SFHL) failed to abrogate MOF response (Figure 5.5 B), which is in line with the more promiscuous phenotype of the ancestral protein. Intriguingly, making the equivalent site mutations horizontally between MR and GR (MR-H853L/L857S and GR-L647H/S651L) not only failed to enable MOF activation in MR, but also abrogated activation by cortisol and dexamethasone in GR (Figure 5.5 D, E). Mutations at these residues during the evolution of glucocorticoid receptors were previously identified to be destabilizing to GRs, contributing to the low affinity but high selectivity of modern GRs to endogenous glucocorticoids (20). Here, disruption of this site in GR fully destabilized the active receptor during cortisol and dexamethasone binding. MOF, in contrast, expands the LBP to make additional hydrophobic contacts offered by the furoate ring (Figure 5.6), and is able to stabilize the active conformation, albeit at a much lower potency than in wild-type GR (Figure 5.5 D, E).

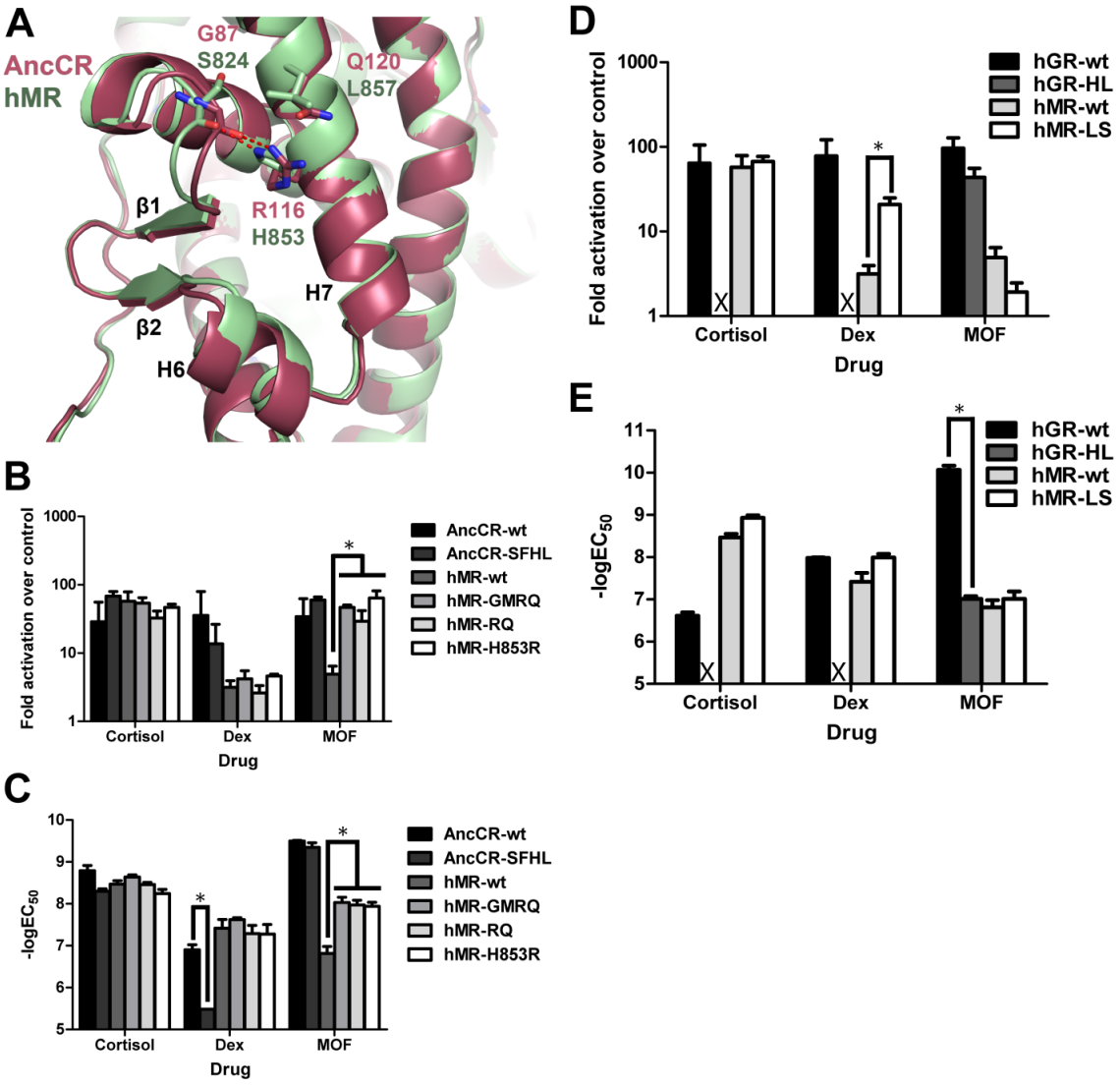

Figure 5.5. Distal residues control corticosteroid specificity.

(A) Key residues preceding β-sheet 1 and helix 7 in AncCR and MR were cross-mutated. (B) Fold activation and (C) potency of was measured via dual luciferase reporter gene activation in transiently transfected CHO-K1 cells. The same residues in GR and MR were cross-mutated, and (D) fold activation and (E) potency was measured via dual luciferase reporter gene activation in transiently transfected CHO-K1 cells. Mean ± S.E.M. shown, n = 3. Statistical analyses were performed using two-factor ANOVA with Tukey HSD post-hoc tests used for individual comparisons; comparisons found to be statistically significant to $p < 0.05$ are marked (*).

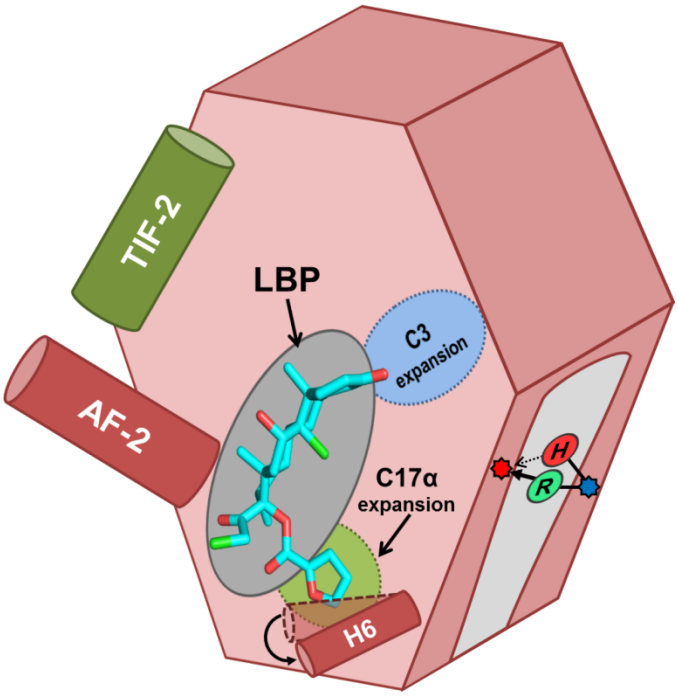

Figure 5.6. Cartoon schematic summarizing the relevant features of the SR LBP that dictate MOF activation.

Activation occurs when a ligand (*e.g.* MOF, cyan) binds to the LBP, stabilizing the AF-2 helix (dark red) to allow for coactivator binding (*e.g.* TIF-2, dark green) and subsequent transcriptional control. In glucocorticoid and progesterone receptors, the LBP can expand to accommodate steroids that are substituted at carbons 3 (blue) or 17α (light green). In the mineralocorticoid receptor, we identified a single site outside of the LBP that can toggle MOF agonism vs. antagonism, ostensibly by forming a bridge between H7 (blue dot) and the H5-β1 loop (red dot). In wild-type MR, H853 makes a weak hydrogen bond that cannot support MOF activation (red H); the historical substitution to a positively-charged arginine (green R) strengthens this interaction, restoring activation.

We anticipate that the findings produced by this study will be applicable to ligands that protrude into to extrasteroidal binding regions within the LBP. Furthermore, the finding that ligand specificity is strongly influenced by structural features that lie well outside of the LBP must be taken into consideration during the development of future drugs. The fact that these sites could not be identified using extant proteins highlights the power of using AGR to identify the obscure conserved structural mechanisms that support activation via endogenous vs. synthetic ligands that may be exploited by selective therapeutics

Discussion

We have successfully adapted AGR to shed light on the structural mechanisms of drug selectivity for SRs. Our approach combines structural and evolutionary biology to overcome many of the obstacles that frequently hinder protein research using modern proteins. It is well known that function shifting amino acid changes are not tolerated well in modern proteins, because most proteins are only moderately stable (7; 15; 16). They display a narrow thermal window of activity dictated by the effects of natural selection on both thermal and kinetic stability (15) and by the accumulation of neutral mutations over evolutionary time (7). A fine balance is necessary to allow small perturbations or signals, such as ligand binding, to functionally alter protein structure: while too little stability prevents proper protein folding, too much stability prevents a receptor from adopting an active conformation in response to stimuli within the host organism. We are therefore limited by the effects of both natural selection and neutral drift, as we are left with mesophilic proteins to use for structure-function analysis. This is exemplified in the SR family, and in particular with modern GRs, which are notoriously difficult to manipulate under laboratory conditions (23; 29). Furthermore, modern proteins have accumulated millions of years of neutral mutations that make it difficult to identify functionally important amino acid residues, as well as restrictive mutations that can further prohibit mutational analysis.

Workarounds to these problems are limited, and frequently involve the incorporation of stabilizing mutations. While this approach does improve the stability of modern proteins, including GR (23; 29), mutations such as these may alter the way ligands interact with their target receptors. As a result, the behavior of these mutants may not accurately mirror the behavior of the wild-type proteins. In contrast, ancestral proteins are subjected to rigorous testing during the reconstruction process to ensure their behavior is consistent with the behavior of other proteins within their phylogeny. (*e.g.* that the structural mechanisms for activation is conserved). Ancestral proteins are inherently more tolerant to mutation and may serve as ideal models in which to study structure activity relationships for moderately stable eukaryotic proteins (8; 22; 23). Even when the resurrection of an entire protein is not feasible, the insertion of ancestral residues in modern proteins can increase stability, enhance adaptability and tolerance to mutations (30). In addition, we have found that ancestral proteins tend to be more promiscuous to synthetic ligands or drug activation, especially in cases where the ancestral proteins display a more promiscuous phenotype than the extant proteins for endogenous ligands. Thus, resurrected proteins may permit the crystallization and functional analysis of previously intractable complexes due to their enhanced stability and promiscuity.

We show that by mirroring what has been done in evolutionary studies aimed at discovering the structural mechanism that conferred hormone selectivity, ancestral proteins may be used to examine cross-pharmacology among homologous proteins. The advantages of using ancestral proteins to study the structural mechanisms of drug promiscuity lie not only in their enhanced stability, but also in locating the structural features that contribute to differences in ligand recognition. AGR therefore provides an elegant solution to some of the troubling problems that currently interfere with the process of drug design.

References

1. Nagy L, Schwabe JW. 2004. Mechanism of the nuclear receptor molecular switch. *Trends Biochem Sci* 29:317-24

2. Kallenberger BC, Love JD, Chatterjee VK, Schwabe JW. 2003. A dynamic mechanism of nuclear receptor activation and its perturbation in a human disease. *Nat Struct Biol* 10:136-40
3. Remick RA. 1988. Anticholinergic side effects of tricyclic antidepressants and their management. *Prog Neuropsychopharmacol Biol Psychiatry* 12:225-31
4. Decalmer PB, Chatterjee SS, Cruickshank JM, Benson MK, Sterling GM. 1978. Beta-blockers and asthma. *Br Heart J* 40:184-9
5. Hanania NA, Singh S, El-Wali R, Flashner M, Franklin AE, et al. 2008. The safety and effects of the beta-blocker, nadolol, in mild asthma: an open-label pilot study. *Pulm Pharmacol Ther* 21:134-41
6. Roy M, Dumaine R, Brown AM. 1996. HERG, a primary human ventricular target of the nonsedating antihistamine terfenadine. *Circulation* 94:817-23
7. Taverna DM, Goldstein RA. 2002. Why are proteins marginally stable? *Proteins* 46:105-9
8. Bridgham JT, Ortlund EA, Thornton JW. 2009. An epistatic ratchet constrains the direction of glucocorticoid receptor evolution. *Nature* 461:515-9
9. Carson-Jurica MA, Schrader WT, O'Malley BW. 1990. Steroid receptor family: structure and functions. *Endocr Rev* 11:201-20
10. Stahn C, Lowenberg M, Hommes DW, Buttgereit F. 2007. Molecular mechanisms of glucocorticoid action and selective glucocorticoid receptor agonists. *Mol Cell Endocrinol* 275:71-8
11. Madauss KP, Stewart EL, Williams SP. 2007. The evolution of progesterone receptor ligands. *Med Res Rev* 27:374-400
12. Thornton JW. 2001. Evolution of vertebrate steroid receptors from an ancestral estrogen receptor by ligand exploitation and serial genome expansions. *Proc Natl Acad Sci U S A* 98:5671-6
13. Bledsoe RK, Montana VG, Stanley TB, Delves CJ, Apolito CJ, et al. 2002. Crystal structure of the glucocorticoid receptor ligand binding domain reveals a novel mode of receptor dimerization and coactivator recognition. *Cell* 110:93-105
14. Gee AC, Katzenellenbogen JA. 2001. Probing conformational changes in the estrogen receptor: evidence for a partially unfolded intermediate facilitating ligand binding and release. *Mol Endocrinol* 15:421-8
15. Godoy-Ruiz R, Ariza F, Rodriguez-Larrea D, Perez-Jimenez R, Ibarra-Molero B, Sanchez-Ruiz JM. 2006. Natural selection for kinetic stability is a likely origin of correlations between mutational effects on protein energetics and frequencies of amino acid occurrences in sequence alignments. *J Mol Biol* 362:966-78
16. DePristo MA, Weinreich DM, Hartl DL. 2005. Missense meanderings in sequence space: a biophysical view of protein evolution. *Nat Rev Genet* 6:678-87
17. Gaucher EA, Thomson JM, Burgan MF, Benner SA. 2003. Inferring the palaeoenvironment of ancient bacteria on the basis of resurrected proteins. *Nature* 425:285-8
18. Jermann TM, Opitz JG, Stackhouse J, Benner SA. 1995. Reconstructing the evolutionary history of the artiodactyl ribonuclease superfamily. *Nature* 374:57-9
19. Wouters MA, Liu K, Riek P, Husain A. 2003. A despecialization step underlying evolution of a family of serine proteases. *Mol Cell* 12:343-54
20. Carroll SM, Ortlund EA, Thornton JW. 2011. Mechanisms for the evolution of a derived function in the ancestral glucocorticoid receptor. *PLoS Genet* 7:e1002117
21. Carroll SM, Bridgham JT, Thornton JW. 2008. Evolution of hormone signaling in elasmobranchs by exploitation of promiscuous receptors. *Mol Biol Evol* 25:2643-52
22. Ortlund EA, Bridgham JT, Redinbo MR, Thornton JW. 2007. Crystal structure of an ancient protein: evolution by conformational epistasis. *Science* 317:1544-8

23. Bridgham JT, Carroll SM, Thornton JW. 2006. Evolution of hormone-receptor complexity by molecular exploitation. *Science* 312:97-101
24. Tan RA, Corren J. 2008. Mometasone furoate in the management of asthma: a review. *Ther Clin Risk Manag* 4:1201-8
25. Bledsoe RK, Madauss KP, Holt JA, Apolito CJ, Lambert MH, et al. 2005. A ligand-mediated hydrogen bond network required for the activation of the mineralocorticoid receptor. *Journal of Biological Chemistry* 280:31283-93
26. Madauss KP, Deng SJ, Austin RJ, Lambert MH, McLay I, et al. 2004. Progesterone receptor ligand binding pocket flexibility: crystal structures of the norethindrone and mometasone furoate complexes. *J Med Chem* 47:3381-7
27. Nettles KW, Bruning JB, Gil G, O'Neill EE, Nowak J, et al. 2007. Structural plasticity in the oestrogen receptor ligand-binding domain. *EMBO Rep* 8:563-8
28. Biggadike K, Bledsoe RK, Hassell AM, Kirk BE, McLay IM, et al. 2008. X-ray crystal structure of the novel enhanced-affinity glucocorticoid agonist fluticasone furoate in the glucocorticoid receptor-ligand binding domain. *J Med Chem* 51:3349-52
29. Seitz T, Thoma R, Schoch GA, Stihle M, Benz J, et al. 2010. Enhancing the stability and solubility of the glucocorticoid receptor ligand-binding domain by high-throughput library screening. *J Mol Biol* 403:562-77
30. Yamashiro K, Yokobori S, Koikeda S, Yamagishi A. 2010. Improvement of Bacillus circulans beta-amylase activity attained using the ancestral mutation method. *Protein Eng Des Sel* 23:519-28

Chapter 6: Discussion

Conclusions

Phospholipids are a novel class of ligands

While the roles of PLs as structural molecules and reservoirs for intermediates in cell signaling cascades has been recognized for quite some time, the idea that intact PLs themselves serve as regulatory ligands is much more novel. To date, PLs have been identified as modulating ligands in three NRs (SF-1, LRH-1, and PPARα) (1-4), and have been observed bound to USP in X-ray crystal structures (5), though its role as a USP regulatory ligand is yet unknown. In contrast to most PL-binding receptors, which recognize the PL headgroup in the context of a lipid bilayer (6), NRs bind to PLs in a unique orientation, fully engulfing the hydrophobic tails and presenting the headgroup on the surface of the protein (7). The NR LBD makes direct contacts with the PL at the glycerol phosphate, along the whole length of the fatty acyl tails, and in the case of LRH-1—PIP_3, the inositol phosphate head (4; 7; 8). In addition, the size and charge of the head group affects the width of the mouth of the LBP (*i.e.* the space between the H5—β-sheet—H7 loop and the bottom of H3), and the overall electrostatic surface of the LBD, even in the absence of direct interaction with the receptor (as is the case with PCs, PGs, and PEs). Each of these interfaces lies along the allosteric network that links the LBP to the AF-2. Thus, any part of the PL may influence the recruitment of coregulators and affect downstream gene transcription.

PL modular structure permits fine control of coregulator recruitment

All ligands manipulate their cognate receptors by making intermolecular interactions between the chemical structure of the ligand and the fold of the protein. Typical NR ligands, such as steroids, serve as hydrophobic nuclei around which the NR LBD folds, stabilizing the receptor in order to permit the recruitment of coregulator proteins, and subsequent activation or repression of the target gene. Nuance in this system is achieved via subtle chemical differences of the ligand that change the way the ligand interacts with the receptor, affecting either the selectivity of the receptor for the ligand, or the coregulators that are then recruited. For many ligands, these differences are quite minimal, such as a slight modification of a functional group linked to a

conserved structural scaffold. For example, the corticosteroids cortisol and cortisone differ only in the moiety attached to the 11-C, which is a hydroxyl in cortisol and an aldehyde in cortisone. However, this slight modification greatly changes the activity of the ligand – while cortisol is a potent GR agonist in humans, cortisone is inactive.

In contrast, PLs have a modular structure consisting of a glycerophosphate core conjugated to one of five possible head groups (or not, *i.e.* PA), and up to two fatty acyl tails of variable length and saturation (see Chapter 2, Figure 2.1). Unlike the steroid hormones, whose differences arise from the modification of individual atoms and small substituents attached to the steroid scaffold, each component of a PL is itself an active metabolic/signaling molecule (or precursor thereof) in its own right. Therefore, the structure of a PL is an integration of information from multiple cellular pathways into a single molecule. Because each component of a PL could influence the overall stabilization of the LBD, and thus the coregulator that is recruited, the diversity of PLs as a class of ligands could enable the exquisitely fine control of gene regulation via LRH-1.

Evolution of protein structure explains drug selectivity

The ability to target a single receptor over its close relatives is critically important to pharmacological therapy. While endogenous ligands have evolved exquisite selectivity for their cognate receptors over millions of years, synthetic drugs have only existed for several decades, and the researchers who developed them have failed to achieve the same selectivity in that time. Many drugs that are prescribed today cause adverse side effects due to the off-target modulation of receptors closely related to their target. The combination of structural biology, molecular evolution, and ancestral gene resurrection offers a powerful strategy to address this problem. By elucidating the structural evolution of a family of closely-related receptors and experimentally determining the activity of a given ligand at these receptors, one can pinpoint the precise structural changes that occurred during the evolution of the protein family that enabled protein-

ligand selectivity or promiscuity. This approach has been successful in elucidating the structural mechanisms that drive hormone selectivity among the SR family (9-16).

In the present work, we adapted this approach to explain the structural mechanisms that govern the selectivity of the synthetic corticosteroid mometasone furoate (MOF) for GR. We discovered that MOF strongly activated AncCR, the common ancestor to both GR and MR, but that activation was lost during the evolution of AncCR to MR. This is reminiscent of the way the SR family evolved to respond selectively to their endogenous ligands: the ancestral SRs were promiscuous receptors with high structural stability, which over evolutionary time became less stable, ultimately losing facets of their original function in order to selectively respond to a particular subset of the steroid hormones (16). While it would be blatantly false to insinuate that any synthetic drug shaped the evolution of a receptor family, the structural changes that have evolved in response to natural hormones may also have an effect on the recognition of synthetic ligands. Furthermore, the binding of synthetic ligands may be affected by neutral mutations – mutations that did not alter hormone binding but were not deleterious for the evolution of the receptor, and thus continue to exist in the extant paralog.

The combined structural/evolutionary approach offered several advantages over approaches that consider drug promiscuity only in the context of extant receptors. Compared to extant SRs, the ancestral SRs were more stable under laboratory conditions. In addition each protein studied in this work was more tolerant of mutations that interconverted residues with an ancestral state (evolutionarily "vertical" mutations) than those mutations that interconverted residues between two extant members ("horizontal" mutations). The reason for this is that during evolution, an ancestral protein acquires "permissive" mutations in addition to mutations that directly alter its function. Permissive mutations do not affect the function of the protein per se, but change the structural dynamics of the receptor in a manner that permit function-switching mutations to occur (14). Without permissive mutations, a function-switching mutation may render the protein nonfunctional, as reflected in the horizontal mutations made between GR and MR (see

Chapter 5, Figure 5.5), which abrogated GR's recognition of cortisol, its cognate hormone. Furthermore, determining the activity of a drug of interest for the common ancestors to two related proteins provides landmarks that can be used to identify which mutations cause the important structural changes that affect drug activity. While there may be many amino acid changes between two extant receptors, identifying the point in evolutionary time at which the response to a particular ligand changed narrows down the number of amino acid changes to only those that occurred between the two proteins on either side of the functional switch. Knowledge of a protein's structural evolution is therefore indispensable when it comes to truly understanding the structural mechanisms that govern drug selectivity.

The therapeutic potential of novel ligands may be enhanced by exploiting the flexibility of the binding pocket

It is tempting to think of the LBP of a receptor as a discrete cavern within the core of the rigid protein, which can accommodate only those ligands of a specific size, shape, and chemical nature, and which prohibits the binding of any ligand that does not fit these narrow criteria. This understanding is reinforced by structural information observed from X-ray crystal structures of protein-ligand complexes, in which the protein is naturally stabilized by the binding of a ligand, and artificially stabilized further in the context of a crystal lattice. In truth, an unliganded protein in its endogenous environment is a highly dynamic entity that samples multiple conformations, and ligand binding selectively stabilizes the protein in a particular fold that is productive for its biological function. The LBP, therefore, *must* be flexible in order to allow ligand binding to promote structurally and functionally distinct active or inactive conformations.

The flexibility of the NR LBP is perfectly exemplified by the binding of MOF to the SRs. A comparison of the crystal structures of MOF in complex with the receptors it activates – GR and PR – suggested that its selectivity for these receptors was driven by the looser conformation of H6 compared to the rigid helix seen in MR, AR, AncGR1, and AncCR (see Chapter 5, Figures 5.2 and 5.3). However not only did MOF bind to MR, AncGR1, and AncCR, it showed strong

agonist activity on AncGR1 and AncCR, demonstrating that the LBP had the flexibility to expand in order to accommodate the bulky ligand (17). LBP expansion has been seen in other synthetic steroid ligands as well (Figure 6.1). Deacylcortivazol, whose 3'-phenylpyrazolo substituent protrudes past the canonical steroid binding pocket between H5 and the β-sheet when compared to the 3'-ketone seen in natural steroids, nearly doubles the apparent volume of the LBP (18). Mifepristone, another synthetic steroid, also expands the SR LBP via its 11-dimethylaniline substituent, which protrudes between H3 and the AF-H (19-21). Expanding the SR LBP has significant and therapeutically desirable effects. Deacylcortivazol's potency for GR is 40-fold higher than dexamethasone, and 200-fold higher than cortisol (18), and the phenylpyrazolo moiety confers exquisite selectivity for GR over the other SRs (22). Mifepristone, unlike most steroid ligands, dissociates the transactivation and transrepression functions of PR and GR by displacing the AF-H from its canonically active conformation, thereby selectively promoting transrepression pathways (19-21). Were these features to be combined, as a hypothetical example, into a single ligand that selectively targets GR, one may have on their hands a powerful immunosuppressant/anti-inflammatory drug that is relatively safe for systemic administration compared to current glucocorticoids – in other words, a blockbuster. Thinking outside the pocket, so to speak, could therefore enable the discovery of novel drugs that vastly improve the therapeutic profile of NR pharmacology.

Identification of a novel activation function in allosteric communication with the AF-2 surface

Canonically speaking, NR activity is driven by two activation functions: the AF-1, a series of induced helices within the NTD, and the AF-2, a hydrophobic surface on the LBD. There is now ample evidence to indicate the existence of a third, alternate activation function (A-AF) comprising the stretch of amino acids between H5 and H7, including the β-sheet and H6 (Figure 6.2). This region was first identified by our lab in LRH-1 after it was observed that DLPC simultaneously affected the dynamics of the receptor at this site and the AF-2, and that mutations that restrict its flexibility ablate receptor activation (23). Intriguingly, one of these

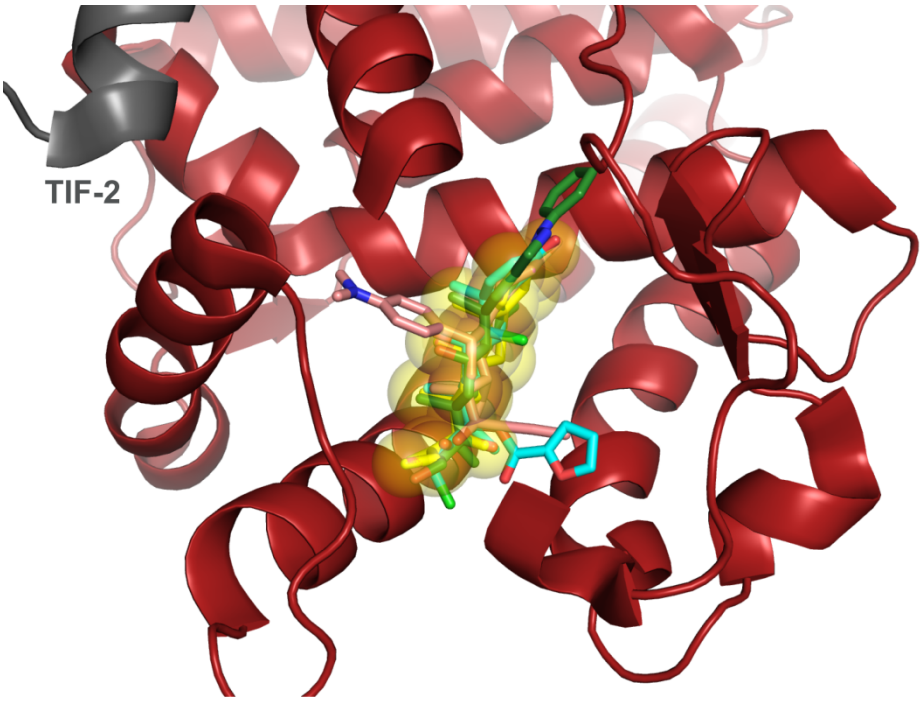

Figure 6.1. Expansion of the SR ligand binding pocket.

Alignment of GR bound to deacylcortivazol (green), MOF (cyan), mifepristone (pink), and cortisol (yellow), demonstrating the protrusion of synthetic ligands outside the canonical steroid binding pocket (yellow spheres around cortisol). To improve visibility of the LBP, the lower section of H3 has been hidden. Structures were drawn from PDBs 3BQD, 2Q1V, 4E2J, and 3H52.

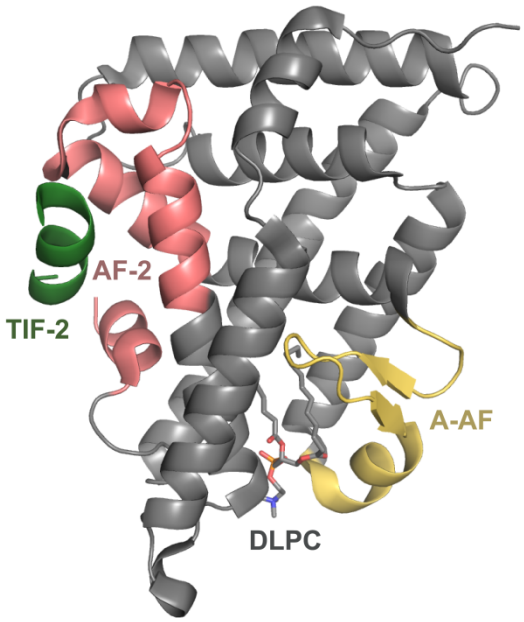

Figure 6.2. Identification of a novel activation function in the NR LBD.

The novel, alternate activation function (A-AF) comprises the region between helices 5-7 and includes the β-sheet and H6 (yellow). This region is in allosteric communication with the canonical AF-2 coregulator binding surface (pink). The A-AF is highlighted here on the structure of LRH-1 in complex with DLPC and TIF-2 (PDB: 4DOS).

mutations was at the C-terminus of H5, at the top of the A-AF in the same location as the mutation in MR that enabled activation by MOF (see Chapter 5, Figure 5.5). Furthermore, the bulky 17α-furoate moiety of MOF protrudes towards H6, at the bottom of the A-AF; the 3-phenylpyrazolo moiety of deacylcortivazol similarly protrudes towards the top of the A-AF in GR (18). In SF-1, IMPK and PTEN catalyze the interconversion of PIP_2 and PIP_3, both of which present the PIP head group at the bottom of this site (4; 24), and increasing or decreasing the expression of IMPK/PTEN strongly affects SF-1 transcriptional activity (24). The structure of LRH-1—PIP_3 has been solved showing PIP_3 in an identical location (8), and it is conceivable that a similar regulatory mechanism could exist in LRH-1. Furthermore, several PPARγ ligands have been shown to interact with this region (25), and mutations in this region enhanced the activity of nitrated linoleic acid, a potent PPARγ agonist (26). Thus, the A-AF may be an important structural region for the NR superfamily as a whole.

The molecular dynamics studies performed on LRH-1 in this work (see Chapter 3) demonstrate the existence of an allosteric network that links the A-AF with the AF-2. This establishes a line of communication between the LBP and the coregulator binding surface, thereby permitting the coordination of ligand binding and coregulator recruitment to drive the overall activity of the receptor. The A-AF is also considerably more accessible to the ligand than the AF-2, as it is itself a part of the mouth of the LBP. Current synthetic ligands that differentially recruit coregulators do so by directly interfering with the AF-2 binding surface (*e.g.* mifepristone), and tend be limited in their capacity to drive NR transactivation. However, novel ligands that instead interact with the A-AF may provide an elegant way of recruiting specific coregulators, thereby selectively promoting the desired therapeutic effect.

Remaining questions and future directions

LRH-1 remains an untapped pharmaceutical target

Its role as a master regulator of lipid and glucose metabolism, cell cycle progression, stem cell differentiation, and proliferation makes LRH-1 a highly attractive pharmaceutical target. Despite this, no FDA approved drugs that target LRH-1 currently exist. The most significant hurdle that prevents the realization of the full potential of LRH-1 as a drug target is a poor understanding of the mechanisms by which ligand binding drives LRH-1 activity. While several small molecule modulators that target LRH-1 are in early development (27-30), knowledge of its modulation by endogenous phospholipids remains elusive. This is problematic, because any attempt to target LRH-1 pharmaceutically would seek to take advantage of its endogenous biology.

We now understand that LRH-1 regulates genes via the differential recruitment of coactivators and corepressors in response to subtle conformational changes induced by PL binding, but still lack knowledge on how endogenous PLs drive this system, and how to manipulate this system using small molecules. Without this insight, novel drugs that target LRH-1 would be invented using the typical discovery pipeline: identification of candidate molecules by their performance in an activity assay (likely as part of a high throughput screen), testing in animal models of disease, and then clinical trials, wherein the safety, efficacy, and adverse effects of the candidate drugs in humans would first be discovered. Due to its role as a master regulator of many physiological processes, any LRH-1 ligand discovered via a basic activity assay may have activity on multiple LRH-1 pathways, increasing the risk of adverse effects. However, with an understanding of the structural mechanisms that drive the differential recruitment of coregulators, one could design potent LRH-1 ligands that specifically target its interaction with the coregulators that promote the desired therapeutic outcome. An ideal goal would therefore be the development of a repertoire of LRH-1 targeting drugs of similar, but not identical properties, each of which modulates a specific subset of LRH-1's target genes via the selective recruitment

of specific coregulator proteins. Furthermore, the fact that LRH-1 responds to lipid species may enable its modulation via dietary adjustments, particularly by careful attention to the intake of dietary fats. Future research on LRH-1 should therefore be focused on the identification of its endogenous regulatory ligands, the structural mechanisms by which its endogenous ligands drive differential coregulator recruitment, the capability of this system to be modulated by dietary fats, and the development of small molecule modulators of LRH-1 for the treatment of LRH-1-linked diseases such as cancer and metabolic disorders.

Evolutionary context enhances our understanding of biological systems

The combination of structural biology with molecular evolution and AGR offers an elegant approach to the elucidation of the structural mechanisms that drive ligand selectivity within a family of evolutionarily-related receptors. Improvements in computing power, gene synthesis technology, and whole genome sequencing have greatly facilitated the reconstruction of whole gene families. Indeed, since 2005, the number of PubMed search results for "ancestral reconstruction" has more than tripled, and has yet to plateau.

In the present work, the combination of AGR and structural biology allowed us to identify an interhelical interaction that controlled the response of the corticosteroid receptors to MOF. In AncCR, the ancestor to GR and MR, this interaction was mediated by a charge-dipole interaction. As the corticosteroid receptors evolved and diverged, this interaction weakened in MR to an electrostatic interaction, and developed into a hydrophobic interaction in GR. This interaction was strong enough in GR to enable activation by MOF, despite the steric bulk added by its bulky 17α furoate moiety, but was weaker in MR, preventing activation. It was only by considering SR-drug interactions through an evolutionary perspective that this interaction was identified; cross mutating this site between the extant GR and MR had no effect on MOF activation, and abrogated recognition of cortisol in GR, thus rendering GR nonfunctional in its endogenous context. Thus, the evolutionary context provided by AGR allowed us to identify a structural feature that would have been overlooked by considering the extant proteins alone.

Novel synthetic glucocorticoids can therefore be designed to take advantage of this structural region to ensure selectivity for GR over MR. Furthermore, this approach could perhaps be used in the future, not only to identify the structural features that drive the selective response to existing drugs, but also to identify structural features between two receptors that both respond to a drug that may be exploited in order to improve drug selectivity, *e.g.* MOF cross-reactivity with PR.

This work focused on the utility of AGR in solving the problem of off-target pharmacology, but the utility of AGR expands to other subjects as well. Recently published work applied AGR and phylogenetics to the development of primary tumors in order to pinpoint the precise gene mutations that enabled their invasion and metastasis, thereby identifying features that can be exploited by targeted therapies (31). The method has been used to study the difference in disease progression in HIV-infected individuals presenting or lacking a mutant allele (CCR5Δ32/Δ32) that confers high resistance to early infection (32), and to elucidate the evolutionary origins of the arbovirus family (33). An intriguing application of this method is in the field of protein engineering, where the increased stability of ancestral proteins relative to their extant descendants is highly beneficial to the large-scale production of robust biologics (34). The potential breadth of application of the evolutionary approach is as large as the number of biological processes that were in any way shaped by evolution, *i.e.* all of them. Thus, by providing insightful evolutionary context to the biological research that already take place, molecular evolution and AGR promises to be a useful adjunct tool in the progression of our understanding of the biology of our world.

Closing remarks: the future of nuclear receptor pharmacology

The ubiquity, diversity, and importance of the NR family have made them important targets for pharmacological treatments using small molecule ligands. While receptor ligands are typically classified as “agonists”, “antagonists”, and “inverse agonists”, this system is inadequate for describing the breadth of possible effects a small molecule can have upon a NR. NR

pharmacology is too complex for this model, because a ligand could activate or prevent the activation of either or both transactivation and transrepression pathways depending upon the coregulators that are recruited in response to its binding. Thus, a ligand that selectively activates a NR's transrepression functions may confusingly be referred to either as an agonist, because it promotes the activity of the receptor, or an antagonist, because it represses the transcription of the NR's target genes (*e.g.* mifepristone) (20). A more sophisticated approach to the classification of NR ligands would therefore be based upon their effect on the recruitment of specific coregulators.

The selective nuclear receptor modulators (SNuRMs) are a novel class of NR drugs that address this problem. SNuRMs were first identified upon observation of the activity of the ER ligand, tamoxifen, which exhibited agonist activity in bone and uterine tissue, but antagonist activity in breast tissue (35). Slight modifications to the chemical structure changed the tissue specificity of the drug; raloxifene, a closely related compound, is an antagonist in the uterus and breast, but an agonist in bone (35). While most examples of SNuRMs exist within the SR family, selective modulators of VDR, TR, and LXR (see Table 1.1) have also been identified (35). By selectively activating specific subsets of the total functionality of their target NRs, SNuRMs tend to cause fewer adverse reactions in the patient and are therefore much better tolerated than classical NR drugs.

The ability to dissociate transactivation and transrepression is no simple task. Early models of NR activation hypothesized that a dramatic repositioning of the AF-H was responsible for the differential recruitment of coactivators or corepressors, but we now know that the structural mechanisms that drive coregulator recruitment are much more complex. The NR LBD is inherently unstable and samples many similar but subtly different conformations in solution. Ligand binding stabilizes the LBD to permit coregulator recruitment, but variation of the ligand structure causes very subtle conformational changes that affect the affinity of the receptor for different coregulators via allosteric communication between the LBP and the AF-2 surface.

The research presented in this work considered the structural biology of the NRs through a lens that highlighted the evolution of the NR family in order to explain the effect of ligand binding on NR activation. The concept of a NR ligand typically refers to a small molecule that binds to the NR LBP, but strictly speaking, coregulator proteins are ligands too. While this research focused on small molecules, future research should consider the evolution of NR recruitment of coregulators in order to elucidate the structural determinants of coactivator vs corepressor binding, and therefore transactivation and transrepression pathways. To this end, the SR family is a very attractive model system: the five extant SRs are highly valuable as drug targets, their evolution is well researched, and most of the ancestral SRs have already been resurrected. Furthermore, while ER, GR, and PR robustly activate both transactivation and transrepression in response to ligand binding, MR and AR have lost their transrepression functions; thus, the same approach we used in this work to explain MOF selectivity between GR and MR may be used to explain the selective loss of transrepression in MR as MR and GR diverged from AncCR (and, similarly, the loss of transrepression in AR as AR and PR diverged from AncSR3). The information revealed by this approach can then be combined with our existing knowledge of the effects of small molecule binding in order to rationally design novel SNuRMs. With respect to LRH-1, an evolutionary perspective may help to explain the regulation of it and its paralog SF-1 by endogenous PLs, particularly by elucidating the mechanisms of ligand regulation of the common ancestor to the NR5A subfamily. Looking forward, the study of structural evolution will undoubtedly be an invaluable tool for improving our pharmacological repertoire.

References

1. Lee JM, Lee YK, Mamrosh JL, Busby Sa, Griffin PR, et al. 2011. A nuclear-receptor-dependent phosphatidylcholine pathway with antidiabetic effects. *Nature*
2. Chakravarthy MV, Lodhi IJ, Yin L, Malapaka RR, Xu HE, et al. 2009. Identification of a physiologically relevant endogenous ligand for PPARalpha in liver. *Cell* 138:476-88

3. Sablin EP, Blind RD, Krylova IN, Ingraham JG, Cai F, et al. 2009. Structure of SF-1 bound by different phospholipids: evidence for regulatory ligands. *Mol Endocrinol* 23:25-34
4. Blind RD, Sablin EP, Kuchenbecker KM, Chiu HJ, Deacon AM, et al. 2014. The signaling phospholipid PIP3 creates a new interaction surface on the nuclear receptor SF-1. *Proc Natl Acad Sci U S A* 111:15054-9
5. Clayton GM, Peak-Chew SY, Evans RM, Schwabe JW. 2001. The structure of the ultraspiracle ligand-binding domain reveals a nuclear receptor locked in an inactive conformation. *Proc Natl Acad Sci U S A* 98:1549-54
6. Gozani O, Karuman P, Jones DR, Ivanov D, Cha J, et al. 2003. The PHD finger of the chromatin-associated protein ING2 functions as a nuclear phosphoinositide receptor. *Cell* 114:99-111
7. Ingraham Ha, Redinbo MR. 2005. Orphan nuclear receptors adopted by crystallography. *Current opinion in structural biology* 15:708-15
8. Sablin EP, Blind RD, Uthayaruban R, Chiu HJ, Deacon AM, et al. 2015. Structure of Liver Receptor Homolog-1 (NR5A2) with PIP3 hormone bound in the ligand binding pocket. *J Struct Biol* 192:342-8
9. Thornton JW. 2001. Evolution of vertebrate steroid receptors from an ancestral estrogen receptor by ligand exploitation and serial genome expansions. *Proc Natl Acad Sci U S A* 98:5671-6
10. Thornton JW, Need E, Crews D. 2003. Resurrecting the ancestral steroid receptor: ancient origin of estrogen signaling. *Science* 301:1714-7
11. Ortlund EA, Bridgham JT, Redinbo MR, Thornton JW. 2007. Crystal structure of an ancient protein: evolution by conformational epistasis. *Science* 317:1544-8
12. Carroll SM, Bridgham JT, Thornton JW. 2008. Evolution of hormone signaling in elasmobranchs by exploitation of promiscuous receptors. *Mol Biol Evol* 25:2643-52
13. Bridgham JT, Brown JE, Rodriguez-Mari A, Catchen JM, Thornton JW. 2008. Evolution of a new function by degenerative mutation in cephalochordate steroid receptors. *PLoS Genet* 4:e1000191
14. Bridgham JT, Ortlund EA, Thornton JW. 2009. An epistatic ratchet constrains the direction of glucocorticoid receptor evolution. *Nature* 461:515-9
15. Carroll SM, Ortlund EA, Thornton JW. 2011. Mechanisms for the evolution of a derived function in the ancestral glucocorticoid receptor. *PLoS Genet* 7:e1002117
16. Eick GN, Colucci JK, Harms MJ, Ortlund EA, Thornton JW. 2012. Evolution of minimal specificity and promiscuity in steroid hormone receptors. *PLoS Genet* 8:e1003072
17. Kohn JA, Deshpande K, Ortlund EA. 2012. Deciphering modern glucocorticoid cross-pharmacology using ancestral corticosteroid receptors. *J Biol Chem* 287:16267-75
18. Suino-Powell K, Xu Y, Zhang C, Tao YG, Tolbert WD, et al. 2008. Doubling the size of the glucocorticoid receptor ligand binding pocket by deacylcortivazol. *Mol Cell Biol* 28:1915-23
19. Schoch GA, D'Arcy B, Stihle M, Burger D, Bar D, et al. 2010. Molecular switch in the glucocorticoid receptor: active and passive antagonist conformations. *J Mol Biol* 395:568-77
20. Kauppi B, Jakob C, Farnegardh M, Yang J, Ahola H, et al. 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. *J Biol Chem* 278:22748-54
21. Raaijmakers HC, Versteegh JE, Uitdehaag JC. 2009. The X-ray structure of RU486 bound to the progesterone receptor in a destabilized agonistic conformation. *J Biol Chem* 284:19572-9

22. Yoshikawa N, Makino Y, Okamoto K, Morimoto C, Makino I, Tanaka H. 2002. Distinct interaction of cortivazol with the ligand binding domain confers glucocorticoid receptor specificity: cortivazol is a specific ligand for the glucocorticoid receptor. *J Biol Chem* 277:5529-40
23. Musille PM, Pathak MC, Lauer JL, Hudson WH, Griffin PR, Ortlund EA. 2012. Antidiabetic phospholipid-nuclear receptor complex reveals the mechanism for phospholipid-driven gene regulation. *Nat Struct Mol Biol* 19:532-7, S1-2
24. Blind RD, Suzawa M, Ingraham HA. 2012. Direct modification and activation of a nuclear receptor-PIP(2) complex by the inositol lipid kinase IPMK. *Sci Signal* 5:ra44
25. Nettles KW. 2008. Insights into PPARgamma from structures with endogenous and covalently bound ligands. *Nat Struct Mol Biol* 15:893-5
26. Li Y, Zhang J, Schopfer FJ, Martynowski D, Garcia-Barrio MT, et al. 2008. Molecular recognition of nitrated fatty acids by PPAR gamma. *Nat Struct Mol Biol* 15:865-7
27. Busby S, Nuhant P, Cameron M, Mercer BA, Hodder P, et al. 2010. Discovery of Inverse Agonists for the Liver Receptor Homologue-1 (LRH1; NR5A2). In *Probe Reports from the NIH Molecular Libraries Program*. Bethesda (MD). Number of.
28. Benod C, Carlsson J, Uthayaruban R, Hwang P, Irwin JJ, et al. 2013. Structure-based discovery of antagonists of nuclear receptor LRH-1. *J Biol Chem* 288:19830-44
29. Whitby RJ, Dixon S, Maloney PR, Delerive P, Goodwin BJ, et al. 2006. Identification of small molecule agonists of the orphan nuclear receptors liver receptor homolog-1 and steroidogenic factor-1. *Journal of medicinal chemistry* 49:6652-5
30. Whitby RJ, Stec J, Blind RD, Dixon S, Leesnitzer LM, et al. 2011. Small Molecule Agonists of the Orphan Nuclear Receptors Steroidogenic Factor-1 (SF-1, NR5A1) and Liver Receptor Homologue-1 (LRH-1, NR5A2). *Journal of medicinal chemistry* 1
31. Zhao ZM, Zhao B, Bai Y, Iamarino A, Gaffney SG, et al. 2016. Early and multiple origins of metastatic lineages within primary tumors. *Proc Natl Acad Sci U S A* 113:2140-5
32. Le AQ, Taylor J, Dong W, McCloskey R, Woods C, et al. 2015. Differential evolution of a CXCR4-using HIV-1 strain in CCR5wt/wt and CCR532/32 hosts revealed by longitudinal deep sequencing and phylogenetic reconstruction. *Sci Rep* 5:17607
33. Marklewitz M, Zirkel F, Kurth A, Drosten C, Junglen S. 2015. Evolutionary and phenotypic analysis of live virus isolates suggests arthropod origin of a pathogenic RNA virus family. *Proc Natl Acad Sci U S A* 112:7536-41
34. Wijma HJ, Floor RJ, Janssen DB. 2013. Structure- and sequence-analysis inspired engineering of proteins for enhanced thermostability. *Curr Opin Struct Biol* 23:588-94
35. Burris TP, Solt LA, Wang Y, Crumbley C, Banerjee S, et al. 2013. Nuclear receptors and their selective pharmacologic modulators. *Pharmacol Rev* 65:710-78