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.

_____ _____

Suzanne G. Mays Date

LRH-1 agonist development: tapping the therapeutic potential of an orphan nuclear hormone receptor

By

Suzanne G. Mays
Doctor of Philosophy

Molecular and Systems Pharmacology
Graduate Division of Biological and Biomedical Sciences

_____
Eric A. Ortlund, Ph.D.
Advisor

_____
John W. Calvert, Ph.D.
Committee Member

_____
Christine M. Dunham, Ph.D.
Committee Member

_____
John R. Hepler, Ph.D.
Committee Member

_____
Nathan T. Jui, Ph.D.
Committee Member

Accepted:

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

_____
Date

LRH-1 agonist development: tapping the therapeutic potential of an orphan nuclear hormone receptor

By

Suzanne G. Mays
B.S., Kennesaw State University, 2008

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 Molecular and Systems Pharmacology
2018

Abstract

LRH-1 agonist development: tapping the therapeutic potential of an orphan nuclear hormone receptor

By

Suzanne G. Mays

Liver Receptor Homolog-1 (LRH-1) is an orphan nuclear hormone receptor that controls bile acid, cholesterol, glucose, and one-carbon homeostasis. Until recently, LRH-1 was thought to be insensitive to ligands and intractable as a therapeutic target. Discoveries that LRH-1 is activated by certain phosphatidylcholines (PCs) and is regulated by availability of dietary PCs has given rise to the intriguing hypothesis that LRH-1 senses PCs as a measure of nutrient availability in order to direct an appropriate metabolic response. The discovery that LRH-1 is ligand-regulated also opens the door for pharmaceutical targeting of this receptor. In this work, I address challenges with development of LRH-1 synthetic modulators using an approach that integrates structural biology with biophysical, biochemical, and cellular studies. This work provides the first detailed investigation into mechanisms used by synthetic molecules to switch LRH-1 into the active state. Insights gained from mechanistic studies led to the development of two classes of new agonists. The first class exploits a polar region within the hydrophobic binding pocket, producing a new lead molecule that is 100 times more potent than the previous best agonist. The second class was inspired by knowledge of PC-driven activation of LRH-1, yielding highly active agonists designed to act as PC mimetics. Together, these studies have provided powerful tools to probe LRH-1 biology that also have potential as therapeutic agents for metabolic diseases.

LRH-1 agonist development: tapping the therapeutic potential of an orphan nuclear hormone receptor

By

Suzanne G. Mays
B.S., Kennesaw State University, 2008

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 Molecular and Systems Pharmacology
2018

Table of Contents

List of Figures

List of Tables

List of Abbreviations

(1S,3aS)-5-hexyl-4-phenyl-3a-(1-phenylvinyl)-1,2,3,3a,6,6a-hexahydropentalen-1-ol)	RR-RJW100, RJW100, *exo*-RJW100
3- ((3-Cholamidopropyl)dimethylammonio)-1-Propanesulfonic Acid	CHAPS
6-[4-(3-chlorophenyl) piperazin-1-yl]-3-cyclohexyl-1H-pyrimidine-2,4-dione	SR-1848
Acetonitrile	MeCN
Acetyl CoA Carboxylase Alpha	ACACA
Acetyl CoA Carboxylase Beta	ACC-2
Activating Transcription Factor 2	ATF2
Activation Function Helix	AF-H
Activation Function Surface	AFS
Alternative Activation Function Surface (LRH-1)	AF-B
Androgen Receptor	AR
ATP Binding Cassette Subfamily G Member 5	ABCG5
ATP Binding Cassette Subfamily G Member 8	ABCG8
Bile Salt Export Pump	BSEP, ABCB11
C-terminal extension	CTE

Cardiovascular disease	CVD
Cytochrome P450 family 7 subfamily A member 1	CYP7A1
Cytochrome P450 family 8 subfamily B member 1	CYP8B1
Differential scanning fluorimetry	DSF
Diisopropylethylamine	DIPEA
Dilauroylphosphatidylcholine	DLPC
Dimethylaminopyridine	DMAP
Dithiothreitol	DTT
Diundecanoylphosphatidylcholine	DUPC
DNA binding domain	DBD
Endoplasmic Reticulum	ER
Estrogen Receptor	EstR
EthyleneDiamine Tetraacetic Acid	EDTA
Farnesoid X Receptor	FXR
Fatty Acid Synthase	FASN
Fetal Bovine Serum	FBS
Fluorescein	FAM
Full-length	FL
GlaxoSmithKline	GSK
Glucocorticoid Receptor	GR
Glucokinase	Gck
Glucose transporter type 4	Glut4
Hepatocyte Nuclear Factor 1	HNF4
Hexahydropentalene	6HP
High density lipoprotein	HDL

Hydrochloric acid	HCl
Hydrogen-deuterium exchange mass spectrometry	HDX-MS
Isopropyl 1-thio-β-galactopyranodside	IPTG
Ligand binding domain	LBD
Liquid Chromatography Mass Spectrometry	LC-MS
Lithium diisopropylamine	LDA
Liver Receptor Homolog-1	LRH-1, NR5A2
Liver X Receptor	LXR
Methoxymethyl	MOM
Minimum Essential Medium	MEM
Mitogen Activated Protein Kinase 1/2	ERK 1/2
Molecular dynamics simulations	MDS
n-butyllithium	nBuLi
N-Methyl morpholine N-oxide	NMO
Nonalcoholic fatty liver disease	NAFLD
Nuclear Magnetic Resonance	NMR
Nuclear Receptor	NR
Nuclear Receptor Coactivator-2	Tif2, NCOR2
Nuclear Receptor Corepressor 1	NCOR1
Oxysterol-binding Protein-Related Protein 3	OSBL3
Particle Mesh Ewald	PME
Paulson-Khand	PK
Peroxisome Proliferator-Activated Gamma Coactivator 1-α	PGC1-α
Phosphate buffered saline	PBS
Phosphatidylinositol 3,4,5-triphosphate	PIP3

Phospholipid	PL
Polo Like Kinase 3	PLK3
POU Class 5 Homeobox 1	POU5F1, OCT4
Preparative high pressure liquid chromatography	Prep-HPLC
Preperative thin layer chromatography	PTLC
Promyelocytic Leukemia Protein	PML
Prospero-Related Homeobox 1	PROX1
Reverse cholesterol transport	RCT
S-Adenosyl methionine	SAM, AdoMet
S-Adenosyl-L-homocysteine	AdoHcy
Scavenger Receptor Class B, Member 1	SCARB1
Silencing Mediator Of Retinoic Acid And Thyroid Hormone Receptor	SMRT
Small Heterodimer Partner	SHP, NR0B2
Sterol Regulatory Element Binding Protein-1c	SREBP-1c
Steroyl CoA Desaturase-1	SCD-1
Surface plasmon resonance	SPR
Tetrabutylammonium fluoride	TBAF
Tetrahydrofuran	THF
Tetrapropylammonium perruthenate	TPAP
Thin layer chromatography	TLC
Tobacco Etch Virus	TEV
Triethylamine	TEA
Type 2 Diabetes Mellitus	T2DM
Wild-type	WT

Chapter 1. Structure and Regulation of the Orphan Nuclear Hormone Receptor, LRH-1

1.1 Nuclear receptors: hormone-regulated transcription factors.

The nuclear receptor (NR) superfamily is a group of ligand-regulated transcription factors that regulate diverse and essential physiological processes, such as embryonic development, response to stress, metabolism, and reproduction (Gronemeyer et al., 2004). NRs respond to hormone ligands to promote or repress transcription of specific sets of genes, often which are functionally related. This allows the response to stimuli or regulation of biological processes to occur in a coordinated manner across multiple organs or cell types within an organism (Gronemeyer et al., 2004).

There are 48 NRs in humans, which are further subdivided into seven subfamilies (numbered 0-6, summarized in Table 1.1). Genes in the NR family are named with the designation "NR" followed by subfamily number and gene number. For example, NR5A2 is in NR subfamily 5A, member number 2. In addition, many NRs have common names, particularly those that have well-known ligands or functions (such as the glucocorticoid receptor, which has a gene name of NR3C1 but is commonly abbreviated "GR"). Structurally, the NR superfamily is defined by a conserved protein fold and modular domain architecture (reviewed recently (Weikum et al., 2018) and described below for the NR that is the focus of this dissertation (Section 1.2)).

Most NRs directly bind DNA to modulate gene expression, recognizing particular DNA response elements (RE). NR can bind DNA either as homodimers (such as GR), as heterodimers with Retinoid X receptor (RXR), or as monomers (Figure 1.1). In addition, the NR family has two noncanonical members (NR0B1 and NR0B2) that do not bind DNA. Rather, they interact with NR to repress transcriptional activity (Weikum et al., 2018).

Transcriptional activity of NRs is influenced by association with hormone ligands. Unlike peptide hormones that bind cell surface receptors, NR ligands are lipophilic and are able to cross cell membranes to reach receptors residing inside of the cell. For example, the steroid hormones are synthesized from the lipid, cholesterol (Payne, 2004). Other examples of NR ligands include fatty acids, oxysterols, and bile acids (Table 1.1)(Gronemeyer et al., 2004). Approximately half of the NR superfamily is orphans, meaning that the endogenous ligands are unknown or are contentious (Table 1.1).

Table 1.1. Human Nuclear Receptors and Cognate Ligands. Table has been adapted from (Gronemeyer et al., 2004)

Common Name	Abbreviation	Nomenclature	Ligand
DSS-AHC critical region on the chromosome, gene 1	DAX1	NR0B1	Orphan
Short heterodimeric partner	SHP	NR0B2	Orphan
Thyroid hormone receptor	TRα	NR1A1	Thyroid hormone
	TRβ	NR1A2	Thyroid hormone
Retinoic acid receptor	RARα	NR1B1	Retinoic acid
	RARβ	NR1B2	Retinoic acid
	RARγ	NR1B3	Retinoic acid
Peroxisome proliferator-activated receptor	PPARα	NR1C1	Fatty acids
	PPARβ	NR1C2	Fatty acids
	PPARDγ	NR1C3	Fatty acids
Reverse erbA	Rev-erbα	NR1D1	Orphan
	Rev-erbβ	NR1D2	Orphan
RAR-related orphan receptor	RORα	NR1F1	Cholesterol
	RORβ	NR1F2	Retinoic acid
	RORγ	NR1F3	Retinoic acid
Liver X receptor	LXRα	NR1H3	Oxysterols
	LXRβ	NR1H2	Oxysterols
Farnesoid X receptor	FXRα	NR1H4	Bile acids
	FXRβ	NR1H5	Orphan
Vitamin D receptor	VDR	NR1I1	1,25-dihydroxy vitamin D3
Pregnane X receptor	PXR	NR1I2	Xenobiotics
Constitutive androstane receptor	CAR	NR1I3	Xenobiotics
Human nuclear factor 4	HNF4α	NR2A1	Orphan
	HNF4β	NR2A2	Orphan
Retinoid X receptor	RXRα	NR2B1	Retinoic acid
	RXRβ	NR2B2	Retinoic acid
	RXRγ	NR2B3	Retinoic acid
Testis receptor	TR2	NR2C1	Orphan
	TR4	NR2C2	Orphan
Tailless	TLL	NR2E2	Orphan
Photoreceptor-specific nuclear receptor	PNR	NR2E3	Orphan
Chicken ovalbumic upstream promoter-transcription factor	COUP-TFI	NR2F1	Orphan
	COUP-TFII	NR2F2	Orphan
ErbA2-related gene-2	EAR2	NR2F6	Orphan
Estrogen receptor	ERα	NR3A1	Estradiol-17β
	ERβ	NR3A2	Estradiol-17β
Estrogen receptor-related receptor	ERRα	NR3B1	Orphan
	ERRβ	NR3B2	Orphan
	ERRγ	NR3B3	Orphan
Glucocorticoid receptor	GR	NR3C1	Cortisol
Mineralocorticoid receptor	MR	NR3C2	Aldosterone
Progesterone receptor	PR	NR3C3	Progesterone
Androgen receptor	AR	MR3C4	Testosterone
NGF-induced factor B	NGFIB	NR4A1	Orphan
Nur related factor 1	NURR1	NR4A2	Orphan
Neuron-derived orphan receptor 1	NOR1	NR4A3	Orphan
Steroidogenic factor 1	SF-1	NR5A1	Orphan
Liver receptor homolog 1	LRH-1	NR5A2	Orphan
Germ cell nuclear factor	GCNF	NR6A1	Orphan

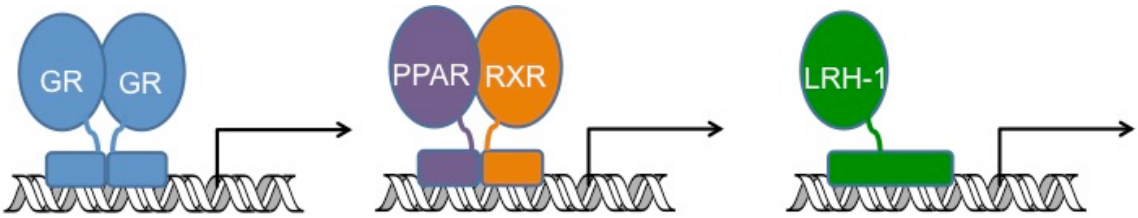


Figure 1.1. Nuclear Receptors bind DNA as dimers or monomers. Recognition of DNA response elements by NRs can involve homodimerization (*left)* as with the glucocorticoid receptor (GR) or heterodimerization with Retinoid X receptor (RXR), as is the case with peroxisome proliferator-activated receptors (PPAR, *middle panel).* Some NR, such as Liver Receptor Homolog-1 (LRH-1), bind DNA as a monomer (*right*).

1.2 Liver Receptor Homolog-1

Liver Receptor Homolog-1 (LRH-1, NR5A2) is an orphan Nuclear Hormone Receptor (NR) that is highly expressed in the liver, intestine, exocrine pancreas, and skeletal muscle, as well as in certain female reproductive organs, such as the ovaries(Fayard, 2004). It is also expressed widely within the brain(Grgurevic, 2005) and has been recently detected in other tissues and cells such as macrophages(Lefevre et al., 2015), prostate(Xiao, 2018), and T-cells(Schwaderer et al., 2017). During development, LRH-1 is required for maintenance of pluripotency by regulating POU Class 5 Homeobox 1 (POU5F1, OCT4) expression(Gu et al., 2005) and plays a vital role in organogenesis of the pancreas and the liver(Nissim et al., 2016). For this reason, knockout of Lrh-1 in mice is embryonically lethal, although inducible global knockdown in adulthood is tolerated(Holmstrom et al., 2011) and many organ- and cell type-specific knockouts are viable(Lee et al., 2011; Lefevre et al., 2015; Mataki et al., 2007; Schwaderer et al., 2017; Venteclef et al., 2010). In adult organisms, LRH-1 regulates cellular proliferation(Bayrer et al., 2015; Lin et al., 2014), local steroid production in the gut(Coste; Fernandez-Marcos et al., 2011; Mueller et al., 2006), hepatic acute phase response(Venteclef et al., 2006), and nutrient metabolism (discussed in detail below). LRH-1 exerts the majority of these biological effects by binding DNA and modulating transcription.

LRH-1 shares the characteristic NR modular composition (Figure 1.2), featuring a highly conserved, zinc-finger-containing DNA binding domain (DBD) and a Ligand Binding Domain (LBD) that is comprised mainly of alpha helices, arranged in a four-layer “sandwich” configuration. While the majority of NRs dimerize on double-stranded DNA and recognize pairs of six-nucleotide sequences (arranged either as direct repeats or inverse repeats)(Weikum et al., 2018), LRH-1 binds DNA as a monomer (Solomon et al., 2005). It utilizes a C-terminal extension (CTE) (located at the C-terminal end of the DBD) to recognize a nine-nucleotide consensus sequence of YCAAGGYCR (where “Y” is a pyrimidine and “R” is a purine)(Solomon et al., 2005).

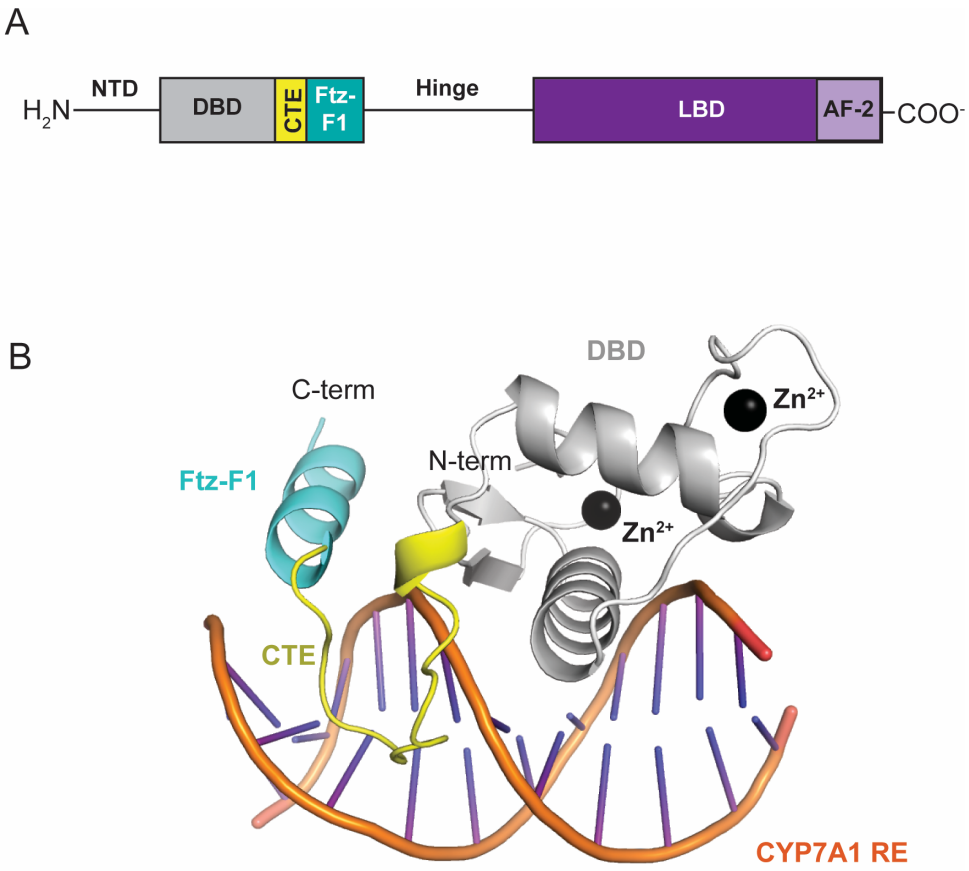


Figure 1.2 LRH-1 Structure. A. Cartoon showing the modular domain structure of LRH-1. It has an unstructured N-terminal domain (NTD) with no known function. The DNA binding domain (DBD) and the C-terminal extension (CTE) interact with DNA, while the Ftz-F1 domain mediates protein-protein interactions. The flexible hinge contains sites for post-translational modifications, and the Ligand Binding Domain (LBD) contains a lipophilic binding pocket (not shown). The Activation Function-2 (AF-2) is a part of the LBD that is important for interaction with coregulator proteins. B. The crystal structure of the LRH-1 DBD bound to its DNA response element (RE) on the human CYP7A1 promoter shows the binding mode and the positioning of the CTE and Ftz-F1 helix. Black spheres represent zinc ions. Colors within the structure correspond to domains shown in panel A. From PBD ID 2A66(Solomon et al., 2005).

A unique structural feature of the NR5A family is the Ftz-F1 domain, located between the CTE and the hinge. This domain is named after the *Drosophila* ortholog of LRH-1, Fushi tarazu factor 1 (Fayard, 2004). Ftz-F1 does not bind DNA, but it plays an important role in Tif2-mediated activation(Solomon et al., 2005). It is also involved in protein-protein interactions, such as with Multiprotein Bridging Factor (MBF-1)(Brendel, 2002). The LRH-1-MBF-1 interaction creates a platform for recruitment of the general transcriptional machinery to drive LRH-1 activation(Brendel, 2002). The binding mode of the LRH-1 DBD, CTE, and Ftz-F1 on the human CYP7A1 promoter is shown in Figure 1.1 (from PDB ID 2A66)(Solomon et al., 2005).

Connecting the DBD and the LBD is a flexible hinge, 112 amino acids in length, that is subject to post translational modifications. Phosphorylation of residues S238 and S243 in the hinge by Mitogen Activated Protein Kinase 1/2 (ERK 1/2) increase LRH-1's transcriptional activity(Lee et al., 2006). The hinge is also SUMOylated by small ubiquitin-like modifier 1 (SUMO-1), which reduces transcriptional activity(Stein et al., 2014; Venteclef et al., 2010; Yang et al., 2009). SUMOylation causes the translocation of LRH-1 to promyelocytic leukemia protein (PML) bodies, which is reversible and is regulated by cAMP signaling(Chalkiadaki and Talianidis, 2005; Yang et al., 2009). Another mechanism through which SUMOylation inhibits transcriptional activity is by promoting association with corepressors, such as Nuclear Receptor Corepressor 1 (NCOR1)(Venteclef et al., 2010), and Prospero-Related Homeobox 1 (PROX1)(Gu et al., 2005; Stein et al., 2014). The LBD, located at the C-terminus of the protein, is comprised mainly of alpha-helices arranged in a four-layered sandwich configuration (Ortlund et al., 2005). The LBD contains a hydrophobic ligand binding pocket and a C-terminal Activation Function-2 (AF-2) subdomain, which serves as the site for recruitment of coactivator and corepressor proteins (collectively called "coregulators"). Coregulators regulate NR transcriptional activity by recruiting histone-modifying factors that control promoter accessibility.

1.3 Regulation of LRH-1 by ligands

Until recently, LRH-1 was thought to be insensitive to ligands, and mechanisms through which it is regulated by ligands are still poorly understood. The endogenous LRH-1 ligand is unknown, but it was hypothesized to be a bile acid or an oxysterol due to its role in bile acid production (Lu et al., 2000). However, exogenous addition of a wide variety of candidate ligands fail to activate LRH-1 above basal levels (Eric Ortlund, personal communication). Together with an empty binding pocket observed in the mouse LRH-1 crystal structure(Sablin et al., 2003), this led to the assumption that LRH-1 is constitutively active. A confounding factor during initial characterization was that recombinant human LRH-1 co-purifies with a variety of phospholipids (PL) from *E. coli* occupying the binding pocket that are difficult to displace(Sablin et al., 2008). In landmark studies, David Moore's group (Baylor, TX) discovered that medium-chained, saturated phosphatidylcholines (PCs), such as dilauoylphosphatidylcholine (PC 12:0/12:0, DLPC) can displace more abundant bacterial PLs from the pocket and activate the receptor above basal levels (Lee et al., 2011). We and others have since shown that LRH-1 can bind a variety of PLs *in vitro* (*e.g.* PCs, phosphatidylglycerols, phosphatidylinositols, phosphatidylethanolamines, and sphingomyelin)(Musille et al., 2015; Sablin et al., 2015); however, only DLPC and other medium-chained saturated PCs have been shown to modulate its transcriptional activity(Bolado-Carrancio et al., 2014; Flynn et al., 2018; Lee et al., 2011).

Our structural studies with DLPC illuminated a unique binding mode and mechanism of action (Figure 1.3). PL interact with LRH-1 with the hydrophobic tails buried inside of the pocket, while the phosphate group engages residues near the mouth of the LRH-1 binding pocket. This is unusual for NR ligands, which typically are completely engulfed in the ligand binding pocket and do not contact residues near the mouth (Musille et al., 2012), Figure 1.3A. We identified the region near the mouth of the LRH-1 pocket as a novel activation function surface (hereafter AF-B) (Figure 1.3B-D). Flexibility in AF-B is required for receptor activation and critical for ligand sensing. In the absence of ligand, AF-B is completely disordered, seen in our crystal structure of

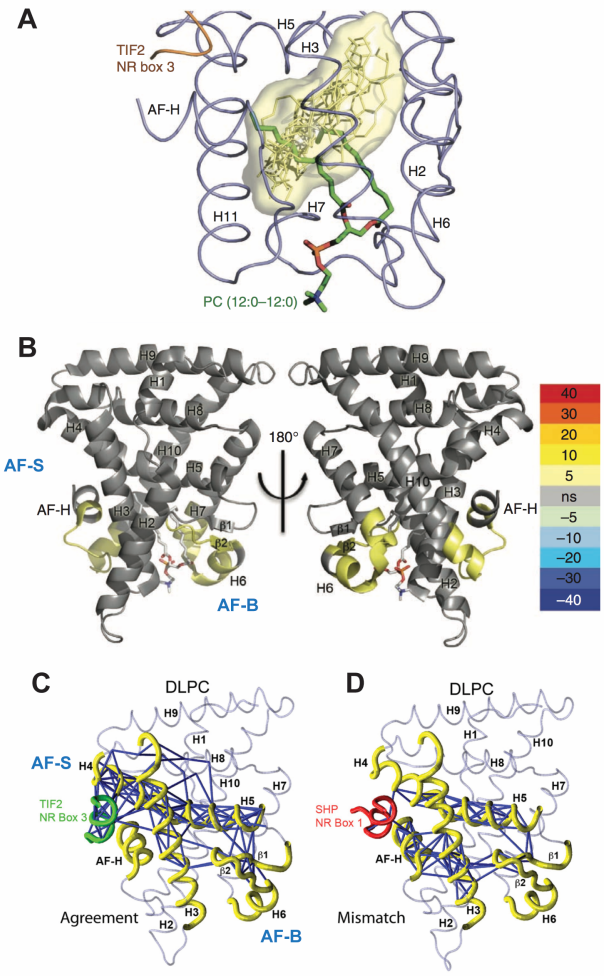


Figure 1.3. The LRH-1-DLPC structures reveals an unusual binding mode and a novel activation function surface. A. Superposition of DLPC (PDB ID 4DOS) with endogenous ligands of other NRs (*i.e.* Retinoic X Receptor (RXR), Vitamin D Receptor (VDR), EstR, Peroxisome Proliferator Activated Receptor (PPAR), FXR and Thyroid Receptor). This highlights the unusual binding mode of DLPC, low in the pocket. B. HDX-MS studies showing selective destabilization (yellow) of the AF-B and AFS regions when DLPC *versus* non-specific *E. coli* PL are bound. C-D. Molecular dynamics simulations show strong communication (blue lines) between AF-B and AFS when DLPC is bound with a coactivator (Tif2) but not a corepressor (SHP). Panels A and B are adapted from reference(Musille et al., 2012). Panels C and D are adapted from reference(Musille et al., 2015).

apo-LRH-1 bound to the corepressor, SHP(Musille et al., 2012). Moreover, motion in AF-B can be used to discriminate between weakly activating bacterial PLs and DLPC. As shown in Figure 1.3B, selective destabilization of AF-B and the AF-H occurs when DLPC, rather than *E. coli* PL, are bound. Molecular dynamics (MD) simulations show striking networks of communication between these two regions, which are strong when DLPC is bound at the same time as a coactivator (Tif2, Figure 1.3C). Conversely, the network is greatly weakened when a DLPC is bound with a corepressor (*e.g.* SHP, Figure 1.3D). Thus, LRH-1 uses this network to communicate ligand status to the AFS, allowing recruitment of appropriate coregulators based on the identity of the ligand.

1.4 Metabolic processes controlled by LRH-1.

LRH-1 controls a diverse program of genes related to nutrient metabolism. This includes regulation of many aspects of lipid, glucose, and one-carbon metabolism.

1.4.1 Bile acid biosynthesis.

Synthesis and mobilization of bile acids is a crucial part of digestion and cholesterol homeostasis. LRH-1 controls bile acid synthesis through direct transcriptional control of Cytochrome P450 family 7 subfamily A member 1 (CYP7A1) and Cytochrome P450 family 8 subfamily B member 1 (CYP8B1). CYP7A1 is the rate-limiting step in bile acid biosynthesis, while CYP8B1 acts downstream of CYP7A1 in the production of cholic acid. Regulation at the CYP7A1 promoter by LRH-1 is quite complex, also involving Farnesoid X Receptor (FXR), Hepatocyte Nuclear Factor 4 (HNF4) and extensive feedback regulation via the corepressor Small Heterodimer Partner (SHP)(Kir et al., 2012; Lu et al., 2000). Reflecting this complexity, knockout of LRH-1 in hepatocytes greatly reduces Cyp8b1 expression but does not affect Cyp7a1. The differential regulation of these genes in the bile acid biosynthetic pathway ultimately alters composition of the bile acid pool, inhibits bile acid reuptake, and reduces lipid absorption from the gut(Mataki et al., 2007). LRH-1 also regulates bile acid secretion from the

liver via transcriptional regulation of the Bile Salt Export Pump (BSEP) transporter(Song et al., 2008).

1.4.2 Reverse cholesterol transport.

Cholesterol homeostasis is maintained, in part, through reverse cholesterol transport (RCT), defined as the transport of cholesterol from peripheral tissues into the liver for excretion. LRH-1 stimulates RCT by direct transcriptional control of Scavenger Receptor Class B, Member 1 (SCARB1) and ATP Binding Cassette Subfamily G Members 5 and 8 (ABCG5 and ABCG8) in the liver(Schoonjans et al., 2002; Stein et al., 2014). SCARB1 is a receptor that recognizes and imports high density lipoprotein (HDL), reducing circulating cholesterol and triglycerides. ABCG5 and 8 are transporters that export cholesterol into the bile. These genes are upregulated by LRH-1 activity and suppressed when LRH-1 is SUMOylated(Stein et al., 2014).

1.4.3 De Novo Lipogenesis.

Activation of LRH-1 represses *de novo* lipogenesis by reducing transcription of Sterol Regulatory Element Binding Protein-1c (SREBP-1c), Fatty Acid Synthase (FASN), Acetyl CoA Carboxylase Beta (ACC-2), and Steroyl CoA Desaturase-1 (SCD-1) in the liver(Lee et al., 2011). Mechanisms through which this occurs are somewhat unclear, since these genes do not contain LRH-1 response elements. Rather, suppression of *de novo* lipogenesis may occur through interaction of LRH-1 with SREBP1 at SREBP1 promoters (Kanayama et al., 2007). Reduction of lipogenesis could improve insulin sensitivity, and, since insulin stimulates SREBP1 transcription, this could create a favorable loop whereby both SREBP mRNA and SREBP-1-controlled transcripts are reduced (model proposed by the Moore group) (Lee et al., 2011; Zhou, 2005).

In addition to the role of LRH-1 suppressing *de novo* lipogenesis, it can also upregulate lipid biosynthesis in certain circumstances. This was demonstrated by the Schoonjans group, which found that ablation of a SUMOylation site on LRH-1 (which prevents SUMO-dependent repression), promotes hepatic steatosis in the context of a high-fat, high sucrose diet(Stein, 2016). This occurs via direct transcriptional upregulation of Oxysterol-binding Protein-Related Protein 3

(OSBL3), which promotes SREBP-1 processing and consequent increases in expression of FASN, SCD-1, and Acetyl CoA Carboxylase Alpha (ACACA)(Stein, 2016).

1.4.4 Endoplasmic-reticulum (ER) stress resolution.

The ER-stress response is a process utilized to relieve cellular stress caused by accumulation of misfolded proteins in the ER. It is initially cytoprotective but becomes cytotoxic when unresolved. In NAFLD, hepatic ER-stress often becomes chronic and cytotoxic(Mamrosh et al., 2014). LRH-1 helps resolve tunicamycin-induced ER-stress via transcriptional upregulation of Polo Like Kinase 3 (PLK3), which phosphorylates Activating Transcription Factor 2 (ATF2). LRH-1-mediated ER-stress resolution is anti-apoptotic and cytoprotective to hepatocytes(Mamrosh et al., 2014).

1.4.5 Glucose homeostasis.

Effects of LRH-1 on glucose metabolism have been studied in the liver and skeletal muscle. Murine hepatic glucokinase (Gck) is a direct transcriptional target of LRH-1, and conditional knockout of Lrh-1 in mouse liver reduces postprandial glycolysis and lipogenesis(Oosterveer et al., 2012). In rat skeletal muscle cells (C2C12 cells), DLPC treatment increased expression of glucose transporter type 4 (Glut4) resulting in increased glucose uptake. DLPC also upregulated several genes related to metabolism of glucose (*i.e* hexokinase 2, muscle glycogen synthase, muscle phosphofructokinase, muscle glycogen phosphorylase) and lipids (*i.e.* Acc-2 and fatty acid binding protein 3). These effects did not occur in cells with low Lrh-1 expression. Although there is a putative Lrh-1 DNA response element at the Glut4 promoter, upregulation of this transporter was not found to be a direct transcriptional effect(Bolado-Carrancio et al., 2014). Because there was an increase in glycolysis and glycogen synthesis with a concomitant decrease in lipogenesis in these cells, the authors hypothesized that Lrh-1 activity promoted a switch from fatty acid to glucose oxidation.

1.4.6 Methyl metabolism.

The one-carbon pool is a source of labile methyl groups that can be used for diverse processes from DNA methylation to lipid synthesis. Lrh-1 impacts one-carbon metabolism in the liver by direct transcriptional control of glycine methyltransferase, which methylates sarcosine to glycine. It also indirectly regulates expression of the Multidrug resistance 2 (Mdr2) flippase (called MDR3 in humans), which transports PLs from hepatocytes into the bile. Inhibition of PC synthesis via methionine and choline depletion (MCD) greatly reduces LRH-1 activity *in vitro*(Wagner et al., 2016). Notably, Lrh-1 liver-specific knockout mice were protected mice from fibrosis and inflammation associated with the MCD diet (although they still developed steatosis). This protection from the more severe aspects of MCD-associated liver disease was attributed to reduction of Lrh-1-mediated flux through the one-carbon pathway, preserving labile methyl groups for more essential processes and maintaining the integrity of cell membranes by protecting the PE/PC ratio(Wagner et al., 2016).

1.5. Potential role as an integrator of metabolism via phosphatidylcholine sensing.

The involvement of LRH-1 in both lipid and methyl metabolism is particularly intriguing in light of the well-established connection between the one-carbon cycle (1CC), PC synthesis, and liver fat accumulation(Walker, 2017). The 1CC provides a means for PC production via the PEMT pathway, in which PC is generated in three successive methylation steps requiring the labile methyl donor, S-adenosylmethionine (SAM) (Figure 1.4). Defective 1CC metabolism is associated with fat accumulation in the liver, both in animal models of NAFLD and in human diseases such as alcoholic steatosis and cystic fibrosis (reviewed recently)(Walker, 2017). Reduced PC levels are linked to fatty liver development both as structural components of cell membranes and as signaling molecules(Furse and de Kroon, 2015; Li et al., 2006; van der Veen et al., 2017). In a recent study, depletion of dietary methionine in bacteria fed to *C. elegans* repressed the bacterial 1CC, leading to reduced PC biosynthesis in the worm. This inhibited the worm Lrh-1 ortholog, causing fat accumulation in a mechanism dependent on hedgehog

signaling(Lin and Wang, 2017). PCs and 1CC metabolism have also been linked to fatty liver via SREBP1. Walker et al. identified a feedback loop whereby SREBP controls expression of several 1CC genes, affecting PC production. In turn, PCs control SREBP1-mediated transcription of lipogeneic genes(Walker et al., 2011). Considering that LRH-1 modulates SREBP signaling(Kanayama et al., 2007; Lee et al., 2011), it is possible that PC-driven LRH-1 signaling could tune SREBP responses in addition to modulating metabolic genes directly. Together, these data have led to the hypothesis that LRH-1 senses PC ligands as a proxy for nutrient status in order to direct an appropriate metabolic response (Figure 1.4). This has broad implications for targeting LRH-1 in metabolic diseases.

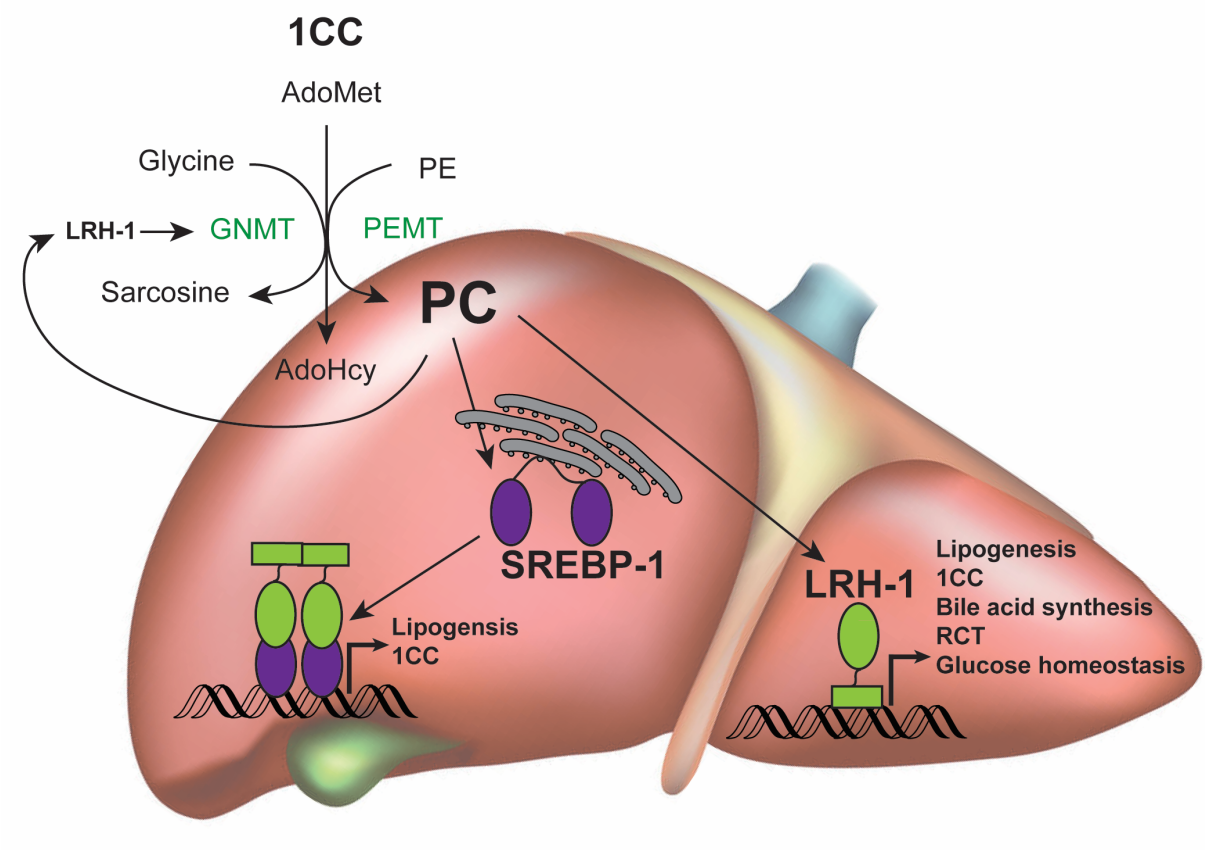


Figure 1.4. Integration of 1CC, PC, and lipogenesis in the liver. Simplified diagram of the 1CC showing production of PC from methylation of PE. PC generation has been linked to SREBP1 processing and transcription of lipogeneic and 1CC genes. LRH-1 is activated by select PCs, mediating expression of genes related to lipid, glucose, and one-carbon metabolism. LRH-1 has also been shown to repress SREBP by direct interactions at SREBP promoters. Abbreviations: PEMT, phosphatidylethanolomine methyltransferase; GNMT, glycine methyltransferase; AdoMet, S-Adenosyl methionine; AdoHcy, S-Adenosyl-L-homocysteine; RCT, reverse cholesterol transport.

Chapter 2: Targeting LRH-1 with Agonists in Metabolic Diseases

2.1 LRH-1 as a therapeutic target for metabolic diseases

Obesity-driven metabolic dysfunction underlies several serious, interrelated disorders, including nonalcoholic fatty liver disease (NAFLD), cardiovascular disease (CVD), and type II diabetes mellitus (T2DM). Nearly 70% of adult Americans are now classified as overweight or obese(National Institute of Diabetes and Digestive and Kidney Diseases, 2010), and this number is likely to increase in the future based on current trends(National Institute of Diabetes and Digestive and Kidney Diseases, 2010). In obese individuals, excessive lipids can overwhelm the storage capacity of adipose tissue, resulting in increased circulating triglycerides and abnormal accumulation of lipid droplets in liver, pancreas, and skeletal muscle(Cusi, 2012). In liver, these droplets are a hallmark of NAFLD, which can progress from steatosis to cirrhosis, causing irreversible liver damage. Even in the mainly asymptomatic early stages of NAFLD, liver fat and impaired lipid metabolism are strongly associated with development of insulin resistance, CVD and diabetes(Bayeva et al., 2013; Firneisz, 2014; Lomonaco, 2016; Lonardo, 2015; Ortiz-Lopez et al., 2012; Savage et al., 2007). Moreover, high levels of triglycerides and dietary carbohydrates contribute to aberrant glucose metabolism, with increases in glycogenolysis and in the release of glucose into circulation. Not surprisingly, numbers of T2DM cases are also rapidly increasing, paralleling the trends for obesity and NAFLD. Nearly 30 million Americans currently have T2DM and 86 million have prediabetes(Centers for Disease Control and Prevention, 2014). Patients with T2DM are more than twice as likely to experience heart attacks, strokes, and death by CD than non-diabetics(Centers for Disease Control and Prevention, 2014). The presence of T2DM also increases the likelihood of NAFLD(Bril, 2016). As an important regulator of lipid and glucose metabolism, LRH-1 has great potential as a novel therapeutic target. Figure 2.1 summarizes how metabolic dysfunction contributes to NAFLD, CVD, and T2DM and the many points at which stimulation of LRH-1 activity could be beneficial.

2.2 Preclinical studies involving LRH-1

Several preclinical studies have studied the effects of LRH-1 in NAFLD and associated diseases. First, acute knockout of LRH-1 in adult mouse liver causes lipid droplet accumulation, which is exacerbated with high fat diet(Miranda et al., 2018). Another study showed that a genetic mutation that prevented LRH-1 inactivation via SUMOylation stimulated the RCT pathway in ApoE mice given a high fat diet (Stein et al., 2014). Mutant mice had greater LRH-1 activity and higher levels of LRH-1-dependent RCT gene expression in the liver. Consequently, cholesterol excretion was enhanced and atherosclerosis was reduced relative to wild-type mice with lower LRH-1 activity(Stein et al., 2014). In mouse models of diet-induced insulin resistance, treatment with the LRH-1 agonist DLPC reduced hepatic fat accumulation, improved glucose tolerance and reduced circulating insulin, triglycerides and free fatty acids(Lee et al., 2011). These anti-diabetic effects were associated with changes in expression of a select subset of LRH-1 target genes involved with lipid metabolism. Importantly, the DLPC-induced effects on health and on gene expression were absent in LRH-1 liver-specific conditional knockout mice, directly implicating LRH-1 in these effects(Lee et al., 2011). Together, these studies have shown that activation of LRH-1 imparts significant effects on lipid and glucose metabolism. The ability to control the LRH-1 metabolic program with small molecule agonists has immense potential for treatment of obesity-related disorders.

2.3 Therapeutic potential beyond NAFLD and diabetes.

In addition to the metabolic effects described above, LRH-1 has potential for treatment of inflammatory bowel disease (IBD) due to its anti-inflammatory actions in the gut. LRH-1 is protective in preclinical models of IBD(Bayrer et al., 2018; Coste). Moreover, LRH-1 antagonists are highly sought for the treatment of several cancer types, including breast, colon,

pancreatic, and liver cancers(Bayrer et al., 2015; Bianco et al., 2014; Chand et al., 2010; Clyne et al., 2004; Lin et al., 2014; Thiruchelvam et al., 2011; Xu, 2016). Aberrant LRH-1 activity

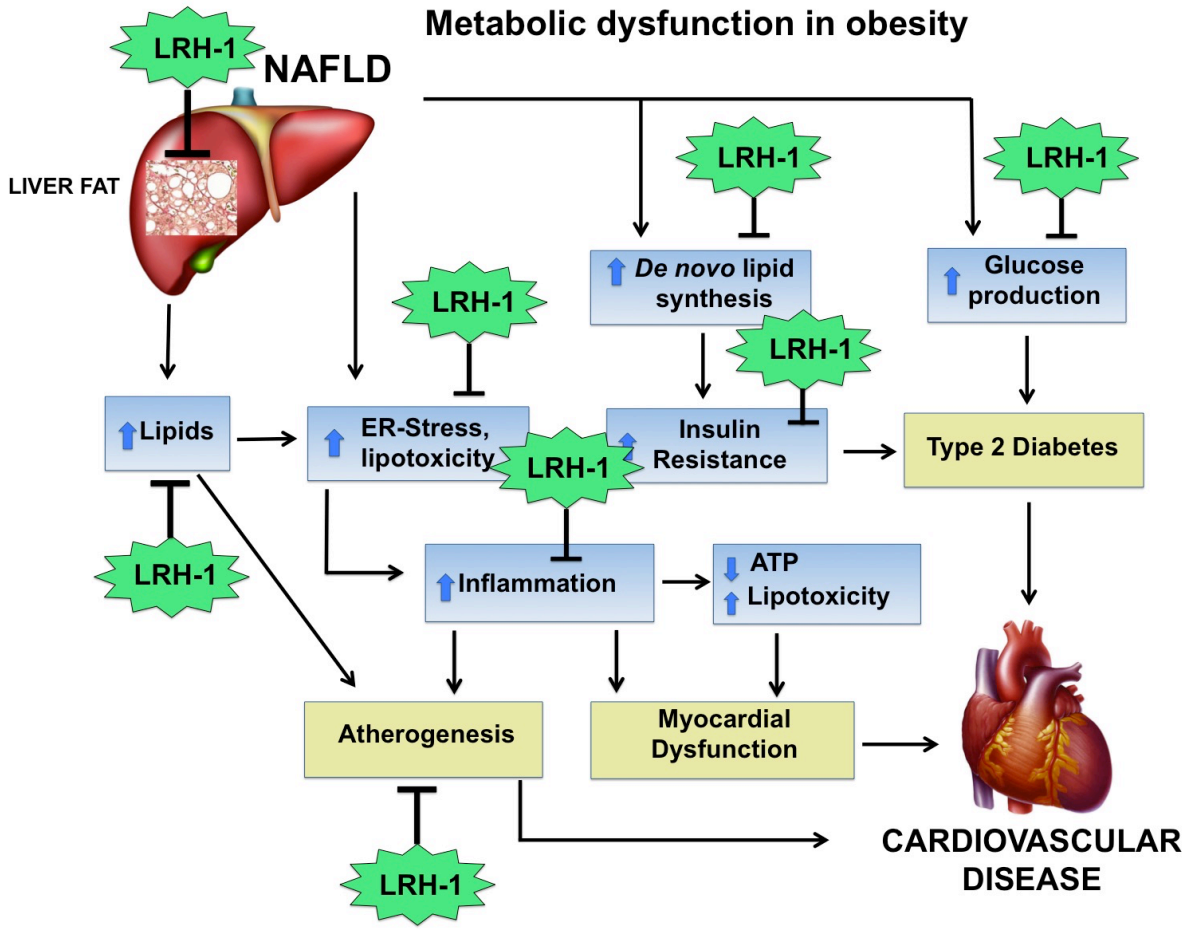


Figure 2.1. Therapeutic potential for LRH-1 agonists in obesity-associated diseases.

promotes cancer growth through estrogen receptor(Bianco et al., 2014; Chand et al., 2010; Clyne et al., 2004; Lai et al., 2013; Thiruchelvam et al., 2011) and β-catenin(Bayrer et al., 2015) signaling, as well as effects on glutamate metabolism(Xu, 2016).

2.4 Previous strategies for discovery of LRH-1 modulators

While discovery that DLPC is an LRH-1 agonist revealed the potential for targeting this receptor, natural phospholipids are not ideal for clinical or research applications due to poor solubility and metabolic instability. For example, orally administered phospholipids are readily hydrolyzed in the gut by pancreatic phospholipase A2 secreted in mammalian bile(Ridgway, 2013). This highlights the need for development of synthetic LRH-1 modulators. Previous strategies for discovery have mainly relied upon high throughput screens, as summarized below.

2.4.1 Antagonists.

LRH-1 antagonists have been highly sought for treatment of certain cancers in which LRH-1 drives proliferation of tumor cells (recently reviewed)(Nadolny and Dong, 2015). In the absence of a LRH-1-antagonist crystal structure, general approaches to discovering LRH-1 inhibitors have often involved *in silico* screening based on estrogen receptor (EstR) antagonists. Benod *et al* generated models for docking based on a phospholipid-bound LRH-1 structure (PDB 1YUC)(Ortlund et al., 2005) combined with features of inactive EstR (from the structure of EstR bound to 4-hydroxytamoxifen, PDB 3ERT)(Benod et al., 2013; Shiau, 1998). Sidechain positions of Leu517 (helix 11) and Asp350 (helix 3) in the LRH-1 structure were altered to imitate positions of analogous ER residues utilized for antagonist binding (Benod et al., 2013). Helix 12 was deleted to model the classical inactive conformation of NRs, in which this helix swings away from helices 3 and 4 in many possible positions. Docking a library of over 5 million compounds from the ChemBridge library into this model led to the discovery of Cpd3 (Figure 2.2). Cpd3 reduced mRNA expression of the LRH-1 target gene G0S2 in HEK293T cells overexpressing LRH-1 and inhibited cancer cell proliferation without causing general toxicity(Benod et al.,

2013). In luciferase reporter assays, Cpd3 inhibited LRH-1 but not closely-related NRs (i.e. Steroidogenic Factor-1 (SF-1), EstR, Androgen Receptor (AR), and Thyroid Receptor Beta (TRβ)). The authors used differential scanning fluorimetry (DSF) and surface plasmon resonance (SPR) to demonstrate direct binding of Cpd3 to the LRH-1 LBD; however, the precise binding mode was not elucidated.

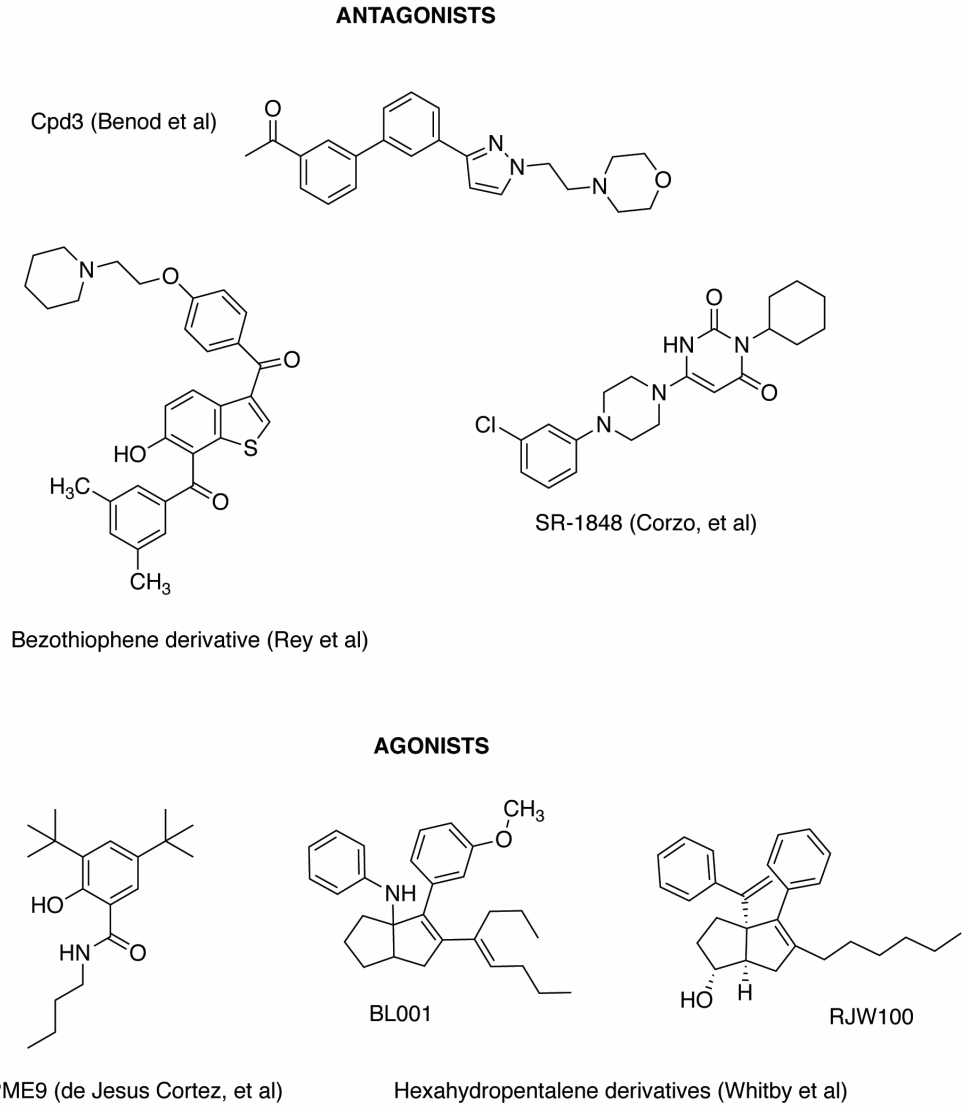


Figure 2.2. Chemical structures of select LRH-1 agonists and antagonists.

Rey et al. took a similar approach to LRH-1 antagonist discovery, using the ER selective modulator, raloxifene, as the template for high-throughput virtual screening of 50 million compounds in the ChemNavigator library(Rey, 2012). Hits from this screen were docked into a LRH-1 crystal structure with a weak agonist: a bacterially-derived phosphatidylglycerol (PDB 1YOK)(Krylova et al., 2005). This work led to the discovery of benzothiopene and derivatives (Figure 2.2), which were predicted to displace LRH-1 helix 12 in a similar manner as ER antagonists. The lead compound from these studies caused dissociation of LRH-1 from the PGC1α coactivator in mammalian 2 hybrid assays and inhibited transcriptional activity in luciferase reporter assays(Rey, 2012).

A third class of LRH-1 antagonist (also described as "inverse agonists") has been described by the Griffin group at Scripps, FL. These antagonists were discovered through high throughput screening for LRH-1-mediated inhibition at the CYP19 (aromatase) promoter(Busby, 2010; Corzo et al., 2015). The best compound from this screen (6-[4-(3-chlorophenyl) piperazin-1-yl]-3-cyclohexyl-1H-pyrimidine-2,4-dione, named SR-1848), repressed activity of endogenous LRH-1 in luciferase reporter assays, decreased expression of two known LRH-1 target genes, Cyclin D1 and Cyclin E1, and inhibited cellular proliferation in an LRH-1-dependent manner(Corzo et al., 2015). Direct binding of SR-1848 to LRH-1 has not been detected; however, this compound promotes dissociation of LRH-1 from DNA and its translocation into the cytosol, which has been proposed as a potential mechanism of action (Corzo et al., 2015).

2.4.2 Agonists.

High throughput screens have also been the primary tool for discovery of LRH-1 agonists. Cortez et al. used a screening approach involving disulfide trapping. An in-house library of disulfide-linked compounds was used to target LRH-1 residue C346 in the ligand binding pocket(de Jesus Cortez et al., 2016). Docking studies with the top hit from this screen predicted a binding mode in which the compound hydroxyl group hydrogen-bonded with residue

M345. Compounds synthesized around this scaffold led to the discovery of PME8 and PME9, which increased expression of a LRH-1 transcriptional target gene (CYP24A1) in HepG2 hepatocytes. Knockdown of LRH-1 ablated this activity, suggesting that this effect was LRH-1-dependent.

Richard Whitby and GlaxoSmithKline discovered a series of *cis*-bicyclo[3.3.0]-octene derivatives that act as LRH-1 agonists(Whitby et al., 2006). These were initially discovered in high throughput screens using a FRET-based coactivator association assay. Candidate agonists were evaluated by the ability to recruit the Tif2 coactivator to purified LRH-1 LBD protein. This work led to the discovery of GSK8470 (Figure 2.3), which exhibited a ~600 nM EC_{50} in this assay. The GSK8470-LRH-1 crystal structure revealed the agonist buried deep within the ligand binding pocket, utilizing mainly hydrophobic interactions (the exception was π- π stacking with LRH-1 residue H390 (Figure 2.3)(Whitby et al., 2011). This binding pose was used to design new agonists, resulting in the discovery of RJW100 (Figure 2.2). RJW100 was slightly more potent than GSK8470 in the FRET assay (292 nM EC_{50}). RJW100 also increased expression of mRNA levels of the LRH-1 transcriptional targets SHP, G0/G1 switch regulatory protein 2 (G0S2), and mesenchyme homeobox protein 1 (MEOX1) in HEK293 cells overexpressing LRH-1(Whitby et al., 2011). Limitations of agonists with this scaffold include fairly low potency (micromolar EC_{50}s in cellular activation assay), poor solubility, and a somewhat unclear structure-activity relationship.

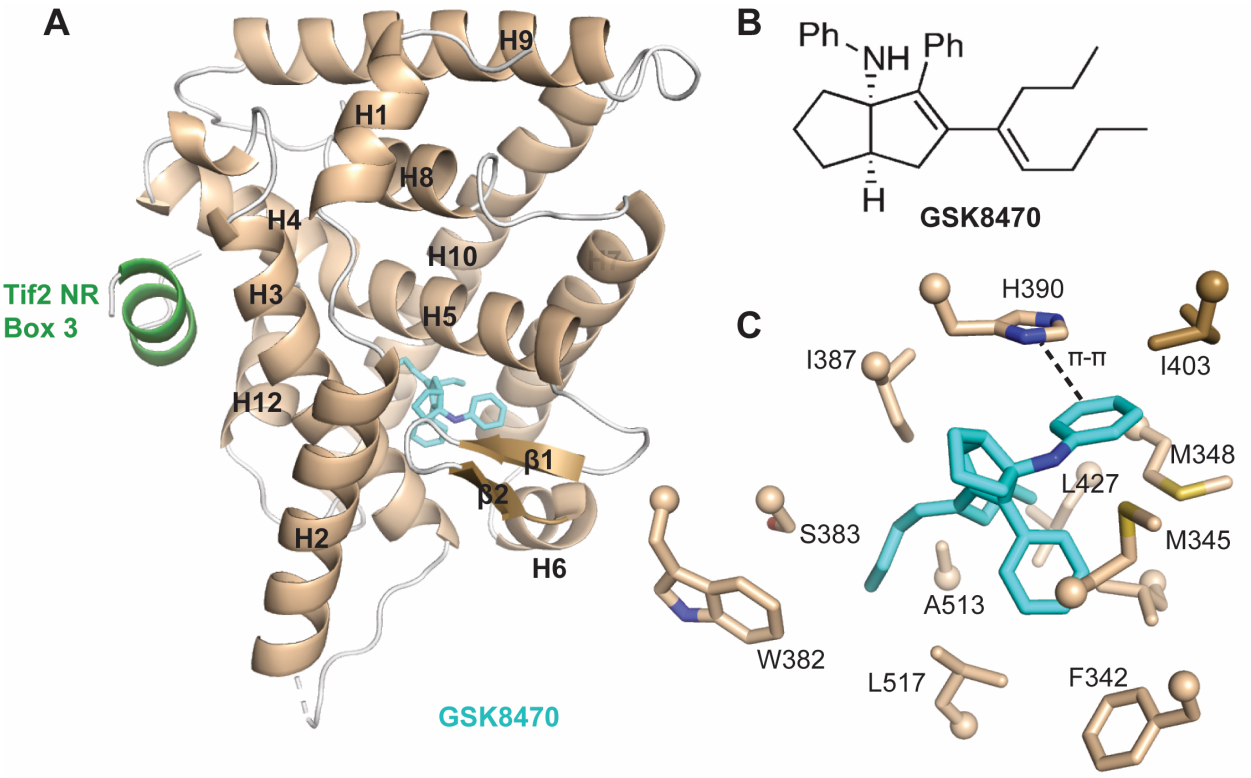


Figure 2.3. X-ray crystal structure of the LRH-1 LBD bound to the agonist GSK8470. A. Overall structure (PDB 3PLZ)(Whitby et al., 2011). The Tif2 coactivator peptide is shown in green. GSK8470 (cyan) is fully engulfed in the pocket. B. Chemical structure of GSK8470. C. Close view of the binding mode of the agonists, with interacting residues shown as sticks. Sidechains are colored according to atom type: N, blue; S, yellow; O, red.

2.5 Need for mechanistic studies to guide the design of LRH-1 modulators

While much effort has been devoted to discovering LRH-1 modulators using high-throughput screens, little is known about how synthetic molecules interact with LRH-1 and drive its activation or inhibition. We address this knowledge gap here by undertaking the first studies to explore mechanisms of action of LRH-1 modulators in detail. Preliminary differential scanning fluorimetry (DSF) experiments with a variety of experimental compounds in our lab (Figure 2.4) led us to focus on agonists with the RJW100 scaffold. RJW100 imparted the greatest stabilization of the LRH-1 LBD compared to a panel of experimental agonists, suggesting that it was binding the LRH-1 LBD and was likely displacing PLs in the pocket. We also expected that the stability of the complex would be favorable for crystallization studies. Using a combination of x-ray crystallography, in-cell luciferase reporter assays, and computational modeling, and other structural and biochemical experiments, we sought to understand agonist structural features important for activity. We also sought to understand similarities and differences in how synthetic agonists *versus* phospholipid agonists switch LRH-1 into the active state. Information gained from these studies was used to design and evaluate novel LRH-1 agonists.

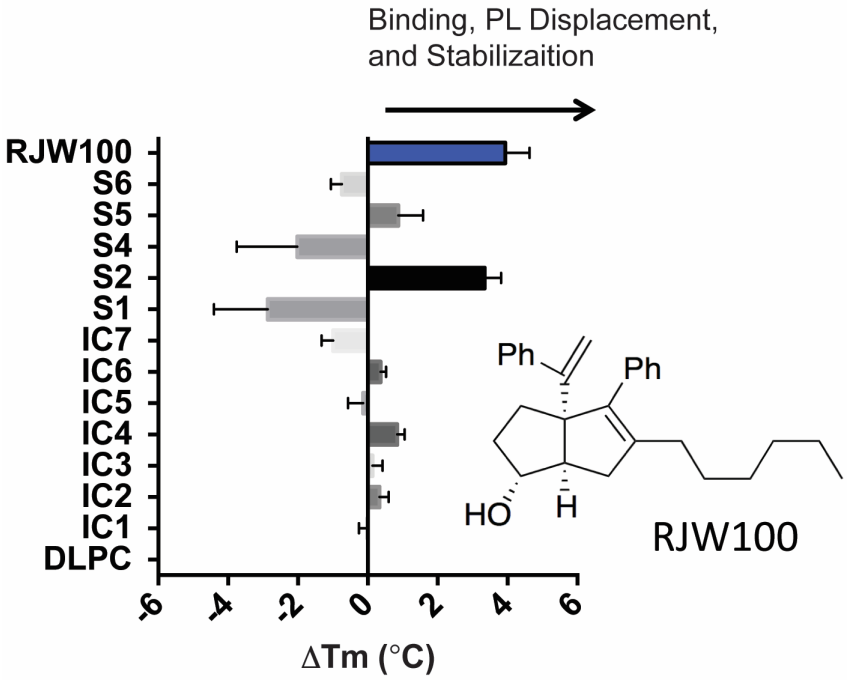


Figure 2.4. Preliminary DSF studies with experimental agonists. Results are presented as Tm difference relative to the LRH-1-DLPC complex. Results are presented as Tm difference relative to the LRH-1-DLPC complex. Each bar represents mean +/- SEM, for three biological replicates conducted in triplicate. Experiment was conducted using purified LRH-1 LBD in the presence of 20-fold molar excess of each ligand. A detailed description of the DSF method is provided in Chapter 3.

Chapter 3. Crystal Structures of LRH-1 with Synthetic Agonists Reveal a Novel Mechanism of Receptor Activation

3.1 Abstract.

Liver receptor homolog 1 (NR5A2, LRH-1) is an orphan nuclear hormone receptor that regulates diverse biological processes, including metabolism, proliferation, and the resolution of endoplasmic reticulum stress. While preclinical and cellular studies demonstrate that LRH-1 has great potential as a therapeutic target for metabolic diseases and cancer, development of LRH-1 modulators has been difficult. Recently, systematic modifications to one of the few known chemical scaffolds capable of activating LRH-1 failed to improve efficacy substantially. Moreover, mechanisms through which LRH-1 is activated by synthetic ligands are entirely unknown. Here, we use x-ray crystallography and other structural methods to explore conformational changes and receptor-ligand interactions associated with LRH-1 activation by a set of related agonists. Unlike phospholipid (PL) LRH-1 ligands, these agonists bind deep in the pocket and do not interact with residues near the mouth, nor do they expand the pocket like PLs. Unexpectedly, two closely related agonists with similar efficacies (GSK8470 and RJW100) exhibit completely different binding modes. The dramatic repositioning is influenced by a differential ability to establish stable, face-to-face π - π -stacking with LRH-1 residue H390, as well as by a novel polar interaction mediated by the RJW100 hydroxyl group. The differing binding modes result in distinct mechanisms of action for the two agonists. Finally, we identify a network of conserved water molecules near the ligand-binding site that are important for activation by both agonists. This work reveals a previously unappreciated complexity associated with LRH-1 agonist development and offers insights into rational design strategies.

This chapter is adapted from the manuscript:
Mays SG, Okafor CD, Whitby RJ, Goswami D, Stec J, Flynn AR, Dugan MC, Jui NT, Griffin PR, Ortlund EA. Crystal Structures of the Nuclear Receptor, Liver Receptor Homolog 1, Bound to Synthetic Agonists. *J Biol Chem* 2016; 291(49): 25281-25291.

3.2 Introduction.

Liver Receptor Homolog 1 (LRH-1; NR5A2) is a nuclear hormone receptor (NR) that controls expression of a diverse set of genes important both in normal physiology and disease. In addition to a vital role during development (Gu et al., 2005; Wagner et al., 2010), LRH-1 regulates many genes related to metabolism, proliferation, and cell survival. In the liver, LRH-1 regulates bile acid biosynthesis (Lu et al., 2000) and reverse cholesterol transport (Schoonjans et al., 2002; Stein et al., 2014), affecting hepatic and circulating cholesterol levels. Glucose metabolism is also regulated by LRH-1 at several points, including GLUT-4-mediated transport (Bolado-Carrancio et al., 2014) and glucose phosphorylation, the latter of which is essential for proper postprandial glucose sensing, flux through glycolysis and glycogenesis pathways, and *de novo* lipogenesis (Oosterveer et al., 2012). LRH-1 is a key mediator of the cell stress response through control of genes involved in the hepatic acute phase response (Venteclef et al., 2006), and in the cytoprotective resolution of endoplasmic reticulum stress (Mamrosh et al., 2014). Additionally, LRH-1 can be aberrantly overexpressed in certain cancers and can promote tumor growth through estrogen receptor and β-catenin signaling (Bayrer et al., 2015; Bianco et al., 2014; Chand et al., 2010; Clyne et al., 2004; Lai et al., 2013; Lin et al., 2014; Thiruchelvam et al., 2011).

Considering the breadth and significance of these physiological effects, LRH-1 modulators are highly desired as potential therapeutic agents. Chemical modulators would also be extremely useful as tools to dissect complex or temporal aspects of LRH-1 biology. However, development of LRH-1-targeted compounds has been challenging, in part due to a lipophilic binding pocket that becomes occupied with bacterial phospholipids (PL) in recombinant protein. Very few small molecules are able to displace these PL in library screens. Moreover, ligand-mediated regulation of LRH-1 is poorly understood. Endogenous ligands for LRH-1 are

unknown, but exogenous administration of dilauroylphosphatidylcholine (DLPC, PC 12:0/12:0) activates LRH-1 and has profound anti-diabetic effects *in vivo*, which are absent in a liver-specific LRH-1 knockout mouse (Lee et al., 2011). In addition to PCs, the signaling PL phosphatidylinositol 3,4,5-triphosphate (PIP3) binds LRH-1 (Krylova et al., 2005; Sablin et al., 2015), although downstream effects of this interaction have yet to be determined.

Typically, NRs are activated by a ligand-induced conformational change, which promotes recruitment of coactivator proteins to the activation function surface (AFS) in the ligand binding domain (LBD) to drive transcription. Our structural studies with DLPC have shown that, contrary to the canonical model of NR activation, LRH-1 relies on small conformational fluctuations to recruit co-activator or co-repressor proteins. These occur mainly in the AFS (comprised of portions of H3, H4, and the AF-H in the LBD), as well as in the H6/β-sheet region at a distal portion of the LBD (Musille et al., 2015; Musille et al., 2012). Flexibility in in the H6/β-sheet region is required for activation by PLs (Musille et al., 2012). Mechanisms through which LRH-1 is activated by synthetic ligands have not been explored but are likely quite different, given the differing structural composition of synthetic *versus* PL ligands.

There are very few known chemical scaffolds capable of activating LRH-1 above basal levels, the best studied of which are the *cis*-bicyclo[3.3.0]-octenes discovered by Whitby (Whitby et al., 2006; Whitby et al., 2011). The first compound described with this scaffold, named GSK8470 (Figure 3.1), was somewhat effective but was acid labile (Whitby et al., 2011). Substitution of the aniline group improved compound stability, and the GSK8470-LRH-1 crystal structure provided the basis for an extensive structure-activity relationship study (Whitby et al., 2011). One of the major objectives of this study was to introduce functional groups near select polar residues within the predominantly hydrophobic pocket. The new lead compound produced from this study, named RJW100, contains an *exo* hydroxyl group at the 1-position of the pentalene scaffold (indicated in red in Figure 3.1), intended to interact with LRH-1 residues H390

or R393. On the other hand, a diastereomer with *endo* stereochemistry (previously known as 24-endo, Figure 3.1) was not predicted to be able to make these interactions due to the alternative

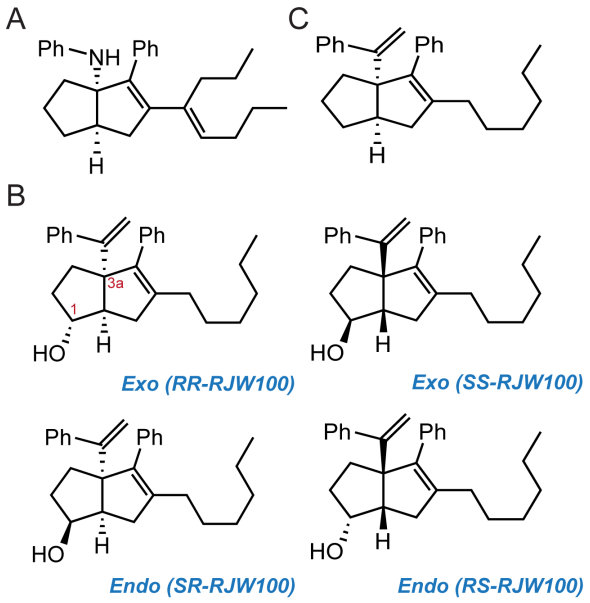


Figure 3.1. Chemical structures of LRH-1 agonists. A. GSK8470, the parent compound. B. RJW100 enantiomers. C. RJW100 analog lacking the hydroxyl group (named 18a), assayed in Figure 3.8.

conformation of the hydroxyl oxygen. The *endo* derivative was less active in biochemical assays, seeming to support this hypothesis. However, RJW100 was not much more potent or effective than GSK8470, and the study did not illuminate strategies for further improvement. In this work, we present crystal structures of RJW100 and its *endo* diastereomer bound to LRH-1. We demonstrate that these compounds bind quite differently than PLs and have distinct effects on protein dynamics compared to DLPC. Unexpectedly, these agonists also bind quite differently than the very closely related compound, GSK8470. We identify receptor-ligand interactions driving the repositioning and show that particular interactions are important for LRH-1 activation. These findings provide the first description of mechanisms involved in LRH-1 activation by synthetic molecules.

3.3 Results.

3.3.1. The Crystal Structure of RJW100 Bound to LRH-1.

To understand how RJW100 interacts with LRH-1 and affects receptor conformation, we determined the X-ray crystal structure of LRH-1 LBD bound to the agonist and to a fragment of the coactivator, Tif2, to a resolution of 1.85 Å (Table 1 and Figure 3.2A). Although the RJW100 used for crystallization was a racemic mixture of two *exo* stereoisomers (Figure 3.1), the electron density in the structure unambiguously indicates that a single enantiomer is bound (Figure 3.2B). The bound isomer has R stereochemistry at both the 1- (hydroxyl-substituted) and 3a- (styrene-substituted) positions (hereafter RR-RJW100, Figure 3.1). The ligand is bound at a single site deep in the binding pocket and is fully engulfed within it. This binding mode is markedly different than that of the PL ligands, DLPC and PIP3, which extend lower in the pocket with the headgroups protruding into the solvent (seen by superposition with PDBs 4DOS and 4RWV, respectively, Figure 3.2C-D). PL ligands also increase the pocket volume and width compared to RJW100. For example, the mouth of the pocket is ~3 Å wider and nearly 40% larger in volume when DLPC is bound *versus* RJW100 (Figure 3.2E and Table 2). This effect appears to be mainly

due to a shift of H6, which swings away from the mouth of the pocket in the DLPC structure by approximately 3 Å (Figure 3.2E). The direction and magnitude of the H6 movement is similar in published LRH-1-PL structures: comparison of four published human LRH-1-PL structures shows an average H6 shift of 3.0 +/- 0.2 Å relative to LRH-1 in the apo state or when synthetic ligands are bound (mean +/- SEM, Figure 3.2F). Although these structures exhibit diverse types of crystal packing, the movement of H6 appears to be related to whether the ligand is a PL or small molecule and not to crystal form or packing contacts. It likely occurs to avoid stearic clashes with the PL headgroup. Notably, the H6/β sheet region has been recently identified as a site through which PL ligands allosterically communicate with the activation function surface (AFS) to modulate LRH-1 activity (Musille et al., 2015; Musille et al., 2012). The fact that the synthetic agonists do not displace H6 relative to apo receptor suggests that they utilize a different mechanism for receptor activation.

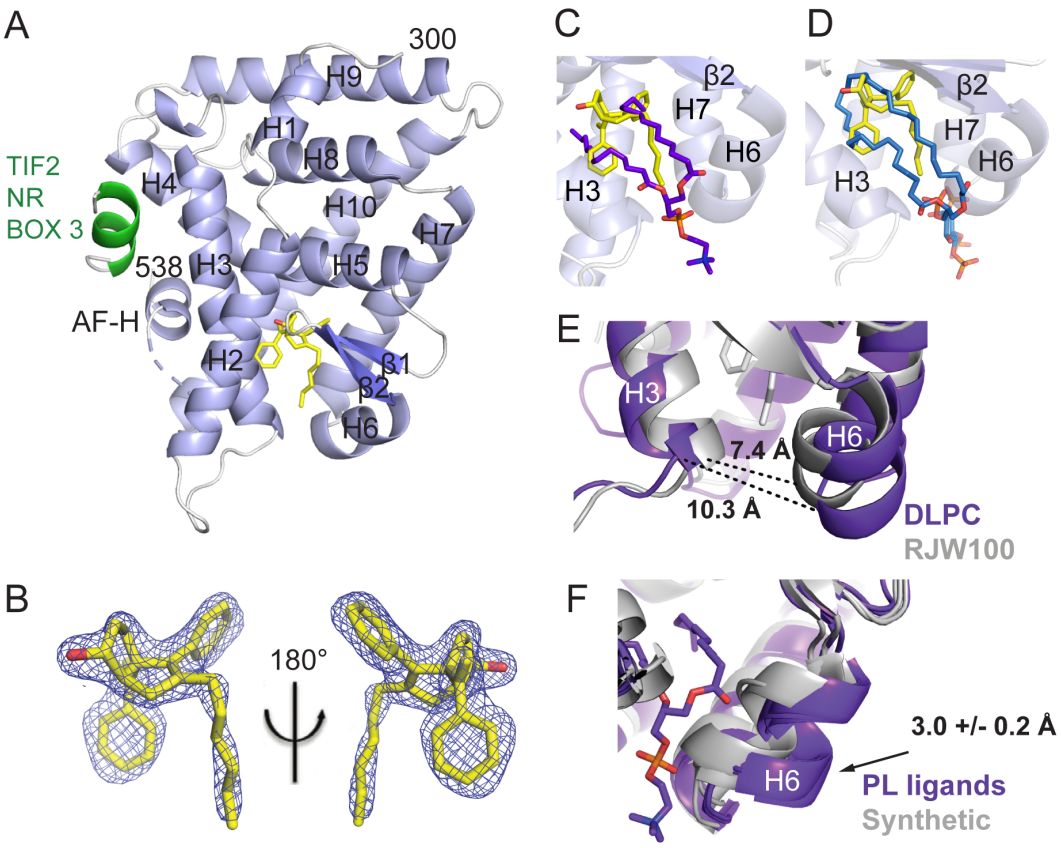


Figure 3.2. Crystal structure of LRH-1 ligand binding domain bound to RR-RJW100 and a fragment of the Tif2 coactivator. A. Overall structure, with α-helices shown in light blue and β-sheets in slate. The Tif2 peptide (green) is bound at the AFS. The ligand (yellow) is bound at a single site in the binding pocket. Dotted line, region of disorder in the protein backbone that could not be modeled. B. Omit map (F_O-F_C, contoured at 2.5 σ) showing that a single enantiomer of RJW100 is bound in the structure. C and D. Superposition of RR-RJW100 (yellow) with the ligand coordinates from (C) DLPC (purple, PDB 4DOS) or (D) PIP3 (blue, PDB 4RWV) show the very different binding mode of RR-RJW100 compared to the PL ligands. E. DLPC expands the width at the mouth of the pocket by approximately 3 Å compared to RR-RJW100. The width was measured from T341 to N419 (alpha carbons). F. Superposition of four PL-bound LRH-1 crystal structures (PDBs 4DOS (Musille et al., 2012), 1YUC (Ortlund et al., 2005), 4RWV (Sablin et al., 2015), and 4PLE (Musille et al., 2015), purple) and three structures of LRH-1

bound to synthetic agonists (PDB 3PLZ (Whitby et al., 2011) and the two structures from this paper, grey) showing the shift of helix 6 by the PL ligands. One PL ligand is shown to illustrate that the shift of H6 is likely due to a stearic clash with the PL headgroup. The number in the top right corner indicates the average distance that H6 shifts when PLs are bound relative to LRH-1 in the apo state or when synthetic ligands are bound (+/- SEM, determined by measuring distances between S418 alpha carbons). The apo LRH-1 structure used for measurements was PDB 4PLD (Musille et al., 2015).

Table 3.1. X-ray Data Collection and Refinement Statistics: LRH-1—RR-RJW100—Tif2 and LRH-1—SR-RJW100—Tif2

Values in parentheses indicate highest resolution shell.

Data collection	LRH-1 + RR-RJW100 +Tif2	LRH-1 + SR-RJW100 + Tif2
Space group	P422[a]	$P4_32_12$
Cell dimensions		
a, *b*, *c* (Å)	46.4 46.4, 220.5	46.3, 46.3, 220.0
α, β, γ (°)	90, 90, 90	90, 90, 90
Resolution (Å)	50−1.85 (1.92−1.85)	50−1.93 (2.00−1.93)
R_{pim}	0.03 (0.22)	0.02 (0.231)
$I / \sigma I$	28.5 (1.98)	33.5 (2.06)
*CC*1/2 in highest shell	0.878	0.861
Completeness (%)	96.8 (86.1)	99.7 (97.9)
Redundancy	8.9 (6.8)	11.1 (8.2)
Refinement		
Resolution (Å)	1.85	1.93
No. of reflections	21026	19073
R_{work}/R_{free} (%)	19.8/24.1	20.5/22.8
No. of atoms		
Protein	4038	4062
Water	92	51
B-factors		
Protein	38.1	46.9
Ligand	34.8	59.2
Water	36.8	43.5
Root mean square deviations		
Bond lengths (Å)	0.004	0.004
Bond angles (°)	0.578	0.588
Ramachandran favored (%)	98	98
Ramachandran outliers (%)	0	0
PDB accession code	5L11	5SYZ

[a] Data were processed in P422 but refined in $P4_32_12$.

Table 3.2. Binding Pocket Dimensions of LRH-1 in Various Liganded States

PDB code	Ligand	Width	Volume
		Å	*Å*^{3}
5SYZ	SR-RJW100	7.2	1083.0
3PLZ	GSK8470	7.3	957.6
5L11	RR-RJW100	7.4	1061.1
4PLD	Apo	8.0	965.8
4PLE	*E. coli* lipids	10.0	1368.7
4DOS	DLPC	10.3	1553.9
4RWV	PIP_3	12.0	1233.3

3.3.2 RJW100 Selectively Destabilizes Components of the AFS Relative to DLPC.

The overall LRH1-RJW100 structure depicts the AF-H in the active conformation and the Tif2 coactivator peptide bound at the AFS, as expected. However, there are a few indications that the active state may not be fully stabilized. There is substantial disorder in the loop connecting Helix 10 to the AF-H, and three residues within this loop cannot be modeled (dotted line in Fig. 2A). This loop is not disordered in published structures of LRH-1 bound to DLPC or GSK8470, and this does not appear to be related to favorable crystal packing in those structures (not shown). A second site of disorder in our structure occurs in the sidechain of residue E534 (not shown), which plays a critical role in securing coactivators via charge clamp (Li et al., 2003).

The disorder in the vicinity of the AFS in our structure suggests that this region is mobile when RJW100 is bound. To test this explicitly, we utilized solution-based hydrogen-deuterium exchange mass spectrometry (HDX-MS) to compare dynamics of purified LRH-1 LBD bound to enantiomerically pure RR-RJW100 or to DLPC. As shown in Figure 3.3A, RR-RJW100 destabilizes the bottom of H10 and the pre-AF-H loop *versus* DLPC, although the AF-H itself is slightly stabilized. Additional destabilization by RR-RJW100 occurs at the top part of H3 (part of the AFS), and in β2 (part of the alternate activation function region responsible for allosteric communication with the AFS, (Musille et al., 2015; Musille et al., 2012)). Interestingly, this localized destabilization occurs even though RJW100 increases overall thermostability of LRH-1 compared to DLPC (Figure 3.3B). This finding may indicate room for improvement in agonist design: compounds that provide a stable surface for coactivator recruitment would likely be more potent activators.

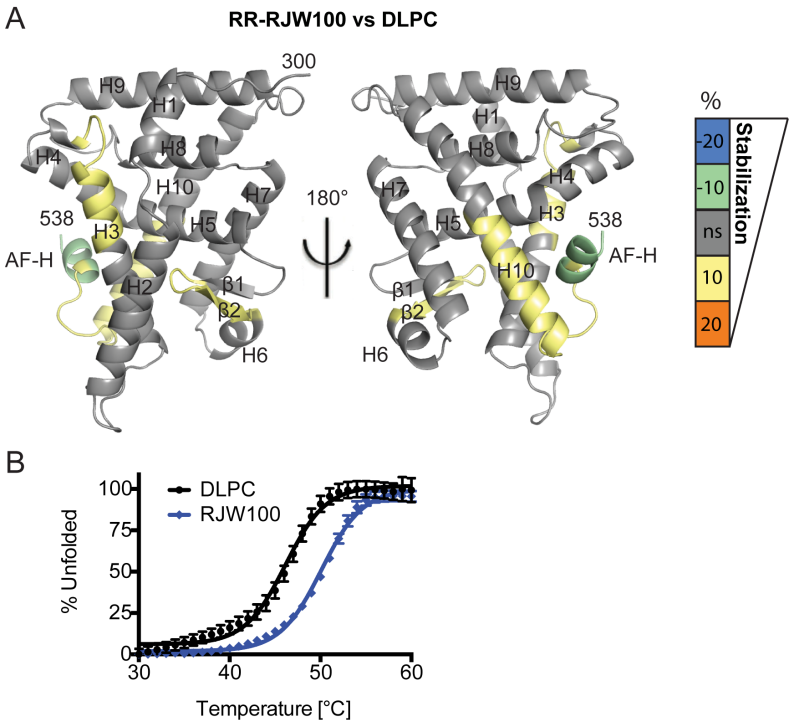


Figure 3.3. Differential effects on LRH-1 dynamics when RR-RJW100 is bound versus DLPC. A. HDX was used to probe differential effects on protein dynamics by the two ligands. The scale refers to the difference in percent deuterium incorporation for RR-RJW100-bound LRH-1 minus DLPC-bound LRH-1). The scale reflects the difference in average percent deuterium incorporation for RR-RJW100- *versus* DLPC- bound LRH-1 LBD. For example, negative numbers reflect slower deuterium incorporation (less motion) for RR-RJW100 *versus* DLPC. Results are mapped onto PDB 4DOS (Musille et al., 2012). B. DSF curves showing that RJW100 increases overall LRH-1 thermostability compared to DLPC. Each point represents the mean +/- SEM of values for three independent experiments, each conducted in triplicate.

3.3.3 Dramatic Repositioning of RJW100 Compared to a Closely Related Synthetic Agonist.

Perhaps the most striking observation from our structure comes from comparison with GSK8470-bound LRH-1 (PBD 3PLZ). Overall protein conformation is highly similar: the largest movement occurs in the bottom of H3, which moves in the direction of H6 (by 2 Å in the RR-RJW100 structure and by 4 Å in the GSK8470 structure relative to Apo-LRH-1, Figure 3.4A). However, there is a substantial difference in the positioning of these agonists within the binding pocket. Although GSK8470 and RR-RJW100 bind in the same vicinity, they are rotated nearly 180° from one another. The bicyclic rings at the cores of each molecule are perpendicular to each other, causing the tails to be pointed in opposite directions (Figure 3.4B). Notably, the rationale for adding a hydroxyl group in the 1 position on this scaffold was to promote an interaction with a "polar patch," consisting of residues R393 and H390 in an otherwise hydrophobic pocket (Whitby et al., 2011). This interaction was predicted based on the position of the ligand in the LRH-1-GSK8470 structure; however, the actual position of RR-RJW100 in the pocket places the hydroxyl group over 6 Å away from these residues (Figure 3.4C). Such a radically different binding mode for closely related molecules was unexpected, and a propensity to rotate within the pocket may contribute to difficulties improving agonist activity by modification of the GSK8470 scaffold.

3.3.4 The crystal structure of LRH-1 bound to a RJW100 diastereomer.

To further explore the effect of agonist structure on ligand binding mode, we determined the 1.93 Å crystal structure of LRH-1 bound to *endo*-RJW100, also in complex with the Tif2 coactivator peptide (Table 1 and Figure 3.5A). Overall protein conformation is highly similar to that of the RR-RJW100 (*exo)* structure, and there is disorder at the same portion of the protein backbone within the pre-AFH loop (Figure 3.5A).

Like *exo*-RJW100, *endo*-RJW100 consists of a mixture of two enantiomers, only one of which is bound in the crystal structure. The presence of a single enantiomer is quite clear from the electron density surrounding the ligand, although the density for the ligand tail is much weaker

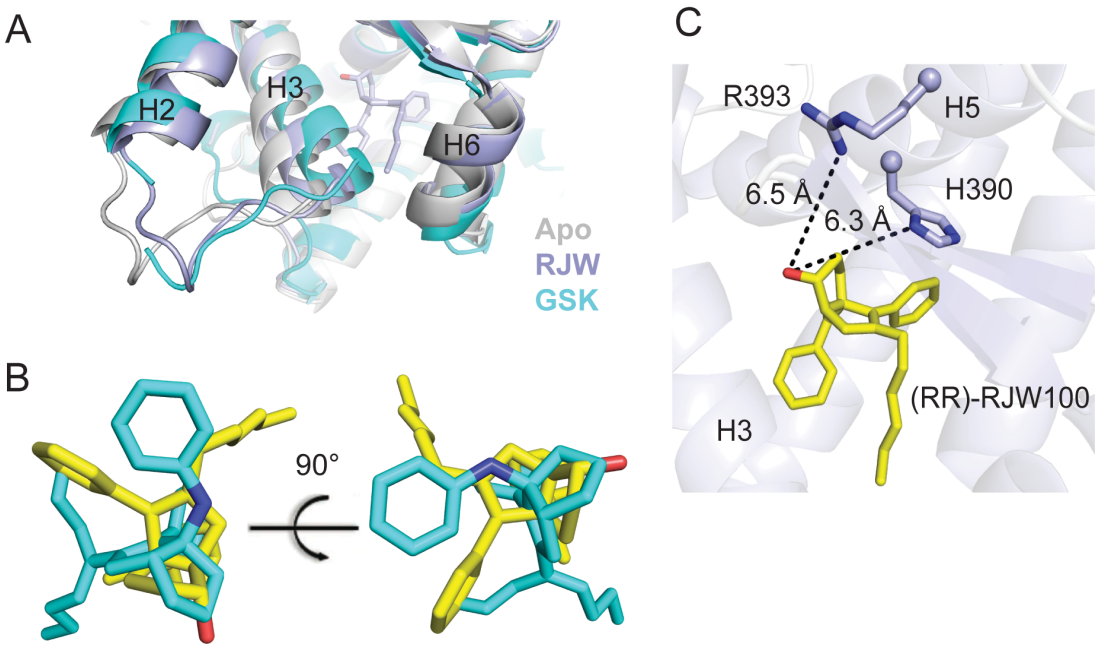


Figure 3.4 A very different binding mode of RR-RJW100 compared to a closely related synthetic agonist. A. Close-view of the bottom of the receptor, showing the shift in H3 induced by both synthetic agonists compared to Apo LRH-1. B. Superposition of coordinates for GSK8470 (cyan, from PBD 3PLZ) and RR-RJW100 (yellow). C. The RR-RJW100 hydroxyl group was predicted to interact with residues H390 and R393, but it is over 6 Å away from these residues in our structure.

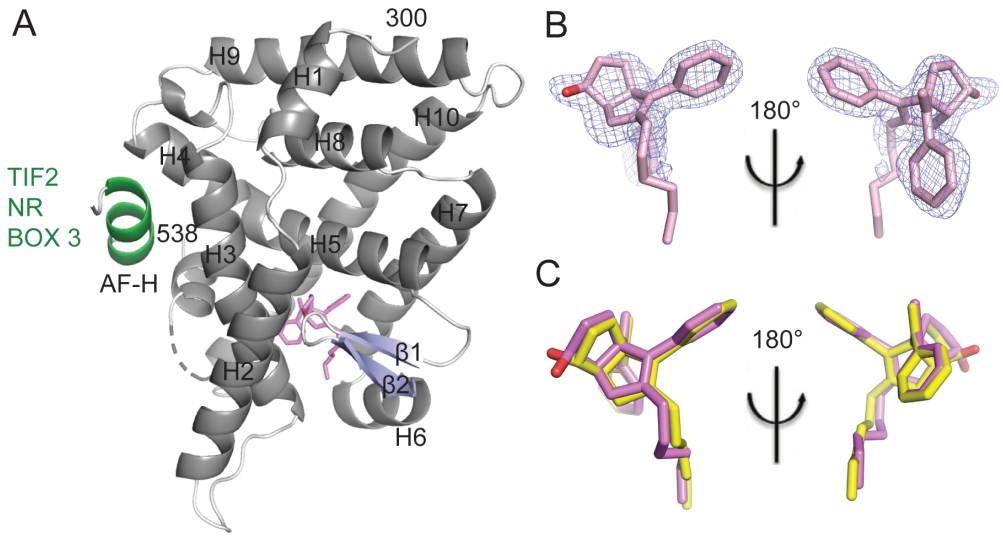


Figure 3.5. Crystal structure of LRH-1 bound to an RJW100 enantiomer with endo stereochemistry. A. The overall structure of the LRH-1 ligand binding domain with SR-RJW100 (violet) and a peptide derived from the Tif2 coactivator (green). *Dotted line,* region of disorder in the protein backbone that could not be modeled. B. Omit map (F_O-F_C, contoured at 2.5 σ) showing that a single enantiomer of RJW100 is bound in the structure. C. Superposition of SR-RJW100 with RR-RJW100 (yellow) showing a very similar position in the binding pocket.

than in the RR-RJW100 structure (Figure 3.5B). The bound enantiomer has R stereochemistry at the styrene moiety, as in the *exo*-RJW100 structure, but the hydroxyl group is pointed in the opposite direction (designated SR-RJW100, Figure 3.3.1). Superposition of the ligand coordinates from the RR-RJW100 and SR-RJW100 structures reveals nearly identical positioning, with the exception of the hydroxyl group (Figure 3.5C). However, a key difference is seen in the ligand B-factors. For RR-RJW100, the average ligand B-factor is 34.8, which is slightly less than the average protein B (38.1, Table 1). In contrast, the ligand B-factor for SR-RJW100 is much higher compared to the protein B-factor (59.2 *versus* 46.9, Table 1). The higher average ligand B-factors do not arise solely from the disordered atoms in the ligand tail: for example, the B-factor for the hydroxyl oxygen of SR-RJW100 is 59. This is an indication of more atomic motion for SR-RJW100 compared to RR-RJW100, suggesting that the ligand has a reduced ability to engage in stabilizing intermolecular interactions in the pocket.

3.3.5 Discovery of a Novel LRH-1 Interaction Mediated by the RJW100 Hydroxyl Group.

Protein-ligand interactions made by GSK8470 and the RJW100 stereoisomers were examined to gain insight into factors influencing ligand-binding mode. A close view of the LRH-1 binding pocket reveals that RR-RJW100 makes several hydrophobic contacts, many of which are also made by GSK8470 (shown in cyan, Figure 3.6A). Additionally, RR-RJW100 makes several unique contacts (shown in grey, Figure 3.6A). Many of these unique contacts are also hydrophobic; however, the RR-RJW100 hydroxyl group forms an indirect polar contact with residue T352 via a water molecule. A portion of the electron density map is shown in Figure 3.6A to emphasize the strong evidence for this interaction. SR-RJW100 also interacts with T352 through the same water molecule, despite the differing conformation of the hydroxyl group (Figure 3.6B). The position of the SR-RJW100 hydroxyl group also permits a second water-mediated hydrogen-bonding interaction with the backbone nitrogen of residue V406 (Figure 3.6B). Although the interaction with T352 is indirect, we observed that the water molecule

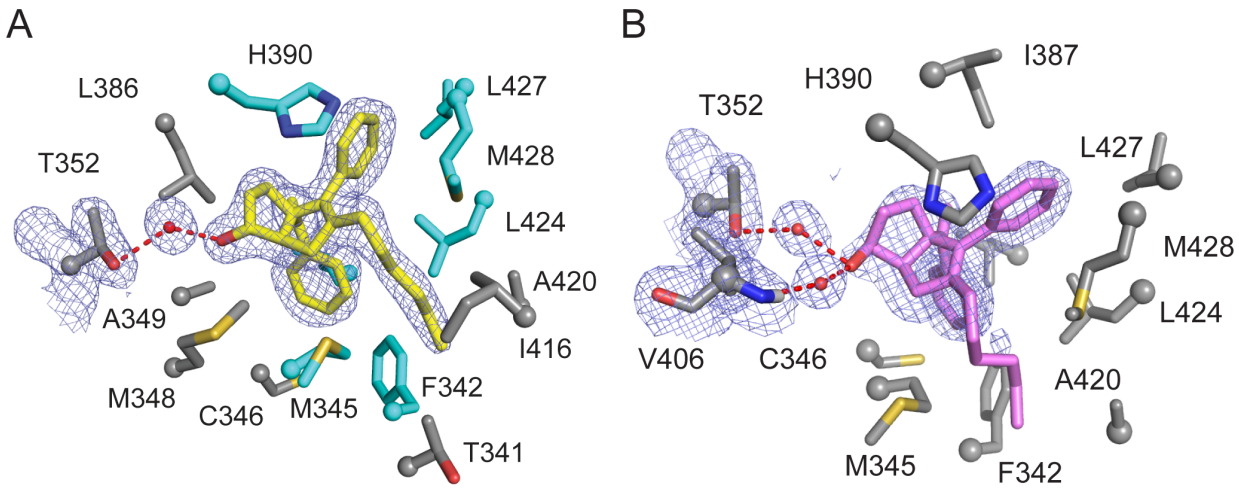


Figure 3.6. Residues interacting with LRH-1 agonists. Close views of the binding pockets from the structures of LRH-1 bound to (*A)* RR-RJW100 or (*B)* SR-RJW100, depicting sidechains of amino acid residues that interact with each ligand. In *A,* residues that also interact with GSK8470 are shown in cyan, while unique interactions made by RR-RJW100 are shown in grey. Portions of the electron density maps are shown to highlight the interactions with T352 through water (F_O-F_C, contoured to 1σ).

involved is part of a network of waters found in every LRH-1 crystal structure in the same location (for examples see PBDs 4DOS, 1YUC, 3PLZ, and 4DOR, as well as Figure 3.7 in this paper). Additional support for modeling water at these sites in the pocket comes from *B*-factors of ligating atoms (Table 3.3). Thus, this water network appears to be a conserved feature of the binding pocket and may play a role in receptor function or stability. To test the hypothesis that the OH-water-T352 interaction was influencing ligand positioning, we analyzed the stability of this bond using molecular dynamics simulations (MDS). Throughout each simulation (200 ns), the four conserved networked water molecules remained in the same positions (if a particular water molecule occasionally left, it was immediately replaced with another in the same location). Residue T352 maintained a hydrogen bond with the water molecule for 100% of each simulation, regardless of which ligand was bound. Additionally, both RR-RJW100 and SR-RJW100 maintained hydrogen bonding with the water molecule for the majority of the simulations (53.7% of the time for RR-RJW100 and 64.4% of the time for SR-RJW100). When residue T352 was mutated to valine in MDS, the time spent interacting with the T352-coordinated water molecule was drastically reduced (22.9% and 0.5% when RR-RJW100 and SR-RJW100 were bound, respectively), demonstrating that this mutation likely disrupts this water-mediated interaction made by these ligands.

3.3.6 Differences in π-π-Stacking with Residue H390 among LRH-1 Agonists.

π-π-Stacking of GSK8470 with residue H390 has been previously described and is hypothesized to be critical for activation of LRH-1 by synthetic compounds (Lalit et al., 2013; Whitby et al., 2011). The RJW100 diastereomers also engage in π-π-stacking with H390, but with some key differences. The π-π-stacking is face-to-face for GSK8470 and edge-to-face for the RJW100 isomers (Figure 3.7). Additionally, by virtue of the very different orientations in the binding pocket, the agonists do not use analogous phenyl rings for π-π-stacking: GSK8470 uses the aniline group, whereas the RJW100 isomers use the adjacent phenyl substituent. Moreover,

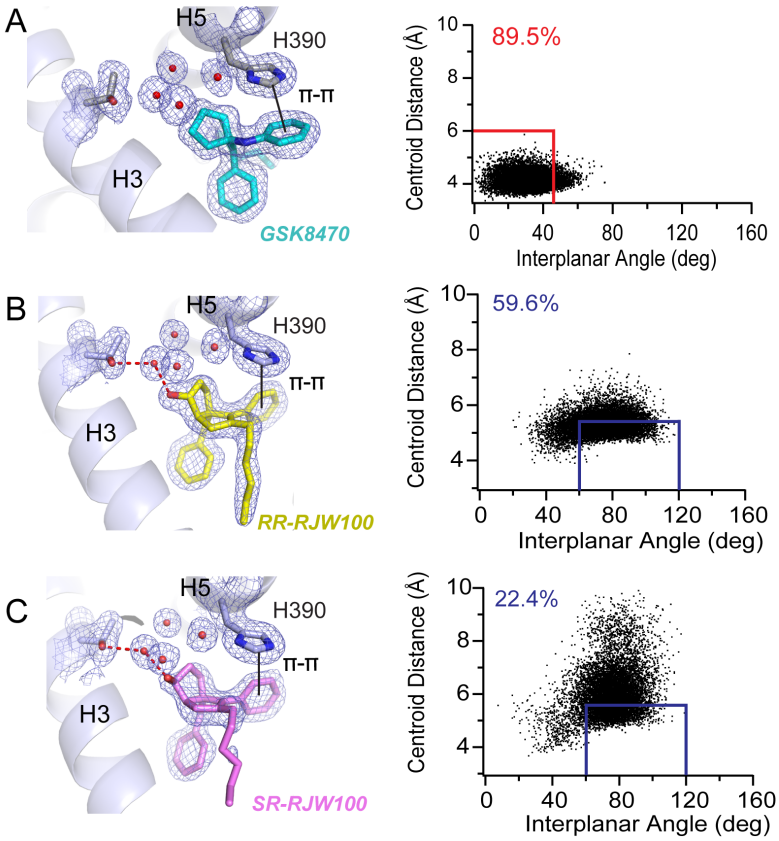


Figure 3.7. π-π-stacking with residue H390 differs among agonists. A-C. *Left,* views of the different types of π-π stacking utilized by *(A)* GSK8470, *(B)* RR-RJW100, and *(C)* SR-RJW100. *Right,* MDS monitoring the distances between ring centroids (x-axis) and angle between the ring planes (y-axis) for the ligand phenyl group and H390 at each time increment of the 200 ns MDS. The red and blue boxes indicate when face-to-face and edge-to-face π-π stacking occurred, respectively. The numbers in the top right corners indicate the percentage of time spent π-π stacking during the MDS.

Table 3.3. B-factors for waters near the ligand and their ligating atoms.

W indicates water; O indicates oxygen; N indicates nitrogen; NE indicates the ϵ nitrogen; and OG1 indicates the γ1 oxygen.

Water	Ligating atom 1	Ligating atom 2	Ligating atom 3	Ligating atom 4[a]
PDB 5L11				
W1 (22.5)	RJW,O (30.8)	Thr-352, OG1 (22.9)	W2 (23.3)	W5 (22.8)
W2 (23.2)	Val-406,N (24.2)	W1 (22.5)	W9 (24.3)	
W5 (22.8)	Leu-386,O (25.2)	Asp-389, OD2 (28.6)	W1 (22.5)	
W9 (24.3)	His-390, ND1 (23.1)	Arg-393, NH1 (28.6)	W2 (23.3)	
PDB 5SYZ				
W17 (34.5)	ENO,O (58.8)	Thr-352, OG1 (28.9)	W16 (25.9)	W4 (32.5)
W16 (25.9)	Leu-386,O (26.3)	Asp-389, OD2 (34.3)	W17	
W3 (32.0)	His-390, ND1 (30.7)	Arg-393, NH1 (33.0)	Phe-404,O (31.9)	W4
W4 (32.5)	ENO,O (58.8)	Val-406,N (32.8)	W3	
PDB 3PLZ, chain A				
W41 (19.5)	Thr-352, OG1 (14.4)	Arg-393, NH2 (21.5)	W16 (18.1)	W13
W16 (18.1)	Asp-389, OD2 (19.4)	Arg-393, NH2 (21.5)	W41 (19.4)	
W13(21.2)	Val-406,N (20.4)	W41	W31	
W31 (21.1)	Phe-404,O (19.2)	Arg-393, NE (22.0)	His-390, ND1 (15.6)	W13
PDB 3PLZ, chain B				
W25 (15.9)	Thr-352, OG1 (11.7)	W14 (13.8)	W129 (14.2)	
W14 (13.8)	Leu-386,O (13.8)	Asp-389, OD2 (18.7)	W25 (15.9)	W52 (18.1)
W129 (14.2)	Val-406,N (20.4)	W25 (15.9)	W52 (18.1)	
W52 (18.1)	Phe-404,O (13.6)	Arg-393, NH2 (22.7)	His-390, ND1 (12.1)	

[a] When a fourth ligating atom is identified, it is likely that the waters intermittently interact with the four atoms.

MDS demonstrate ligand-dependent differences in the stability of this interaction. For this latter analysis, time spend in π-π-stacking was quantified over the course of the 200 ns MDS. Face-to-face π-π-stacking was defined as a distance between ring centroids of < 6 A, an angle between ring planes < 45°, and an angle between the centroid-centroid vector and one plane < 60° (Zhang et al., 2007). Edge-to-face stacking was defined as a distance of < 5.5 Å between ring centroids and an angle between the ring planes between 60° and 120° (Schrodinger, 2016; Zhang et al., 2007). Applying these criteria as appropriate for the type of π-π-stacking made by each ligand revealed that GSK8470 maintained π-π-stacking with H390 for most of the simulation (89.5% of the time, Figure 3.7A). The edge-to-face π-π-stacking made by RR-RJW100 was also fairly stable, maintained for 59.6% of the MDS time (Figure 3.7B). In contrast, SR-RJW100 made this interaction much less frequently (22.4% of the time) and exhibited much more motion than RR-RJW100 (Figure 3.7C). These observations are consistent with the relatively high B-factors seen for SR-RJW100 in the crystal structure.

3.3.7 Role of T352 and H390 in LRH-1 Activation by Synthetic Agonists.

The importance of the T352 and H390 interactions for binding and activation of LRH-1 by the agonists was investigated using mutagenesis. Binding and stabilization of LRH-1 were detected using DSF. While a T352V mutation (designed to remove the water-mediated hydrogen bond with bound ligands) had little effect on the overall thermostability of DLPC-bound LRH-1, it completely abrogated the stabilizing effect of RR-RJW100 and SR-RJW100 (Figure 3.8A). Likewise, disrupting this interaction by using a RJW100 analog lacking the hydroxyl group (named 18a) prevented the positive Tm shift in wild-type (WT) LRH-1 (Figure 3.8B). GSK8470 did not affect the melting profile of LRH-1 in WT or T352V protein, supporting the notion that the hydroxyl group is important for stabilizing the protein-ligand complex.

The T352 interaction was also found to be important for LRH-1 activation by small molecule agonists. Compound 18a, lacking the hydroxyl group and unable to make this interaction, was an extremely poor LRH-1 activator in luciferase reporter assays (Figure 3.8C). *Endo*-RJW100 was

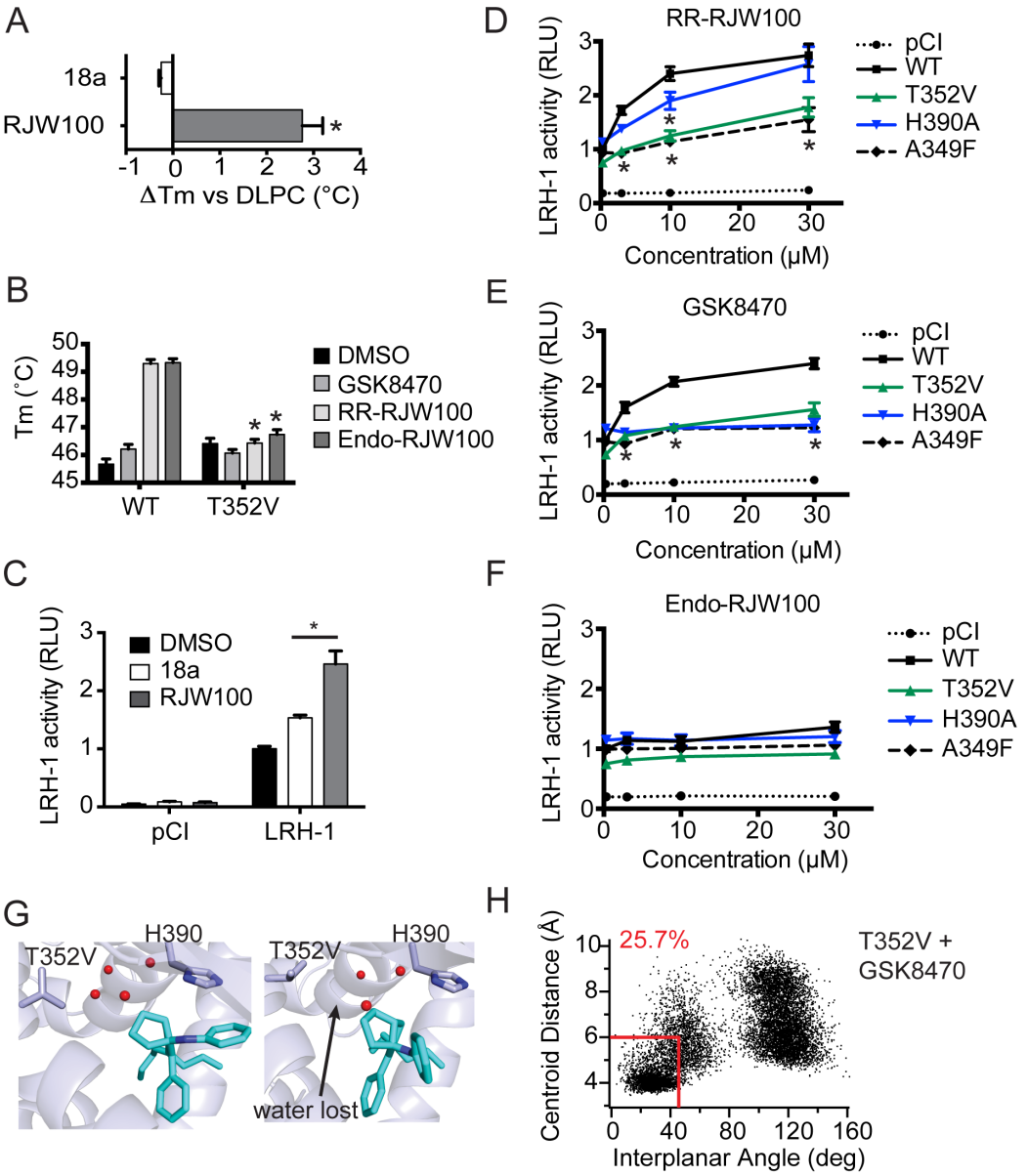


Figure 3.8. Importance of protein-ligand interactions on ligand binding and activity.

A. The analog 18a, lacking the hydroxyl group, does not stabilize wild-type LRH-1 in DSF assays. B. Introduction of the T352V mutation to LRH-1 ablates the stabilizing effects of RR-RJW00 and Endo-RJW100. C. Compound 18a, which lacks a hydroxyl group, is a significantly weaker agonist in luciferase reporter assays. D-F. Luciferase reporter assays measuring LRH-1 activity, using the SHP-luc reporter. Values have been normalized to constitutive Renilla luciferase signal and are presented as fold change versus wild-type LRH-1 + DMSO. The A349F

mutation introduces a bulky aromatic side chain, which blocks the binding pocket and prevents binding of synthetic ligands (Benod et al., 2013). This was used as a negative control. G. Snapshots from MDS using T352V LRH-1 with GSK8470 bound. H. Plot of distances between ring centroids (x-axis) and angle between the ring planes (y-axis) for the GSK8470 phenyl group and H390 at each time increment of the 200 ns with T352V LRH-1 (as described for Figure 7). For A-F, each bar (or point, for panels D-F) represents the mean +/- SEM for three independent experiments, each conducted in triplicate. * $p < 0.05$ (significance was determined by two-way ANOVA followed by Tukey's multiple comparisons test).

also a weak agonist, as previously reported (Whitby et al., 2011), although statistically significant activation was achieved at the highest dose with WT LRH-1 (~1.4-fold over DMSO, Figure 3.8F). RR-RJW100 and GSK8470 were equally effective toward WT LRH-1: both increased activity by approximately 2.5-fold compared to DMSO at the highest dose, and both had EC_{50}s of around 4 μM (Figure 3.8D-E). Notably, the T352V mutation greatly reduced the ability of RR-RJW100 to activate LRH-1 compared to with WT protein, while not significantly affecting baseline activity (Figure 3.8D). Unexpectedly, this mutation similarly attenuated activation by GSK8470, perhaps suggesting a broader role for this residue (or perhaps for the water network it coordinates) in ligand-mediated activation. Indeed, introduction of a T352V mutation to GSK8470-bound LRH-1 in MDS disrupts the water network, causing complete displacement of the water molecule typically coordinated by T352 (Figure 3.8G). The T352V mutation also significantly reduces the amount of time GSK8470 spends π-π-stacking with H390 (25.7% *versus* 89.5% of the simulation, Figures 3.8H and 3.7A). Both the destabilization of the water network and the disruption of stable H390 π-π-stacking by the T352V mutation could contribute to the observed loss of activity for GSK8470 in the context of this mutation.

While the T352V mutation resulted in a loss of activity for both RR-RJW100 and GSK8470, mutating H390 to alanine had a different effect on LRH-1 activation depending on the agonist involved. GSK8470 was completely unable to activate H390A-LRH-1, but this mutation had little to no effect on RR-RJW100-mediated activation (Figure 3.8D-E). This differential reliance on H390 for activation is consistent with the observation that GSK8470 interacts with H390 more stably than RR-RJW100 in MDS. This also provides evidence that RR-RJW100 must utilize a different mechanism of action than GSK8470 for LRH-1 activation.

3.3.8 Capacity for productive π-π-stacking with residue H390 influences agonist positioning. Considering the stable nature of the π-π-stacking GSK8470 with residue H390, reasons that RJW100 would abandon this strong interaction for an entirely different binding mode were unclear. To answer this question, we artificially re-oriented RR-RJW100 in our structure to be

aligned with GSK8470, inspected the fit, and looked at predicted interactions using MDS. While this repositioning did not produce any obvious clashes, the planarity of the styrene moiety restricted its mobility. This constraint prevents the rotation necessary for ideal face-to-face π-π-stacking with H390 (Figure 3.9A). Although MDS using this re-oriented ligand demonstrated fairly stable face-to-face π-π-stacking for re-oriented RR-RJW100, the interaction was less stable than GSK8470 (present 71.1% of the time *versus* 89.5% for GSK8470, $p < 0.0001$, Figure 3.9B). There was increased variability in both ring centroid distances and angle between the rings involved in this interaction for repositioned RR-RJW100 (mean centroid distances +/- standard deviations were 5.1 +/- 0.6 Å *vs* 4.2 +/- 0.3 Å for GSK8470, and mean angles were 40 +/- 20 *vs* 30 +/- 10 degrees for GSK8470). This variability indicates more motion of the rings involved in the *π-π*-stacking for the repositioned ligand. Another interesting observation with this MDS pertained to the water network in the ligand-binding pocket. The hydroxyl group of the repositioned ligand was initially within hydrogen-bonding distance of the T352-coordinated water molecule. Unlike native RR-RJW100, however, the hydroxyl group in the repositioned ligand was also within hydrogen- bonding distance of the three other water molecules in the immediate vicinity (Figure 3.9A, right panel). During the MDS, the water network was very quickly disrupted, with a complete loss one of the four conserved water molecules near the ligand hydroxyl group. One of the remaining three waters moved to coordinate T352, while the ligand hydroxyl group alternated between interactions with each of the three waters without stably interacting with a single water molecule (data not shown). Thus, the water-mediated contact of the RR-RJW100 hydroxyl group with T352 is not maintained for the reoriented ligand. Moreover, the loss of a typically conserved water molecule supports the idea that the ligand-binding pocket is more dynamic with the artificially repositioned ligand. Therefore, it appears that an impaired ability of the styrene phenyl ring to interact with H390, combined with a favorable interaction mediated by the hydroxyl group with the T352-coordinated water, is

responsible for the dramatically different position adopted by RJW100 compared to a structurally very similar agonist.

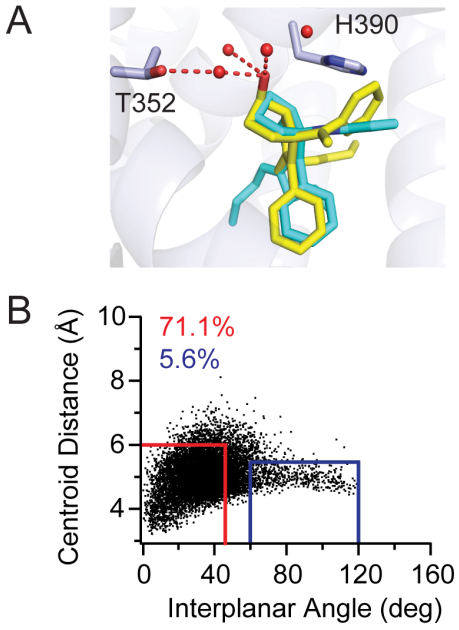


Figure 3.9. Impaired π-π stacking of the RR-RJW100 styrene group in MDS. A. RR-RJW100 (yellow) was artificially aligned with GSK8470 (cyan) for MDS to study its ability to π-π-stack with residue H390. Red dotted lines indicate water molecules that are within hydrogen bonding distances of the ligand or residue T352. B. Plot of distances between ring centroids (x-axis) and angle between the ring planes (y-axis) for the ligand phenyl group and H390 at each time increment of the 200 ns MDS for the styrene phenyl ring of repositioned RR-RJW100 and residue H390 over the course of the simulation. The red and blue boxes indicate when face-to-face and edge-to-face π-π stacking occurred, respectively. The numbers in the top right corners indicate the percentage of time π-π stacking during the MDS.

3.4 Discussion.

Although LRH-1 synthetic modulators are highly sought as pharmacological tools and as potential therapeutic agents, a limited understanding of ligand characteristics important for binding and activating LRH-1 has impeded agonist development. This work represents the first detailed exploration of structural mechanisms governing regulation of LRH-1 by synthetic ligands. Relative to the PL LRH-1 agonist, DLPC, the current best agonist (RR-RJW100) constricts the binding pocket and destabilizes portions of the AFS (Figure 3.3.1-2). In future studies, it will be interesting to investigate the causes of this latter effect, since stabilization of the AFS may facilitate co-activator binding, leading to greater potency or efficacy. Alternatively, analogs designed to enhance the AFS destabilization may be effective antagonists or inverse agonists.

In a previous study, RJW100 was the most effective of a large series of GSK8470 derivatives but still only modestly increased LRH-1 activation (Whitby et al., 2011). Indeed, we find that these two agonists are statistically indistinguishable in luciferase reporter assays measuring LRH-1 activity (Figure 3.8D). Given the similarities in structures and efficacies for these ligands, we expected them to utilize similar mechanisms of action; however, this is not the case. Our crystal structure reveals a dramatically different binding mode for RR-RJW100 compared to GSK8470 (Figure 3.4). While this was surprising, it is not unreasonable, considering that LRH-1 has a very large hydrophobic binding pocket and that these agonists are also quite hydrophobic, filling only 37% of the available space (excluding waters). It is possible that many of the GSK8470 analogs investigated in the previous SAR study adopt a variety of different conformations. This seems to be the case in our docking studies with these ligands: multiple, very different binding modes with similar energies are predicted (data not shown). Importantly, however, the repositioning of RR-RJW100 in our structure appears to be driven by particular interactions, since SR-RJW100 assumes a very similar pose (Figure 3.5). This occurs despite the fact that the SR derivative

exhibits signs of motion in our crystal structure, with significant disorder in the tail of the ligand and higher relative B-factors than RR-RJW100.

A major factor driving repositioning of the RJW100 enantiomers was the hydrogen bonding interaction made by the hydroxyl group. Although the contact with residue T352 is indirect, it is mediated by a water molecule that is part of a network of waters found in every published LRH-1 crystal structure (with the exception of PDB 4DOR, in which a major portion of the ligand binding pocket is disordered (Musille et al., 2012)). The existence of conserved water molecules, as well as their participation in ligand binding has been described (Barillari, 2007; Klebe, 2015; Ogata, 2002). Thus, this interaction could serve as an anchor point to secure the compound in a predictable orientation, enabling the targeting of desired parts of the binding pocket via strategic addition of substituents to the ligand's scaffold. Moreover, replacing the RJW100 hydroxyl group with a larger polar moiety may allow direct contact with T352, leading to a stronger interaction. This strategy is being actively explored in our laboratory.

The role of the T352 interaction in LRH-1 activation by RR-RJW100 was demonstrated through the marked loss of activation by this compound when this residue was mutated (Figure 3.8). In addition, a RJW100 analog lacking a hydroxyl group and thus unable to make this was a poor activator. Unexpectedly, the T352V mutation also resulted in a loss of activity for GSK8470, although this compound does not interact with the T352-coordinated water molecule. However, we show that the T352V mutation weakens GSK8470's interaction with H390, perhaps via destabilization of the conserved water network (Figure 3.8G-H). This could be responsible for the loss of activity of GSK8470 when T352 is mutated.

It has been hypothesized in *in silico* studies that π-π-stacking with residue H390 is critical for activation of LRH-1 by this ligand class (Lalit et al., 2013); however, this had not been explicitly tested. We find that this is the case for GSK8470, which stably interacts with H390 via face-to-face π-π-stacking. Interestingly, while mutation of H390 to alanine ablated LRH-1 activation by GSK8470, it had no effect on RR-RJW100-mediated activation (Figure 3.8). We show that the

RR-RJW100 interaction with H390 is much less stable than that of GSK8470 and is mediated by a different phenyl ring (Figure 3.7). Substitution of the GSK8470 aniline group with the styrene appears to have had the unexpected effect of making face-to-face stacking with H390 less favorable (Figure 3.9). This, combined with the favorable, water-mediated interaction with T352, influences the positioning of RJW100.

Together, these findings reveal that the interaction of small molecule agonists with LRH-1 is more complex than originally supposed. Not only do these agonists affect receptor conformation differently than PL ligands, but they also exhibit an unexpected variability in binding modes. This work has uncovered some of the molecular interactions responsible for both positioning and activation of two very similar agonists, which provide insights into strategies to improve the design of LRH-1-targeted compounds.

3.5 Materials and Methods.

3.5.1 Materials and reagents—GSK8470, RJW100, and analogs were synthesized as previously described (Whitby et al., 2006; Whitby et al., 2011). RR-RJW100 was separated from SS-RJW100 by chiral preparative chromatography (Diacel OD-H column, (Stec, 2010)). *Endo*-RJW100 is a diastereoisomer of RJW100, previously referred to as "24-endo" (Whitby et al., 2011). In this paper, we use "SR-RJW100" and "RS-RJW100" to refer to the enantiomers of *endo*-RJW100 (Figure 3.3.1). pCI empty vector was purchased from Promega. The SHP-luc and Renilla reporters, as well as pCI LRH-1, have been previously described (Musille et al., 2012). The vector for His-tagged tobacco etch virus (TEV) was a gift from John Tesmer (University of Texas at Austin). The pMSC7 (LIC_HIS) vector was provided by John Sondek (University of North Carolina at Chapel Hill). The Tif2 NR Box 3 peptide was purchased from RS Synthesis (Louisville, KY). DNA oligonucleotide primers were synthesized by IDT (Coralville, IA).

3.5.2 Protein expression and purification—LRH-1 LBD (residues 299-541) in the pMSC7 vector was expressed and purified as previously described (Musille et al., 2015). Briefly, protein was

expressed in BL21(DE3) pLysS *E. coli* by induction with IPTG (1 mM) for 4 hr at 30°C. Protein was purified by nickel affinity chromatography. Protein used for DSF experiments was incubated with DLPC (five-fold molar excess) for four hours at room temperature, and then repurified by size-exclusion into an assay buffer of 20 mM Tris-HCl, pH 7.5, 150 mM NaCl, and 5% glycerol. Protein used for crystallization was incubated with TEV protease to cleave the His tag. The cleaved protein was then separated from the His tag and TEV by a second round of nickel affinity chromatography. To make protein-ligand complexes, protein was incubated with ligands overnight (10-fold molar excess) and repurified by size-exclusion, using a final buffer of 100 mM ammonium acetate, pH 7.4, 150 mM sodium chloride, 1 mM DTT, 1 mM EDTA, and 2 mM CHAPS.

3.5.3 Crystallization—Protein-ligand complexes were incubated with a peptide derived from human Tif2 NR Box 3 (+H_3N-KENALLRYLLDKDDT-CO_2-) at four-fold molar excess for two hours at room temperature and then concentrated to 6.5 mg/ml. A crystallant of 0.05 M sodium acetate, pH 4.6, 5-11% PEG 4000 and 0-10% glycerol was used. Crystals were grown by hanging drop vapor diffusion in drops containing 1 µl protein and 1 µl crystallant, at a temperature of 18-20°C.

3.5.4 Structure determination—Crystals were flash frozen in liquid nitrogen, using a cryopreservative consisting of crystallant plus 30% glycerol. Data were collected remotely from the South East Regional Collaborative Access Team at the Advanced Photon Source, 22ID beamline (Argonne National Laboratories, Chicago, IL). Data were processed and scaled using HKL2000 (Otwinowski, 1997) and phased by molecular replacement using Phaser-MR (Phenix (Adams et al., 2010)). For the RR-RJW100 structure, PDB 3PLZ (Whitby et al., 2011) was used as the search model, with the ligand and a portion of the bottom of the receptor omitted. For the SR-RJW100 structure, the search model was the RR-RJW100 structure with the ligand omitted. Model building and refinement were conducted with Coot (Emsley and Cowtan, 2004) and

phenix.refine (Adams et al., 2010), respectively. Figures were constructed using Pymol (Schrödinger, LLC) (Schrodinger, 2010).

3.5.5 Structure analysis—Dimensions of the binding pocket in the presence of various ligands were calculated using CastP software (Binkowski, 2003). Ligplot+ was used to identify residues interacting with the ligands (Laskowski and Swindells, 2011).

3.5.6 Mutagenesis—Mutations were introduced to pMSC7 and pCI LRH-1 constructs using the Quikchange site-directed mutagenesis kit (Strategene). Constructs were sequence-verified prior to use.

3.5.7 Cell culture—HeLa cells were purchased from Atlantic Type Culture Collection and grown in phenol red-free MEMα media (CellGro) supplemented with 10% charcoal-stripped fetal bovine serum (Atlanta Biologicals). Cells were maintained using standard culture conditions.

3.5.8 Reporter gene assays—HeLa cells were seeded at a density of 10,000 cells per well in white-walled, clear-bottomed 96-well culture plates. The next day, cells were transfected with LRH-1 and reporters, using Fugene HD (Roche) at a ratio of 5:2 Fugene:DNA. The transfected plasmids included full-length LRH-1 in a pCI vector (5 ng/ well), and a SHP-luc reporter, encoding the LRH-1 response element and surrounding sequence from the SHP promoter cloned upstream of firefly luciferase in the pGL3 basic vector (50 ng/ well). Cells were also co-transfected with a constitutive *Renilla* luciferase reporter (utilizing the CMV promoter), which was used for normalization of firefly signal (1 ng/ well). Control cells received pCI empty vector at 5 ng/ well in place of LRH-1-pCI. Following an overnight transfection, cells were treated with agonists for 24 hours at the concentrations indicated in the figure legends. Agonists were dissolved in DMSO and then diluted into media, with a final concentration of 0.3% DMSO in all wells. Luciferase signal was quantified using the DualGlo kit (Promega). Experiments were conducted at least three times in triplicate.

3.5.9 Hydrogen-deuterium exchange mass spectrometry—Purified LRH-1 LBD protein (His tag removed) was incubated with of a five-fold molar excess DLPC or synthetic agonist overnight at

4°C. Protein-ligand complexes were then re-purified by size exclusion to remove displaced phospholipids and unbound ligands. An additional bolus of agonist or DLPC (5-fold molar excess) was added to the complexes prior to analysis by HDX. The assay buffer consisted of 20 mM Tris HCl, pH 7.4, 150 mM NaCl, and 5% glycerol. Solution-phase amide HDX was carried out with a fully automated system as described previously (Goswami, 2013) Briefly, 5 µl of protein was diluted to 25 µl with D_2O-containing HDX buffer and incubated at 25 °C for 10 s, 30 s, 60 s, 900 s or 3,600 s. Following on exchange, back exchange was minimized and the protein was denatured by dilution to 50 µL in a low pH and low temperature buffer containing 0.1% (v/v) trifluoroacetic acid in 5 M urea (held at 1 °C). Samples were then passed across an immobilized pepsin column (prepared in house) at 50 µl min-1 (0.1% v/v TFA, 15 °C); the resulting peptides were trapped on a C8 trap cartridge (Hypersil Gold, Thermo Fisher). Peptides were then gradient-eluted (4% (w/v) CH_3CN to 40% (w/v) CH_3CN, 0.3% (w/v) formic acid over 5 min, 2 °C) across a 1 mm × 50 mm C18 HPLC column (Hypersil Gold, Thermo Fisher) and electrosprayed directly into an Orbitrap mass spectrometer (Q Exactive, Thermo Fisher). Data were processed with in-house software (Pascal et al., 2012) and visualized with PyMOL (Schrödinger, LLC, (Schrodinger, 2010)). To measure the difference in exchange rates, we calculated the average percent deuterium uptake for hLRH-1 LBD – RR-RJW100 complexes following 10, 30, 60, 900 and 3,600 s of on exchange. From this value, we subtracted the average percent deuterium uptake measured for the DLPC-hLRH-1 LBD complex. Negative perturbation values indicate exchange rates are slower for these regions within the RR-RJW100-LRH-1 complex relative to DLPC-bound LRH-1.

3.5.10 Differential scanning fluorimetry (DSF)—Purified LRH-1 LBD-His protein (0.2 mg/ml) was incubated overnight with 50 µM of each compound at 4°C. The final DMSO concentration in the reactions was 1%. SYPRO orange dye (Invitrogen) was then added at a 1:1000 dilution. Reactions were heated at a rate of 0.5°C per minute, using a StepOne Plus Real Time PCR

System (ThermoFisher). Fluorescence was recorded at every degree using the ROX filter (602 nm). Data were analyzed by first subtracting baseline fluorescence (ligands + SYPRO with no protein) and then fitting the curves using the Bolzman equation (GraphPad Prism, v6) to determine the Tm.

3.5.11 Model construction for molecular dynamics simulations—Three crystal structures of LRH-1 LBD in complex with Tif2 were used to construct models for the simulations. These were (1) PDB 3PLZ, chains B and D (GSK8470 ligand), (2) PDB 5L11 (RR-RJW100 ligand), and (3) PDB 5SYZ (SR-RJW100 ligand). For consistency, the structures were modified at the N- and C-termini so that all contained residues 300-540 of LRH-1 and residues 742-751 of the Tif2 peptide. Missing residues within this protein sequence were added, as well as missing protein sidechains (3PLZ). The T352V mutation was introduced by mutating the sequence of WT LRH-1 in each of these structures. Finally, in a separate simulation, RR-RJW100 ligand was artificially reoriented in WT LRH-1 to be aligned with GSK8470, allowing a face-to-face π-π-stacking interaction of LRH-1 residue H390 with the aniline ring of RR-RJW100. Seven complexes in total were used in molecular dynamics simulations.

3.5.12 Molecular dynamics simulations—The complexes were solvated in an octahedral box of TIP3P water with a 10 Å buffer around the protein complex. Na^+ and Cl^- ions were added to neutralize the protein and achieve physiological conditions. All systems were set up using xleap in AmberTools (Case and Kollman, 2012) with the parm99-bsc0 forcefield (Pérez et al., 2007). Parameters for all ligands (GSK8470, RR-RJW100, and SR-RJW100) were obtained using Antechamber (Wang et al., 2001) in AmberTools. All minimizations and simulations were performed with Amber14 (Case et al., 2014). Systems were minimized with 5000 steps of steepest decent followed by 5000 steps of conjugate gradient minimization with 500 kcal/mol·$Å^2$ restraints on all atoms. Restraints were removed from all atoms excluding the atoms in both the ligand and the Tif2 peptide, and the previous minimization was repeated. The systems were heated from 0 to 300 K using a 100-ps run with constant volume periodic boundaries and 5

kcal/mol·Å^2 restraints on all protein and ligand atoms. Twelve ns of MD equilibration was performed with 10 kcal/mol·Å^2 restraints on protein and ligand atoms using the NPT ensemble. Restraints were reduced to 1 kcal/mol·Å^2 for an additional 10 ns of MD equilibration. Then restraints were removed and 200 ns production simulations were performed for each system in the NPT ensemble. A 2-fs timestep was used and all bonds between heavy atoms and hydrogens were fixed with the SHAKE algorithm (Ryckaert et al., 1977). A cut-off distance of 10 Å was used to evaluate long-range electrostatics with Particle Mesh Ewald (PME) and for van der Waals forces. 10,000 evenly spaced frames were taken from each simulation for analysis. Analysis was performed with the CPPTRAJ module (Roe and Cheatham III, 2013) of AmberTools. Statistical significance of differences in time π-π-stacking with residue H390 for GSK8470 *versus* the re-oriented RR-RJW100 was assessed using a Chi Squared test with Yate's correction (GraphPad Prism, v6).

3.6 Acknowledgements

This work was supported in part by NIH T32-GM008602 (SGM), NIH F31DK111171 (SGM), R01DK095750 (EAO) and an Emory Catalyst Grant to EAO and NTJ. RJW and JS thank GlaxoSmithKline for generous funding. The authors thank Sally Bloodworth (University of Southhampton, UK) for assistance with synthetic chemistry, as well as Bradley Kossmann and Dr. Ivaylo Ivanov (Georgia State University) for helpful discussions about MDS and Dr. Kay Diederichs (Universität Konstanz) for advice in model refinement.

Chapter 4. Structure and Dynamics of the Liver Receptor Homolog 1–PGC1α Complex

4.1 Abstract.

Peroxisome proliferator-activated gamma coactivator 1-α (PGC1α) regulates energy metabolism by directly interacting with transcription factors to modulate gene expression. Among the PGC1α binding partners is Liver receptor homolog 1 (LRH-1; NR5A2), an orphan nuclear hormone receptor that controls lipid and glucose homeostasis. Although PGC1α is known to bind and activate LRH-1, mechanisms through which PGC1α changes LRH-1 conformation to drive transcription are unknown. Here, we used biochemical and structural methods to interrogate the LRH-1-PGC1α complex. Purified, full-length LRH-1, as well as isolated ligand binding domain, bound to PGC1α with higher affinity than to the coactivator, Nuclear Receptor Coactivator-2 (Tif2) in coregulator peptide recruitment assays. We present the first crystal structure of the LRH-1-PGC1α complex, which depicts several hydrophobic contacts and a strong charge clamp at the interface between these partners. In molecular dynamics simulations, PGC1α induced correlated atomic motion throughout the entire LRH-1 activation function surface, which was dependent on charge clamp formation. In contrast, Tif2 induced weaker signaling at the activation function surface than PGC1α but promoted allosteric signaling from the Helix 6/β-sheet region of LRH-1 to the activation function surface. These studies are the first to probe mechanisms underlying the LRH-1-PGC1α interaction and may illuminate strategies for selective therapeutic targeting of PGC1α-dependent LRH-1 signaling pathways.

This chapter was adapted from the manuscript:
Mays SG, Okafor CD, Tuntland ML, Whitby RJ, Dharmarajan V, Stec J, Griffin PR, and Ortlund EA. Structure and Dynamics of the Liver Receptor Homolog-1-PGC1α Complex, *Mol Pharamacol* 2017; 92:1-11.

4.2 Introduction.

Liver Receptor Homolog-1 (LRH-1) is an orphan nuclear receptor (NR) that acts as an important regulator of lipid and glucose metabolism. It is highly expressed in liver, where it controls bile acid biosynthesis (Lu et al., 2000), *de novo* lipogenesis (Lee et al., 2011), and reverse cholesterol transport (Schoonjans et al., 2002; Stein et al., 2014). Notably, activation of LRH-1 in obese mice improves glucose tolerance and insulin resistance (Lee et al., 2011), as well as reduces atherosclerosis formation (Stein et al., 2014). LRH-1 also plays key roles in the resolution of hepatic endoplasmic reticulum stress (Mamrosh et al., 2014) and maintenance of the one-carbon pool (Wagner et al., 2016), which are both critical for metabolic homeostasis and cell survival. Glucose transport, metabolism, and capture are regulated by LRH-1 via control of proteins such as the GLUT4 transporter in skeletal muscle and glucokinase in the liver (Bolado-Carrancio et al., 2014; Oosterveer et al., 2012). On the other hand, aberrant activation of LRH-1 drives tumorigenesis and tumor cell proliferation in several cancer types (Bayrer et al., 2015; Bianco et al., 2014; Chand et al., 2010; Lin et al., 2014; Nadolny and Dong, 2015; Thiruchelvam et al., 2011; Xu, 2016). Because of this vital transcriptional program, LRH-1 is garnering attention as a new therapeutic target for treatment of diseases such as nonalcoholic fatty liver disease, diabetes, and cancer.

As with other NRs, transcriptional activity of LRH-1 relies upon associations with coregulators, a diverse family of proteins that act as chromatin-remodeling factors (or which recruit such factors) to control promoter accessibility. Coregulator interactions typically occur at the NR activation function surface (AFS), located in the ligand binding domain (LBD). When the AFS is in the active conformation (*e.g.* upon binding of an activating ligand), coactivators bind to a canonical cleft within this surface via a helical LXXLL motif to drive NR activity (where "L" is leucine and "X" is any amino acid). Corepressors inhibit NR activity by binding to an alternative conformation of the AFS using an extended leucine-rich motif. In addition, atypical corepressors

(such as small heterodimer partner, SHP) utilize a LXXLL-containing helix to compete with coactivators for binding the active AFS, resulting in suppression of NR activity. Interestingly, a phospholipid LRH-1 agonist with antidiabetic effects (Lee et al., 2011) completely ablates SHP binding while preserving the ability to bind coactivators (Musille et al., 2012). This effect is mediated through ligand-driven allosteric communication between the AFS and a distal portion of the LBD, which is now considered to be an "alternate" AFS (hereafter "AF-B") (Musille et al., 2015; Musille et al., 2012).

Among the LRH-1 coregulators is peroxisome proliferator gamma coactivator 1-α (PGC1α), which plays a critical role in energy homeostasis by interacting with a variety of NRs and other transcription factors (Finck, 2006; Lin et al., 2005). PGC1α is expressed in tissues with high demand for energy, such as heart, skeletal muscle, and brown adipose tissue, as well as in the liver. It is expressed at low basal levels but is highly inducible upon certain stimuli, such as cold exposure, during exercise, and signaling from AMPK, SIRT, and cAMP (Finck, 2006). Downstream effects of PGC1α activation include stimulation of mitochondrial biogenesis and increased rates of cellular respiration (Finck, 2006). In the liver, PGC1α plays a major role in oxidative metabolism, including the control of bile acid production (Lin et al., 2005). PGC1α has several overlapping biological roles with LRH-1, including in disease states such as obesity, diabetes, and cancer (Lin et al., 2005), and the ability of PGC1α to act as an LRH-1 coactivator has been documented in several studies. The two proteins directly interact via the PGC1α NR box 2 and the LRH-1 AFS (Shin, 2008). Overexpression of PGC1α enhances LRH-1 activity at the aromatase and SHP promoters in luciferase reporter assays,(Safi, 2005) and PGC1α drives expression CYP7A1 in hepatocytes (Shin, 2008). On the other hand, the similar PGC1β isoform cannot activate LRH-1 (Shin, 2008). Interestingly, PGC1α appears to be better able to discriminate between LRH-1 ligand-bound states than the Nuclear Receptor Coactivator-2 (Tif2) coactivator, since PGC1α is unable to bind apo-LRH-1 *in vitro*, while Tif2 binds both apo- and

agonist-bound receptor (Musille et al., 2012). The discrete tissue expression of PGC1α and its inducible nature has made it an attractive therapeutic target; indeed, disruption of the LRH-1-PGC1α interaction has been proposed as a strategy to achieve tissue-specific inhibition of aromatase production in breast cancer (Safi, 2005). However, mechanisms through which PGC1α drives LRH-1 activation have not been delineated. Here, we present the first crystal structure of LRH-1 bound to PGC1α, allowing the visualization of the interface between these two partners. PGC1α bound LRH-1 with high affinity and induced strong communication within the LRH-1 AFS. Unlike Tif2, PGC1α did not induce conformational changes to AF-B or promote allosteric signaling from AF-B to the AFS, suggesting that the two coregulators utilize distinct mechanisms to activate LRH-1.

4.3 Results.

4.3.1. Coregulator binding affinities for full-length LRH-1

Previously published LRH-1 binding assays have utilized isolated LBD, due to difficulties purifying full-length (FL) protein. We have recently developed a method for FL-LRH-1 purification and used this protein to investigate binding of fluorescein (FAM)-conjugated coregulator peptides by fluorescence polarization. To stabilize the receptor and to provide a biologically relevant context, the protein was co-purified with a portion of the CYP7A1 promoter containing the LRH-1 binding site (Weikum et al., 2016). The protein was purified from *E. coli* and contained a variety of bacterial phospholipids in the binding pocket (previously shown to act as weak activators) (Musille et al., 2015; Musille et al., 2012). We also determined affinities for various coregulators when FL-LRH-1 was bound to the agonist, RJW100. Binding curves from these experiments are shown in Figure 4.1, and K_D values are summarized in Table 1. When *E. coli* PL occupied the binding pocket, PGC1α and SHP bound FL-LRH-1 with higher affinity than Tif2 (~500 nM *versus* 2.3 μM). The corepressor Silencing Mediator Of Retinoic Acid And Thyroid Hormone Receptor (SMRT) bound with much lower affinity (>20 μM). The addition of

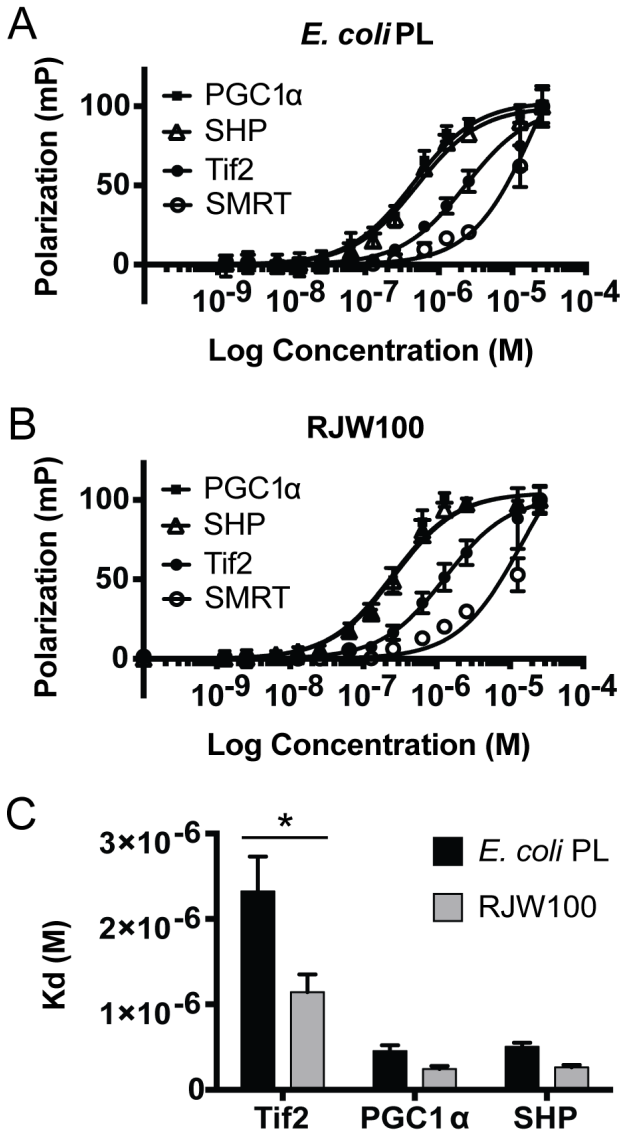


Figure 4.1. Coregulator binding affinities for the LRH-1-RJW100 complex. Fluorescence polarization was used to determine binding affinities of LRH-1 for various coregulators. Full-length (FL) LRH-1 bound to PL from *E. coli* (A) or to RJW100 (B) was titrated in the presence of FAM-labeled coregulator peptides. C. RJW100 increases the affinity of LRH-1 for Tif2. *$p<0.05$

the LRH-1 agonist RJW100 to the receptor increased affinity for Tif2 by around 2-fold ($p<0.05$, Figure 4.1C). Affinities for PGC1α and SHP trended higher in the presence of RJW100 (Table 1), but the differences were not statistically significant. As with native protein, PGC1α and SHP bound the LRH-1-RJW100 complex with higher affinity than Tif2 (1.1 +/- 0.2 μM for Tif2 *versus* 240 +/- 40 nM for PGC1α, Table 4.1). To guide our crystallization efforts and to provide some comparisons to previous studies, we also determined the affinities of the coregulators for purified LRH-1 LBD with or without RJW100. The K_D values were all higher than with FL protein, but they were proportionally the same (*e.g.,* PGC1α bound with higher affinity than Tif2, Table 1). Together, these studies demonstrate the relatively high affinity binding of PGC1α for LRH-1 and suggest that using isolated LBD for crystallography would provide an adequate model of this high affinity interaction.

Table 4.1. Affinities of LRH-1 for Various Coregulators

Values represent the mean ± S.E.M. of the K_D (μM) calculated from three separate experiments conducted in triplicate. Values in parentheses represent the goodness of fit (R^2).

	LBD		FL LRH-1	
	E. coli PL	RJW100	*E. coli* PL	RJW100
Tif2	8 ± 1 (0.99)	3.4 ± 0.3 (0.99)	2.3 ± 0.4 (0.93)	1.1 ± 0.2 (0.92)
PGC1α	0.8 ± 0.1 (0.98)	0.5 ± 0.1 (0.96)	0.45 ± 0.07 (0.95)	0.24 ± 0.04 (0.95)
SHP	0.9 ± 0.1 (0.99)	0.6 ± 0.1 (0.98)	0.50 ± 0.05 (0.99)	0.26 ± 0.03 (0.97)
SMRT	60 ± 30 (0.96)	25 ± 4 (0.99)	22 ± 9 (0.91)	20 ± 8 (0.91)

4.3.2. Crystal structure of LRH-1 with PGC1α

To visualize the LRH-1-PGC1α interaction surface, we determined the crystal structure of the LRH-1 LBD, bound to a fragment of PGC1α, to a resolution of 1.95 Å (Figure 4.2A, Table 2). The agonist RJW100, used to aid crystallization and to model the active state, is clearly bound in the pocket based on the surrounding electron density (Figure 4.2B). The ligand adopts a similar position as in our previous structure (PDB 5L11) (Mays et al., 2016), where the LRH-1-RJW100 complex was crystallized with the Tif2 coactivator rather than PGC1α (Figure 4.2C). Overall LRH-1 conformation is not greatly changed when PGC1α, rather than Tif2, is bound (RMSD = 0.5 Å). In the PGC1α structure, the loop connecting helix 2 to helix 3 is highly disordered and cannot be modeled (dotted line in Figure 4.2A); however, this region tends to be mobile when not stabilized by crystal contacts (such as the stabilization of this region that occurs in PDB 5L11). A second region of disorder occurs within the loop between helices 8 and 9, which prevented modeling of two residues (K462 and N463). This region also tends to be disordered when not stabilized by crystal contacts (for example, see PBD 4DOS (Musille et al., 2012)).

PGC1α is bound at the expected site at the AFS, a surface formed by portions of helices 3 and 4 and the AF-H (Figure 4.2). Although electron density is strong for the LXXLL consensus sequence and immediately adjacent amino acids, residues on either side of this sequence are disordered and cannot be modeled (*i.e.* residues 740-742 at the N-terminus and 752-753 at the C-terminus). Several PGC1α leucine sidechains fit within the AFS and make hydrophobic interactions with the receptor (Figure 4.3). We also examined the structure for a direct electrostatic interaction similar to the asparagine-lysine contact found to be important for the high affinity interaction between PPARλ-PGC1α (Li, 2008). LRH-1 has an aspartate residue (D372) that is in an analogous position to the PPARλ asparagine; however, it is positioned 5.5 Å away from the nearest PGC1α lysine sidechain and does not appear to be interacting (Figure 4.3). On

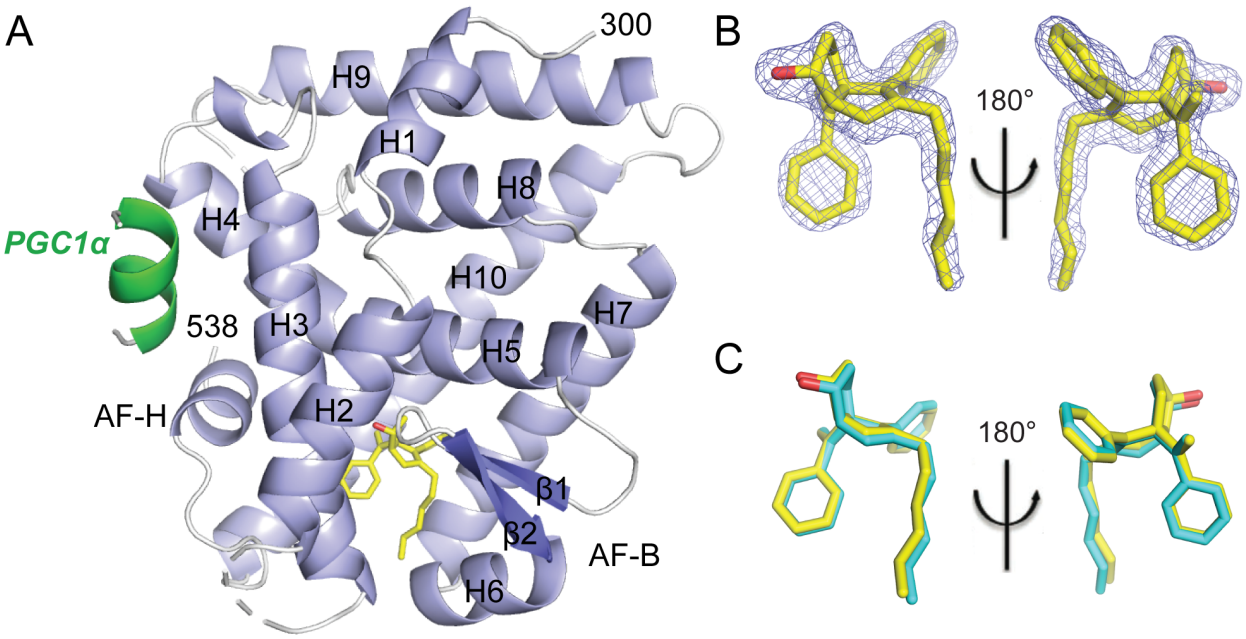


Figure 4.2. Crystal Structure of LRH-1-RJW100 with the PGC1α co-activator. A. Overall structure, with α-helices shown in light blue, β-sheets in slate, and loops in white. The ligand is shown in yellow. PGC1α (green) is bound at the AFS. Dotted lines indicate regions that could not be modeled. B. Electron density map surrounding the RJW100 ligand in the LRH-1-PGC1α complex (F_O-F_C omit map contoured to 3.0 σ). C. Superposition of the ligand with PDB 5L11 (LRH-1-RJW100-Tif2) showing that it assumes a similar conformation in both structures.

Table 4.2. X-ray Data Collection and Refinement Statistics: LRH-1—RJW100—PGC1α

Values in parentheses indicate highest resolution shell.

Data Collection	LRH-1–RJW100–PGC1α
Space group	$P2_12_12$
Cell dimensions	
a, *b*, *c* (Å)	66.2, 84.0, 45.4
α, β, γ (°)	90, 90, 90
Resolution (Å)	50-1.95 (2.02-1.95)
R_{pim}	0.04 (0.18)
$I/\sigma I$	17.6 (2.1)
CC1/2 in highest shell	0.915
Completeness (%)	96.8 (83.8)
Redundancy	6.1 (5.0)
Refinement	
Resolution (Å)	1.95
No. reflections	18122
R_{work} / R_{free} (%)	19.9 / 22.7
No. atoms	
Protein	3847
Water	58
B-factors (Å^2)	
Protein	43.6
Ligand	35.2
Water	42.7
R.m.s. deviations	
Bond lengths (Å)	0.003
Bond angles (°)	0.489
Ramachandran favored (%)	97
Ramachandran outliers (%)	0
PDB accession code	5UNJ

R.m.s., root mean square.

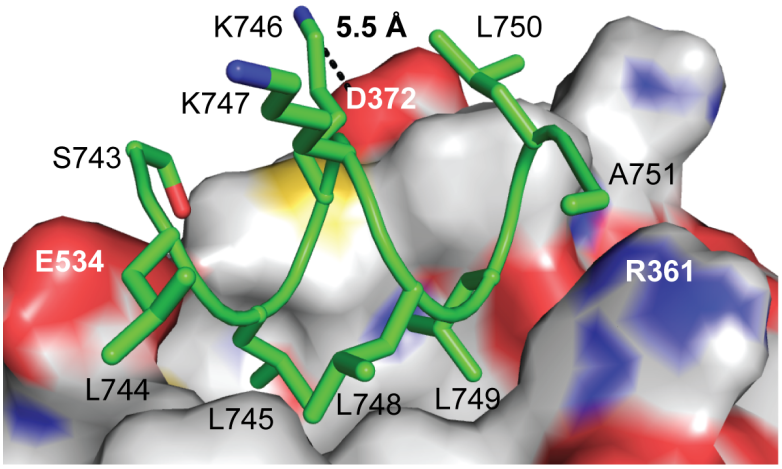


Figure 4.3. Coregulator-LRH-1 binding interface. Molecular surface showing the interaction of PGC1α (green helix) with LRH-1. Surface has been colored by atom type (carbon, grey; oxygen, red; nitrogen, blue; and sulfur yellow).

the other hand, LRH-1-Tif2 structures (both with *E. coli* PL and RJW100 bound) depict a direct contact with D372 via the sidechain of residue R746 (seen in PDB 4PLE and 5L11, respectively, not shown) (Mays et al., 2016; Musille et al., 2015). This contact therefore does not appear to be correlated with the higher binding affinity of PGC1α for LRH-1 observed in our biochemical assay.

4.3.3. PGC1α strengthens the coactivator charge clamp and communication within the LRH-1 AFS

An important driving force behind binding of coactivators and atypical corepressors to NRs is a charge clamp that neutralizes the helix dipole of the coregulator and secures it to the binding cleft in the AFS (Li et al., 2003). LRH-1 utilizes residues R361 and E534 to form the charge clamp, with an arginine substituted for the canonical lysine used by most NRs. We have shown that the presence of this charge clamp is closely associated with the strength of LRH-1 activation. For example, when LRH-1 is bound to weak PL activators, E534 is swung away from the Tif2 coactivator (>5 Å away), and the charge clamp is incomplete (Musille et al., 2015). However, when LRH-1 is bound to a specific and stronger PL agonist (dilauroylphosphatidylcholine, DLPC), E534 makes direct contact with Tif2 (Musille et al., 2015). In the LRH-1-PGC1α structure, the electron density provides strong evidence for the interaction with E354: the sidechain engages backbone amide nitrogens of PGC1α residues L744 and L745 (Figure 4.4A). Residue R361 has somewhat weaker electron density than E534 but clearly interacts with PGC1α residue A751 via NH1 and L749 via NH2 (Figure 4.4A). In contrast, while the LRH-1-RJW100-Tif2 structure depicts similar interactions with R361, residue E534 has very weak electron density for the sidechain (Figure 4.4B). This weaker density is not related to resolution, since the Tif2 structure was determined at higher resolution than the PGC1α structure (1.85 Å *versus* 1.95 Å). The disorder thus suggests that the E534 sidechain is mobile when Tif2 is bound and that the charge clamp is incomplete. Supporting this idea, E534 maintains interactions with PGC1α for 93% of the time during 500 ns molecular dynamics simulations

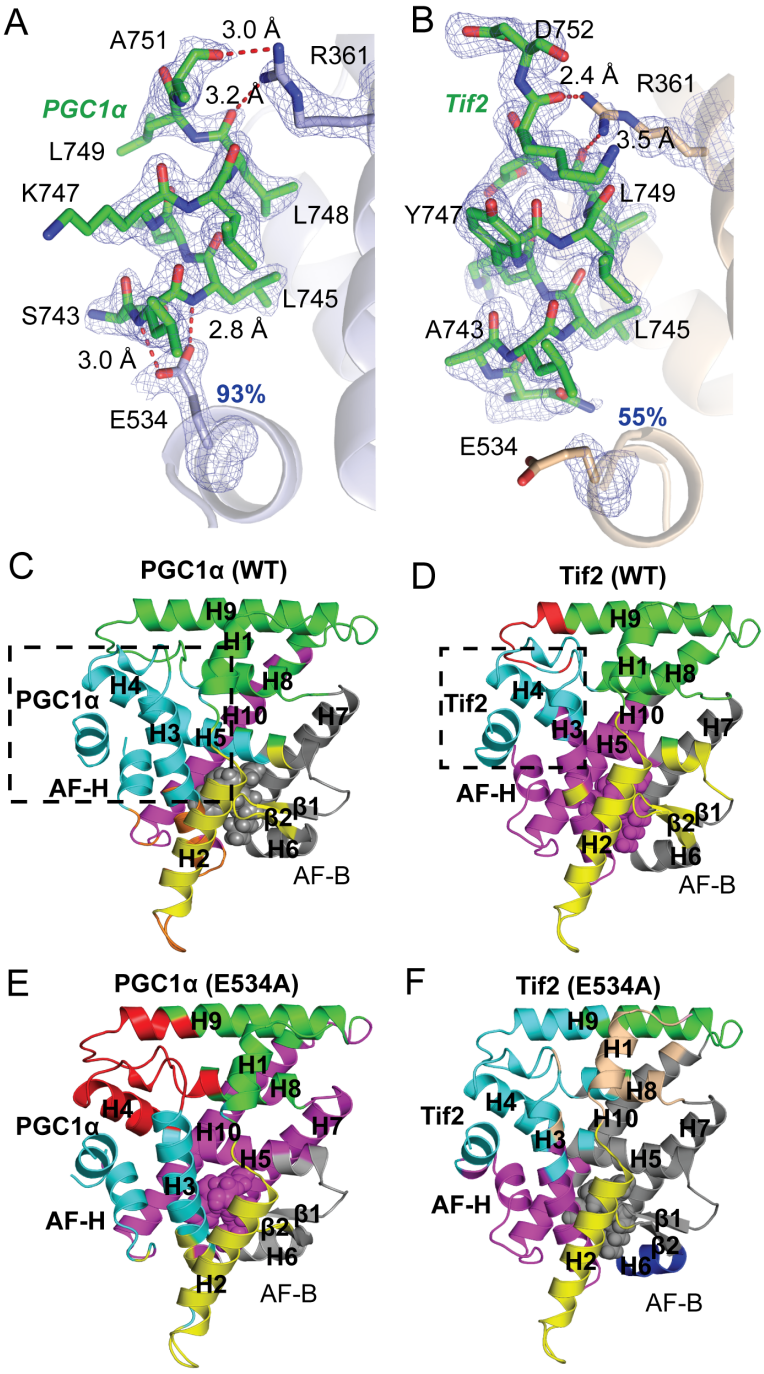


Figure 4.4. A strong charge clamp occurs at the LRH-1-PGC1α interface. A-B. Electron density maps showing strong evidence for placement of *(A)* PGC1α and *(B)* Tif2 (from PBB 5L11). Maps show F_O-F_C density contoured to 2.5 σ. Numbers in blue show the percent of time

that glutamate 534 interacted with each coregulator in 500 ns molecular dynamics simulations (MDS). C-F. Community analysis of MDS for either *(C)* PGC1α with wild type (WT) LRH-1 *(D)* Tif2 with WT LRH-1 (from PBD 5L11(Mays et al., 2016)), *(E)* PGC1α with mutant (E534A) LRH-1, or *(F)* Tif2 with mutant (E534) LRH-1. *Dotted lines* in *C* and *D* indicate the size and composition of the AFS community for WT LRH-1 in the presence of each coregulator.

(MDS) *versus* only 55% of the time with Tif2. Interactions with R361 are maintained ~77% of the time for both coregulators in these simulations.

The stability of the E534-corregulator interaction is associated with correlated motion of residues within the LRH-1 AFS, indicative of strong communication in this region. This was determined using community analysis of the MDS (Figure 4.4), which clusters residues into "communities" comprising residues that exhibit the greatest degree of correlated motion with one another. When PGC1α is bound, the coregulator and the entire LRH-1 AFS cluster as a single community that extends into helix 5 (Figure 4.4C). However, the Tif2-bound AFS is split into two communities, the first comprising helix 4 and Tif2, and the second containing helix 3 and the AF-H (Figure 4.4D). Mutation of residue E534 to alanine splits the AFS into two communities in the presence of PGC1α while not greatly affecting the AFS communities in the Tif2 structure (Figure 4.4E-F). These findings imply that PGC1α induces coordinated motion within the AFS, which is dependent, at least in part, on a stable interaction with LRH-1 residue E534.

4.3.4. Differential effects of PGC1α and Tif2 on LRH-1 allosteric communication

In addition to the AFS, the helix 6 (H6)/ β-sheet region flanking the lower part of the binding pocket is important for LRH-1 activation and has been termed an alternative AFS (AF-B) (Musille et al., 2012). The PL LRH-1 agonist, DLPC, induces flexibility in AF-B (Musille et al., 2012) and promotes communication through the receptor from AF-B to the AFS (Musille et al., 2015). This communication is weaker when DLPC is bound in the presence of the corepressor SHP rather than Tif2, leading to the hypothesis that the AF-B transmits information about ligand status to the AFS to promote recruitment of appropriate corregulators (Musille et al., 2015). To understand how PGC1α affects this allosteric network, we used amide hydrogen deuterium exchange mass spectrometry (HDX-MS) and MDS to compare changes in protein dynamics induced by PGC1α and Tif2. In both experiments, the agonist RJW100 was bound in the pocket to model the activated state. In the HDX-MS studies, less deuterium incorporation was seen in the

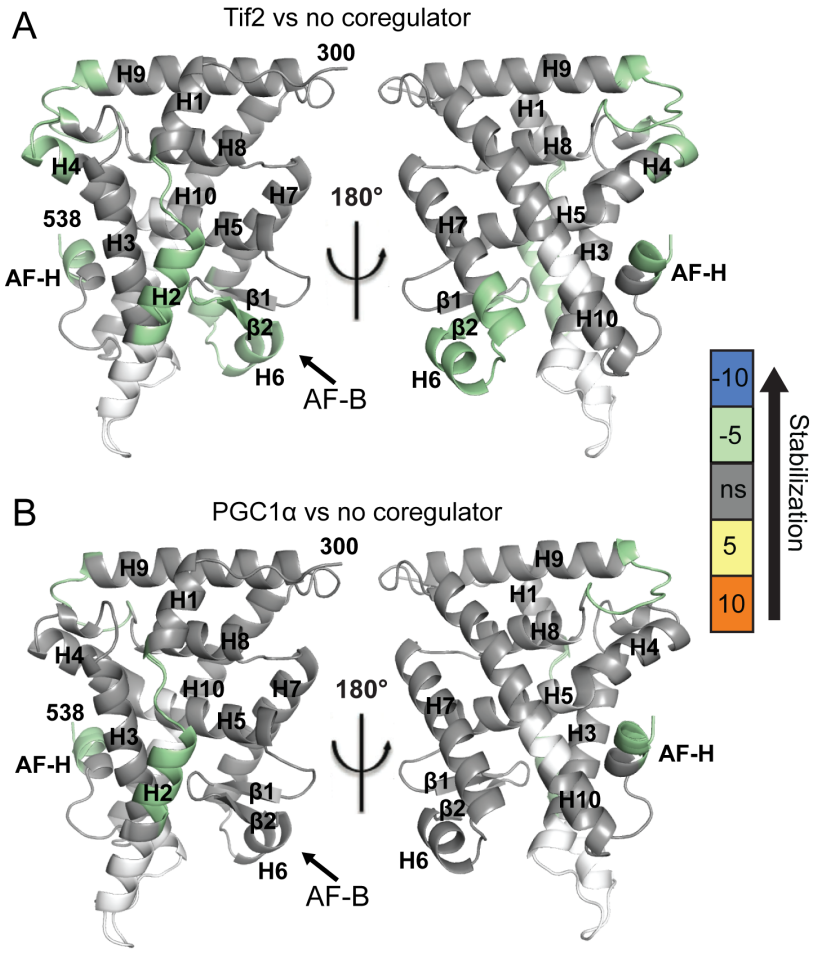


Figure 4.5. Stabilization of the alternate AFS occurs upon Tif2 binding. Hydrogen deuterium exchange mass spectrometry was used to identify effects of PGC1α (A) and Tif2 (B) on LRH-1 dynamics. Each cartoon is a map of differential deuterium incorporation of LRH-1 + coregulator versus LRH-1 only. The scale indicates the difference in percentage of deuterium incorporation: for example, negative numbers reflect slower deuterium incorporation (less motion) for coregulator versus no coregulator. Results are mapped onto PBD 4DOS(Musille et al., 2012).

AF-H when either PGC1α or Tif2 were bound *versus* no coregulator (Figure 4.5A-B). This reduced motion at the site of coregulator binding was expected. However, the Tif2-bound receptor also exhibited reduced flexibility in AF-B, including H6, β2, and the bottom of H7 that was not seen for the LRH-1-PGC1α complex (Figure 4.5A-B).

Communication between AF-B and the AFS for each LRH-1-coregulator complex was determined by analysis of correlated motion between these regions in the MDS studies. Cross-correlation matrices were used to rank the degree of correlation between each pair of residues in the protein complexes on a scale of -1 to 1 (where -1 is perfect anti-correlation (opposite motion), 0 is no correlation, and 1 is perfect correlation) (Musille et al., 2015). Figure 4.6A-B depicts these values as a heatmap, in which correlated motion is red and anti-correlated motion is blue. For the PGC1α complex, one of the largest regions of correlated motion is seen between residues 350 and 400 (Figure 4.6A). Correlation in this vicinity is also seen for the Tif2 complex, but it is larger, extending to residue 420 (Figure 4.6B). Mapping these regions onto LRH-1 shows that both coregulators induce correlated motion across helix 5 to helix 3, but Tif2 extends this correlation network into AF-B (Figure 4.6C).

Strength of communication between AF-B and the AF-H was further assessed by examination of the suboptimal paths between these sites. For this analysis, each Cα is defined as a "node," and the communication between each node is called an "edge." Edges are weighted by correlation, such that correlation is inversely proportional to edge weight(Musille et al., 2015). Communication between two distant nodes can occur through thousands of possible paths through the receptor, and the path for which the sum of the edges is lowest is called the shortest, or optimal, path. The optimal path and a set of suboptimal paths having the shortest lengths are thought to convey the greatest amount of communication between two distant nodes(Musille et al., 2015; Sethi et al., 2009b). For our analysis, we considered the shortest 1000 suboptimal paths between β2 and the AF-H for each LRH-1-corregulator complex. The routes taken by the paths

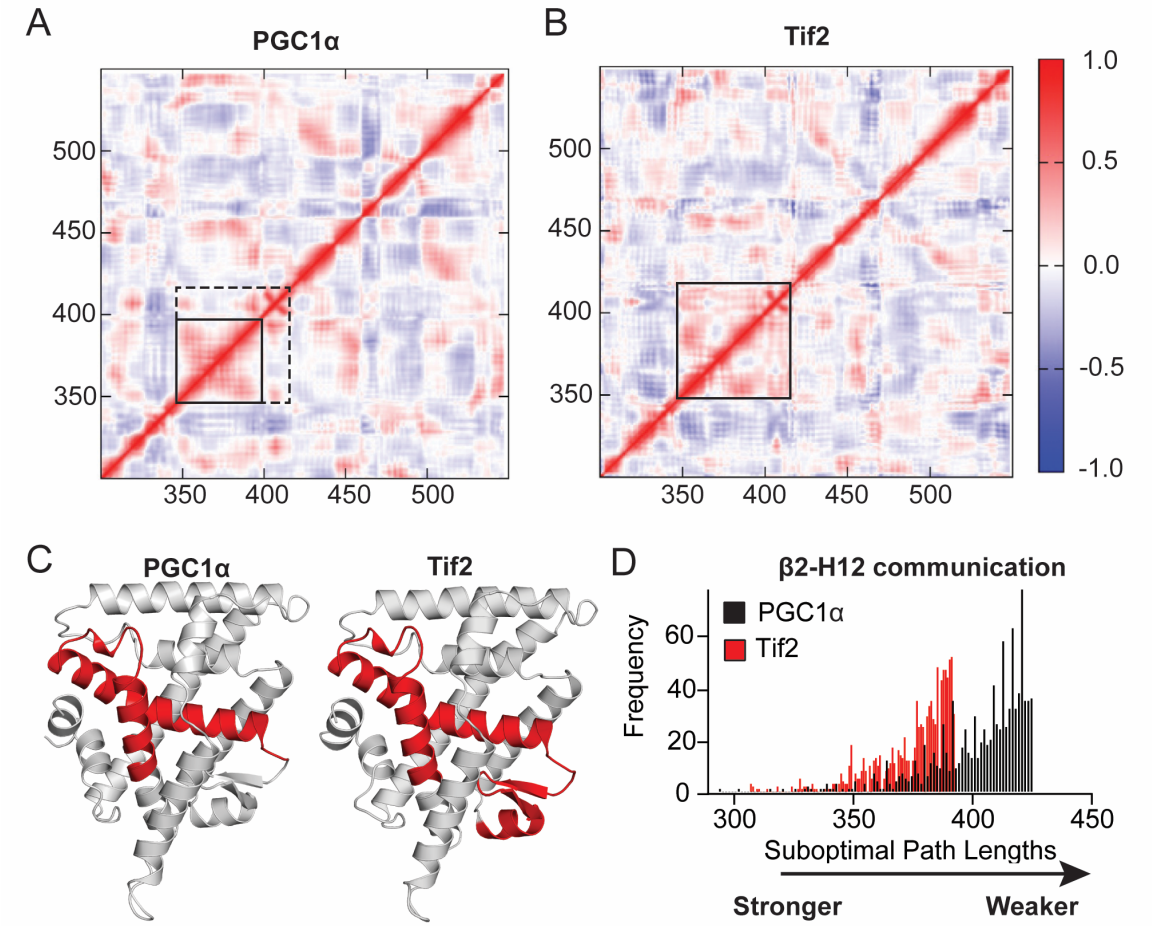


Figure 4.6. Extended communication from AF-B to the canonical AFS with Tif2 but not PGC1α. A-B. Covariance matrices for the MDS simulations with the LRH-1 bound to either *(A)* PGC1α or *(B)* Tif2. *Solid squares* surround the large region of correlated motion described in the text. The square with the dotted line shows the region of correlation in the Tif2 simulation that is not seen in the PGC1α simulation. C. Shown in red are the regions identified in panels *A* and *B* mapped onto LRH-1 for each simulation. D. The shortest 1000 suboptimal paths between LRH-1 β-sheet 2 in AF-B and the AFS for the PGC1α and Tif2 simulations, plotted as a histogram.

were not substantially different in the presence of either coregulator: they traversed H5 and through H3 to the AF-H (not shown). However, the strength of the communication was significantly weaker when PGC1α was bound compared to Tif2. This is seen in the plot of suboptimal path lengths, which shows shorter lengths when Tif2 is bound (Figure 4.6D). Together with the HDX-MS results, these findings indicate that Tif2 affects LRH-1 conformation at AF-B and promotes communication between AF-B and the AFS, whereas PGC1α acts primarily at the AFS.

4.3.5. Ligand participation in allosteric signaling within the LRH-1-Tif2 complex.

We have previously identified a water-mediated interaction with residue T352 as being important for LRH-1 activation by the RJW100 agonist(Mays et al., 2016). In the LRH-1-PGC1α structure, the ligand is oriented very similarly as in the LRH-1-Tif2 complex (Figure 4.2), and it makes contact with the T352-coordinated water via the hydroxyl group (Figure 4.7A). However, this interaction is predicted to be much less stable in the presence of PGC1α *versus* Tif2 (occurring 33% and 68% of the time, respectively, over the course of 500 ns MDS). Introduction of threonine to valine mutation at position 352 in the MDS had very little effect on suboptimal path length between AF-B and the AFS for the LRH-1-PGC1α complex (Figure 4.7B). In contrast, a dramatic lengthening of the suboptimal paths occurred for the T352V mutant in the LRH-1-Tif2 complex (Figure 4.7B). This is consistent with our observations that

4.4 Discussion.

Regulation of NR activity is complex, involving a dynamic interplay of ligand binding, posttranslational modifications, and coregulator associations. Study of LRH-1 regulation is particularly challenging, since the endogenous ligand for this receptor is unknown, and since only a few coregulators have been crystallized with LRH-1 previously (*i.e.* Tif2 (Mays et al., 2016; Musille et al., 2015; Musille et al., 2012), SHP(Li et al., 2005; Ortlund et al., 2005), and DAX(Sablin et al., 2008)). Tif2 is the only coactivator among these, which has greatly limited the

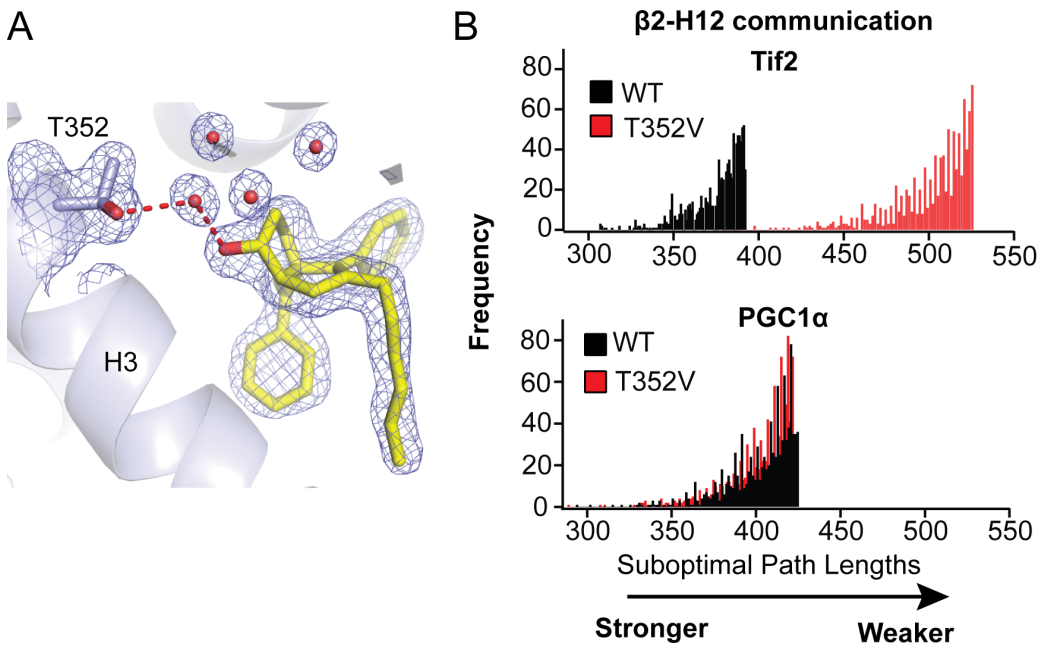


Figure 4.7. Contribution of the ligand to allosteric signaling between two LRH-1 activation surfaces. A. Electron density map ($2F_O$-F_C, contoured to 1σ) showing evidence for the water-mediated interaction of the RJW100 ligand with LRH-1. B-C. Histogram showing the distribution of 1000 shortest suboptimal paths for wild-type (WT) or T352V LRH-1 when bound to either (B) PGC1a or (C) Tif2.

ability to investigate LRH-1-coactivator interactions. The LRH-1-PGC1α structure thus makes a significant contribution to our understanding of its regulation by coactivators. Notably, PGC1α interacts stably with LRH-1 residue E534. This interaction is clearly seen in the electron density maps, confirmed by MDS, and required for the large region of coordinated motion within the LRH-1 AFS and with the PGC1α peptide (Figure 4.4). In contrast, Tif2 did not interact with E534 in the RJW100-LRH-1 crystal structure and engaged this residue approximately half of the time compared to PGC1α during MDS. Unfortunately, only a short stretch of the PGC1α peptide could be modeled in our structure, which may have prevented the identification of other important interactions. In particular, a PGC1α proline residue had been hypothesized to be important for LRH-1 binding, as is the case for the LRH-1-SHP complex (Ortlund et al., 2005). However, we were not able to model this proline residue due to weak electron density. A natural extension of this work would be to investigate the interaction surface of LRH-1 with a larger portion of PGC1α. Although the degree of intrinsic disorder of PGC1α would likely make crystallization of the complex difficult, HDX-MS experiments with ERRγ and a large domain PGC1α have been successful in the past(Devarakonda et al., 2011). This approach could be a useful way to study the effect of PGC1α on LRH-1 conformation (and vice versa), as well as the role particular residues play at the interface of the two partners.

In addition to the differential effects at the AFS, PGC1α and Tif2 had strikingly different effects on LRH-1 dynamics at AF-B, seen in both solution-based and *in silico* experiments. HDX-MS showed no significant effect on AF-B when PGC1α was bound, whereas Tif2 caused reduced flexibility at this site. Moreover, MDS showed a greater strength of intramolecular communication from AF-B to the AF-H in the presence of Tif2 (Figure 4.6). Collectively, we have identified two distinct mechanisms through which coregulators alter LRH-1 conformation in the presence of the same LRH-1 agonist. The reason for the differing mechanisms of activation is unclear, but a possible explanation lies in the fact that Tif2 is expressed at a fairly constant level,

while PGC1α is highly inducible. Tif2 has a relatively low binding affinity for LRH-1 (Figure 4.1), and it is plausible that recruitment of Tif2 is driven mainly by availability of an activating ligand. On the other hand, PGC1α-driven activation likely originates from upstream signaling pathways causing, for example, cAMP generation and consequent PGC1α production. Strong action of PGC1α at the LRH-1 AFS could then promote a receptor conformational change that favors agonist binding and transcriptional activation. This could thus serve as a platform through which LRH-1 could drive alternative transcriptional programs in response to specific stressors. Although additional work is needed to support this hypothesis, the idea that Tif2-mediated activation is more reliant on signaling from the ligand is supported by the fact that communication from AF-B to the AFS is greatly weakened upon mutation of a RJW100 contact critical for activation of LRH-1 by this ligand (Mays et al., 2016), (Figure 4.7). Strong communication between AF-B and the ligand is a hallmark of activating ligands, as shown in our previous publications (Mays et al., 2016; Musille et al., 2015; Musille et al., 2012). The identification of separate mechanisms of action of two major LRH-1 coactivators has potential to be exploited for selective targeting of desired LRH-1 signaling pathways as a novel therapeutic strategy for the treatment of metabolic diseases and cancer.

4.5 Materials and Methods.

4.5.1 Materials and Reagents

FAM-labeled coregulator peptides were synthesized by RS synthesis (SMRT, Tif2, and SHP) or purchased from Thermofisher Scientific (PGC1α) (sequences listed below). Unlabeled PGC1α peptide corresponding to NR Box 2, used for crystallization and HDX, was purchased from RS synthesis. Unlabeled Tif2 peptide corresponding to NR Box 3, used for HDX, was purchased from RS Synthesis (all peptide sequences are listed below). Enantiomerically pure RR-RJW100 (1*S*,3a*S*)-5-hexyl-4-phenyl-3a-(1-phenylvinyl)-1,2,3,3a,6,6a-hexahydropentalen-1-ol) was synthesized as described (Mays et al., 2016; Stec, 2010; Whitby et al., 2011). For simplicity,

this compound will be referred to as "RJW100," a term previously used to describe a racemic mixture of RJW100 enantiomers(Mays et al., 2016; Whitby et al., 2011). The vector for His-tagged Tobacco Etch Virus (TEV) protease was a gift from John Tesmer (University of Texas at Austin). The pMSC7 vector was provided by John Sondek (University of North Carolina at Chapel Hill).

4.5.2 Protein purification

Purification of the LRH-1 LBD (residues 299-541) from the pMSC7 vector and was carried out as in previous studies (Mays et al., 2016). Briefly, protein expression was induced in *E. coli* BL21 PLysS *E. coli* with 1 mM isopropyl 1-thio-β-galactopyranodside for 4 h at 30 °C. Following purification via nickel affinity chromatography, the protein was cleaved from the His tag using TEV protease. Cleaved protein was incubated with RJW100 overnight (10-fold molar excess) and then repurified by size exclusion chromatography. The crystallization buffer consisted of 100 mM ammonium acetate, pH 7.4, 150 mM sodium chloride, 1 mM DTT, 1 mM EDTA, and 2 mM CHAPS. For full-length (FL) LRH-1 purification, cell growth and fusion protein preparation were carried out as previously described(Weikum et al., 2016). To obtain pure hLRH-1-CYP7A1 promoter complex, dsDNA was added to fusion protein at 1.2-fold molar excess. The complex was incubated with TEV protease overnight at 4 °C for His-SUMO tag cleavage. Samples were cleared of precipitate and loaded to a gel filtration column for purification. Fractions containing the complex were pooled, concentrated and frozen in liquid N_2. Aliquots were stored at -80 °C for future use. Purity was assessed by SDS-PAGE and Coomassie staining and was found to be ~95% pure. Protein concentration was determined using the BCA assay (Pierce) prior to use in coregulator assays.

4.5.3 Fluorescence polarization coregulator binding assays

FL LRH-1 (co-purified with a portion of the CYP7A1 promoter) (Weikum et al., 2016) was incubated with RJW100 (10-fold molar excess) or an equal volume of DMSO overnight at

4°C. Binding of the ligand was verified using differential scanning fluorimetry, as described (Mays et al., 2016) (data not shown). The complex was serially diluted in assay buffer (20 mM HEPES, pH 7.4, 150 mM NaCl) in black-walled 384 well plates. Coregulator peptides, labeled at the N-terminus with fluorescein (FAM) were then added to a final concentration of 50 nM. Fluorescence polarization was measured on a BioTek Neo plate reader. Sequences of peptides used were as follows: Tif2 NR Box 3 ($^{+}H_3N$ -PVSPKKKENALLRYLLDKDDT-CO_2^-); PGC1α NR Box 2 $^{+}H_3N$- EEPSLLKKLLLAPA -CO_2^-); SHP NR Box 1 ($^{+}H_3N$ -QGAASRPAILYALLSSSLK-CO_2^-), SMRT ($^{+}H_3N$-TNMGLEAIIRKALMGKYDQW-CO_2^-). Assays were conducted three times in triplicate, using three separate preparations of protein. The exception was the *E. coli* PL-SHP complex, which was assayed twice in triplicate. Data were combined and fitted to a single site equilibrium binding equation with GraphPad Prism software to determine K_D. Significance of difference in K_D values was determined using 2-way ANOVA followed by Sidak's multiple comparisons test. P-values < 0.05 were considered significant.

4.5.4 Crystallization.

Protein ligand complexes were concentrated to 6.5 mg/ml and incubated with a peptide from PGC1α NR Box 2 at four-fold molar excess. Crystals were grown by hanging drop vapor diffusion at room temperature, using 1 µl protein and 1 µl crystallant (0.05 M sodium acetate, pH 4.6, 14% PEG 4000, and 15-21% glycerol) per drop.

4.5.5 Structure determination

Crystals were flash frozen in liquid nitrogen, using a cryoprotectant of crystallant with 30% glycerol. Data was collected remotely from the Argonne National Laboratory (South East Regional Collaborative Access Team), using the 22ID beamline. Data were processed using HKL2000 (Otwinowski, 1997). The structure was phased by molecular replacement using Phaser in Phenix (Adams et al., 2010), with PBD 5L11 (ligand and coactivator removed) used as the

search model. The model was refined using Phenix.refine (Adams et al., 2010), and figures were generated using Pymol (Schrodinger, 2010).

4.5.6 Hydrogen-deuterium exchange (HDX) mass spectrometry

Purified LRH-1 LBD (His tag removed) was complexed with RJW100 by incubation overnight (ten-fold molar excess) and then repurified by size exclusion into a buffer of 20 mM Tris-HCl, pH 7.5, 150 mM NaCl, and 5% glycerol. Additional RJW100 was added to the sized complex to ensure the receptor was saturated with ligand (five-fold molar excess). Solution-phase amide HDX experiments were carried out as described previously (Feng et al., 2016) using a fully automated system, in which sample handling was done using CTC HTS Twin PAL robots (LEAP Technologies) housed inside a 4 °C cabinet. In parallel reactions, 10 μM of the LRH1 LBD-RJW100 complex was premixed with four-fold molar excesses of peptides derived from either PGC1α ($^{+}H_3N$- EEPSLLKKLLLAPA-CO_2^-) or Tif2 ($^{+}NH_3$-KENALLRYLLDKDDT-CO_2^-). LRH-1-RJW100 with no coregulator added (designated "apo" for these studies) was used for comparison. The LRH-1-RJW100-coregulator complexes were allowed to form on ice and then subjected to HDX analysis. For the differential HDX experiments, 5 μl aliquots of 10 μM apo-LRH1 or LRH1-peptide complexes were mixed with 20 μl of D_2O-containing HDX buffer (20 mM Tris-HCl, pH 7.5, 150 mM NaCl, 5% glycerol) and incubated for a range of exchange times from 10 sec to 1 hr before quenching the deuterium exchange with an acidic quench solution (5M Urea, 50 mM TCEP, and 1% v/v TFA pH 2.4). Protease digestion was performed in-line with chromatography using an immobilized pepsin column. Mass spectra were acquired on a Q Exactive hybrid quadrupole-Orbitrap mass spectrometer (Thermo Scientific) and peptide identification from the MSMS data was done using mascot. HDX experiments for each pairwise comparison (apo vs. Tif2-bound LRH1 LBD or apo vs. PGC1α -bound LRH1 LBD) were run separately under the same conditions and percent deuterium exchange values for peptide isotopic

envelopes at each time point were calculated and processed using the HDX Workbench software (Pascal et al., 2012).

4.6.7 Model construction for molecular dynamics simulations

Six LRH-1 LBD complexes were prepared for molecular dynamics simulations, all containing the RJW100 ligand in the binding pocket. The first set was constructed from PBD 5L11 (with the Tif2 peptide bound). These included (i) wild-type (WT), (ii) T352V, and E534A protein. The second set was constructed from PBD 5UNJ (with the PGC1α peptide bound): (iv) WT, (v) T352V, and (vi) E534A. All mutations were introduced *in silico* to the referenced structure coordinates. For consistency, all structures contained LRH-1 residues 300-540, and missing residues (*i.e.* that could not be modeled in the structures) were added to the models used in the simulations.

4.6.8 Molecular dynamics simulations

The complexes were solvated in an octahedral box of TIP3PB water with a 10 Å buffer around the protein complex. Na^+ and Cl^- ions were added to neutralize the protein and achieve physiological conditions. All systems were set up using xleap in AmberTools (v15) (Case and Kollman, 2012) with the parm99-bsc0 forcefield (Pérez et al., 2007). Parameters for the RJW100 ligand were obtained using Antechamber (Wang et al., 2001) in AmberTools. All minimizations and simulations were performed with Amber14 (Case et al., 2014). Systems were minimized with 5000 steps of steepest decent followed by 5000 steps of conjugate gradient minimization with 500 kcal/mol·$Å^2$ restraints on all atoms. Restraints were removed from all atoms excluding the atoms in both the ligand and the Tif2 and PGC1α peptides, and the previous minimization was repeated. The systems were heated from 0 to 300 K using a 100-ps run with constant volume periodic boundaries and 5 kcal/mol·$Å^2$ restraints on all protein and ligand atoms. MD equilibration was performed for 12 ns with 10 kcal/mol·$Å^2$ restraints on protein and ligand atoms using the NPT ensemble. Restraints were reduced to 1 kcal/mol·$Å^2$ for an additional 10 ns of MD equilibration.

Then restraints were removed and 500 ns production simulations were performed for each system in the NPT ensemble. A 2-fs timestep was used and all bonds between heavy atoms and hydrogens were fixed with the SHAKE algorithm (Ryckaert et al., 1977). A cut-off distance of 10 Å was used to evaluate long-range electrostatics with Particle Mesh Ewald (PME) and for van der Waals forces. 25,000 evenly spaced frames were taken from each simulation for analysis. Structural averaging and analysis were performed with the CPPTRAJ module(Roe and Cheatham III, 2013) of AmberTools. The NetworkView plugin (Sethi et al., 2009b) in VMD (Humphrey et al., 1996) and the Carma program (Glykos, 2006a) were used to produce dynamical networks for each system. Briefly, networks are constructed by defining all protein C-α atoms as nodes, using Cartesian covariance to measure communication within the network. Pairs of nodes that reside within a 4.5 A cutoff for > 75% of the simulation are connected via an edge. Edge weights are inversely proportional to the covariance between the nodes. Networks are resolved into communities, *i.e.* a group of nodes with correlated motions. Communities are generated using the Girvan-Newman algorithm. The minimum number of communities possible were generated while maintaining at least 98% maximum modularity (Newman, 2006). Suboptimal paths between the AF-B and AFS regions were identified using the Floyd-Warshall algorithm (Floyd, 1962a). Suboptimal path analyses were performed using Carma and the subopt program in NetworkView. Cross-correlation matrices for C-α atoms in each system were computed with Carma.

4.6 Acknowledgments

The authors thank Sally Bloodworth (University of Southampton, UK) for assistance with synthetic chemistry.

Chapter 5. Discovery of the First Low Nanomolar LRH-1 Agonist Through Structure-Guided Design

5.1 Abstract

As a key regulator of metabolism and inflammation, the orphan nuclear hormone receptor, Liver Receptor Homolog-1 (LRH-1), has potential as a therapeutic target for obesity-associated metabolic diseases. Discovery of LRH-1 modulators has been difficult, in part due to the tendency for synthetic compounds to bind unpredictably within the lipophilic binding pocket. Using an iterative, structure-guided approach, we exploit a newly-discovered polar interaction to lock agonists in a consistent orientation. This enabled the discovery of the first low nanomolar LRH-1 agonist, one hundred times more potent than the previous best modulator. We demonstrate that the new agonist is active *in vivo* and elucidate a novel mechanism of action, which relies upon specific polar interactions deep in the LRH-1 binding pocket. These studies constitute major progress in developing LRH-1 modulators with potential clinical utility.

5.2 Introduction

Liver Receptor Homolog 1 (LRH-1; NR5A2) is a nuclear hormone receptor (NR) that is highly expressed in the liver and tissues of endodermal origin. It is indispensable during embryonic development, where it plays a role in maintenance of pluripotency(Gu et al., 2005), as well as in the development of the liver and pancreas(Nissim et al., 2016). In adults, LRH-1 controls a diverse transcriptional program related to metabolism, inflammation, and cellular proliferation. Metabolic effects include bile acid biosynthesis(Lu et al., 2000), reverse cholesterol transport(Schoonjans et al., 2002; Stein et al., 2014), *de novo* lipogenesis(Lee et al., 2011; Stein, 2016), and glucose phosphorylation and transport(Bolado-Carrancio et al., 2014; Oosterveer et al., 2012). This ability to modulate lipid and glucose metabolism suggests therapeutic potential for LRH-1 agonists in metabolic diseases such as non-alcoholic fatty liver disease, type II diabetes, and cardiovascular disease. Indeed, obese mice treated with the phospholipid LRH-1 agonist dilauroylphosphatidylcholine (DLPC; PC 12:0/12:0) exhibit improved glucose tolerance and insulin sensitivity, as well as lower hepatic and circulating triglycerides(Lee et al., 2011). These anti-diabetic effects occur in an LRH-1-dependent manner and have been primarily attributed to a reduction of *de novo* lipogenesis(Lee et al., 2011). While these findings demonstrate a tremendous potential for LRH-1 as a therapeutic target, natural PLs have poor pharmacological properties, such as low potency and poor aqueous solubility. Small molecule agonists are highly sought; however, the large and highly lipophilic LRH-1 binding pocket has been extremely challenging to target.

Currently, the most promising LRH-1 agonists are certain *cis*-bicyclo[3.3.0]-octene derivatives discovered by Whitby and colleagues(Whitby et al., 2006; Whitby et al., 2011). The best of this class, named RJW100, was discovered as a part of an extensive synthetic effort to improve acid stability and efficacy of a related compound, GSK8470(Whitby et al., 2011) (Figure 5.1a). Recently, we determined the crystal structure of LRH-1 bound to RJW100 and made the

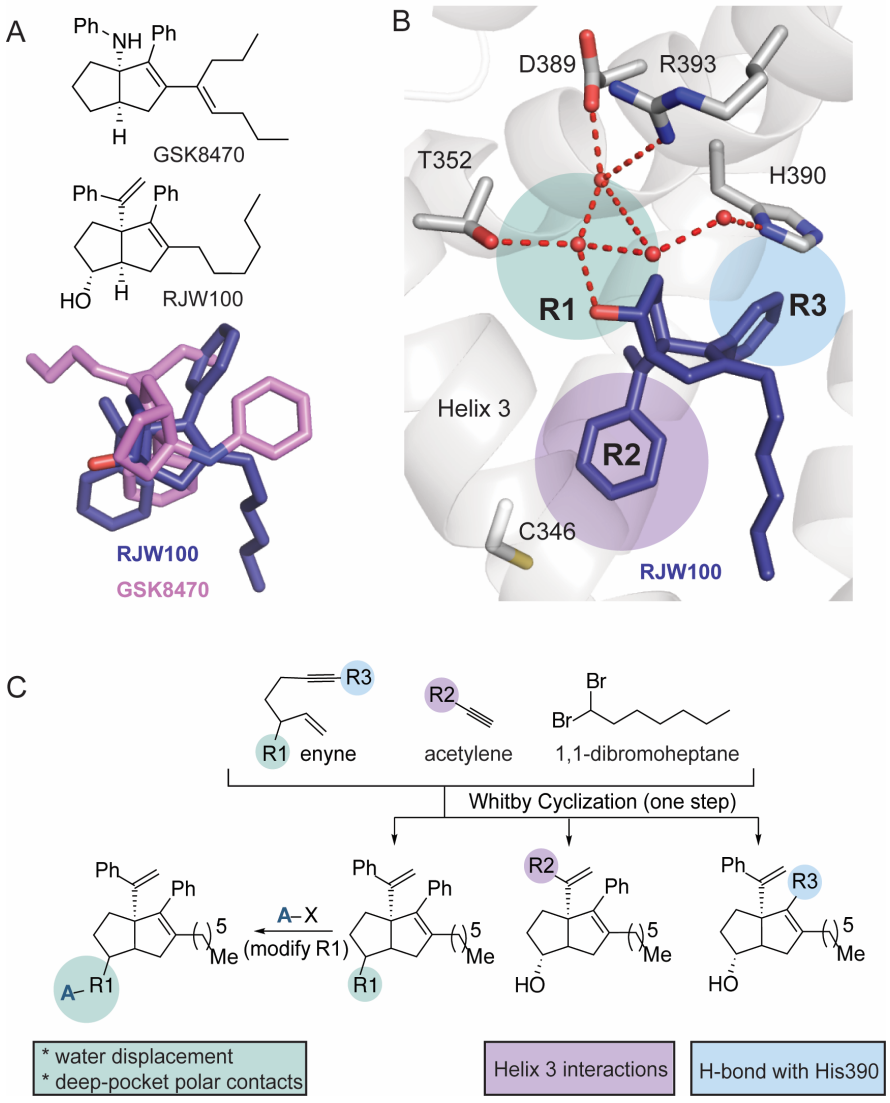


Figure 5.1. Structure-based design of LRH-1 agonists. A. *Top,* Chemical structures of the agonists GSK8470 and RJW100. *Bottom,* Superposition of GSK8470 and RJW100 (from PDB 3PLZ and 5L11, respectively) show the very different binding modes for these similar agonists. B. RJW100 interacts with LRH-1 residue T352 via water. The four water molecules shown coordinate a group of polar residues deep in the binding pocket. The colored circles indicate the space targeted by modifications to the RJW100 scaffold. C. Overview of synthetic strategy used to generate the novel agonists.

surprising discovery that it exhibits a completely different binding mode than GSK8470. The bicyclic cores of the two agonists are perpendicular to each other (Figure 5.1a), resulting in differing mechanisms of action(Mays et al., 2016). The apparent tendency for ligands in this class to rotate within the pocket is thus a confounding factor in agonist design.

In the LRH-1-RJW100 crystal structure, the ligand hydroxyl group contacts a network of water molecules in the deep part of the ligand binding pocket (Figure 5.1b). This water network is conserved in LRH-1 structures and coordinates a small group of polar residues (*e.g.* T352, H390, and R393) in an otherwise predominantly hydrophobic pocket. Using a RJW100 analog lacking a hydroxyl group or a LRH-1 T352V mutation, we demonstrated that this interaction is required for RJW100-mediated activation of LRH-1(Mays et al., 2016). Moreover, the *endo* disastereomer of RJW100 adopts a nearly identical pose as RJW100 and makes the same water-mediated contact with T352, supporting the idea that this interaction contributes to ligand orientation(Mays et al., 2016). As the basis for these studies, we hypothesized that this polar interaction could anchor the ligand core, enabling more predictable targeting of desired parts of the pocket. We therefore synthesized and evaluated novel compounds around the RJW100 scaffold with the primary aim of strengthening deep-pocket polar contacts. This systematic approach has revealed agonist features important for potency and efficacy, including a mid-sized polar moiety substituted for the RJW100 hydroxyl group, *endo* stereochemistry, and a linker containing a hydrogen bond donor. The best new agonist has an EC_{50} of 15 nM, two orders of magnitude lower than RJW100, and it modulates expression of LRH-1-controlled genes in guinea pig liver *in vivo.* We present three crystal structures of LRH-1 bound to novel agonists, which depict a consistent position of the bicyclic cores and modified polar groups projecting into the deep part of the pocket as designed. Using a combination of structural and biochemical approaches, we show that potency is linked to specific deep-pocket interactions, increased allosteric signaling to the LRH-1 activation function surface, and stronger coactivator

recruitment. This breakthrough in LRH-1 agonist development is a crucial step in developing potential new treatments for metabolic diseases.

5.3 Results

5.3.1 Locking the agonist in place with polar interactions.

Our recent structural studies have revealed a tendency for highly similar LRH-1 synthetic agonists to bind unpredictably within the hydrophobic binding pocket, which has presented a challenge for improving agonist design in a rational manner(Mays et al., 2016). We observed that the agonist RJW100 makes a water-mediated contact with residue T352 in the deep part of the LRH-1 binding pocket that is critical for binding and activation(Mays et al., 2016) and reasoned that strengthening this contact may secure synthetic compounds in a consistent orientation. We therefore synthesized RJW100 analogs with bulkier polar groups in place of the RJW100 hydroxyl (R1, Figure 5.1b), aiming to displace bridging waters and to generate direct interactions with T352 or other nearby polar residues (this deep-pocket polar region near T352 is hereafter abbreviated "DPP"). In parallel, we synthesized compounds designed to interact with DPP by (1) modifying the internal styrene to promote direct hydrogen bonding with LRH-1 residue H390 or (2) modifying the external styrene to promote interactions with residues in helix 3 (R3 and R2 on the RJW100 scaffold, respectively, Figure 5.1b). To generate the novel agonists, we utilized and expanded on Whitby's work with zirconecene catalysis(Thomas, 2006), using an enyne, an alkyne, and 1,1-dibromoheptane to generate all-carbon bridgehead 1-alken-2-yl substituted systems with varying functionality at positions R1, R2, and R3 (Figure 5.1c). Full synthetic methods will be described in the dissertation of Autumn Flynn (Emory Department of Chemistry, Jui lab).

We evaluated the new compounds using differential scanning fluorimetry (DSF), since entropic gain from displacement of buried water molecules or favorable energetics from bond formation would result in global stabilization of the LRH-1-agonist complex. As previously

observed(Mays et al., 2016), RJW100 stabilizes the LRH-1 ligand binding domain (LBD) by around 3 ºC relative to a PL ligand (Figure 5.2a). While the styrene-modified compounds (**1-15)** destabilize the receptor relative to RJW100 (Figure 5.2a) and tend to be poor activators (Figure 5.2S), substitution of a sulfamate for the hydroxyl group at R1 is highly stabilizing (Figure 5.2a). Both the *endo*-sulfamate (**16**) and the *exo*-sulfamate (**17**) greatly increase the T_m of the LRH-1-ligand complex, with particularly strong stabilization seen by the *endo* diasteromer (T_m increase of 8.2 ºC by **16**, Figure 2a). This strong stabilization is likely indicative of water displacement, possibly accompanied by new hydrogen bond formation.

To investigate how the sulfamate modification affects ligand binding mode and receptor conformation, we determined the X-ray crystal structure of **16** (the *endo* diastereomer) bound to the LRH-1 LBD to a resolution of 2.0 Å (Table 1, Figure 5.2b). The complex was crystallized with a fragment of the coactivator protein, Transcriptional Intermediary Factor 2 (Tif2) to stabilize the complex and to model the active state. Tif2 is bound, as expected, at the LRH-1 activation function surface (AFS), located at the interface of helices 3, 4, and 12 (Figure 2b). Overall protein conformation does not differ greatly from the LRH-1-RJW100 structure (RMSD = 0.1 Å relative to PBD 5L11). The binding pocket assumes the contracted conformation seen in previous LRH-1-synthetic agonist structures, smaller in volume and narrower at the mouth of the pocket than when LRH-1 is bound to phospholipid ligands(Mays et al., 2016). Strong electron density surrounding the ligand clearly indicates the position of the bicyclic core and phenyl rings (Figure 5.2c), which display a nearly identical conformation as RJW100 (Figure 5.2d). The sulfamate moiety is well-defined by strong, tetrahedral electron density: it protrudes into the DPP as intended, filling the space typically occupied by two conserved water molecules (Figure 5.2d). The sulfamate directly interacts with residue T352 and makes several other polar contacts within the DPP (Figure 5.2e). These include water-mediated interactions with residues H390 and R393, both predicted to be important for receptor activation(Lalit et al., 2013). The *exo* sulfamate (**17**)

Table 5.1: X-ray data collection and refinement statistics: LRH bound to compounds 16, 33, or 25 and to the Tif2 coregulator.

Data collection	LRH-1—**16**—Tif2	LRH-1—**33**—Tif2	LRH-1—**25**—Tif2
Space group	$P4_32_12$	$P4_32_12$	$P4_32_12$
Cell dimensions			
a, *b*, *c* (Å)	46.5, 46.5, 221.0	46.7, 46.7, 218.0	46.7, 46.7, 222.7
α,β,γ (°)	90, 90, 90	90, 90, 90	90, 90, 90
Resolution (Å)	50 – 2.00 (2.07-2.00)	50 – 2.23 (2.31-2.23)	50 – 2.20 (2.28-2.20)
R_{pim}	0.06 (0.52)	0.07 (0.46)	0.04 (0.31)
$I / \sigma I$	21.3 (1.72)	8.9 (3.2)	18.5 (1.6)
$CC_{1/2}$ in highest shell	0.596	0.976	0.697
Completeness (%)	99.9 (100.0)	97.3 (86.5)	96.6 (87.9)
Redundancy	11.2 (6.8)	16.6 (12.5)	21.1 (13.0)
Refinement			
Resolution (Å)	2.00	2.23	2.20
No. reflections	17346	12205	13216
R_{work} / R_{free} (%)	20.6/ 24.5	23.2/26.9	20.2/23.7
No. atoms			
Protein	4038	4098	4077
Water	71	24	31
Ligand	68	69	69
B-factors			
Protein	44.8	60.5	56.8
Ligand	53.6	66.1	65.8
Water	44.0	52.7	51.5
R.m.s. deviations			
Bond lengths (Å)	0.002	0.002	0.003
Bond angles (°)	0.504	0.474	0.511
Ramachandran favored (%)	97.6	98.0	97.9
Ramachandran outliers (%)	0.4	0.0	0.0
PDB accession code	tbd	tbd	tbd

Values in parenthesis indicate highest resolution shell

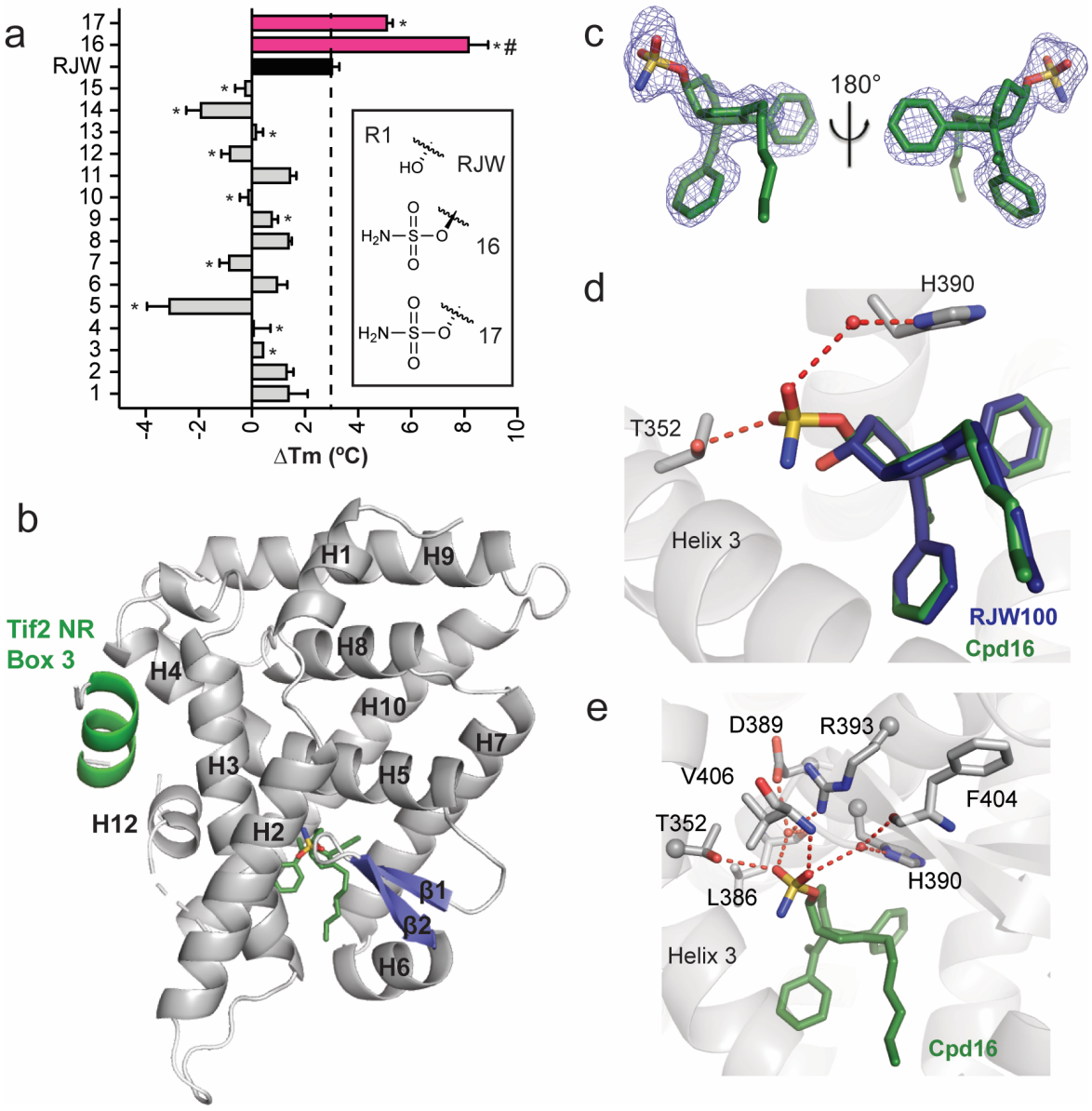


Figure 5.2. The endo-sulfamate stabilizes LRH-1 and directly interacts with T352. A. DSF showing destabilization of LRH-1 relative to RJW100 by compounds modified at R2 and R3 (**1-15**) and the stabilizing effect of the *endo*-sulfamide (**16**) and *exo*-sulfamide (**17**). Values shown are Tm differences relative to the LRH-1-DLPC complex. Significance of differences was determined using One-Way ANOVA with Sidak's multiple comparisons test. *, $p < 0.05$ compared to RJW100; #, significantly more stabilizing than **16** ($p < 0.05$). *Inset,* chemical structure of the R1 group for the sulfamates and RJW100. *Dotted line* indicates the degree of

stabilization by RJW100. B. Overall x-ray crystal structure of the LRH-1 LBD bound to **16.** The complex was co-crystallized with a fragment of the Tif2 coactivator (green). H, helix. The dotted line indicates a region of the structure that was disordered and could not be modeled. C. Omit map showing the binding mode of **16** (2Fo-Fc, contoured to 2.5 σ). D. Superposition of **16** (dark green) and RJW100 (deep blue) showing a consistent conformation of the agonist cores and a direct interaction with T352. E. Multiple polar interactions are made by **16** in the deep part of the LRH-1 binding pocket.

may not be ideally positioned to make these additional contacts (as seen by following the trajectory of the *exo*-RJW100 hydroxyl group, Figure 5.2d), likely explaining the increased stabilization of LRH-1 when bound to **16** *versus* **17** in DSF experiments (Figure 5.2a). Together, these results provide strong support to the hypothesis that interactions in the DPP prevent major repositioning of the ligand and demonstrate the potential for strategic targeting of specific sites in the pocket by expanding the R1 polar group.

5.3.2. Limitations of the sulfamate agonist.

Although **16** is positioned in the DPP as expected, we found that it does not significantly improve potency or efficacy over RJW100 in cellular luciferase reporter assays (EC_{50} = 900 nM for **16** *versus* 1.5 μM for RJW100, Figure 5.3a). Poor solubility or cell permeability could play a role in this effect, but there is also evidence of mechanistic differences between **16** and RJW100**.** LRH-1 activation by **16** is not significantly affected by either a T352V or H390A mutation (Figure 5.3b). This is quite different than RJW100, which is completely dependent on the T352 interaction to activate LRH-1(Mays et al., 2016), and it suggests that **16** does not require stable interactions with these DPP residues as a part of its mechanism of action. Indeed, the LRH-1-**16** structure provides a few indications that the ligand may be somewhat mobile. Ligand B-factors are higher than the average protein B-factors, including the atoms in the sulfamate group (average B for atoms in the sulfamate is 70.9 *versus* 44.9 for protein atoms, ratio of 1.6). The presence of small areas of positive electron density near the sulfamate (not shown) also supports the idea that this group may be mobile and not forming stable interactions in the DPP.

To discover how **16** and RJW100 differentially alter LRH-1 dynamics to drive receptor activation, we investigated how they affect LRH-1 conformation in solution using hydrogen-deuterium exchange mass spectrometry (HDX). The most significant changes are seen at or near the LRH-1 activation function surface (AFS). Relative to RJW100, **16** destabilizes LRH-1 helix

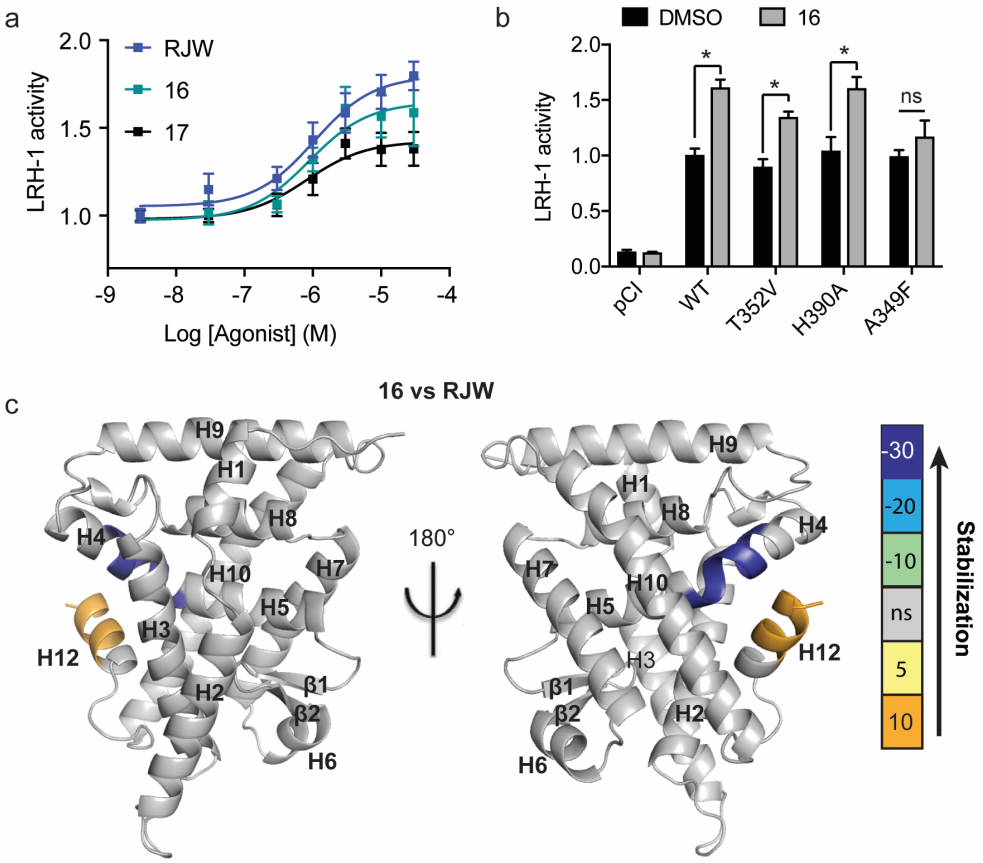


Figure 5.3. Limitations of the sulfamate agonist. A. Luciferase reporter assay in Hela cells comparing the sulfamate agonists to the parent compound, RJW100. Each point represents the mean and SEM of three experiments conducted in triplicate. Activity is reported as fold over baseline (DMSO only). B. Luciferase reporter assay in Hela cells overexpression wild-type (WT), mutant LRH-1, or the pCI empty vector as a control. The A349F mutant occludes the DPP, preventing activation by these compounds. This was used as a negative control. Each point represents the mean and SEM of three experiments conducted in triplicate. Activity is reported as fold over baseline (DMSO only). C. Differential HDX data comparing **16** to RJW100 showing stabilization of a portion of helix 4 and destabilization of helix 12 by **16**. Color bar indicates the percent difference in deuterium uptake when **16** is bound compared to RJW100. Data are mapped onto PBD 4DOS with three residues modeled in at the c-terminus to reflect peptide coverage in the HDX study.

12, a critical part of the AFS that undergoes conformational fluctuations to favor either coactivator or corepressor binding(Musille et al., 2015) (Figure 5.3c). At the same time, it strongly stabilizes a portion of helix 4 near the AFS: 35% less deuterium is incorporated there when **16** is bound *versus* RJW100 (Figure 5.3c). The stabilized region is near the ligand binding site but does not overlap with it, suggesting that the effect is allosteric (perhaps compensating for the destabilized helix 12). Thus, while **16** is well-positioned to interact in the DPP, mobility of the sulfamate group and destabilization of an important part of the AFS may limit the ability to activate LRH-1.

5.3.3 Discovery of a low nanomolar LRH-1 agonist by enhancing polar interactions.

To improve cellular activity of the agonists, we sought to optimize the size, geometry, and stereochemistry of the R1 group targeting the DPP. We therefore synthesized and evaluated compounds with R1 modifications ranging from small and linear to much bulkier polar groups, including hydrogen bond donors and acceptors and *endo* and *exo* diastereomers (compounds **18-39**, Figure 5.4a). Overall, there is a striking correlation between potency in luciferase reporter assays and LRH-1 stabilization by DSF for these compounds, where lower EC_{50}s are associated with higher T_m values ($p = 0.0009$, Figure 5.4b). This correlation provides a direct link between cellular activity and receptor stabilization and suggests that improved potency is due to specific polar interactions mediated by the R1 group.

The compounds display a wide range of potencies and efficacies in luciferase reporter assays (Figure 5.4c, 5.4S), enriching our understanding of agonist properties that drive LRH-1 activation. Both size and stereochemistry of the R1 group are important. Mid-sized, polar groups, mainly tetrahedral in geometry, tend to increase potency relative to RJW100 (pink boxes in Figure 5.4a). Compounds with small R1 groups (**18-28)** or very bulky R1 groups (**38-39**) are less potent or inactive (Figure 5.4S). The close relationship between R1 size, agonist potency, and LRH-1 stabilization is evident looking at DSF results, where a strong peak in stabilization

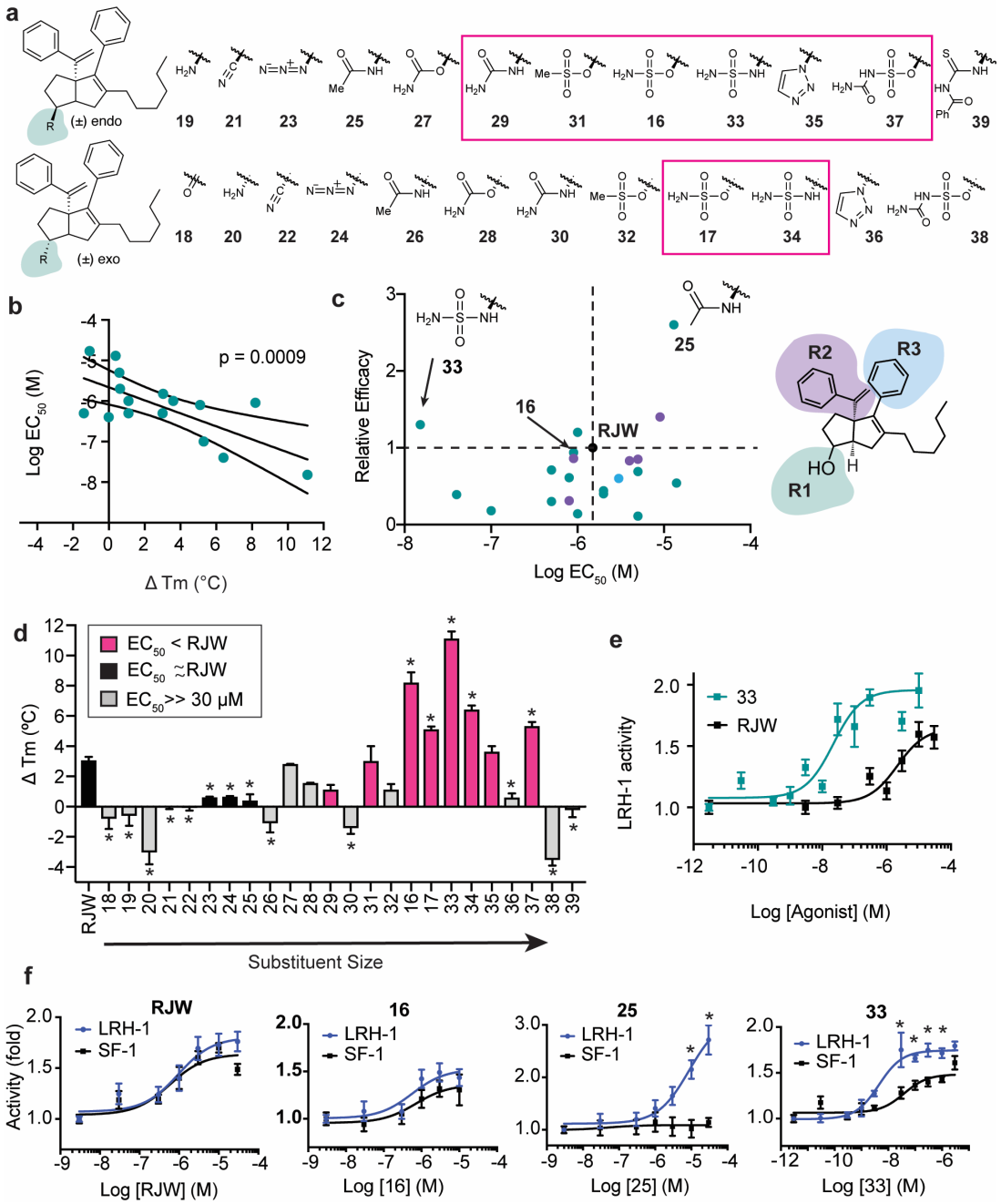


Figure 5.4. Optimization of the R1 modification improves potency by two orders of magnitude. A. Chemical structures of the compounds **19-39** (R1 groups shown). B. Scatter plot showing the correlation between Tm shift in DSF assay (x-axis) and EC_{50} from luciferase reporter assays (y-axis) for compounds **19-39**. Data were analyzed by linear regression. Curved lines represent the 95% confidence interval. C. Scatter plot comparing potency (EC_{50}) and efficacy

relative to RJW100 (relative efficacy) for all compounds (**1-39**) for which EC_{50} values could be calculated. Results are color-coded by site of modification (as shown in the schematic on the right). The relative potencies and efficacies of compounds **16, 25,** and **33** are indicated with arrows. Relative efficacy was calculated as described in the online methods section. D. Size and stereochemistry of the R1 modification affects global LRH-1 stabilization in DSF assays. Each bar represents three experiments conducted in triplicate. *, $p< 0.05$ for Tm change *versus* RJW100. Significance was assessed by One-Way ANOVA followed by Dunnett's multiple comparison's test. E. Dose response curves comparing **33** and RJW100 in luciferase reporter assays. Each point represents the mean +/- SEM for three experiments conducted in triplicate. F. Dose response curves comparing activation of LRH-1 and SF-1 by select compounds. Significance of differences in activities of each compound toward for LRH-1 *versus* SF-1 was determined by 2-Way ANOVA followed by Sidak's multiple comparison's test. *, $p< 0.05$.

occurs for the mid-sized compounds (Figure 5.4e). Interestingly, RJW100 is much more active (and stabilizing) than **18-24** despite having a small R1 (a hydroxyl group). The fact that **18-24** are inactive suggests that the water-mediated T352 interaction requires a hydrogen bond acceptor with nonlinear geometry. Another strong trend among the data is that *endo* diastereomers tend to be better activators (and more stabilizing) than the corresponding *exo* disatereomers. For example, the *endo*-triazole (**35**) has an EC_{50} of 1 μM, while the *exo*-triazole (**36)** is minimally active at concentrations up to 30 μM (Figure 5.4S). Similar trends are seen for the *endo vs exo* versions of the sulfamides, mesylates, and acetamides (Figure 5.4S). By far the most stabilizing and potent agonist evaluated is the *endo* sulfamide (**33**). With an EC_{50} of 15.7 +/- 0.8 nM, compound **33** is two orders of magnitude more potent than RJW100, and it is also a stronger activator (Figure 5.4d-e). This is the first discovery of a low nanomolar LRH-1 modulator, a leap forward in developing agonists for this challenging target. Also notable is the selectivity of **33** and the *endo* acetamide (**25**) for LRH-1 over the closest LRH-1 homolog, Steroidogenic Factor-1 (SF-1, Figure 5.4f). Compound **33** is less active against SF-1 than LRH-1 at several doses, whereas **25** strongly activates LRH-1 and has not activity against SF-1 in the dosage range tested (Figure 5.4f). In contrast, RJW100 and **16** equally activate both receptors (Figure 5.4f).

5.3.4 Activation of LRH-1 in vivo.

The discovery of the first highly potent LRH-1 agonist provides the opportunity to elucidate ligand-regulated transcriptional pathways controlled by LRH-1 *in vivo*. Mice and rats are not ideal models to evaluate LRH-1 agonists, since the rodent receptor contains sequence differences in the LBD that alter ligand binding(Musille et al., 2013). Key sequence changes span helix 6-7, an important allosteric site utilized for ligand-mediated activation of the human receptor(Mays et al., 2017; Musille et al., 2015; Musille et al., 2012) (Figure 5.5a). We therefore chose guinea pigs for these studies, which have identical protein sequence as human LRH-1 in the helix 6-7 loop and at the ligand binding site (protein sequences are 97% identical overall). To

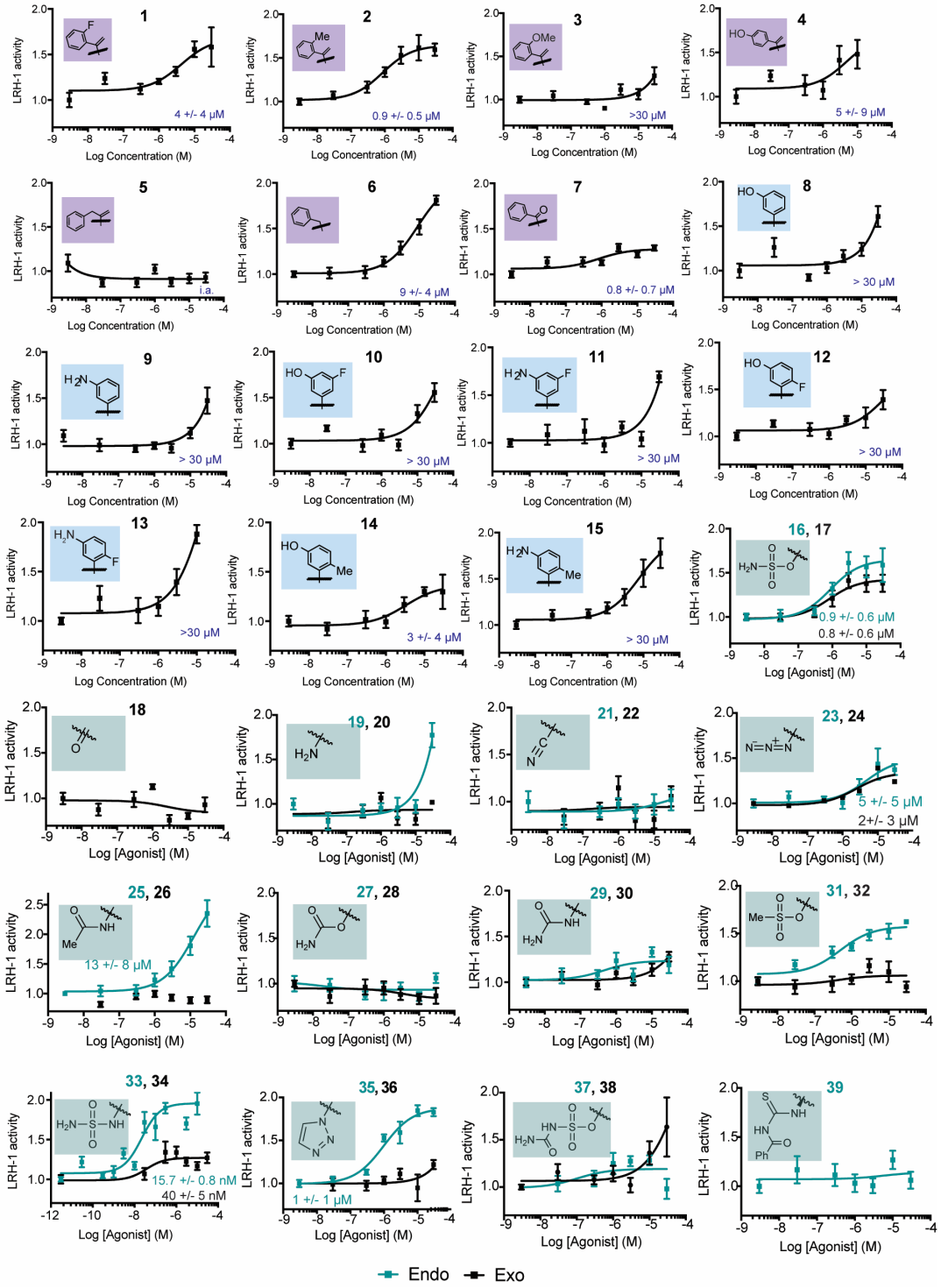


Figure 5.4S. Summary of Data from Luciferase Reporter Assay for Compounds 1-39.

Insets, chemical structures of the modified groups. Colored boxes indicate the site of modification: purple, R2; blue, R3; teal, R1.

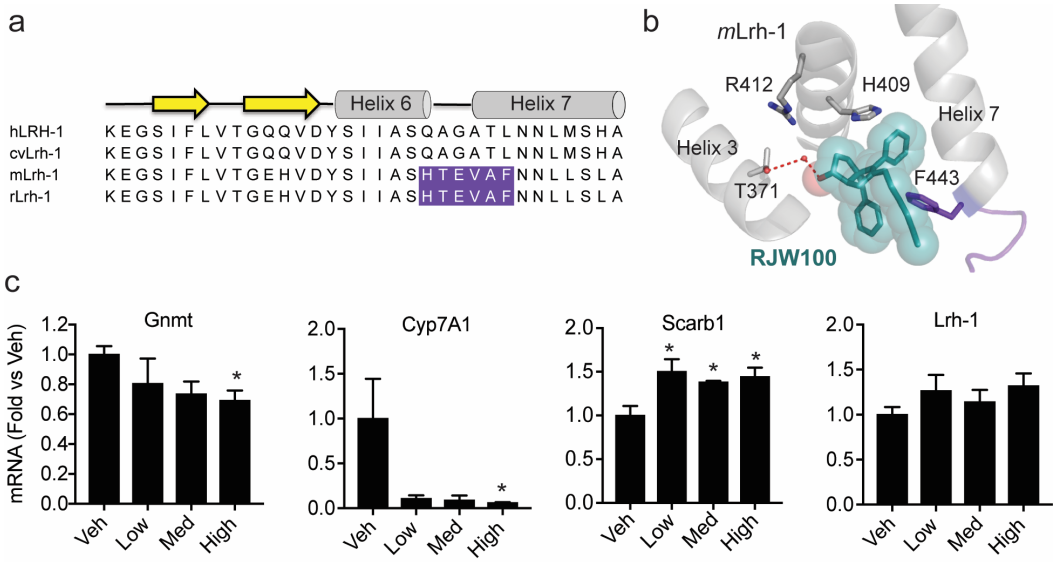


Figure 5.5. Compound 33 activates Lrh-1 in vivo. A. Choice of guinea pigs for this experiment was guided by the comparison of the helix 6-helix 7 region of LRH-1. Sequence differences in mouse (m) and rat (r) Lrh-1 are shown. Human and guinea pig (cv) Lrh-1 are identical in this key region and X% identical overall. B. Model of RJW100 from PBD 5L11 superposed onto mLrh-1, showing a predicted clash with residue Phe443 in the mouse receptor. C. Guinea pigs were treated with intraperitoneal injections of vehicle (Veh) or **33** at 0.001 mg/kg (Low), 0.01 mg/kg (Med), or 0.1 mg/kg (High). Liver tissue was analyzed by qRT-PCR for changes in expression of known LRH-1 transcriptional targets. Each bar represents the mean +/- SEM for 5-10 animals in each group. Significant of differences was determined using the Kruskal-Wailis test followed by Dunn's multiple comparison's test. *, $p < 0.05$.

assess agonist activity, guinea pigs were treated with **33** or **16** by a single intraperitoneal injection, and changes in expression of known LRH-1 transcriptional target genes in the liver were measured by qRT-PCR. To increase to likelihood of measuring direct transcriptional targets and not secondary effects, we chose an acute treatment period (8 hours). Whereas compound **16** did not affect Lrh-1 activity at doses up to 100 mg/ kg (data not shown), a single treatment of **33** at 0.01-0.1 mg/ kg was sufficient to alter expression of Lrh-1 transcriptional targets involved in lipid and methyl metabolism, including Cytochrome P450 Family 7 Subfamily A Member 1 (Cyp7a1), glycine methyltransferase (Gnmt), and Scavenger Receptor 1 (Scarb1) (Figure 5.5b). Lrh-1 mRNA levels were not significantly changed by the treatment in these experiments (Figure 5.5b). This is a substantially lower dosage range than used in previous *in vivo* studies with LRH-1 agonists (which have required 10-100 mg/ kg)(Cobo-Vuilleumier et al., 2018; Lee et al., 2011). Though more work is needed to thoroughly characterize the activation profile of **33**, these studies provide the first evidence of *in vivo* activity of a highly potent agonist at very low dosage levels.

5.3.5 Structural basis for LRH-1 binding and activation by compound 33.

The dramatic increase in potency for **33** compared to **16** is driven by replacement of oxygen with nitrogen in the R1 linker, since this is the only difference between the two. Remarkably, this effect appears to be a generalizable: a nitrogen-containing linker improves potency relative to an oxygen linker for several sets of compounds that differ only at this site (Figure 5.6a). To investigate the role of the R1 linker in agonist activity and to gain insights into mechanisms contributing to the potency of **33**, we determined two x-ray crystal structures of the LRH-1 LBD bound to agonists with nitrogen-containing linkers: **33** and **25** (Table 5.1). Functionally, **25** is somewhat less potent than RJW100 (EC_{50} = 13 μM), but it exhibits high levels of activity at 30 μM (Figure 5.4). The two structures were determined to resolutions of 2.2-2.3 Å and do not exhibit obvious changes in overall protein conformation compared the LRH-1-RJW100 complex (overall RMSD for both are within 0.2 Å of LRH-1-RJW100). The positions of the ligands are

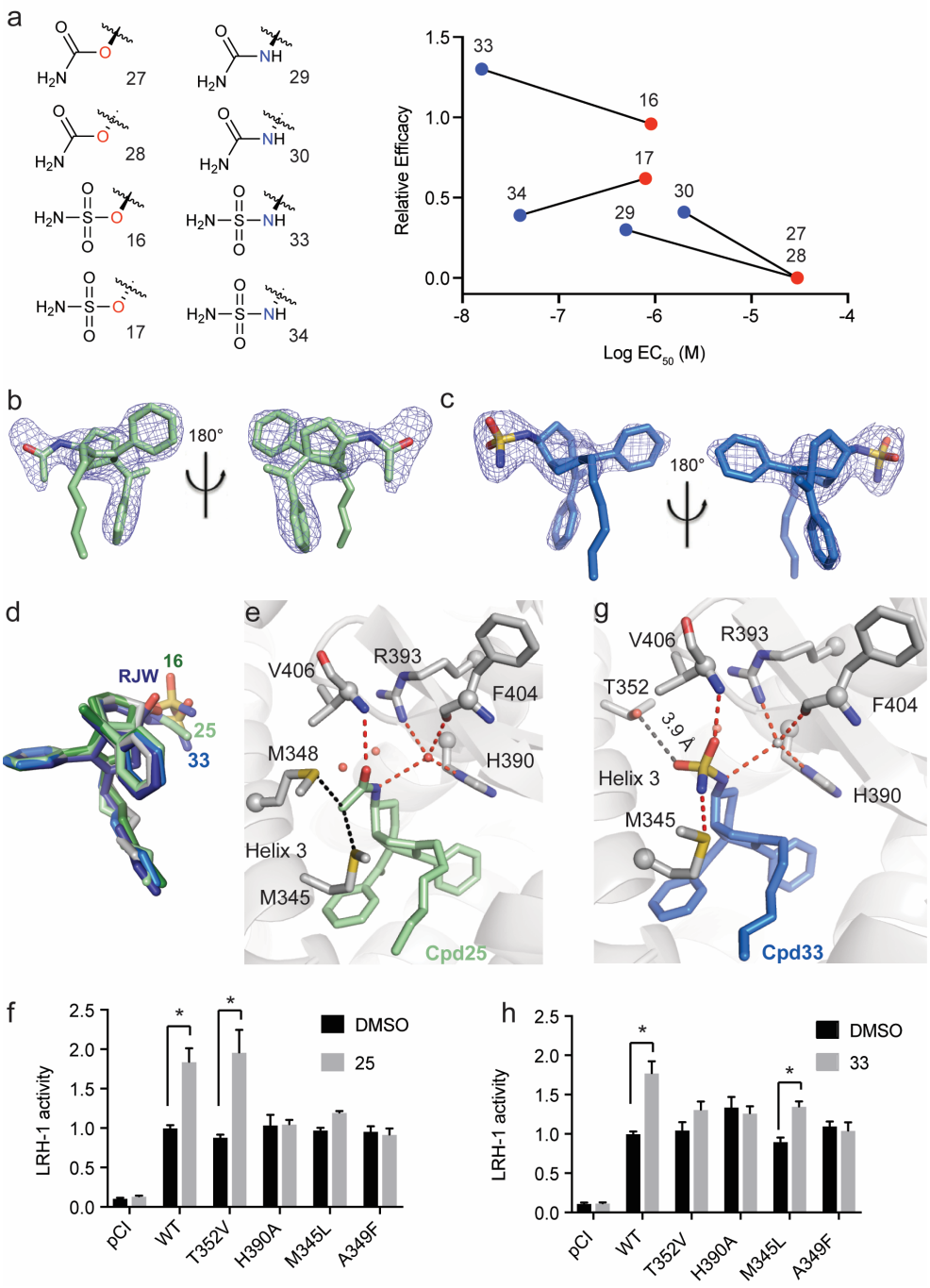


Figure 5.6. Structural basis for improved potency of compound 33. A. An NH linker in the R1 group improves potency. Comparison of potencies and efficacies for four sets of compounds that are identical except for the presence of a R1 linker containing an oxygen (red dots) or nitrogen (blue dots). Compounds that had no activity at doses up to 30 µM are (**27** and **28**) are

plotted as having EC_{50}s of 30 μM and Relative Efficacy of 0 for illustrative purposes. B-C. Omit maps showing the binding modes of **25** (*B)* and **33** (*C*). Maps are F_O-F_C, contoured at 2.5σ. D. Superposition of ligands from four crystal structures showing a consisted position of the cores of the modified agonists compared to RJW100. E-F. Close view of LRH-1 binding pocket with either **25** (*E)* or **33** (*F)*, featuring interactions made by the modified R1 groups. G-H. Results from luciferase reporter assays showing how the interactions made by the agonists observed in the crystal structures affect LRH-1 activity. Each bar represents the mean +/- SEM for three independent experiments conducted in triplicate. Cells were treated with 10 μM **25** or 0.3 μM **33** for 24 hours (concentrations were chosen based on agonist EC_{50} toward WT LRH-1). *, $p< 0.05$ by Two-Way ANOVA followed by Dunnett's multiple comparisons test.

well-defined by the electron density (Figure 5.6b-c). As with the LRH-1-**16** structure, the ligand tails are disordered. This may be a general feature of *endo* agonists with this scaffold, since the tail of *endo*-RJW100 is also disordered (whereas *exo*-RJW100 has strong electron density surrounding the tail)(Mays et al., 2016). Superposition of RJW100, **16, 25,** and **33** from each structure show an extremely consistent orientation of the agonists' cores, with slight variation in the positions of the R1 headgroups (Figure 5.6d).

A close view of the ligand binding pocket in the LRH-1-**25** structure (Figure 5.6e) shows the R1 group interacting with the backbone amide of residue Val406 via the carbonyl oxygen (also seen with **16** and **33**). The R1 acetamide methyl group makes hydrophobic interactions with sidechains of two methionine residues on helix 3 (M345 and M348). We were particularly interested in potential interactions made by the NH linker and observed that it participates in water-mediated polar contacts with residues H390 and R393. These interactions appear to be very important, since **25** is completely unable to activate a LRH-1-H390A mutant (Figure 6f). In contrast, the R1 acetamide is not well-positioned to interact with the water coordinating T352, and the T352V mutation has no effect on LRH-1 activity (Figure 5.6f).

In the LRH-1-**33** structure, the R1 group is shifted slightly compared to **16**, amounting to a ~1 Å difference in positions of the sulfur atoms and a 2.1 Å difference in positions of the R1 nitrogen atoms for these agonists. This shift places the sulfamide within hydrogen bonding distance of M345 but not T352 (which is 3.9 Å away, Figure 5.6g). As with **16,** ligand B-factors suggest that the R1 group may be sampling more than one conformation. Since **33** is unable to activate a LRH-1-T352V mutant but is not significantly affected by a M345L mutation (Figure 5.6h), it is likely that it interacts with T352 at least transiently and that the M345 interaction is less important.

Like **25**, the NH linker of **33** is positioned to make water-mediated contacts with H390**;** however, we have not modeled a bridging water molecule due to weak electron density. The

weak density for the water molecule is likely a consequence of poor crystallographic order, since very few waters could be modeled in this structure (24 total, unusual for a 2.3 Å structure). However, the fact that **33** is unable to activate a LRH-1-H390A mutant supports the idea that **33** makes this interaction and that it is functionally important (Figure 5.6h). Compound **16,** with an oxygen linker, is positioned to interact with the water molecule coordinating H390 (both with the linker and a sulfonyl in the sulfamate group, Figure 5.2d-e). However, since **16** activates the LRH-1-H390A mutant equally well as wild-type LRH-1 (Figure 5.3b), it does not utilize the H390 interaction for activation. Together, these results suggest that activation of LRH-1 via the H390 interaction requires an interaction with a hydrogen bond donor and provide a potential mechanism through which the NH linker increases agonist potency.

5.3.6 Compound 33 stabilizes the AFS, strengthens allosteric signaling, and promotes coactivator recruitment.

To understand how the binding mode of **33** leads to LRH-1 activation, we investigated how it affects LRH-1 conformation and allostery using structural and computational approaches. LRH-1 activation by phospholipids and RJW100 involves allosteric signaling from the helix 6/ β-sheet area (AF-B) to the AFS[15,16]. Communication between these two regions is thought to convey ligand status to the AFS to direct recruitment of appropriate coregulators. Using HDX, we show that **33** and RJW100 differentially affect the conformation of AF-B. Relative to RJW100, **33** destabilizes N-terminal portion of helix 7 and stabilizes the loop between helices 6 and 7 (Figure 7a). Rigidification of the loop between these helices may induce pressure to unwind helix 7, which could explain this pattern of motion. In addition to these changes at AF-B**, 33** strongly stabilizes a portion of helix 4 near the AFS (as seen with **16)**; however, unlike with **16,** there is no concomitant destabilization of helix 12 (Figure 5.7a). A direct comparison of **33** and **16** shows clear stabilization of helix 12 by **33** (Figure 5.7b).

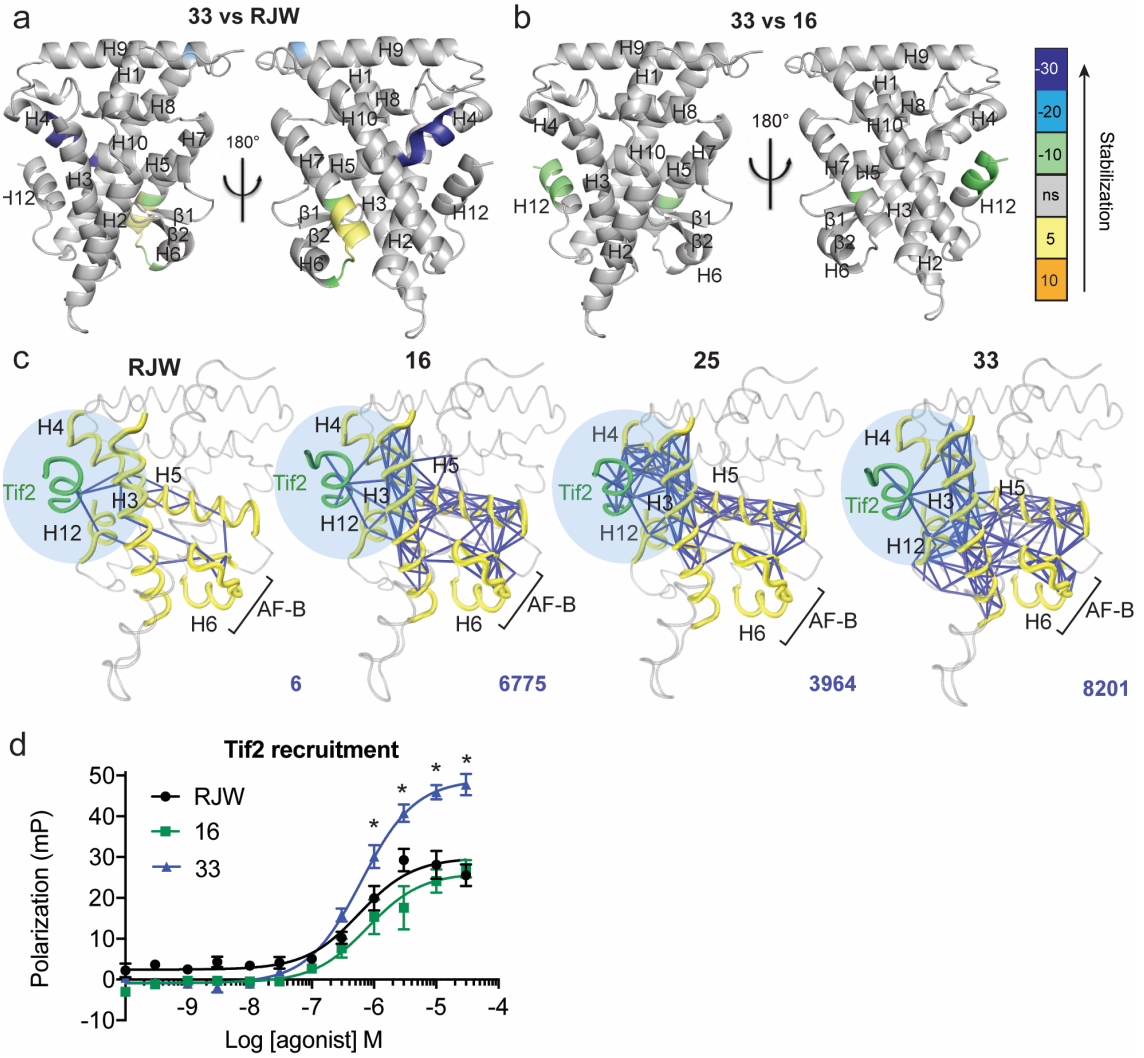


Figure 5.7. Compound 33 promotes allosteric communication to the AFS and coactivator recruitment. A-B. Differential HDX comparing **33** to RJW100 (*A)* or to **16** (*B)*. Color bar indicates the percent difference in deuterium uptake when **33** is bound compared to RJW100 or **16**. C. MDS results showing the strongest suboptimal paths (blue lines) between AF-B and the Tif2 coactivator (green) when the indicated agonists are bound. The AFS is highlighted in light blue. Numbers in blue in the bottom right corner of each graphic indicate the total number of strong paths found in the analysis. D. Compound 33 promotes recruitment of the Tif2 coactivator to purified LRH-1 LBD in a fluorescence polarization-based binding assay.

The conformational changes induced by **33** suggest that it may increase communication between AF-B and AFS. To quantify and model the strength of agonist-driven communication between these two sites, we conducted 1 μs molecular dynamics simulations (MDS) using the new crystal structures as starting models. Residues undergoing correlated motions in proteins are believed to facilitate allosteric coupling between distant sites(Gasper et al., 2012; Sethi et al., 2009a; VanWart et al., 2012). Communication paths can traverse thousands of possible routes through the receptor, and the chains of residues with the strongest patterns of correlated motion—the optimal path and a subset of suboptimal paths—are thought to convey the most information(Bowerman and Wereszczynski, 2016; Van Wart et al., 2014). We therefore constructed dynamical networks of LRH-1-agonist complexes, using calculated covariance to weight the strength of communication between pairs of residues. We then identified the strongest suboptimal paths facilitating communication between AF-B and Tif2 coactivator (bound at the AFS). There is a striking increase in the number of paths when **16**, **25**, and **33** are bound compared to RJW100, with **33** exhibiting the strongest communication between these sites (Figure 5.7c). There are also significant differences in the directionality of the paths promoted by each agonist. Although all paths traverse helix 5, indicating that correlated motion is induced in this region, compounds **16**, **25**, and **33** also induce strong communication along helix 3. Compound **33** also induces highly interconnected communication within the AFS and the Tif2 coactivator, including significant involvement of helix 12. Helix 12 is notably excluded from the paths when the other agonists are bound. This finding is consistent with the HDX data, which show stabilization of helix 12 by compound **33**.

As expected from the data showing stabilization of the AFS and increased allosteric communication to the site of coactivator binding, **33** enhances coactivator association. We show this using a fluorescence-polarization-based coregulator binding assay to quantify binding of the Tif2 coactivator to purified LRH-1 LBD in the presence of increasing concentrations of RJW100,

16, or **33**. All agonists dose-dependently recruit fluorescein-labeled Tif2 peptide to LRH-1 and exhibit similar EC_{50}s (50% of maximum Tif2 binding occurs with ~600-700 nM agonist, Figure 5.7d). Each curve reaches a well-defined plateau that indicates the maximum response with saturating concentrations of agonist; however, curve maxima are lower for RJW100 and **16** than **33** by 50-60%. Lower quantities of coactivator binding at saturating ligand concentrations is a characteristic of partial agonists, which by definition only activate a portion of the receptors they occupy. Although the endogenous ligand or other full agonist for LRH-1 has not been defined for comparison, **33** behaves more like a full agonist than **16** and RJW100. Therefore, we have elucidated a novel mechanism of action utilized by **33**, whereby specific interactions by the sulfamide and R1 linker promote allosteric signaling to the AFS, stabilizing the site of coactivator interaction and increasing Tif2 association.

5.4 Discussion

LRH-1 is an attractive therapeutic target for diabetes and other metabolic disorders. It is downregulated in the liver in human NAFLD patients *versus* healthy individuals(Sahini and Borlak, 2016). Acute knockdown of Lrh-1 in mouse hepatocytes *in vivo* causes formation of large lipid droplets in the liver that is exacerbated by high fat diet, suggesting a direct link between fatty liver disease and Lrh-1 expression(Miranda et al., 2018). Further, activation of Lrh-1 by DLPC in models of diet-induced obesity reduces hepatic and circulating lipids and has profound antidiabetic effects that are Lrh-1-dependent(Lee et al., 2011). Selective activation of Lrh-1 (by ablating Prox1-mediated repression) improves cholesterol homeostasis in obese mice, reducing formation of atherosclerotic lesions(Stein et al., 2014). Because of its role in steroid production in the gut, LRH-1 agonists also have potential in inflammatory disorders such as inflammatory bowel disease, demonstrated in preclinical studies(Bayrer et al., 2018; Coste).

While this therapeutic potential is widely recognized, LRH-1 has been notoriously difficult to target with synthetic modulators. Agonists with the hexahydropenalene

scaffold(Whitby et al., 2006; Whitby et al., 2011) (such as RJW100) are promising and have been used in several studies to probe LRH-1 biology(Cobo-Vuilleumier et al., 2018; Lee et al., 2017; Mamrosh et al., 2014; Venteclef et al., 2010; Zhang et al., 2011). However, we have shown that small modifications to this scaffold can greatly affect binding mode(Mays et al., 2016). This has confounded previous attempts to improve agonist potency and efficacy(Whitby et al., 2011). By exploiting a novel polar interaction in the LRH-1 DPP, we have overcome this challenge and have made substantial progress in agonist development. Systematic variation of the R1 group of the RJW100 scaffold revealed a robust structure-activity relationship: increased potency is associated with global receptor stabilization by DSF and by mid-sized, polar R1 substituents with *endo* stereochemistry (Figure 5.4). In addition, the composition of the R1 group, particularly the linker, is critical for activity. This is exemplified through the comparison of **16** and **33**, which differ only at the R1 linker. Whereas the R1 group of **16** is somewhat mobile and does not stably interact with DPP residues (Figure 5.3), **33** utilizes interactions with both T352 and H390 to activate LRH-1, the latter of which is likely mediated by the linker nitrogen (Figure 5.6). This novel binding mode leads to a distinct mechanism of action for **33** compared to similar, less potent compounds, inducing conformational changes at AF-B, stabilizing helix 12, and increasing coactivator association (Figure 6). Results from MDS support the idea that **33** promotes very strong allostery to the AFS, evidenced in the strong communication between the AF-B and the AFS predicted to occur when **33** is bound compared to less potent agonists (Figure 5.7).

With three separate crystal structures, we demonstrated that polar modifications at the RJW100 R1 group do not cause major repositioning of the scaffold (Figure 5.2, 5.6), supporting our hypothesis that this polar group acts as an important anchor point to secure the scaffold. This finding was not only key to the success of the current study, but it will also greatly benefit future work. The ability to anchor the scaffold consistently provides an opportunity to tune for additional desired effects, such as solubility or selectivity. Moreover, the trajectory of the alkyl

"tails" of these molecules is amenable for introduction of modifications that could engage residues near the mouth of the pocket in a PL-like manner(Musille et al., 2012; Sablin et al., 2015). Initial studies in this vein have been fruitful, leading to the discovery of highly active compounds(Flynn et al., 2018). Finally, the establishment of a predictable binding mode may open avenues for antagonist design; for example, by modifying the scaffold to promote displacement of helix 12 and recruitment of corepressors. This approach has been successful for other nuclear receptors(Raaijmakers et al., 2009; Shiau, 1998) and could generate LRH-1 antagonists useful as therapeutics for certain cancers in which LRH-1 is aberrantly active(Bayrer et al., 2015; Chand et al., 2010; Lai et al., 2013; Lin et al., 2014; Nadolny and Dong, 2015; Safi, 2005; Thiruchelvam et al., 2011). This is an active area of research in our laboratory.

In conclusion, a systematic, structure-guided approach has resulted in the discovery of the first low nanomolar LRH-1 agonist and elucidated a novel mechanism of action. This agonist has great potential as a tool to uncover novel aspects of LRH-1 biology and as a therapeutic for obesity-associated metabolic diseases. Equally importantly, the discovery of elements that stabilize the orientation of the hexahydropentalene scaffold and drive activation of LRH-1 is invaluable for understanding ligand-regulation of this receptor and for future design of LRH-1 modulators.

5.5 Materials and Methods

5.5.1 Chemical Synthesis of Cpd33

The synthesis and characterization of key target compound **33** is outlined below. Detailed synthetic procedures and characterization data for all new compounds will be provided in the dissertation of Autumn Flynn, Emory Department of Chemistry, Jui lab.

Exo 5-hexyl-4-phenyl-3a-(1-pheylvinyl)-1,2,3,3,3a,6,6a-hexahydropoentalen-1-yl methanesulfonate (**32**): A solution of RJW100 *exo* (122.5 mg, 0.3 mmol, 1.0 equiv) in

dichloromethane was treated with methanesulfonyl chloride (5.0 equiv), then triethylamine (5.0 equiv) The reaction mixture was allowed to stir 1 h before concentrating and purifying on silica in 30% EtOAc/hexanes eluent. (**32** *exo*: 139 mg, >99% yield) ^{1}H NMR (500 MHz, $CDCl_3$) δ 7.37 – 7.25 (m, 8H), 7.27 – 7.19 (m, 2H), 5.11 (d, *J* = 1.3 Hz, 1H), 5.01 (d, *J* = 1.3 Hz, 1H), 4.83 (d, *J* = 4.0 Hz, 1H), 2.95 (s, 3H), 2.63 (d, *J* = 9.2 Hz, 1H), 2.41 (dd, *J* = 17.4, 9.5 Hz, 1H), 2.14 (dd, *J* = 17.5, 2.0 Hz, 1H), 2.11 – 1.98 (m, 4H), 1.90 – 1.75 (m, 2H), 1.40 – 1.31 (m, 2H), 1.32 – 1.17 (m, 6H), 0.87 (t, *J* = 7.1 Hz, 3H).

Endo 5-hexyl-4-phenyl-3a-(1-phenylvinyl)-1,2,3,3a,6,6a-hexahydropentalen-1-amine **(23):** A solution of **32** (57 mg, 0.12 mmol, 1.0 equiv) in DMF was treated with sodium azide (10.0 equiv) and the reaction was stirred 16 h at 80 °C behind a blast shield. The solution was allowed to cool to room temperature and poured over water and extracted with EtOAc three times. The combined organic layers were washed with water and brine, dried over $MgSO_4$, and concentrated. The reaction mixture was purified on silica in 0-10% EtOAc/hexanes eluent. (**23** *endo*: 117.6 mg, 95% yield) (Note: inversion of stereochemistry). (Warning: caution must be exercised when handling organic and inorganic azides for their toxicity and instability. Aqueous layers were basified and disposed of appropriately). ^{1}H NMR (600 MHz, $CDCl_3$) δ 7.36 – 7.26 (m, 8H), 7.23 – 7.18 (m, 2H), 5.10 (d, *J* = 1.3 Hz, 1H), 4.94 (d, *J* = 1.3 Hz, 1H), 3.87 (ddd, *J* = 10.5, 8.8, 5.9 Hz, 1H), 2.62 – 2.51 (m, 2H), 2.16 – 2.01 (m, 4H), 1.97 – 1.88 (m, 1H), 1.79 (ddd, *J* = 12.4, 5.9, 1.8 Hz, 1H), 1.71 (td, *J* = 12.4, 5.2 Hz, 1H), 1.67 – 1.59 (m, 1H), 1.40 (p, *J* = 7.5 Hz, 2H), 1.31 – 1.19 (m, 5H), 0.87 (t, *J* = 7.2 Hz, 3H).

Endo 5-hexyl-4-phenyl-3a-(1-phenylvinyl)-1,2,3,3a,6,6a-hexahydropentalen-1-amine **(19):** Under nitrogen, a solution of **23** (54 mg, 0.13 mmol, 1.0 equiv) in anhydrous Et_2O was cooled to 0 °C and treated dropwise with $LiAlH_4$ (4.0M in Et_2O, 10.0 equiv). The reaction was stirred at ambient temperature until the reaction was complete by TLC (ca. 1 h). The reaction was cooled to 0 °C, diluted with anhydrous Et_2O, and slowly treated with water (1mL/g $LiAlH_4$). Excess 4 M NaOH

was added slowly and the solution was extracted with EtOAc three times. The combined organic layers were washed with Rochelle's salt and brine, dried over $MgSO_4$, and concentrated. The crude oil was purified by silica gel chromatography in 50% EtOAc/Hexanes eluent (1% triethylamine) to afford the title compounds as colorless oils. (**19** *endo*: 47.9 mg, 95% yield). ^{1}H NMR (600 MHz, $CDCl_3$) δ 7.37 – 7.19 (m, 10H), 5.08 (d, *J* = 1.4 Hz, 1H), 4.94 (d, *J* = 1.5 Hz, 1H), 3.30 (ddd, *J* = 11.0, 8.8, 5.7 Hz, 1H), 2.48 (d, *J* = 17.4 Hz, 1H), 2.42 (t, *J* = 9.0 Hz, 1H), 2.12 – 2.00 (m, 2H), 1.83 – 1.78 (m, 1H), 1.73 – 1.68 (m, 2H), 1.46 – 1.37 (m, 2H), 1.35 – 1.20 (m, 8H), 0.88 (t, *J* = 7.1 Hz, 3H).

endo 5-hexyl-4-phenyl-3a-(1-phenylvinyl)-1,2,3,3a,6,6a-hexahydropentalen-1-yl sulfamide **(33)** A solution of **19** (30 mg, 0.08 mmol, 1.1 equiv) in DCM was treated with triethylamine (2.0 equiv.) and solution of 2-oxo-1,3-oxazolidine-3-sulfonyl chloride (0.5 M in DCM, 1.0 equiv) (prepared according to the procedure of Borghese *et a*l(Borghese, 2006)). The reaction was stirred at room temperature for 3 h then concentrated. The residue was treated with ammonia (0.5 M in dioxane, 1.5 equiv) and triethylamine (3.0 equiv). The solution was heated in a sealed tube at 85°C for 16 h behind a blast shield. After cooling to ambient temperature, the reaction was diluted with 3:3:94 MeOH:Et_3N:EtOAc and passed through a pad of silica. The eluent was concentrated, and the crude oil was purified on silica in 20-30% EtOAc/hexanes eluent to afford the title compound as a colorless oil. (**33** *endo*: 21.6 mg, 60% yield). ^{1}H NMR (600 MHz, $CDCl_3$) δ 7.33 – 7.23 (m, 8H), 7.20 – 7.17 (m, 2H), 5.09 (d, *J* = 1.3 Hz, 1H), 4.96 (d, *J* = 1.3 Hz, 1H), 4.44 (s, 2H), 4.36 (d, *J* = 8.0 Hz, 1H), 3.84 – 3.77 (m, 1H), 2.62 (td, *J* = 8.9, 2.0 Hz, 1H), 2.38 (dd, *J* = 17.5, 2.0 Hz, 1H), 2.20 – 2.13 (m, 1H), 2.08 – 2.04 (m, 2H), 2.00 – 1.95 (m, 1H), 1.74 – 1.70 (m, 2H), 1.50 – 1.43 (m, 1H), 1.42 – 1.16 (m, 8H), 0.86 (t, *J* = 7.1 Hz, 3H). ^{13}C NMR (126 MHz, $CDCl_3$) δ 154.1, 143.6, 142.8, 139.3, 136.6, 129.6, 127.8, 127.7, 126.9, 126.8, 115.5, 68.8, 57.2, 47.4, 35.4, 32.3, 32.0, 31.6, 29.8, 29.5, 27.9, 22.6, 14.1. LRMS (ESI, APCI) *m/z*: calc'd for $C_{28}H_{37}N_2O_2S$ $[M+H]^+$ 465.7, found 464.8

5.5.2 Biology: materials and reagents.

pCI empty vector was purchased from Promega. The SHP-luc and *Renilla* reporters, as well as pCI LRH-1, have been previously described(Musille et al., 2012). The vector for His-tagged tobacco etch virus (TEV) was a gift from John Tesmer (University of Texas at Austin). The pMSC7 (LIC-HIS) vector was provided by John Sondek (University of North Carolina at Chapel Hill). The Tif2 NR Box 3 peptide was purchased from RS Synthesis. DNA oligonucleotide primers were synthesized by Integrated DNA Technologies.

5.5.3 Protein purification.

Purification of human LRH-1 ligand binding domain (residues 300-537) in a pMCSG7 expression vector was performed as described(Mays et al., 2016). Briefly, protein was expressed in BL21 PLysS *E. coli*, using 1 mM IPTG for 4 hours (30°C) to induce expression. Protein was purified by nickel affinity chromatography. For DSF assays, protein eluted from the nickel column was exchanged with DLPC (5-fold molar excess overnight at 4 °C), followed by repurification by size exclusion to remove displaced lipids. The assay buffer was 20 mM Tris-HCl, pH 7.5, 150 mM sodium chloride, and 5% glycerol. Cleaved LRH-1 was then incubated with ligands overnight at 4 °C prior to repurification by size exclusion, using the same assay buffer as for DSF. Protein used for crystallography was prepared as for coregulator recruitment, except that it was sized into a buffer of 100 mM ammonium acetate (pH 7.5), 150 mM sodium chloride, 1 mM DTT, 1 mM EDTA, and 2 mM CHAPS.

5.5.4 Differential scanning fluorimetry (DSF).

DSF assays were conducted on a StepOne Plus thermocycler as previously described(Mays et al., 2017; Mays et al., 2016). Briefly, aliquots of purified LRH-1 LBD protein (0.2 mg/ ml) were incubated with saturating concentrations of ligand overnight at 4 °C. Protein-ligand complexes were heated in the presence of SYPRO orange dye at a rate of 0.5 degree/ minute. Complexes were excited at 488 nm, and fluorescence emissions at each degree Celsius

were measured using the ROX filter (~600 nm). Tm values were calculated using the Bolzmann equation in GraphPad Prism, v7.

5.5.5 Crystallography.

Compounds **16**, **33**, or **25** were incubated with purified LRH-1 LBD (His tag removed) at 5-fold molar excess overnight at 4°C. The complexes were re-purified by size exclusion chromatography into the crystallization buffer (see above). Protein was concentrated to 5-6 mg/ml and combined with a peptide from human Tif2 NR box 3 (H_3N-KENALLRYLLDKDDT-CO_2) at four-fold molar excess. Crystals were generated by hanging drop vapor diffusion at 18 °C, using a crystallant of 0.05 M sodium acetate (pH 4.6), 5-11% PEG 4000, and 0-10% glycerol. Crystals of **25** with LRH-1 were generated by microseeding, using RJW100-LRH-1 crystals as the seed stocks (crystals used for seeding were grown as described)(Mays et al., 2016).

5.5.6 Structure Determination.

Crystals were flash-frozen in liquid nitrogen, using a cryoprotectant of crystallant plus 30% glycerol. Diffraction data were collected remotely from Argonne National Laboratory, Southeast Regional Collaborative Access Team, Beamline 22ID. Data were processed and scaled using HKL2000(Otwinowski, 1997). Structures were phased by molecular replacement using Phenix(Adams et al., 2010), with PBD 5L11 used as the search model. The structure was refined using phenix.refine(Adams et al., 2010) and Coot(Emsley and Cowtan, 2004), with some additional refinement done using the PDB Redo web server(Joosten et al., 2014).

5.5.7 Tissue culture.

Hela cells were purchased from Atlantic Type Culture Collection and cultured in phenol red-free MEMα media supplemented with 10% charcoal-dextran-stripped fetal bovine serum. Cells were maintained under standard culture conditions.

5.5.8 Reporter gene assays.

Hela cells were reverse-transfected with three vectors: (1) full-length, human LRH-1 in a pCI vector, (2) a firefly reporter (pGL3 Basic) with a portion of the SHP promoter cloned upstream of the firefly luciferase gene, and (3) a constitutively active vector expressing *Renilla* luciferase under control of the CMV promoter. Transfections utilized the Fugene HD transfection reagent at a ratio of 5 µl per 2 µg DNA. To perform the reverse transfections, cells were trypsinized, combined with the transfection mixture, and plated at densities of 7,500 cells per well in white-walled 96-well plates. The following day, cells were treated with each compound (or DMSO control) for 24 hours. In most cases, six points in the concentration range of 0.03 – 30 µM were used (exceptions noted in figures), with a final DMSO concentration of 0.3% in all wells. Luciferase expression was measured using the DualGlo Kit (Promega). Firefly luciferase signal was normalized to *Renilla* luciferase signal in each well. EC_{50} values were calculated using three-parameter curve-fitting (GraphPad Prism, v.7). Assays were conducted in triplicate with at least two independent biological replicates. Experiments involving SF-1 activation were conducted in an identical manner, except full-length human SF-1 (in a pcDNA3.1+ vector) was overexpressed instead of LRH-1. Significance of differences in luminescence signal for LRH-1 *versus* SF-1 promoted by particular agonists was determined using two-way ANOVA followed by Sidak's multiple comparisons test.

5.5.9 Calculation of Relative Efficacy (RE).

This value was calculated from curve-fitting to data from luciferase reporter assays. To compare the maximum activities of the new compounds to RJW100, we used the formula $(Max_{new\_cpd} - Min_{new\_cpd}) / (Max_{RJW100} - Min_{RJW100})$, where "Max" and "Min" denote the dose response curve maximum and minimum, respectively. A RE of 0 indicates a completely inactive compound, a value of 1 indicates equal activity to RJW100, and values above 1 indicate greater activity.

5.5.10 Mutagenesis.

Mutations were introduced to LRH-1 in the pCI vector using the Quikchange Lightning site-directed mutagenesis kit (Ambion). Constructs were sequenced prior to use in reporter gene assays as described above.

5.5.11 Model Construction for Molecular Dynamics Simulations.

Four LRH-1 LBD complexes were prepared for molecular dynamics simulations. 1) LRH-1-Tif2-RJW100 (PDB 5L11), 2) LRH-1-Tif-2-**16**. 3LRH-1-Tif2-**25**, LRH-1-Tif2-**33**. For consistency, all structures contained LRH-1 residues 300–540. Missing residues (i.e., that could not be modeled in the structures) were added to the models used in the simulations.

5.5.12 Molecular Dynamics Simulations.

The complexes were solvated in an octahedral box of TIP3P water with a 10-Å buffer around the protein complex. Na^+ and Cl^- ions were added to neutralize the protein and achieve physiological buffer conditions. All systems were set up using the xleap tool in AmberTools17(Case et al., 2017) with the ff14SB forcefield(Maier et al., 2015). Parameters for the agonist ligands 33, 16 and 25 were obtained using Antechamber(Wang et al., 2001) also in AmberTools17. All minimizations and simulations were performed with Amber16(Case et al., 2017). Systems were minimized with 5000 steps of steepest decent followed by 5000 steps of conjugate gradient minimization with 500-kcal/mol·$Å^2$ restraints on all solute atoms. Restraints were removed excluding the atoms in both the ligand and the Tif2 peptide, and the previous minimization was repeated. This minimization was repeated with restraints lowered to 100-kcal/mol·$Å^2$. Finally, all restraints were removed for a last minimization step. The systems were heated from 0 to 300 K using a 100-ps run with constant volume periodic boundaries and 5-kcal/mol·$Å^2$ restraints on all protein and ligand atoms. MD equilibration was performed for 12 ns with 10-kcal/mol·$Å^2$ restraints on Tif2 peptide and ligand atoms using the NPT ensemble. Restraints were reduced to 1 kcal/mol·$Å^2$ for an additional 10 ns of MD equilibration. Then, restraints were removed, and 1000-ns production simulations were performed for each system in

the NPT ensemble. A 2-fs time step was used with all bonds between heavy atoms and hydrogens fixed with the SHAKE algorithm(Ryckaert et al., 1977). A cutoff distance of 10 Å was used to evaluate long-range electrostatics with particle mesh Ewald and for van der Waals forces. Fifty thousand evenly spaced frames were taken from each simulation for analysis, using the CPPTRAJ module(Roe and Cheatham III, 2013) of AmberTools. The NetworkView plugin(Sethi et al., 2009a) in VMD(Humphrey et al., 1996) and the Carma program(Glykos, 2006b) were used to produce dynamic networks for each system. In brief, networks are constructed by defining all protein C-α atoms as nodes, using Cartesian covariance to measure communication within the network. Pairs of nodes that reside within a 4.5-Å cutoff for 75% of the simulation are connected via an edge. Edge weights are inversely proportional to the covariance between the nodes. Networks were constructed using 500 ns of the MDS trajectories, to enable direct comparison with our previous LRH-1-RJW MDS(Mays et al., 2017). Suboptimal paths between the AF-B and Tif2 peptide were identified using the Floyd-Warshall algorithm(Floyd, 1962b). Suboptimal path analyses were performed using Carma and the subopt program in NetworkView. Cross-correlation matrices for C-α atoms in each system were computed with Carma.

5.5.13 Coregulator Recruitment Assays.

Synthetic agonists were titrated in the presence of purified LRH-1 LBD protein (2 μM) and a fluorescein (FAM)-labeled peptide corresponding to the Tif2 NR box 3 (FAM-H3N-PVSPKKKENALLRYLLDKDDT-CO2-) (50 nM). Protein and probe concentrations were determined from preliminary experiments titrating LRH-1 protein with no ligand added in the presence of FAM-Tif2 (2 μM was slightly above the Tif2 Kd in these experiments). Tif2 binding was detected by fluorescence polarization, using a BioTek Neo plate reader. Assays were conducted three times in triplicate, using two separate protein preparations. Significance of differences in Tif2 association at each dose was determined using two-way ANOVA followed by Tukey's multiple comparisons test.

5.5.14 Animal Studies.

All experimental protocols were approved by the Institute for Animal Care and Use Committee at T3 Laboratories and conformed to the Guide for the Care and Use of Laboratory Animals, published by the National Institutes of Health (NIH Publication No. 86-23, revised 1996), and with federal and state regulations. Female Hartley guinea pigs (6 weeks of age) were purchased from Charles River Laboratories. Guinea pigs were group housed and were allowed to access water *ad libitum* in an animal facility with a 12-h light–dark cycle. For treatment, **33** or **16** was dissolved in 100% ethanol and then diluted in sunflower oil to achieve solutions containing a final concentration of 5% ethanol. Guinea pigs (n = 5-10) were administered **33** (0.001 mg/kg, 0.01 mg/kg, and 0.1 mg/kg) or **16** (1 mg/kg, 10 mg/kg, and 100 mg/kg). A separate group was administered vehicle alone (5% ethanol in sunflower oil). Treatments were delivered via a single intraperitoneal injection in a volume of 1 ml. Guinea pigs were sacrificed 8 hours after treatment by cardiac puncture under anesthesia (ketamine/ xylazine, 50/5 mg/kg) and euthanized by inhalation of carbon dioxide. Liver samples were quickly extracted, washed with ice-cold PBS, snap frozen in liquid nitrogen, and stored at −80°C.

5.5.15 Gene expression analysis.

RNA was isolated from guinea pig livers using the RiboPure kit according to the manufacturer's instructions (Ambion). RNA quality was assessed by A260/280 and A260/230 values as well as presence of intact ribosomal RNA by agarose gel electrophoresis. RNA (1.5 µg) was reverse-transcribed using the High Capacity cDNA Reverse Transcription Kit (Applied Biosystems). Resulting cDNA was analyzed by qRT-PCR on a StepOne Plus Thermocycler, using SYBR green dye to visualize PCR amplification. Differences in gene expression between groups were determined using the ΔΔCt method, using Tata-box binding protein (Tbp) for normalization(Lindqvist, 2013). Primers for qPCR were designed with the help of the

PrimerBLAST tool(Ye, 2012) and exhibited amplification efficiencies between 90-108%. Primer sequences were as follows:

Cyp7a1 forward	5'- AGT GAG AAG AAG GAA AAT AGG CG -3'
Cyp7a1 reverse	5'- TGA GGA ACT CCA GAG GGT TG -3
Tbp forward	5'- ACT TGA CCT AAA GAC AAT TGC ACT TC -3'
Tbp reverse	5'- CAG CAA ACC GCT TGG GAT TA -3'
Nr5a2 forward	5'- GAA GCT GGA AGC CGT AAG GG -3'
Nr5a2 reverse	5'- TAA GTC CGT TGG CTC GGA TG -3'
Scarb1 forward	5'- ATG TCC TCA ATC CTG ACG AGG T -3'
Scarb1 reverse	5'- CGT GTC ATT GTT GTG GAA GGA GA -3'
Gnmt forward	5'- CAT GGA TGC CAG CGA GTG TT -3'
Gnmt reverse	5'- CTT CAG CAT CTT GTC ACT GGC -3'

5.5.16 Hydrogen-deuterium exchange mass spectrometry.

Following cleavage of the His tag from purified LRH-1 LBD with TEV protease as described above, the protein was further purified by size exclusion chromatography into a buffer of phosphate buffered saline (pH 7.5) plus 5% glycerol. Protein purity exceeded 98% by Coomassie staining. Protein-ligand complexes were prepared by adding each ligand at 5-fold molar excess to 2 mg/ml protein and incubating overnight at 4 °C. Complexes were centrifuged to remove any aggregates prior to analysis by HDX-MS. HDX-MS was conducted using Waters' UPLC HDX system coupled with a Q-Tof Premier mass spectrometer (Waters Corp, Milford, MA). Protein-ligand complexes were diluted 1:7 (v/v) into labeling buffer (protein buffer containing D_2O instead of water) via an autosampler. Labeling took place at 20°C for time periods of 0, 10, 100, 1,000, and 10,000 seconds prior to quenching in with equal volume of precooled quenching buffer (100 mM phosphate, 0.5 M tris)2-carboxyethl)phosphine, 0.8% formic acid, and 2% acetonitrile, pH 2.5, 1°C). After quenching, samples were applied to a Waters enzymate pepsin column (2.1 x 30 mm). Peptides from the pepsin column were separated in-line on a Waters Acuity UPLC BEH C18 column (1.7 µM, 1.0 x 100 mm) at a flow of 40 µl/ min for 12 minutes (8-40% linear gradient, mobile phase: 0.1% formic acid in acetonitrile) at 1°C. The mass spectrometer was operated with the electrospray ionization source in positive ion

mode, and the data were acquired in elevated-energy mass spectrometry mode. For internal calibration, a reference lock-mass of Glu-Fibrinopeptide (Sigma-Aldrich, St Louis, MO) was acquired along with each sample data collection. Peptides were identified by comparison to human LRH-1 protein sequence using the ProteinLynx Global SERVER (version 3.02). HDX data were processed in DynamX (version 3.0). Mass assignment for each peptide at 0 seconds of exchange was checked manually, and any assignment with a mass deviation > 0.2 Da was removed. Hydrogen-deuterium exchange protection was quantified by comparison of hydrogen exchange profiles at different time points. Peptide coverage was 99.2% with a redundancy of 6.68 for this experiment (Figure 5.8).

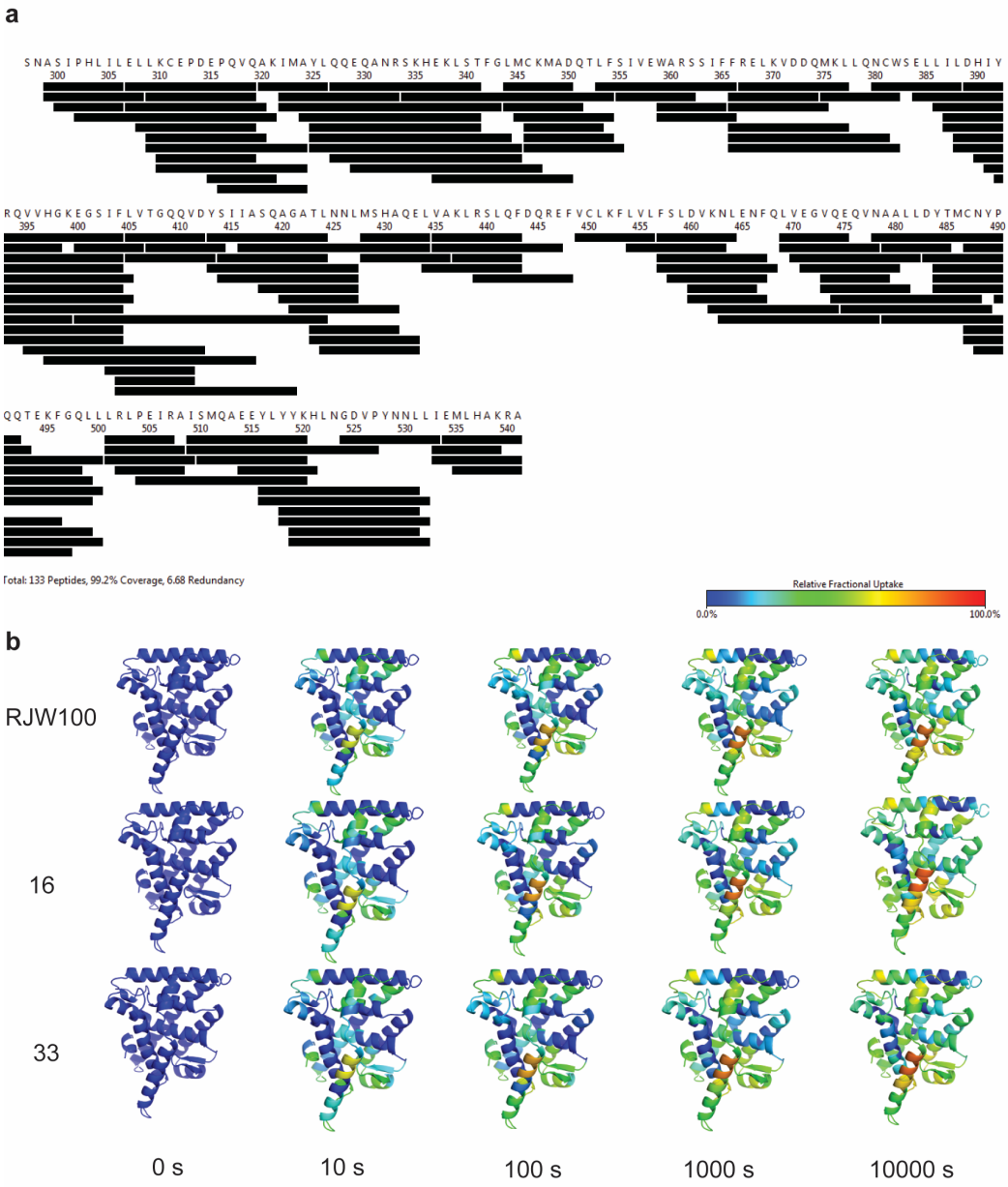


Figure 5.8. HDX Data. A. Peptide coverage map. B. Deuterium uptake over time for each complex.

5.6 Acknowledgements

We thank the beamline staff at Argonne National Laboratory, South East Regional Collaborative Team, Beamline 22ID for support during remote data collection. We are also grateful to the Emory HDX Core, particularly Dr. Wei Dang and Dr. Renhao Li, for assistance with HDX experiments. This work was supported in part by the National Institutes of Health under the following awards: T32GM008602 (S.G.M.), F31DK111171 (S.G.M.), R01DK095750 (E.A.O.), R01DK114213 (E.A.O., N.T.J., J.W.C), and the Emory Catalyst Award (E.A.O., N.T.J).

I would also like to acknowledge the significant contributions of several people to the work presented in this chapter. Autumn R. Flynn, Jeffery L. Cornelison, Michael Dugan, and Nathan T. Jui made invaluable contributions to the design of the compounds, as well as to the development of synthetic methods and compound synthesisis. C. Denise Okafor performed molecular dynamics simulations and analyzed data. Heather N. Donaldson conducted some of the DSF studies. Elizabeth J. Millings, Rohini Polavarapu, and John W. Calvert designed and executed the guinea pig studies and extracted RNA from the liver tissue.

Chapter 6. Development of Synthetic Phospholipid Mimetics as LRH-1 Agonists

6.1 Abstract

The orphan nuclear receptor Liver Receptor Homolog-1 (LRH-1) is an emerging drug target for metabolic disorders. The most effective known LRH-1 modulators are medium-chained phosphatidylcholines or synthetic hexahydropentalene compounds, which interact with different portions of the ligand binding pocket and activate LRH-1 through different mechanisms. Guided by crystallographic data, we combined aspects both ligand classes into a single scaffold to generate "phospholipid mimics" that are significantly more potent and activating than natural phosphatidylcholines. These compounds are valuable tools to study phospholipid-driven regulation of LRH-1, circumventing the poor pharmacological properties inherent to natural lipids.

This chapter was adapted from the manuscript:
Flynn AR, Mays SG, Ortlund EA, and Jui NT. Development of Phospholipid Mimics as Effective LRH-1 Agonists. *ACS Med Chem Lett* 2018; 9: 1051-1056.

6.2 Introduction

Liver Receptor Homolog-1 (LRH-1, NR5A2) is an orphan nuclear hormone receptor (NR) that serves as an important regulator of lipid and glucose metabolism(Stein and Schoonjans, 2015). LRH-1 activation increases bile acid biosynthesis and reverse cholesterol transport, diminishes the hepatic acute phase response, and helps resolve endoplasmic reticulum (ER) stress(Mamrosh et al., 2014; Stein and Schoonjans, 2015). Because of these key roles in metabolism regulation, LRH-1 has emerged as a promising drug target for metabolic diseases like type II diabetes and nonalcoholic fatty liver disease (NAFLD).

While endogenous LRH-1 ligands are unknown, the receptor binds a variety of phospholipids (PLs) *in vitro*, which presumably contributes to a low level of constitutive activity(Krylova et al., 2005; Ortlund et al., 2005; Sablin et al., 2015). However, a subset of phosphatidycholines (PCs) can activate LRH-1 above basal levels. Specifically, certain medium-chained, saturated PCs (*i.e.* diundecanoylphosphatidylcholine (DUPC, PC 11:0/11:0) and dilauroylphosphatidylcholine (DLPC, PC 12:0/12:0)) activate LRH-1 *in vitro* and *in vivo* when administered exogenously(Lee et al., 2011). In two models of obesity-induced insulin resistance, DLPC treatment decreased hepatic and circulating lipids, increased bile acid production, and improved insulin sensitivity and glucose homeostasis. These effects were entirely LRH-1 dependent, as they did not occur in liver-specific LRH-1 knockout mice(Lee et al., 2011).

Our recent structural studies with DLPC revealed an unusual ligand binding mode. Unlike typical NR ligands, which are buried deep in the ligand binding pocket, DLPC binds low in the pocket, bridging the mouth of the receptor and leaving unoccupied space behind its alkyl tails(Musille et al., 2012). DLPC binding is mediated by polar interactions of the phosphate headgroup with a cluster of residues at the mouth of the LRH-1 binding pocket (*i.e*. G421, Y516, and K520), as well as by hydrophobic interactions of the lipid tails. We demonstrated that DLPC induces specific conformational changes within the LRH-1 ligand binding domain (LBD) that

promotes allosteric signaling from the ligand binding pocket to the activation function surface (AFS), the site of coregulator interaction (Fig. 6.1)(Musille et al., 2015; Musille et al., 2012). DLPC is able to promote recruitment of specific coactivators, suggesting that this signaling network is used to communicate ligand status to the AFS(Musille et al., 2012).

Although DLPC-mediated anti-diabetic effects are striking, very high concentrations of this agonist are required to activate LRH-1 (100 μM *in vitro* and 100 mg/kg twice daily *in vivo*)(Lee et al., 2011). Further, natural PCs are not ideal pharmacological agents, since they are constantly remodeled(MacDonald, 1991), prone to hydrolysis(Ridgway, 2013), and frequently incorporated in membranes where they may impact biological processes influenced by membrane composition and fluidity. Potent and chemically stable small molecules that mimic phospholipid-driven LRH-1 activation could provide powerful tools for further evaluation of LRH-1 biology. Therefore, we have synthesized a set of "PL-mimics" with the goal of improving potency while maintaining PL-like contacts near the mouth of the pocket. To do this, we incorporated phosphorylcholine moieties or isoteres into the scaffold of the synthetic LRH-1 agonist, RJW100.

RJW100 is the most potent of a series of cis-bicyclo[3.3.0]-octene derivatives discovered by Whitby(Whitby et al., 2011) and is one of the few scaffolds shown to bind and activate LRH-1. Agonists of this class bind deep within the LRH-1 pocket, with the predominantly hydrophic scaffold fully engulfed in the lipophilic pocket(Mays et al., 2016; Whitby et al., 2011). In our LRH-1-RJW100 structure (PDB 5L11), we discovered a novel binding pose of the ligand, primarily driven by a water-mediated interaction between the RJW100 *exo*-hydroxyl group and a threonine residue (T352) that is deep in the pocket. An edge-to-face π-π stacking interaction between residue H390 and the RJW100 hexahydropentalene (6HP) phenyl, although not essential for activity, also contributes to the ligand's binding orientation. Interestingly, superimposition of RJW100 and DLPC reveals direct intersection of the hexyl tail of RJW100 and the Sn1-undecane tail of DLPC, presenting an opportunity to extend the RJW100 tail to access residues near the

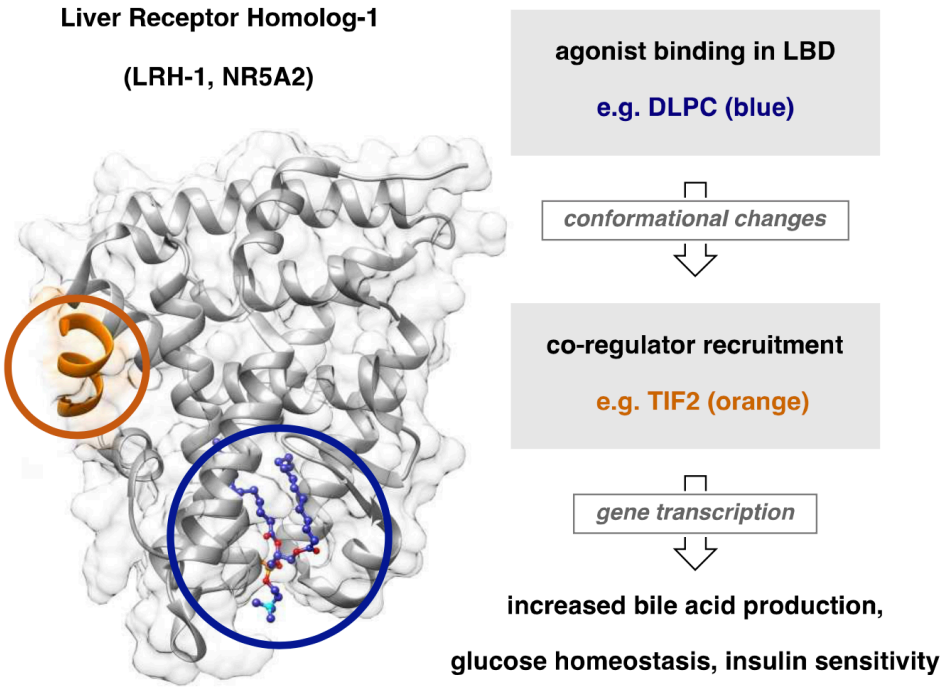


Figure 6.1. Structure of LRH-1 bound to PC agonist. Allosteric activation of LRH-1 drives coactivator recruitment and transcriptional activity, resulting in improved metabolic function.

mouth of the pocket (Figure 6.2). To combine binding elements used by these agonists, we designed a series of hybrid compounds with three distinct regions: the 6HP core, modular linkers, and varying polar groups to engage the PC-binding cluster of amino acids (Figure 6.2). This structure-guided design has the potential to improve potency while maintaining PL-like allostery, thereby generating tools that could be used to offer new insights into LRH-1 biology. Here, we show that this approach resulted in the development of novel LRH-1 agonists with significant improvements over either of the individual scaffolds.

6.3 Results

6.3.1 Synthesis and evaluation of the "full" phospholipid hybrid.

To generate the targeted PL mimics, we utilized the synthetic design shown in Figure 6.3. Employing Whitby's carbene-interrupted Pauson-Khand (PK) reaction, zirconocene-mediated reaction of silyl ether-containing 1,1-dibromoalkanes (**3**), enynes (**4**), and phenylacetylene would deliver the 6HP-containing primary alcohols **5**, after desilylation. We developed this specific approach because it would allow for systematic variation of the alkyl linker length and the primary alcohols could be easily manipulated to a range of polar head groups. This retrosynthetic disconnection is ideal, in part, because it utilizes readily accessible precursors of similar complexity (**3** and **4**) in the convergent 6HP-forming step.

Whitby showed that, although *exo* and *endo* diastereomers of RJW100 are roughly equipotent, the *exo* diastereomer is a more effective agonist for LRH-1(Whitby et al., 2011). Moreover, we have found that *endo*-RJW100 is a poor LRH-1 activator in cellular assays(Mays et al., 2016). We sought reaction conditions to selectively produce the *exo* isomer. Similar to other PK reactions, Whitby's system operates through an early transition state, where diastereoselectivity is dictated by the conformational preference of the corresponding enyne starting materials(Stec, 2010). Indeed, in a study of diastereoselectivity in the cobalt-mediated PK reaction of enyne structures similar to **4**, Carretero showed that *exo*:*endo* diastereomeric ratios

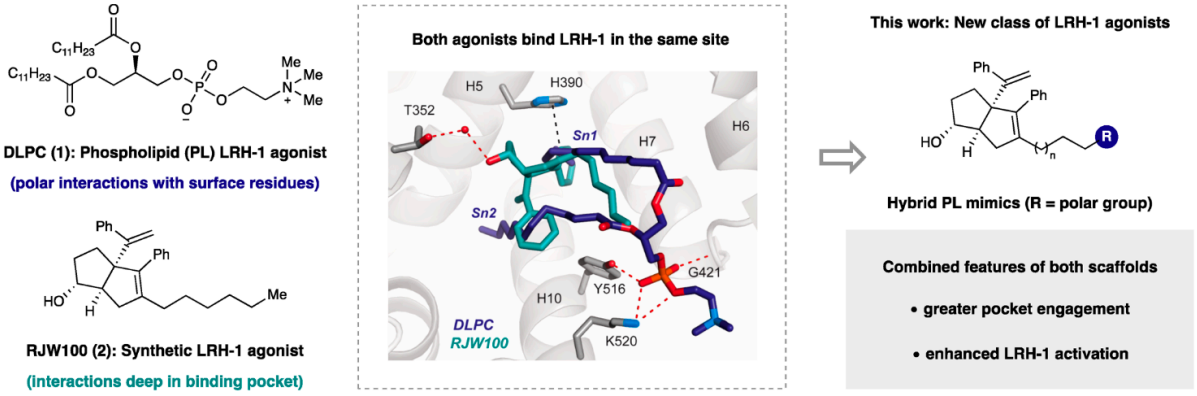


Figure 6.2. Design of hybrid phospholipid mimics. Two different agonist classes (*left)* were combined into a single scaffold (*right).* Middle panel shows hydrogen bonding and pi-pi stacking interactions made by RJW100 and DLPC. In addition, the agonists utilize multiple hydrophobic interactions (omitted for clarity).

can be correlated to the coupling constants between H_β and H_γ (shown in the bottom of Figure 6.3)(Adrio, 2001). We prepared a small collection of enynes **4** with different alcohol protecting groups and found that the *exo*-selectivity increased with relative population of pre-*exo* conformer (as indicated by $J_{H\beta, H\gamma}$), where a methoxymethyl (MOM) ether reliably gave the desired *exo*-6HP diastereomer in 7.2:1 d.r. (Figure 6.3).

In addition to providing diastereocontrol, MOM protection to give **4d** was strategically enabling due to its orthogonal reactivity to silyl protecting groups, which allows for chemoselective manipulations of the primary alcohol. Using *exo*-selective conditions for the construsction of 6HP functionalized alcohols, we prepared a series of intermediates with alkyl linkers of varied length (**5a**–**5h**, n = 1–8). Assembly of the linker library **3** was accomplished by sequential mono-silylation of terminal diols and

oxidation to the corresponding ω-siloxy aldehydes. Subsequent geminal dibromide formation was performed using triphenylphosphite in the presence of bromine (**3a**–**3h**, 30–75% overall yield from terminal diols). Beginning with 4-pentyn-1-ol, **4d** preparation was straightforward via successive Sonogashira coupling, Swern oxidation, vinyl Grignard addition, and MOM protection (72% yield overall). Both of these components (accessible on multigram scale) participated in Whitby PK and subsequent silyl deprotection to give the requisite hybrid precursors **5a**–**5h**.

To examine the effect of combining PC binding elements with those of the 6HP scaffold, we designed a hybrid compound where the RJW100 core was appended to a glycerophosphorylcholine headgroup (Scheme 6.1). Analysis of the crystallographic data described above indicated that this could be achieved via extension of the RJW100 alkyl chain (along the trajectory of the sn1 lipid tail) along with truncation of the sn2 tail (to avoid disruption of 6HP binding). Our synthetic efforts began with alcohol **5e** to provide the seven required methylene units between the 6HP core and acylglycerol (Scheme 6.1). Ley-Griffith oxidation of **5e** efficiently yielded terminal carboxylic acid **6**, which participated in a regioselective cobalt-

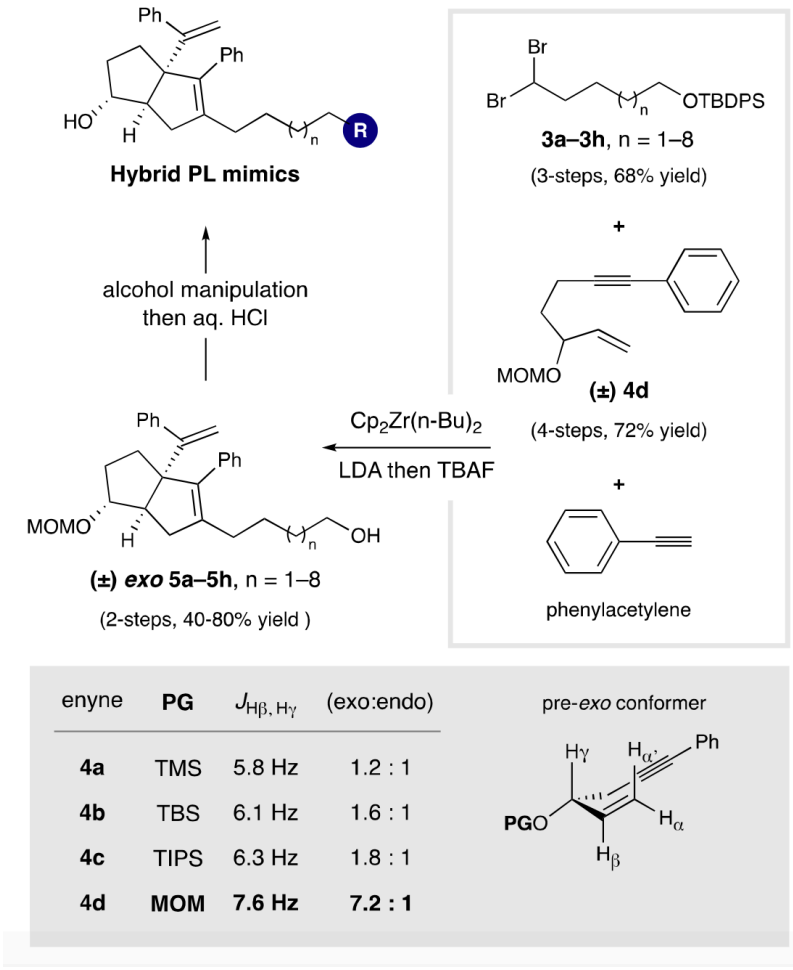
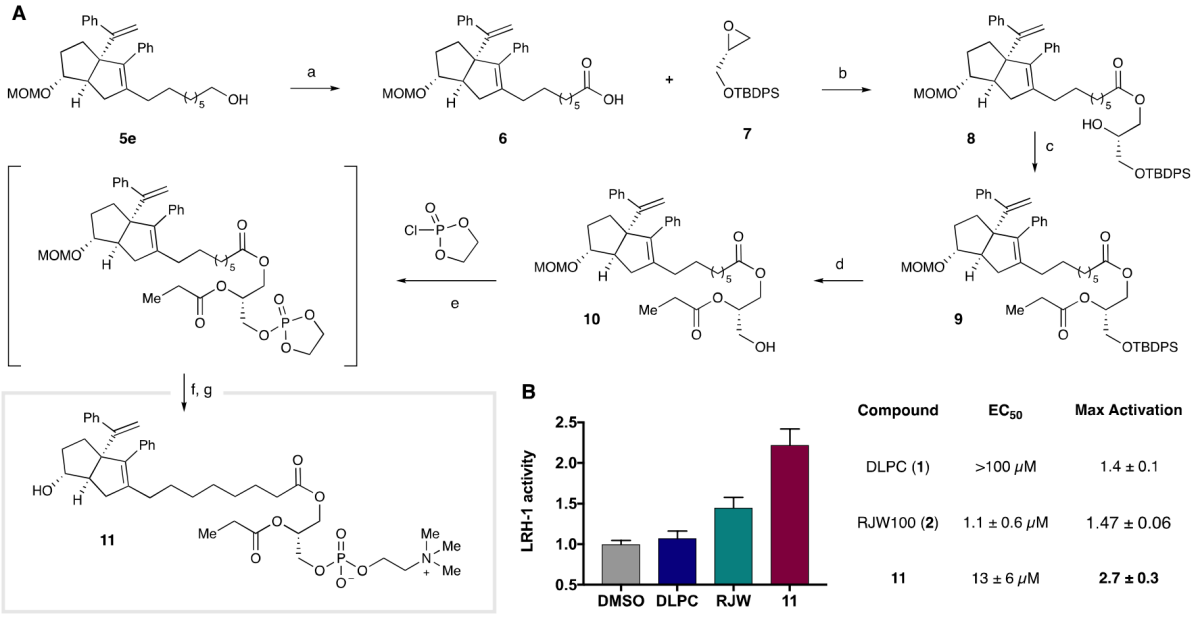


Figure 6.3. Synthetic approach to hybrid compounds. The choice of protecting group allow enrichment of the *exo* diastereomer.



Scheme 6.1. Synthesis and evaluation of compound 11. A. Reagents and conditions: (a) TPAP, NMO, H_2O, MeCN, 23 °C; 1 h, 85% yield. (b) i. Co(salen), Et_2O, 23 °C; 1 h; ii. DIPEA, neat, 60 °C, 16 h, 24% yield. (c) propionylchloride, DMAP, THF, 23 °C; 30 min, 77% yield. (d) TBAF, THF, 23 °C; 16 h, 76% yield. (e) 2-Chloro- 1,3,2-dioxaphospholane 2-oxide, TEA, toluene, 23 °C; 4 h. (f) trimethylamine, MeCN, 90 °C, 16 h. (g) conc. aq. HCl, MeCN, 23 °C; 10 min, purified by preparative HPLC (4% isolated yield over three steps e–g). B. Ligand-driven LRH-1 activity using a Luciferase reporter assay. Each bar represents mean +/- SEM for two experiments conducted in triplicate, showing fold activation *versus* DMSO for a single dose of 30 μM. Table: maximum activation values for each ligand (fold *vs* vehicle) at 30 μM. EC_{50} values were calculated from curve-fitting from dose titration experiments.

mediated opening of epoxide **7**[28] to give ester **8** as a single isomer. Acylation of **8** with propionyl chloride afforded the advanced intermediate **9**, which was reacted with tetrabutylammonium fluoride to unveil the free primary alcohol **10**. Terminal phosphorylcholine construction commenced with reaction of **10** with 2-chloro-1,3,2-dioxaphospholane 2-oxide in toluene. After filtration and concentration of the reaction mixture, the crude cyclic phospholane was opened with trimethylamine. Finally, acidic cleavage of the MOM-ether and isolation by preparative HPLC gave the hexahydropentalene-based phospholipid **11**.

We evaluated the efficacy of **11** using a luciferase reporter assay, which measures LRH-1 activity in intact (HeLa) cells. This assay utilizes a reporter plasmid in which the gene for luciferase is cloned downstream of a LRH-1 response element. In this way, luciferase is generated in response to LRH-1 transactivation. The overlaid dose-response curves for DLPC, RJW100, and **11** are shown in Scheme 6.1, where dose-dependent increases in luminescence was observed in all cases. Compound **11** was significantly more potent than DLPC (EC_{50} of 13 ± 6 μM vs. >100 μM for DLPC). The low potency of DLPC in this assay is consistent with previous reports(Lee et al., 2011). RJW100 was more potent than **11**, with an EC_{50} of 1.1 ± 0.6 μM; however, **11** induced higher levels of LRH-1 activation than either DLPC or RJW100 in these experiments (Scheme 6.1).

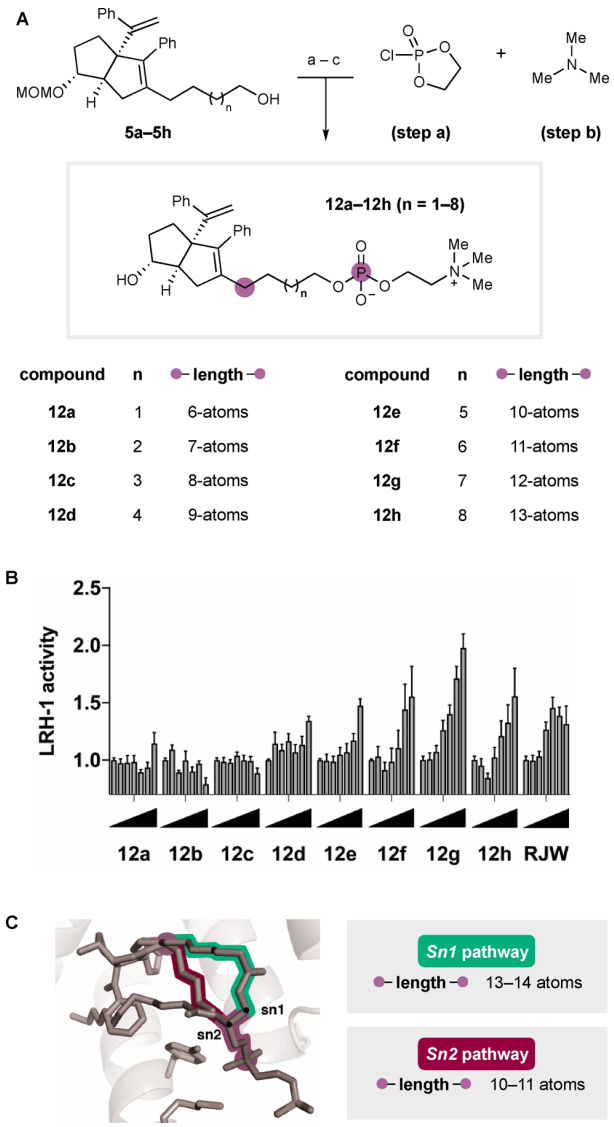
6.3.2 Scaffold simplification improves both synthetic accessibility and agonist activity.

We sought to simplify the scaffold to remove the hydrolytic liabilities inherent to the glyceryl esters while retaining the 6HP core and phosphorylcholine elements. We were also interested in determining the optimal distance between these binding elements. In the overlaid ligand poses of **1** and **2**, the linear measured distance between the first methylene unit in the hexyl tail off the 6HP core and the DLPC phosphate is 10.0 Å. However, the ligand binding pocket is flexible, and we have observed that is can change shapes to accommodate either DLPC or RJW100. We anticipated that hybrid ligand would be anchored at two sites in the pocket (by 6HP

and polar group), although hydrophobic interactions within the pocket could enforce a nonlinear trajectory of the hybrid tail. This underscores the importance of systematic evaluation of linker length. We therefore synthesized and evaluated compounds **12a**–**12h** comprised of the 6HP core, phosphorylcholine polar group, and alkyl linker with systematically varied lengths.

The synthesis of **12a**–**12h** is summarized in Scheme 6.2A. Beginning with primary alcohols **5a**–**5h**, transformation to their corresponding cyclic phospholanes followed by ring opening with trimethylamine afforded the crude MOM-protected phosphorylcholines. Direct cleavage of the MOM group with aqueous hydrochloric acid gave rise to the compounds **12** with linkers between 6- and 13-atoms long. The improvement to LRH-1 activity observed with the hybrid **11** was retained with the structural simplification. Scheme 6.2B shows LRH-1 activity in the presence of these ligands, where each vertical bar represents activity at a given dose of each compound (concentration increases from 0–30 μM (left to right), represented by triangles). Agonism was clearly related to linker length; compounds with shorter linkers (**12a**–**12c**, n = 1–4) were inactive in luciferase reporter assays, while compounds with longer linkers (**12f-12h**) induced robust, dose-dependent activation of the receptor (Scheme 6.2). The optimal alkyl linker length is 12-atoms, as demonstrated by the activity of compound **12g**. The EC_{50}s for RJW100 and **12g** were similar; however, **12g** exhibited significantly higher maximum LRH-1 activity. Visual examination of the plot in Scheme 6.2B reveals that ligand-driven LRH-1 activity is diminished on either side of the optimal linker length. As illustrated in Scheme 6.2C, this optimal linker length is consistent with the alkyl tail following the trajectory of the Sn2 acyl chain of DLPC.

In addition to the phosphorylcholines, we synthesized and evaluated a set of carboxylic acid isosteres (**13a-13h)**. These compounds are chemically stable, synthetically accessible, and the tunable nature of carboxylic acid pKa values could potentially offer a handle for further optimization(Ballatore, 2013; Rye, 2005). This was preferable to phosphates, which can be



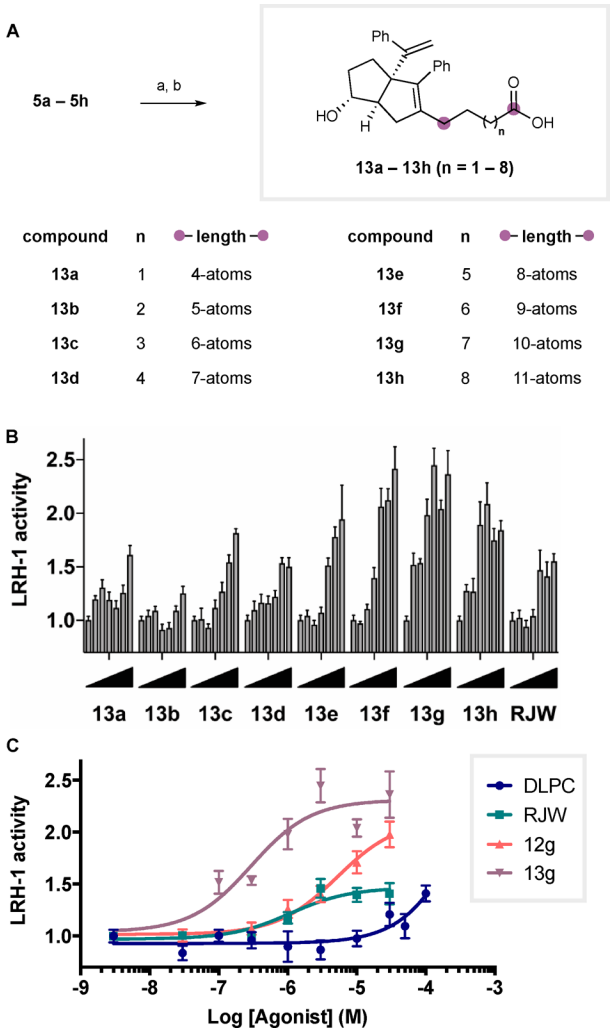
Scheme 6.2. Synthesis and evaluation of Simplified 6HP-Phosphorylcholine Hybrids. A. Reagents and conditions: (a) 2-Chloro-1,3,2- dioxaphospholane 2-oxide, TEA, toluene, 23 °C; 4 h. (b) trimethylamine, MeCN, 90 °C, 16 h. (c) conc. aq. HCl, MeCN, 23 °C; 10 min. 12–49% over three steps and after purification by preparative HPLC. B. Luciferase reporter assay. Each bar represents mean +/- SEM fold change in LRH-1 activity by each compound (given at concentrations of 0, 0.03, 0.3, 1.0, 3.0, 10, and 30 μM, left to right). C. Overlay of DLPC and RJW100 in LRH- 1 LBD. Number of atoms in linker is counted between the 6HP core and PC binding element.

poorly cell permeant and subject to degradation by cellular phosphatases(Rye, 2005). Construction of the terminal carboxylate series 13a–13h was straightforward using our collection of primary alcohols 5a–5h. As summarized in Scheme 6.3A, Ley-Griffith oxidation of each parent alcohol 5a–5h followed by MOM deprotection gave the desired carboxylic acid hybrids. Again, we observed highly pronounced length-dependent LRH-1 activation, where compounds with longer linkers were more effective (activity of each compound by dose is visually depicted in Scheme 6.3B). The most efficacious agonist 13g (6HP-CA), displays both improved efficacy and potency relative to RJW100 (EC_{50} = 0.4 ± 0.2 μM). In contrast to the phosphorylcholine compounds, the carboxylic acid hybrids have sigmoidal activation curves in this concentration range. A comparison of 13g to 12g, RJW100, and DLPC is shown in Scheme 6.3C, and data for key compounds are summarized in Table 1. In addition, introduction of the carboxylate improves aqueous solubility, a persistent challenge to overcome with ligands of the 6HP scaffold. Compound 13g contains 10-atoms between the 6HP core and carboxylate functional group. As with the most effective phosphorylcholine derivative, this is most consistent with the intersection of the DLPC sn2 pathway.

6.4 Conclusions

These studies have described the development of a new and highly effective class of LRH-1 agonists. Guided by structural insights, we have incorporated the most important binding elements from two discrete ligand classes into a single hybrid scaffold. The full 6HP phospholipid hybrid (11) displayed significantly higher LRH-1 activation than either of the parent structures (RJW100 and DLPC). Simplification of this scaffold and systematic evaluation of various alkyl tail lengths revealed a clear relationship between linker length and LRH-1 activation. We identified the carboxylate 6HP-CA (13g) that acts as an attractive phosphate isostere. Efforts to investigate the binding pose of these ligands for further development, probe

their mechanism of activation, evaluate and optimize physiochemical properties, and assess their activity *in vivo* are currently underway in our laboratories.



Scheme 6.3. Synthesis and Evaluation of Simplified 6HP- Carboxylic Acid Hybrids. A. Reagents and conditions: (a) TPAP, NMO, H_2O, MeCN, 23 °C; 1 h. (b) conc. aq. HCl, MeCN, 23 °C, 10 min. 40–85% yield over two steps. B. Luciferase reporter assay showing LRH-1 activity (dosing as described in Scheme 6.2). C. Dose-response curves of indicated ligands. Each point is the mean +/- SEM for two independent experiments conducted in triplicate.

Table 6.1. Summarized Activity Data for the Best Hybrid PL Mimics

Compound	EC_{50} (μM)	Max LRH-1 activation[a]
RJW100 (**1**)	1.1 ± 0.6	1.47 ± 0.06
11	13 ± 6	2.7 ± 0.3
12g	5 ± 2	2.1 ± 0.1
6HP-CA (**13g**)	0.4 ± 0.4	2.3 ± 0.2

[a]Maximum LRH-1 activation was calculated at the top of the resulting dose-response curves, as a fold increase in LRH-1 activity over baseline (DMSO).

6.5 Materials and Methods

6.5.1 Chemical synthesis

The synthesis and characterization of key target compound **6HP-CA**(**13g**) is outlined below; detailed syntheses of all compounds will be presented in the dissertation of Autumn Flynn, who played a lead role in compound design and in devising synthetic routes to all compounds presented herein. Autumn is a member of Nathan Jui's laboratory in the Department of Chemistry, Emory University.

(5-(methoxymethoxy)hept-6-en-1-yn-1-yl)benzene (4d): The preceding alcohol (7-phenylhept-1-en-6-yn-3-ol) was synthesized by known procedures and matched spectral data reported in literature. Under nitrogen, 7-phenylhept-1-en-6-yn-3-ol (1.0 equiv.) was dissolved in DCM, followed by diisopropylethyl amine (1.25 equiv.) Chloromethyl methyl ether (1.5 equiv.) was added and the reaction mixture was stirred at 30 °C for 4 h. The reaction mixture was cooled to room temperature before being poured onto water, washed with 1M HCl, and extracted with DCM. The organic layers were dried with $MgSO_4$, filtered, and concentrated before being subjected to silica gel chromatography in 5% EtOAc/Hex, (3.5 g, 86%). ^{1}H NMR (600 MHz, $CDCl_3$) δ 7.42 – 7.35 (m, 2H), 7.32 – 7.22 (m, 3H), 5.70 (ddd, *J* = 17.6, 10.4, 7.6 Hz, 1H), 5.27 (dt, *J* = 17.1, 1.4 Hz, 1H), 5.22 (dt, *J* = 10.4, 1.3 Hz, 1H), 4.73 (d, *J* = 6.8 Hz, 1H), 4.57 (d, *J* = 6.7 Hz, 1H), 4.22 (td, *J* = 7.8, 5.3 Hz, 1H), 3.40 (d, *J* = 1.5 Hz, 3H), 2.60 – 2.45 (m, 2H), 1.95 – 1.84 (m, 1H), 1.84 – 1.76 (m, 1H).^{13}C NMR (151 MHz, $CDCl_3$) δ 137.7, 131.5, 128.2, 127.6, 123.9, 117.7, 93.8, 89.4, 81.0, 75.8, 55.5, 34.4, 15.6.

10-(6-*exo*-(methoxymethoxy)-3-phenyl-3a-(1-phenylvinyl)-1,3a,4,5,6,6a-hexahydropentalen-2-yl)decan-1-ol (5g). Bis(cyclopentadienyl)zirconium(IV) dichloride (2.47 mmol, 1.2 equiv) was dried by azeotroping with benzene four times before being placed under nitrogen, dissolved in dry, degassed tetrahydrofuran (THF, approx. 15 mL) and cooled to -78 °C in a dry ice/acetone bath. The resulting solution of zirconecene dichloride was treated with n-BuLi (5.0 mmol, 2.4

equiv) to form a clear, light yellow solution and allowed to stir. After 30 minutes, azeotroped (5-(methoxymethoxy)hept-6-en-1-yn-1-yl)benzene **4d** (2.02 mmol, 1.0 equiv) in dry, degassed THF (approx. 8 mL) was added portionwise to afford a pink-orange solution, and the reaction mixture was held at -78 °C for 30 minutes before allowed to warm to room temperature and stirred 2.5 hours. The reaction mixture was then cooled to -78 °C and azeotroped **3g** (2.18 mmol, 1.1 equiv.) was added in dry, degassed THF (approx. 8 mL). Freshly prepared lithium diisopropylamine (LDA, 1.0 M, 2.2 mmol, 1.1 equiv) was added at -78 °C and stirred for 15 minutes. Freshly prepared lithium phenylacetylide (7.2 mmol, 3.6 equiv, in approx. 10 mL dry, degassed THF) was added to the reaction mixture dropwise). The resulting dark reddish-brown solution was stirred at -78 °C for 1.5 hours. The reaction was then quenched with methanol (10 mL) and saturated aqueous sodium bicarbonate (20 mL) and allowed to warm to room temperature to form a light yellow slurry. The slurry was poured over water and extracted with ethyl acetate four times. The combined organic layers were washed with brine, dried with $MgSO_4$, and concentrated *in vacuo*. The resulting yellow oil was carried on without further purification. The crude oil was dissolved in THF in a round bottom flask charged with a stir bar and open to air. Tetrabutylammonium fluoride (TBAF, approximately 2 equiv of enyne starting material) was added. The solution rapidly darkened and was allowed to stir at 23 °C for 16 h. After reaction completion, the reaction mixture was concentrated and to purified on silica gel in 20% EtOAc/Hex) to afford **5g** as a yellow oil (584.5 mg, 58% over 2 steps). ^{1}H NMR (500 MHz, $CDCl_3$) δ 7.41 – 7.17 (m, 10H), 5.07 (d, J = 1.4 Hz, 1H), 5.02 (d, J = 1.5 Hz, 1H), 4.64 – 4.56 (m, 2H), 3.81 (p, J = 2.1 Hz, 1H), 3.62 (d, J = 6.6 Hz, 2H), 3.32 (s, 3H), 2.44 (dq, J = 9.2, 1.7 Hz, 1H), 2.35 (dd, J = 16.9, 9.2 Hz, 1H), 2.11 – 2.00 (m, 4H), 1.81 – 1.63 (m, 4H), 1.57 (m, 2H), 1.40 – 1.19 (m, 13H). ^{13}C NMR (126 MHz, $CDCl_3$) δ 154.5, 144.1, 141.3, 139.2, 137.5, 129.6, 127.8, 127.6, 126.62, 126.56,, 114.9, 94.7, 86.7, 69.1, 63.0, 55.2, 52.7, 40.5, 32.8, 32.4, 31.4, 29.7, 29.6, 29.5, 29.4, 29.3, 27.8, 25.7. LRMS (ESI, APCI) *m/z*: calc'd for $C_{33}H_{43}O_2$ $[M\text{-}CH_3O]^+$ 471.3, found 470.9.

10-(6-*exo*-(methoxymethoxy)-3-phenyl-3a-(1-phenylvinyl)-1,3a,4,5,6,6a-hexahydropentalen-2-yl)decanoic acid (S1g). In a scintillation vial equipped with stir bar, **5g** (1.16 mmol, 1.0 equiv) was dissolved in acetonitrile (approx. 10 mL). Tetrapropylammonium perruthenate (TPAP, 0.1 mmol, 0.1 equiv), N-Methylmorpholine-N-Oxide (NMO, 7.1 mmol, 6 equiv), and water (7.2 mmol, 6 equiv) were added. The reaction mixture was allowed to stir until complete by TLC and LCMS, 1-16 h. Upon reaction completion, the reaction mixture was concentrated and purified on silica gel in 20-50% EtOAc/Hex (0.1% acetic acid) to afford **S1g** as a clear, colorless oil (498.2 mg, 88%). ^{1}H NMR (500 MHz, $CDCl_3$) δ 7.38 – 7.18 (m, 10H), 5.05 (d, *J* = 1.4 Hz, 1H), 5.01 (d, *J* = 1.5 Hz, 1H), 4.61 – 4.59 (m, 2H), 3.80 (s, 1H), 3.32 (s, 3H), 2.42 (d, *J* = 9.2, 1.7 Hz, 1H), 2.38 – 2.29 (m, 3H), 2.11 – 1.97 (m, 4H), 1.70 – 1.59 (m, 5H), 1.37 – 1.18 (m, 13H).^{13}C NMR (126 MHz, $CDCl_3$) δ 179.4, 154.5, 144.1, 141.3, 139.2, 137.5, 129.6, 127.8, 127.6, 126.61, 126.56, 114.9, 94.7, 86.7, 69.1, 55.1, 52.7, 40.5, 34.0, 32.4, 31.4, 29.7, 29.6, 29.30, 29.26, 29.2, 29.0, 27.8, 24.7. LRMS (ESI, APCI) *m/z*: calc'd for $C_{34}H_{43}O_4$ $[M-H]^-$ 515.3, found 515.1. Calc'd for $C_{33}H_{41}O_3$ $[M-CH_3O]^+$ 485.3, found 484.9.

10-(6-*exo*-hydroxy-3-phenyl-3a-(1-phenylvinyl)-1,3a,4,5,6,6a-hexahydropentalen-2-yl)decanoic acid (6HP-CA; 13g). In a scintillation vial equipped with a stir bar, **S1g** (0.97 mmol, 1.0 equiv) was dissolved in acetonitrile (5 mL). Three drops of concentrated aq. HCl was added to the reaction mixture, which was allowed to stir until complete by TLC and LCMS (5 min). The resulting solution was concentrated and purified on a plug of silica in 50% EtOAc/Hex (0.1% Acetic Acid) to afford the title compound **6HP-CA (13g)** as clear, colorless oil. (498.2 mg, >95%). ^{1}H NMR (500 MHz, $CDCl_3$) δ 7.39 – 7.15 (m, 10H), 5.07 (d, *J* = 1.6 Hz, 1H), 4.99 (d, *J* = 1.6 Hz, 1H), 3.95 (s, 1H), 2.41 – 2.24 (m, 5H), 2.15 – 1.94 (m, 5H), 1.75 – 1.50 (m, 6H), 1.42 – 1.09 (m, 9H). ^{13}C NMR (126 MHz, $CDCl_3$) δ 179.4, 154.6, 144.2, 141.2, 139.1, 137.4, 129.7, 127.74, 127.71, 127.6, 126.7, 126.6, 115.0, 82.1, 69.3, 55.7, 40.3, 33.98, 33.94, 32.1, 29.7, 29.6, 29.30, 29.28, 29.2, 29.0, 27.8, 24.7. LRMS (ESI, APCI) *m/z*: calc'd for $C_{32}H_{39}O_3$ $[M-H]^-$ 471.3,

found 471.3. Calc'd for $C_{32}H_{39}O_2$ $[M\text{-}OH]^+$ 455.3, found 454.8. HRMS (ESI) *m/z*: calc'd for $C_{32}H_{39}O_3$ $[M\text{-}H]^-$ 471.2882, found 471.2905. Purity was determined to be 98.5% t_R = 3.6 min (as single *exo* diastereomer) by HPLC using a linear gradient of water and 0.1% formic acid (A) and acetonitrile and 0.1% formic acid (B); t=0 min, 30% B, t=4 min, 99% B (held for 1 min) then 50% B (held for 1 min).

6.5.2 Reporter gene assays.

We performed cellular assays with HeLa cells which were reverse-transfected with three vectors: (1) full-length, human LRH-1 in a PCl vector, (2) a firefly reporter (pGL3 Basic) with a portion of the SHP promoter cloned upstream of the firefly luciferase gene, and (3) a constitutively active vector expressing Renilla luciferase under control of the cytomegalovirus (CMV) promoter. The following day, cells were treated with each compound or vehicle control for 24 hours (in the case of all synthetic compounds, vehicle was DMSO). In most cases, six points in the concentration range of 0.03—30 μM were used, with a final DMSO concentration of 0.3% in all wells. Luciferase expression was measured using the DualGlo Kit (Promega). Firefly luciferase signal was normalized to Renilla luciferase signal in each well. EC_{50} values were calculated using three-parameter curve-fitting (GraphPad Prism, v.7). Assays were conducted in triplicate with at least two independent biological replicates. DLPC was purchased from Avanti Polar Lipids (Alabaster, AL) and dissolved in ethanol for delivery to cells (with a final concentration of 0.3% ethanol in each well). A concentration range of 0-100 μM was used. Fold change in LRH-1 activation for DLPC-treated cells was calculated relative to response in cells treated with 0.3% ethanol vehicle alone.

6.6 Acknowledgements

This work was supported in part by the National Institutes of Health under the following awards: T32GM008602 (S.G.M.), F31DK111171 (S.G.M.), R01DK095750 (E.A.O.), R01DK114213 (E.A.O., N.T.J.), and the Emory Catalyst Award (E.A.O., N.T.J).

Chapter 7. Discussion and Future Directions

7.1 Discussion

The discovery that DLPC mediates profound anti-diabetic effects via direct binding and activation of LRH-1 has revolutionized the way investigators think about this orphan receptor and has spurred efforts to develop synthetic modulators. High throughput screens have only identified a few synthetic scaffolds that bind LRH-1, and rational design has been challenging due to a poor understanding of how synthetic molecules switch this receptor into the active state. The work presented here has provided the first detailed investigation of mechanisms used by LRH-1 small molecule modulators. We reasoned that understanding mechanisms of action would inform rational design efforts and may offer more general insights into how this receptor is regulated by ligands.

Our discovery of a novel polar interaction made by the RJW100 hydroxyl group (Chapter 3)(Mays et al., 2016) and insights from our previous structural studies with DLPC(Musille et al., 2015; Musille et al., 2012) provided a structural rationale for agonist design. Previous efforts to target the "polar patch" of residues in the deep part of the LRH-1 binding pocket (DPP) were unsuccessful due to unpredictable binding orientations of the 6HP scaffold(Mays et al., 2016; Whitby et al., 2011). Our extensive use of x-ray crystallography to link binding mode with receptor activation has enabled predictable targeting of the DPP and showed that this approach dramatically improves agonist potency (Chapter 5). The use of in-cell activation assays throughout these studies ensured that we were identifying agonist features important for activation as opposed to merely binding. Cellular assays were also advantageous in that they provided a way to study full-length LRH-1 in the context of an intact cell, which provides information about pharmacological potential. At the same time, using luciferase activation at the SHP promoter as the sole means to evaluate agonist efficacy was a limitation of these studies. Expanding biological analyses will be an important future step in characterizing the new compounds. Nonetheless, the strong structure-activity relationship elucidated from the cellular

studies has been highly informative, particularly when paired with DSF and x-ray crystallography. This iterative, mechanistic approach has led to the discovery of improved LRH-1 agonists that fall into two broad classes based on site of modification: the RJW100 hydroxyl site (R1) or the alkyl tail (R4). The two classes have distinct activity profiles, summarized in Figure 7.1.

7.1.1. Tuning potency through modifications at the RJW100 hydroxyl site

Modification at R1 greatly affected agonist EC_{50}, with several compounds much less potent and several more potent than RJW100 (Figure 7.1). There was a strong correlation between increased potency and global stabilization of the LRH-1-agonist complex by DSF (Chapter 5, Figure 5.4). Structural analyses demonstrated that this relationship likely resulted from displacement of conserved water molecules and formation of hydrogen bonds in the DPP by the modified R1 group (together with associated global protein conformational changes). Interestingly, the R1 interactions observed in the structures did not always contribute to LRH-1 activation. Figure 7.2 compares the role of two key residues involved in LRH-1 activation: T352 and H390. These residues are highlighted with colored circles if they were involved in the activation mechanism for the agonist in each panel. Whereas *exo*-RJW100 was dependent on the water-mediated interaction with T352, *endo*-RJW100 was a weak activator despite making the T352 contact (Chapter 3). Similarly, Cpd16 contacts both T352 and H390, but mutation of T352 or H390 does not significantly impact Cpd16-driven activation of LRH-1 (Chapter 5). In contrast, the very similar Cpd33 is a stronger activator than Cpd16 and utilizes both T352 and H390 to drive LRH-1 activation (Chapter 5). The involvement of H390 in the activation mechanism for the 6HP ligands required either stable, face-to-face π-π stacking with the agonist phenyl ring (as with GSK8470, Chapter 3), or a hydrogen bond donor in the R1 linker (as with Cpds 25 and 33, Chapter 5). Indeed, the incorporation of a NH linker *versus* an oxygen linker in the R1 group was

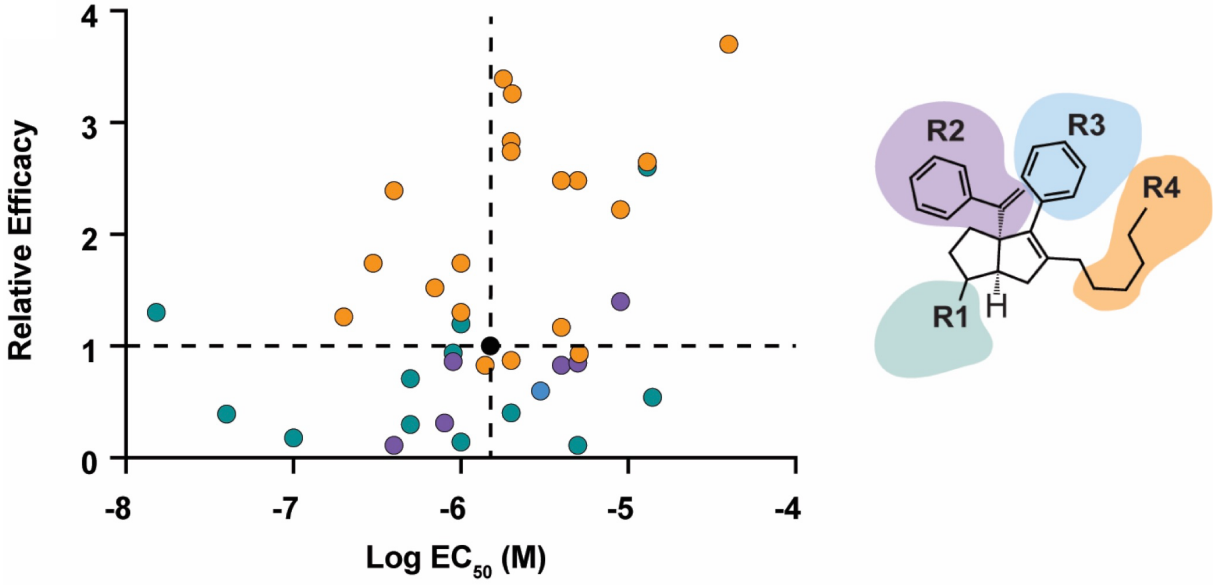


Figure 7.1. Overall distribution of activation data for the new agonists. Dots are color-coded according to the site of modification on the scaffold (see schematic on the right). Compounds modified at R1, R2, or R3 are described in Chapter 5. Compounds modified at R4 are described in Chapter 6. Black dot in the center is RJW100. More potent and effective compounds are distributed in the top left quadrant of the graph. Calculation of relative efficacy is described in Chapter 5.

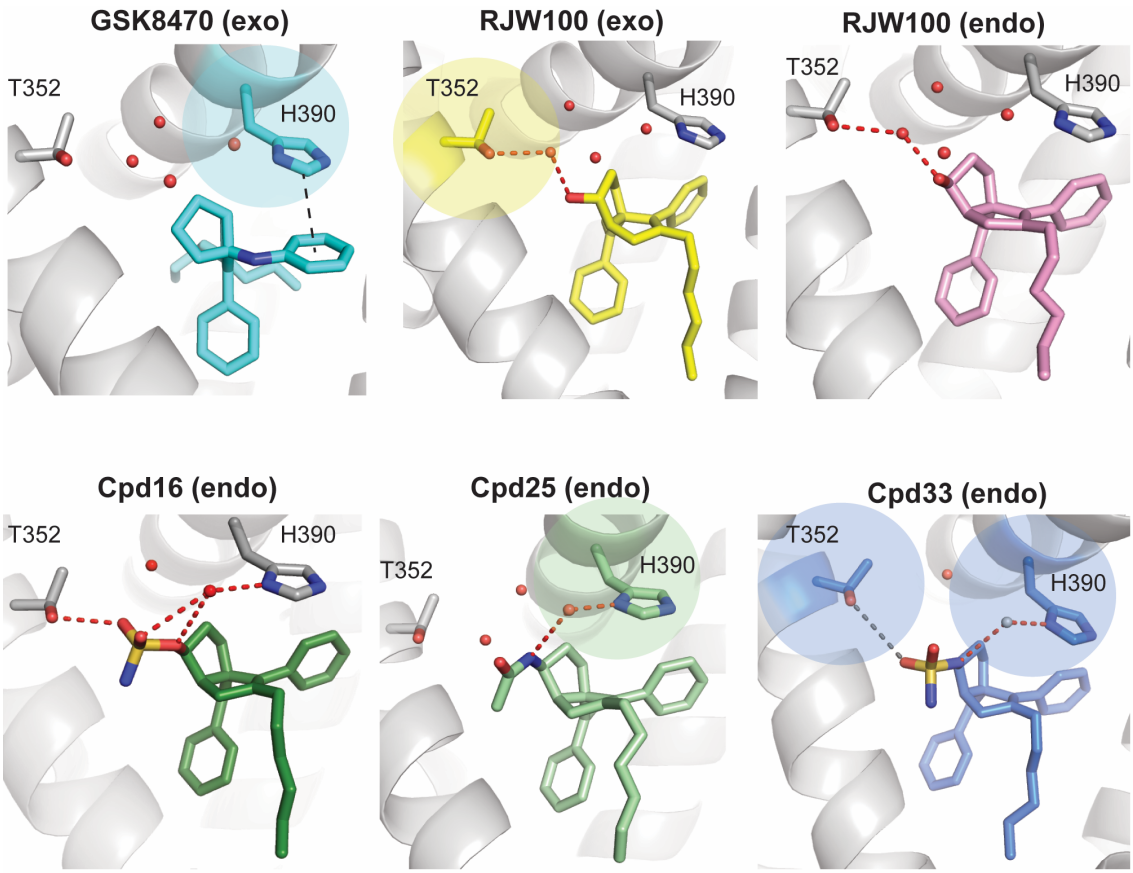


Figure 7.2. Role of key DPP interactions in driving LRH-1 activity by synthetic agonists. View of binding poses of LRH-1 agonists, with interactions made with T352 and H390 shown as sticks and colored by atom type. Residues are highlighted with colored circles if they were involved in the activation mechanism for the agonist in each panel. Importance of particular interactions was determined in luciferase reporter assays with mutant LRH-1 constructs (experiments shown in Chapters 3 and 5).

sufficient to increase potency by two orders of magnitude and to switch the compound from a partial to full agonist (Chapter 5, Figure 5.7).

7.1.2 R4-modified compounds: the "PL mimics"

The second class of agonists was designed based on our observations that the RJW100 "tail" (R4) assumes a similar trajectory to PL tails, providing the opportunity to make hybrid compounds that incorporate features of RJW100 and PLs into a single molecule (Chapter 6). We extended and modified the tail to introduce interactions near the mouth of the pocket in a similar manner to the DLPC phosphate group. Compared to the R1 compounds, the tail-modified agonists had a greater impact on efficacy, while potency was not as strongly affected (Figure 7.1). The length of the alkyl tail was important for activity: tails of 4-5 carbons were completely inactive and tails of 9-10 carbons were highly active (Chapter 6). This trend held true whether the polar group at the end of the tail was a phosphorylcholine, carboxylic acid, or alcohol (Chapter 6 and not shown).

Mechanistic studies with the PL mimics are ongoing; however, preliminary analyses suggest several differences from the R1 class. Whereas the R1 agonists exhibit a highly significant correlation between potency in cellular assays (EC_{50}) and stability of the protein-ligand complex (Chapter 5, Figure 5.4B), no such correlation exists for the PL mimics. Several of the PL mimics stabilize the receptor, but there is no relationship between Tm and either EC_{50} or efficacy (Figure 7.3A-B). However, the plot of Tm change relative to the LRH-1-DLPC complex for this class of agonist reveals an interesting feature: while several of the compounds are stabilizing, the Tm change is close to zero for the 10-carbon analogs (regardless of terminal polar group) (Figure 7.3C). The fact that the 10-carbon PL mimics have similar effects as DLPC on the stability of the LRH-1-ligand complex could indicate that the 10-carbon tail is promoting a DLPC-like conformation of the receptor. This is an attractive hypothesis but not the only possibility. For example, the 10-carbon tail could promote an alternative receptor conformation with similar

thermostability to the LRH-1-DLPC complex. Studies are underway to investigate binding mode and mechanism of action of the analogs with 10-carbon tails, since this chain length is optimal for activating LRH-1 in luciferase reporter assays (Chapter 6)(Flynn et al., 2018).

Preliminary structural studies with PL mimetics reveal that modification of the tail does not impact the binding mode of the bicyclic pentane core of the molecule, styrene rings, or R1 hydroxyl group (Figure 7.3D). However, the R4 tails are highly disordered, suggesting that they are adopting more than one conformation (Figure 7.3D). Composite omit electron density maps and Polder maps (Liebschner et al., 2017) indicate that the tails are predominantly occupying a hydrophobic groove between LRH-1 helices 2 and 3, which provided the rationale for modeling their positions as shown in Figure 7.3D. However, weak density and high B factors for the tails suggest that they may be occupying more than one conformation. Supporting this idea, the sidechain of a methionine residue near the ligand binding site adopts two conformations, one of which is shifted to accommodate the trajectory of the tails toward the helix 2-3 groove (not shown). The other conformation of the methionine is the same as in the LRH-RJW100 structure, which would allow the tails to follow a trajectory toward the mouth of the pocket. The implications of this flexible tail are being explored; however, we have observed that "PL-like" positioning of the tail is incompatible with the crystal form (not shown). Preliminary data suggest that the 9- and 10-carbon phosphorylcholines and carboxylic acids rely, at least in part, on interactions with Y516 and K520 for activation. This idea is supported by data from molecular dynamics simulations (not shown) and luciferase reporter assays with mutant LRH-1 (shown for the carboxylic acids in Figure 7.3E). Therefore, we have generated two classes of agonists with distinct activation profiles and mechanisms of action, which will be extremely useful as probes to better understand LRH-1 biology.

7.1.3. Possibility for hybrid agonists with dually modified R1 and R4 groups?

A logical question may be to ask whether combining the high potency associated with R1 modifications with the greater activity of the R4 modifications would lead to a "hybrid" molecule that possesses both characteristics. However, this strategy has not been successful to date. For example, a hybrid combining the R1 *endo* sulfamide group of Cpd33 and the R4 group from the 10C-acid is not more potent than the 10C-acid (Figure 7.4). Reasons for this effect are unclear and will be explored in future studies.

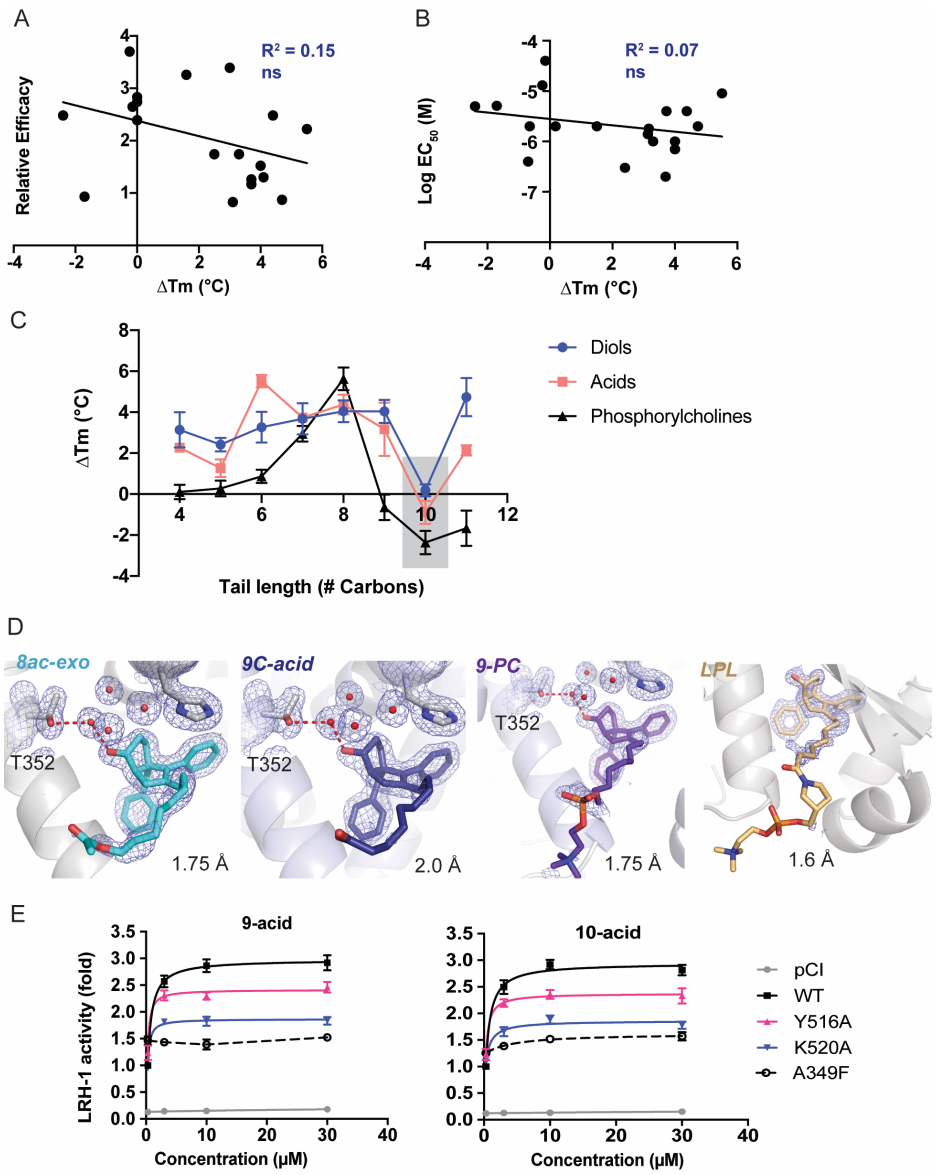


Figure 7.3. Preliminary mechanistic studies with the PL mimics. A-B. The relationship between Tm and EC_{50} (A) or Tm and efficacy (B) is not significant for this class of compounds. In contrast, the R1-modified compounds (Chapter 5) exhibit a highly significant relationship between Tm and EC_{50}, as shown in Figure 5.4B. C. Plot of Tm change versus tail length shows a dip in Tm value for analogs with 10-carbon tails. D. Close view of ligands in crystal structures of the LRH-1 LBD bound to four different R4-modified compounds, all of which exhibit substantial activity in luciferase reporter assays. The electron density maps (2Fo-Fc, contoured to 1σ show a lack of electron density for the ligand tails. E. Luciferase reporter assays showing a

reduction in activity when residues near the DLPC binding site are mutated (Y516 and K520). The A549F mutation occludes the binding pocket and was used as a negative control. WT, wild-type LRH-1.

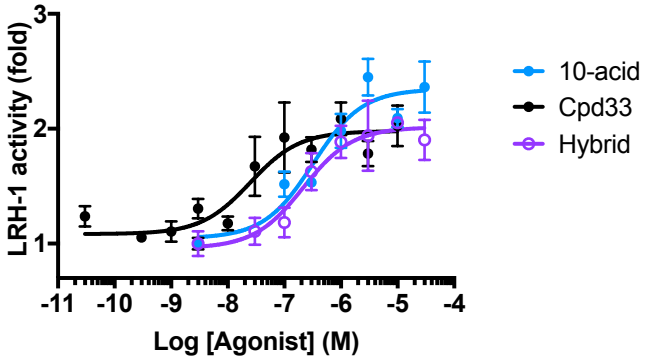


Figure 7.4. Luciferase reporter assay shows no improvement in potency or efficacy for a scaffold combining the best R1 and R4 modifications. Each point represents mean +/- SEM from 1-3 experiments conducted in triplicate.

7.2. Applications for this diverse compound library

This library of compounds with differing properties and activities is a valuable tool to study ligand-mediated activation of LRH-1 and for future design of modulators. Some of the potential uses for this library are highlighted in this section.

7.2.1 High affinity ligand enables the development of a competitive binding assay.

In addition to the therapeutic potential associated with more potent agonists, the discovery of a high affinity LRH-1 compound has several potential biochemical applications. For example, we have used Cpd33 to develop the first competitive binding assay for LRH-1 and SF-1. In this fluorescence-polarization (FP)-based assay a FAM-labeled version of Cpd33 binds purified LRH-1 in a dose-dependent manner with a Kd of 1 nM (Emma D'Agonstino, data not shown). This probe is competed off with both synthetic and PL ligands. The high affinity probe is essential to the success of this assay, since assessment of competitor affinity is limited by probe affinity(Huang, 2003). This probe will be an essential tool for the field: demonstrating direct ligand binding has been extremely challenging, often necessitating alternative approaches (such as monitoring coregulator recruitment as a proxy for ligand binding)(Busby, 2010; Whitby et al., 2011). Our binding assay has been useful to study the PL mimics: for example, we have used it to identify a significant relationship between R4 tail length and Kd. Moreover, using this assay, we discovered that several experimental antagonists (e.g. SR-1848(Busby, 2010) and Cpd3(Benod et al., 2013)) do not bind in the LRH-1 pocket (all binding data generated by Emma D'Agostino, not shown).

7.2.2 Tools to elucidate the ligand-dependent transcriptional program of LRH-

Most of the biological functions attributed to LRH-1 have been identified using gain- and loss-of-function studies (Lefevre et al., 2015; Mataki et al., 2007; Nissim et al., 2016; Oosterveer et al., 2012; Stein et al., 2014; Wagner et al., 2016). LRH-1 is also regulated by post translational modifications and by association with several different coregulators (Stein and Schoonjans,

2015), and it is unclear what portion of LRH-1's transcriptional program is specifically ligand-regulated. This is a fairly complex question and likely involves a dynamic interplay between ligand affinity, coregulator expression, and biological context. It is quite likely that allosteric communication between the ligand and coregulator is important and that different ligands will alter coregulator selectivity. We have seen that the presence of the Tif2 or the PGC1α coactivator dramatically affects receptor allostery when LRH-1 is bound to the RJW100 agonist (Chapter 4). Having a set of agonists with different properties will be useful in elucidating the ligand-driven component of LRH-1 signaling.

Another interesting avenue will be to investigate whether the two classes of agonists promote different gene expression changes. Even within the R1-modified family, we have identified partial agonists and a "full" agonist (Figure 5.7). Aside from differences in potency, Cpds 16 and 33 exert distinct effects on receptor conformation at saturation (Figure 5.6-5.7). Whether a partial or full agonist would be most desirable in a therapeutic setting remains to be determined. For other NRs, partial agonists have dissociative properties that are beneficial. For example, PPAR partial agonists improve insulin resistance with fewer adverse effects (*i.e.* weight gain or heart attacks) than full agonists (Berger, 2005; Liu, 2015) In addition, agonists that bind but do not activate could be used to probe effects of endogenous PL ligands by occupying the pocket and preventing PL-driven effects, somewhat like a dominant negative. RNAseq and ChIPseq, and coregulator profiling experiments with agonists will be useful to understand these questions and are planned for the immediate future.

7.2.3 Computational modeling of ligand-driven LRH-1 activation

We have observed that several of our R1 agonists induce conformational changes at the LRH alternative activation function surface (AF-B, shown in HDX and MDS studies, Chapters 3-5). This is rather remarkable, since they do not directly contact AF-B. A certain bias associated with our computational modeling (and structural analysis) thus far is that we have selectively

chosen the most activating compounds for crystallization. Including compounds that bind but do not activate or are poor activators in this analysis will be very useful to tease out conformational changes specifically associated with activation. New LRH-1-agonist structures would be ideal for this analysis; however, considering that we have established quite a consistent conformation of the agonist cores, we will likely be able to dock agonists *in silico* with confidence. This will enable a quick generation of several models that could be used for MDS to elucidate molecular motions involved in activation. Further, the establishment of a consistent binding mode for this class of ligand, together with the knowledge gained from MDS and HDX with inactive ligands, could be extremely beneficial for design of LRH-1 inhibitors. For example, incorporation of a bulky moiety at position R2 on the scaffold (external styrene) could result in displacement of helix 12 (Figure 7.5). This is a common strategy for NR antagonist design.

7.3. Future directions for evaluating agonists as therapeutic agents.

One of the long-term objectives of this work is to produce LRH-1 agonists useful for treatment of human metabolic diseases. Several studies are needed to assess suitability for the clinic.

7.3.1 LRH-1 Selectivity

We have not extensively explored off-target effects with other NRs, and this will be an important future direction. In particular, the SF-1 binding pocket is highly similar to LRH-1, and activating adrenal SF-1 may cause deleterious side effects if these agonists were to be used in a therapeutic setting. Although Cpd25 (*endo*-acetamide) is highly selective for LRH-1 over SF-1 (Chapter 5), the mechanism through which this selectivity is achieved is unclear. Cpd33 also showed some selectivity, although it exhibited some activity toward SF-1 (Figure 5.6). Both **25** and **33** interact with M345, which is a leucine in SF-1, and we initially hypothesized that this interaction could confer selectivity. However, preliminary mutagenesis studies suggest that this is not the case: a M345L mutation does not significantly affect the ability for these compounds to

activate LRH-1 (Figure 5.6). Likewise, a L265M mutation in SF-1 does not restore the activity of **25** (data not shown). Other closely related NRs should also be explored for potential off-target effects, such as EstR and the PPARs.

7.3.2 Stereospecific effects.

Each agonist in these studies is a racemic mixture of two enantiomers. For example, *exo*-RJW100 consists of a mixture of RR-RJW100 and SS-RJW100 (Figure 3.1). Although the LRH-1 crystals exclusively select for binding of the "RR" enantiomer, it is possible that the other enantiomer plays a biological role. Indeed, in preliminary studies investigating enantiomerically pure versions of RJW100, we have seen that both enantiomers bind LRH-1 and have distinct effects on LRH-1 transactivation in cells and on LRH-1

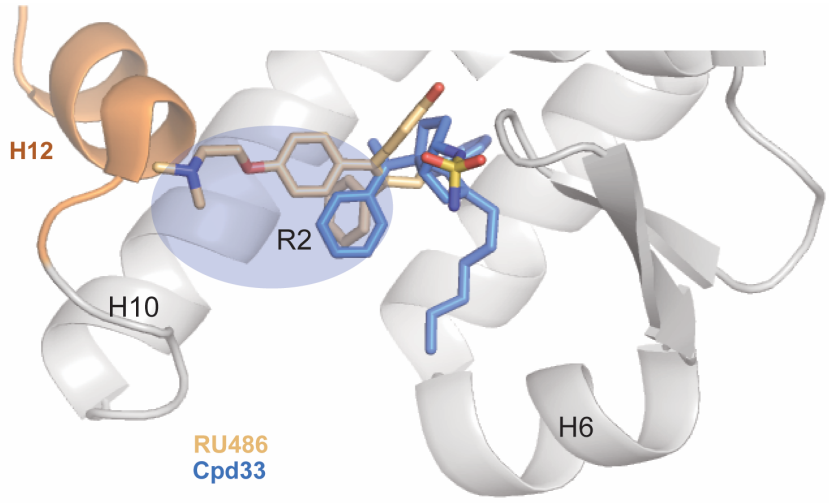


Figure 7.5. Reliable positioning of the 6HP core provides the basis for rational design of LRH-1 antagonists. The external styrene ring of the 6HP compounds (R2, highlighted in blue), is positioned such that bulky moieties could be incorporated to displace LRH-1 helix 12 (H12, orange). Shown is the LRH-1-Cpd33 crystal structure with the progesterone receptor antagonist RU486 superposed from PDB 2GPU (shown in sticks, light orange)(Wang et al., 2006). Cpd33 is shown in blue sticks. Portions of helices 2 and 3 have been omitted for clarity.

protein conformational dynamics in solution (data not shown). Enantiomer-specific effects will be important to elucidate in future studies to determine whether a tricky separation is necessary for biological applications.

7.3.3 Pharmacokinetics and Toxicity

In vivo studies will require a careful study of pharmacokinetics (PK), general toxicity, and potential for drug-drug interactions for the lead compounds. Toward this end, we have conducted preliminary studies to investigate cytochrome P450 (CYP) inhibition by representative agonists from the R1 and R4 classes (Table 7.1). This assay investigated the ability of agonists to inhibit a panel of CYPs commonly involved in drug metabolism.

Additional studies required would be an analysis of absorption, distribution, metabolism, and excretion kinetics, as well as a study oral bioavailability and tissue distribution of the agonists. Interestingly, the 9C-acid exhibited less inhibition of these CYPs than either RJW100 or the *endo*-triazole (an R1-compound (Cpd35 from Chapter 5)).

7.3.4 Efficacy in Disease Models

One of the most exciting potential applications of these compounds is to probe their efficacy in metabolic and inflammatory diseases. Preliminary studies involving an organoid model of inflammatory bowel disease are promising. These studies (conducted by the Moore group, Baylor) involve TNFα -mediated induction of inflammation. Both Cpd33 and the 10C-acid reduced expression of the LRH-1-controlled anti-inflammatory gene Cyp11b1 (Alex Wang, personal communication). Further gene expression and phenotypic analyses are planned, as well as studies for efficacy in *in vivo* models of IBD. Additional disease-related studies planned for the immediate future include: high fat diet-induced obesity (Calvert lab, Emory), lifespan and metabolic studies in worms (Wang lab, Baylor), and pancreatic studies in zebrafish (Goessling Lab, Harvard). Moreover, we are in the process of generating a mouse via CRISPR-Cas9 that involves humanization of the LRH-1 binding pocket. The humanized receptor

Table 7.1. Assays for cytochrome P450 inhibition. The ability for LRH-1 agonists to inhibit a panel of cytochrome P450 (P450) enzymes in human liver microsomes was quantified to assess potential off-target effects. Each column shows a particular P450 and the name of the substrate used to measure activity. Liquid chromatography plus tandem mass spectrometry (LC-MS/MS) was used to quantify formation of metabolites of each drug as a measure of activity. Compounds listed in the left column are positive controls that elicit maximum inhibition of each P450. LRH-1 agonists (blue-shaded rows) were compared to these positive controls. The endo-triazole is referred to as Cpd 35 in Chapter 5 (Figure 5.4). The 9C-acid is described in Chapter 6 (also called Cpd 13g). Assay conducted by Pharmeron, Inc.

Compound	% Inhibition @ 10 μM					
	CYP1A2 (Phenacetin)	CYP2A6 (Coumarin)	CYP2B6 (Bupropion)	CYP2E1 (Chlorzoxazone)	CYP3A4 (Midazolam)	CYP3A4 (Testosterone)
Furafylline	84.46	-	-	-	-	-
Tranylcypromine	-	98.27	-	-	-	-
Quercetin	-	-	-	-	-	-
Sulfaphenazole	-	-	-	-	-	-
N-3-benzylnirvanol	-	-	-	-	-	-
Quinidine	-	-	-	-	-	-
Zinc Diethyl-dithiocarbamate	-	-	-	33.80	-	-
Ketoconazole	-	-	77.02	-	99.43	98.71
Endo-Triazole	20.01	14.03	71.93	0	88.80	80.12
9C-acid	0	5.97	28.95	0	22.17	7.82
RJW100	6.42	6.57	30.21	0	27.14	32.24

responds to our synthetic agonists *in vitro* (not shown) and is designed to be expressed by the endogenous murine Lrh-1 promoter. Moreover, incorporation of an Avi tag to the modified receptor will facilitate pulldown studies to identify endogenous ligands.

7.4 Concluding Remarks

LRH-1 is an important regulator of metabolism, inflammation, pluripotency, and cellular proliferation and has significant pharmacological potential. This work has greatly furthered mechanistic understanding of how LRH-1 is regulated by synthetic agonists and has provided a substantial foundation for further studies. It has also generated greatly improved agonists with potential clinical utility.

References

Adams PD, Afonine PV, Bunkoczi G, Chen VB, Davis IW, Echols N, Headd JJ, Hung LW, Kapral GJ, Grosse-Kunstleve RW, McCoy AJ, Moriarty NW, Oeffner R, Read RJ, Richardson DC, Richardson JS, Terwilliger TC and Zwart PH (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta crystallographica Section D, Biological crystallography* **66**(Pt 2): 213-221.

Adrio J, Rivero, M.R., Carretero, J.C. (2001) Endo-Selective Intramolecular Paulson-Khand Reactions of Gamma-Oxygenated-Alpha,Beta-Unsaturated Phenylsulfones. *Chem - Eur J* **7**(11): 2435-2448.

Ballatore C, Huryn, D.M., Smith, A.B. III (2013) Carboxylic Acid (Bio)Isosteres in Drug Design. *ChemMedChem* **8**: 385-395.

Barillari C, Taylor, J., Viner, R., Essex, J.W. (2007) Cassification of Water Molecules in Protein Binding Sites. *J Am Chem Soc* **129**(9): 2577-2587.

Bayeva M, Sawicki KT and Ardehali H (2013) Taking diabetes to heart--deregulation of myocardial lipid metabolism in diabetic cardiomyopathy. *Journal of the American Heart Association* **2**(6): e000433.

Bayrer JR, Mukkamala S, Sablin EP, Webb P and Fletterick RJ (2015) Silencing LRH-1 in colon cancer cell lines impairs proliferation and alters gene expression programs. *Proceedings of the National Academy of Sciences of the United States of America* **112**(8): 2467-2472.

Bayrer JR, Wang H, Nattiv R, Suzawa M, Escusa HS, Fletterick RJ, Klein OD, Moore DD and Ingraham HA (2018) LRH-1 Mitigates Intestinal Infammatory Disease by Maintaining Epithelial Homeostasis and Cell Survival. *bioRxiv*.

Benod C, Carlsson J, Uthayaruban R, Hwang P, Irwin JJ, Doak AK, Shoichet BK, Sablin EP and Fletterick RJ (2013) Structure-based discovery of antagonists of nuclear receptor LRH-1. *The Journal of biological chemistry* **288**(27): 19830-19844.

Berger JP, Akiyama, T.E., Meinke, P.T., (2005) PPARs: therapeutic targets for metabolic disease. . *Trends Pharmacol Sci* **26**: 244-251.

Bianco S, Brunelle M, Jangal M, Magnani L and Gevry N (2014) LRH-1 governs vital transcriptional programs in endocrine-sensitive and -resistant breast cancer cells. *Cancer research* **74**(7): 2015-2025.

Binkowski TA (2003) CASTp: Computed Atlas of Surface Topography of proteins. *Nucleic acids research* **31**(13): 3352-3355.

Bolado-Carrancio A, Riancho JA, Sainz J and Rodriguez-Rey JC (2014) Activation of nuclear receptor NR5A2 increases Glut4 expression and glucose metabolism in muscle cells. *Biochemical and biophysical research communications* **446**(2): 614-619.

Borghese A, Antoine, L., Van Hoeck, J.P., Merschaert, A. (2006) Mild and Safer Perparative Method for Nonsymmetrical Sulfamides via *N*-Sulfamoyloxazolidinone Derivatives: Electronic Effects Affect the Trassulfamoylation Reactivity. *Org Process Res Dev* **10**(4): 770-775.

Bowerman S and Wereszczynski J (2016) Detecting Allosteric Networks Using Molecular Dynamics Simulation. *Methods in enzymology* **578**: 429-447.

Brendel C, Gelman, L., Auwerx, J. (2002) Multiprotein Bridging Factor-1 (MBF-1) Is a Cofactro for Nuclear Receptors that Regulate Lipid Metabolsim. *Molecular Endocrinology* **16**(6): 1367-1377.

Bril F, Cusi, K. (2016) Nonalchoholic Fatty Liver Disease: The New Complication of Type 2 Diabetes Mellitus. *Endocrinol Metab Clin N Am* **45**: 765-781.

Busby SA, Nuhant, P., Cameron, M., Mercer, B.A., Hodder, P., Roush, W.R. Griffin, P.R. (2010) Discovery of inverse agonists for the liver receptor homologue-1 (LRH1; NR5A2), in *Probe Reports from the Molecular Libraries Program*, Bethesda, MD.

Case D, Babin V, Berryman J, Betz R, Cai Q, Cerutti D, Cheatham Iii T, Darden T, Duke R and Gohlke H (2014) Amber 14.

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

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

Centers for Disease Control and Prevention (2014) National Diabetes Statistics Report: Estimates of Diabetes and Its Burden in the United States, 2014, National Institutes of Health, Atlanta, GA.

Chalkiadaki A and Talianidis I (2005) SUMO-dependent compartmentalization in promyelocytic leukemia protein nuclear bodies prevents the access of LRH-1 to chromatin. *Molecular and cellular biology* **25**(12): 5095-5105.

Chand AL, Herridge KA, Thompson EW and Clyne CD (2010) The orphan nuclear receptor LRH-1 promotes breast cancer motility and invasion. *Endocrine-related cancer* **17**(4): 965-975.

Clyne CD, Kovacic A, Speed CJ, Zhou J, Pezzi V and Simpson ER (2004) Regulation of aromatase expression by the nuclear receptor LRH-1 in adipose tissue. *Molecular and cellular endocrinology* **215**(1-2): 39-44.

Cobo-Vuilleumier N, Lorenzo PI, Rodriguez NG, Herrera Gomez IG, Fuente-Martin E, Lopez-Noriega L, Mellado-Gil JM, Romero-Zerbo SY, Baquie M, Lachaud CC, Stifter K, Perdomo G, Bugliani M, De Tata V, Bosco D, Parnaud G, Pozo D, Hmadcha A, Florido JP, Toscano MG, de Haan P, Schoonjans K, Sanchez Palazon L, Marchetti P, Schirmbeck R, Martin-Montalvo A, Meda P, Soria B, Bermudez-Silva FJ, St-Onge L and Gauthier BR (2018) LRH-1 agonism favours an immune-islet dialogue which protects against diabetes mellitus. *Nature communications* **9**(1): 1488.

Corzo CA, Mari Y, Chang MR, Khan T, Kuruvilla D, Nuhant P, Kumar N, West GM, Duckett DR, Roush WR and Griffin PR (2015) Antiproliferation activity of a small molecule repressor of liver receptor homolog 1. *Molecular pharmacology* **87**(2): 296-304.

Coste A, Dubuquoy, L., Barnouin, R., Annicotte, J-S., Magnier, B. Notti, M., Corazza, N., Antal, M.C., Metzger, D., Desreumaux, P., Brunner, T., Auwerx, J., and Schoonjans, K. LRH-1-mediated glucocorticoid synthesis in enterocytes protects

against infalmmatory bowel disease. *Proceedings of the National Academy of Sciences of the United States of America* **104**(32): 13098-13103.

Cusi K (2012) Role of Obesity and Lipotoxicity in the Development of Nonalcoholic Steatohepatitis: Pathophysiology and Clinical Implications. *Gastroenterology*(142): 711-725.

de Jesus Cortez F, Suzawa M, Irvy S, Bruning JM, Sablin E, Jacobson MP, Fletterick RJ, Ingraham HA and England PM (2016) Disulfide-Trapping Identifies a New, Effective Chemical Probe for Activating the Nuclear Receptor Human LRH-1 (NR5A2). *PLoS One* **11**(7): e0159316.

Devarakonda S, Gupta K, Chalmers MJ, Hunt JF, Griffin PR, Van Duyne GD and Spiegelman BM (2011) Disorder-to-order transition underlies the structural basis for the assembly of a transcriptionally active PGC-1alpha/ERRgamma complex. *Proceedings of the National Academy of Sciences of the United States of America* **108**(46): 18678-18683.

Emsley P and Cowtan K (2004) Coot: model-building tools for molecular graphics. *Acta crystallographica Section D, Biological crystallography* **60**(Pt 12 Pt 1): 2126-2132.

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

Feng L, Dharmarajan V, Serrao E, Hoyte A, Larue RC, Slaughter A, Sharma A, Plumb MR, Kessl JJ, Fuchs JR, Bushman FD, Engelman AN, Griffin PR and Kvaratskhelia M (2016) The Competitive Interplay between Allosteric HIV-1 Integrase Inhibitor BI/D and LEDGF/p75 during the Early Stage of HIV-1 Replication Adversely Affects Inhibitor Potency. *ACS Chem Biol* **11**(5): 1313-1321.

Fernandez-Marcos PJ, Auwerx J and Schoonjans K (2011) Emerging actions of the nuclear receptor LRH-1 in the gut. *Biochim Biophys Acta* **1812**(8): 947-955.

Finck BN, Kelly, D.P. (2006) PGC-1 coactivators: inducible regulators of metabolism in health and disease. *The Journal of clinical investigation* **116**(3): 615-622.

Firneisz G (2014) Non-alcoholic fatty liver disease and type 2 diabetes mellitus: the liver disease of our age? *World J Gastroenterol* **20**(27): 9072-9089.

Floyd RW (1962a) Algorithm 97: Shortest Path. *Communications of the ACM* **5**(6): 345.

Floyd RW (1962b) Algorithm 97: shortest path. *Communications of the ACM* **5**(6): 345.

Flynn AR, Mays SG, Ortlund EA and Jui NT (2018) Development of Hybrid Phospholipid Mimics as Effective Agonists for Liver Receptor Homologue-1. *ACS Medicinal Chemistry Letters* **9**(10): 1051-1056.

Furse S and de Kroon AI (2015) Phosphatidylcholine's functions beyond that of a membrane brick. *Mol Membr Biol* **32**(4): 117-119.

Gasper PM, Fuglestad B, Komives EA, Markwick PRL and McCammon JA (2012) Allosteric networks in thrombin distinguish procoagulant vs. anticoagulant activities. *Proceedings of the National Academy of Sciences* **109**(52): 21216-21222.

Glykos NM (2006a) Carma: a molecular dynamics analysis program. *J Comput Chem* **27**: 1765-1768.

Glykos NM (2006b) Software news and updates carma: A molecular dynamics analysis program. *Journal of Computational Chemistry* **27**(14): 1765-1768.

Goswami D, Devarakonda, S., Chalmers, M.J., Pascal, B.D., Spiegelman, B.M., Griffin, P.R. (2013) Time Window Expansion for HDX Analysis of an Intrinsically Disordered Protein. *J Am Soc Mass Spectrom* **24**: 1584-1592.

Grgurevic N, Tobet, S., Majdic, G. (2005) Widespread expression of liver receptor homolog 1 in mouse brain. *Neuro Endocrinol Lett* **26**(5): 541-547.

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

Gu P, Goodwin B, Chung AC, Xu X, Wheeler DA, Price RR, Galardi C, Peng L, Latour AM, Koller BH, Gossen J, Kliewer SA and Cooney AJ (2005) Orphan nuclear receptor LRH-1 is required to maintain Oct4 expression at the epiblast stage of embryonic development. *Molecular and cellular biology* **25**(9): 3492-3505.

Holmstrom SR, Deering T, Swift GH, Poelwijk FJ, Mangelsdorf DJ, Kliewer SA and MacDonald RJ (2011) LRH-1 and PTF1-L coregulate an exocrine pancreas-specific transcriptional network for digestive function. *Genes & development* **25**(16): 1674-1679.

Huang X (2003) Fluorescence polarization competition assay: the range of resolvable inhibitor potency is limited by the affinity of the fluorescent ligand. *J Biomol Screen* **8**(1): 34-38.

Humphrey W, Dalke A and Schulten K (1996) VMD: visual molecular dynamics. *Journal of molecular graphics* **14**(1): 33-38.

Joosten RP, Long F, Murshudov GN and Perrakis A (2014) The PDB_REDO server for macromolecular structure model optimization. *IUCrJ* **1**(Pt 4): 213-220.

Kanayama T, Arito M, So K, Hachimura S, Inoue J and Sato R (2007) Interaction between Sterol Regulatory Element-binding Proteins and Liver Receptor Homolog-1 Reciprocally Suppresses Their Transcriptional Activities. *Journal of Biological Chemistry* **282**(14): 10290-10298.

Kir S, Zhang Y, Gerard RD, Kliewer SA and Mangelsdorf DJ (2012) Nuclear receptors HNF4alpha and LRH-1 cooperate in regulating Cyp7a1 in vivo. *The Journal of biological chemistry* **287**(49): 41334-41341.

Klebe G (2015) Applying thermodynamic profiling in lead finding and optimization. *Nat Rev Drug Discov* **14**(2): 95-110.

Krylova IN, Sablin EP, Moore J, Xu RX, Waitt GM, MacKay JA, Juzumiene D, Bynum JM, Madauss K, Montana V, Lebedeva L, Suzawa M, Williams JD, Williams SP, Guy RK, Thornton JW, Fletterick RJ, Willson TM and Ingraham HA (2005) Structural analyses reveal phosphatidyl inositols as ligands for the NR5 orphan receptors SF-1 and LRH-1. *Cell* **120**(3): 343-355.

Lai CF, Flach KD, Alexi X, Fox SP, Ottaviani S, Thiruchelvam PT, Kyle FJ, Thomas RS, Launchbury R, Hua H, Callaghan HB, Carroll JS, Charles Coombes R, Zwart W, Buluwela L and Ali S (2013) Co-regulated gene expression by oestrogen receptor alpha and liver receptor homolog-1 is a feature of the oestrogen response in breast cancer cells. *Nucleic acids research* **41**(22): 10228-10240.

Lalit M, Gangwal RP, Dhoke GV, Damre MV, Khandelwal K and Sangamwar AT (2013) A combined pharmacophore modeling, 3D-QSAR and molecular docking

study of substituted bicyclo-[3.3.0]oct-2-enes as liver receptor homolog-1 (LRH-1) agonists. *Journal of Molecular Structure* **1049**: 315-325.

Laskowski RA and Swindells MB (2011) LigPlot+: multiple ligand-protein interaction diagrams for drug discovery. *J Chem Inf Model* **51**(10): 2778-2786.

Lee JM, Lee YK, Mamrosh JL, Busby SA, Griffin PR, Pathak MC, Ortlund EA and Moore DD (2011) A nuclear-receptor-dependent phosphatidylcholine pathway with antidiabetic effects. *Nature* **474**(7352): 506-510.

Lee JS, Bae S, Kang HS, Im SS and Moon YA (2017) Liver receptor homolog-1 regulates mouse superoxide dismutase 2. *Biochemical and biophysical research communications* **489**(3): 299-304.

Lee YK, Choi YH, Chua S, Park YJ and Moore DD (2006) Phosphorylation of the hinge domain of the nuclear hormone receptor LRH-1 stimulates transactivation. *The Journal of biological chemistry* **281**(12): 7850-7855.

Lefevre L, Authier H, Stein S, Majorel C, Couderc B, Dardenne C, Eddine MA, Meunier E, Bernad J, Valentin A, Pipy B, Schoonjans K and Coste A (2015) LRH-1 mediates anti-inflammatory and antifungal phenotype of IL-13-activated macrophages through the PPARgamma ligand synthesis. *Nature communications* **6**: 6801.

Li Y, Choi M, Suino K, Kovach A, Daugherty J, Kliewer SA and Xu HE (2005) Structural and biochemical basis for selective repression of the orphan nuclear receptor liver receptor homolog 1 by small heterodimer partner. *Proceedings of the National Academy of Sciences of the United States of America* **102**(27): 9505-9510.

Li Y, Kovach, A., Suino-Powell, K., Martynowski, D., Xu, H.E. (2008) Structural and Biochemical Basis for the Binding Selectivity of Peroxisome Proliferator-activated Receptor Gamma to PGC-1 Alpha. *The Journal of biological chemistry* **283**(27): 19132-19139.

Li Y, Lambert MH and Xu HE (2003) Activation of Nuclear Receptors. *Structure* **11**(7): 741-746.

Li Z, Agellon LB, Allen TM, Umeda M, Jewell L, Mason A and Vance DE (2006) The ratio of phosphatidylcholine to phosphatidylethanolamine influences membrane integrity and steatohepatitis. *Cell metabolism* **3**(5): 321-331.

Liebschner D, Afonine PV, Moriarty NW, Poon BK, Sobolev OV, Terwilliger TC and Adams PD (2017) Polder maps: improving OMIT maps by excluding bulk solvent. *Acta Crystallogr D Struct Biol* **73**(Pt 2): 148-157.

Lin CJ and Wang MC (2017) Microbial metabolites regulate host lipid metabolism through NR5A-Hedgehog signalling. *Nat Cell Biol* **19**(5): 550-557.

Lin J, Handschin C and Spiegelman BM (2005) Metabolic control through the PGC-1 family of transcription coactivators. *Cell metabolism* **1**(6): 361-370.

Lin Q, Aihara A, Chung W, Li Y, Huang Z, Chen X, Weng S, Carlson RI, Wands JR and Dong X (2014) LRH1 as a driving factor in pancreatic cancer growth. *Cancer letters* **345**(1): 85-90.

Lindqvist AM, D., Word, R.A. (2013) The impact of reference gene selection in quantification of gene expression levels in guinea pig cervical tissues and cells. *BMC Research Notes* **6**(34).

Liu H-J, Zhang, C-Y., Song, F., Xiao, T., Meng, J. Zhang, Q., Liang, C-L., Li, S., Wang, J. Zhang, B., Liu, Y-R., Sun, T., Zhou, H-G. (2015) A Novel Partial Agonist of Peroxisome Proliferator-Activated Receptor γ with Excellent Effect on Insulin Resistance and Type 2 Diabetes. *Journal of Pharmacology and Experimental Therapeutics* **353**(3): 573-581.

Lomonaco RB, F., Portillo-Sanchez, P., Ortiz-Lopez, C., Orsak, B., Biernacki, D., Lo, M., Suman, A., Weber, M.H., Cusi, K. (2016) Metabolic Impact of Nonalcoholic Steatohepatitis in Obese Patients With Type 2 Diabetes. *Diabetes Care* **39**: 632-638.

Lonardo A, Ballestri, S. Marchesini, G. Angulo, P. Loria, P. (2015) Nonalcolholic fatty liver disease: A precursor of the metabolic syndrome. *Digestive and Liver Disease* **47**: 181-190.

Lu T, Makishima M, JJ R, Schoonjans K, Kerr T, Auwerx J and DJ M (2000) Molecular Basis for Feedback Regulation of Bile Acid Synthesis by Nuclear Receptors. *Molecular Cell* **6**: 507-515.

MacDonald JI, Specher, H. (1991) Phospholipid Fatty Acid Remodeling in Mammalian Cells. *Biochim Biophys Acta, Lipids Lipid Metab* **1084**(2): 105-121.

Maier JA, Martinez C, Kasavajhala K, Wickstrom L, Hauser KE and Simmerling C (2015) ff14SB: Improving the accuracy of protein side chain and backbone parameters from ff99SB. *Journal of chemical theory and computation* **11**(8): 3696-3713.

Mamrosh JL, Lee JM, Wagner M, Stambrook PJ, Whitby RJ, Sifers RN, Wu SP, Tsai MJ, Demayo FJ and Moore DD (2014) Nuclear receptor LRH-1/NR5A2 is required and targetable for liver endoplasmic reticulum stress resolution. *eLife* **3**: e01694.

Mataki C, Magnier BC, Houten SM, Annicotte JS, Argmann C, Thomas C, Overmars H, Kulik W, Metzger D, Auwerx J and Schoonjans K (2007) Compromised intestinal lipid absorption in mice with a liver-specific deficiency of liver receptor homolog 1. *Molecular and cellular biology* **27**(23): 8330-8339.

Mays SG, Okafor CD, Tuntland ML, Whitby RJ, Dharmarajan V, Stec J, Griffin PR and Ortlund EA (2017) Structure and Dynamics of the Liver Receptor Homolog 1-PGC1alpha Complex. *Molecular pharmacology* **92**(1): 1-11.

Mays SG, Okafor CD, Whitby RJ, Goswami D, Stec J, Flynn AR, Dugan MC, Jui NT, Griffin PR and Ortlund EA (2016) Crystal Structures of the Nuclear Receptor, Liver Receptor Homolog 1, Bound to Synthetic Agonists. *The Journal of biological chemistry* **291**(49): 25281-25291.

Miranda DA, Krause WC, Cazenave-Gassiot A, Suzawa M, Escusa H, Foo JC, Shihadih DS, Stahl A, Fitch M, Nyangau E, Hellerstein M, Wenk MR, Silver DL and Ingraham HA (2018) LRH-1 regulates hepatic lipid homeostasis and maintains arachidonoyl phospholipid pools critical for phospholipid diversity. *JCI Insight* **3**(5).

Mueller M, Cima I, Noti M, Fuhrer A, Jakob S, Dubuquoy L, Schoonjans K and Brunner T (2006) The nuclear receptor LRH-1 critically regulates extra-adrenal glucocorticoid synthesis in the intestine. *J Exp Med* **203**(9): 2057-2062.

Musille PM, Kossmann BR, Kohn JA, Ivanov I and Ortlund EA (2015) Unexpected Allosteric Network Contributes to LRH-1 Coregulator Selectivity. *The Journal of biological chemistry*.

Musille PM, Pathak M, Lauer JL, Griffin PR and Ortlund EA (2013) Divergent sequence tunes ligand sensitivity in phospholipid-regulated hormone receptors. *The Journal of biological chemistry* **288**(28): 20702-20712.

Musille PM, Pathak MC, Lauer JL, Hudson WH, Griffin PR and Ortlund EA (2012) Antidiabetic phospholipid-nuclear receptor complex reveals the mechanism for phospholipid-driven gene regulation. *Nature structural & molecular biology* **19**(5): 532-537, S531-532.

Nadolny C and Dong X (2015) Liver receptor homolog-1 (LRH-1): a potential therapeutic target for cancer. *Cancer Biol Ther* **16**(7): 997-1004.

National Institute of Diabetes and Digestive and Kidney Diseases (2010) Overweight and Obesity Statistics.

Newman ME (2006) Modularity and community structure in networks. *Proceedings of the National Academy of Sciences of the United States of America* **103**(23): 8577-8582.

Nissim S, Weeks O, Talbot JC, Hedgepeth JW, Wucherpfennig J, Schatzman-Bone S, Swinburne I, Cortes M, Alexa K, Megason S, North TE, Amacher SL and Goessling W (2016) Iterative use of nuclear receptor Nr5a2 regulates multiple stages of liver and pancreas development. *Dev Biol* **418**(1): 108-123.

Ogata K, Wodak, S.J. (2002) Conserved water molecules in MHC class-I molecules and their structural and functional roles. *Protein Engineering* **15**(8): 697-705.

Oosterveer MH, Mataki C, Yamamoto H, Harach T, Moullan N, van Dijk TH, Ayuso E, Bosch F, Postic C, Groen AK, Auwerx J and Schoonjans K (2012) LRH-1-dependent glucose sensing determines intermediary metabolism in liver. *The Journal of clinical investigation* **122**(8): 2817-2826.

Ortiz-Lopez C, Lomonaco R, Orsak B, Finch J, Chang Z, Kochunov VG, Hardies J and Cusi K (2012) Prevalence of prediabetes and diabetes and metabolic profile of patients with nonalcoholic fatty liver disease (NAFLD). *Diabetes Care* **35**(4): 873-878.

Ortlund EA, Lee Y, Solomon IH, Hager JM, Safi R, Choi Y, Guan Z, Tripathy A, Raetz CR, McDonnell DP, Moore DD and Redinbo MR (2005) Modulation of human nuclear receptor LRH-1 activity by phospholipids and SHP. *Nature structural & molecular biology* **12**(4): 357-363.

Otwinowski Z, Minor, W. (1997) Processing of X-Ray Diffraction Data Collected in Oscillation Mode. *Methods in Enzymology* **276**: 307-326.

Pascal BD, Willis S, Lauer JL, Landgraf RR, West GM, Marciano D, Novick S, Goswami D, Chalmers MJ and Griffin PR (2012) HDX workbench: software for the analysis of H/D exchange MS data. *J Am Soc Mass Spectrom* **23**(9): 1512-1521.

Payne A, Hales, DB (2004) Overview of Steroidogenic Enzymes in the Pathway from Cholesterol to Active Steroid Hormones. *Endocrine Reviews* **25**(6): 947-970.

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

Raaijmakers HC, Versteegh JE and Uitdehaag JC (2009) The X-ray structure of RU486 bound to the progesterone receptor in a destabilized agonistic conformation. *The Journal of biological chemistry* **284**(29): 19572-19579.

Rey J, Hu, H., Kyle, F. Lai, C-F., Buluwela, R., Coombes, R.C., Ortlund, E.A., Ali, S., Snyder, J.P., Barrett, A.G.M. (2012) Discovery of a New Class of Liver Receptor Homologue-1 (LRH-1) Antagonists: Virtual Screening, Synthesis, and Biological Evaluation. *Chem Med Chem* **7**(11): 1909-1914.

Ridgway ND (2013) The role of phosphatidylcholine and choline metabolites to cell proliferation and survival. *Critical Reviews in Biochemistry and Molecular Biology* **48**(1): 20-38.

Roe DR and Cheatham III TE (2013) PTRAJ and CPPTRAJ: software for processing and analysis of molecular dynamics trajectory data. *Journal of Chemical Theory and Computation* **9**(7): 3084-3095.

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

Rye CS, Baell, J.B. (2005) Phosphate Isosteres in Medicinal Chemistry. *Curr Med Chem* **12**(26): 3127-3141.

Sablin EP, Blind RD, Uthayaruban R, Chiu HJ, Deacon AM, Das D, Ingraham HA and Fletterick RJ (2015) Structure of Liver Receptor Homolog-1 (NR5A2) with PIP hormone bound in the ligand binding pocket. *Journal of structural biology*.

Sablin EP, Krylova IN, Fletterick RJ and Ingraham HA (2003) Structural Basis for Ligand-Independent Activation of the Orphan Nuclear Receptor LRH-1. *Molecular Cell* **11**(6): 1575-1585.

Sablin EP, Woods A, Krylova IN, Hwang P, Ingraham HA and Fletterick RJ (2008) The structure of corepressor Dax-1 bound to its target nuclear receptor LRH-1. *Proceedings of the National Academy of Sciences of the United States of America* **105**(47): 18390-18395.

Safi R, Kovacic, A., Gaillard, S., Murata, Y., Simpson, E.R., McDonnell, D.P., Clyne, C.D. (2005) Coactivation of Liver Receptor Homologues-1 by Peroxisome Proliferator-Activated Receptor Gamma Coactivator-1 alpha on Aromatase Promoter II and Its Inhibition by Activated Retinoid X Receptor Suggest a Novel Target for Breast-Specific Antiestrogen Therapy. *Cancer research* **65**(24): 11762-11770.

Sahini N and Borlak J (2016) Genomics of human fatty liver disease reveal mechanistically linked lipid droplet-associated gene regulations in bland steatosis and nonalcoholic steatohepatitis. *Transl Res* **177**: 41-69.

Savage DB, Petersen KF and Shulman GI (2007) Disordered lipid metabolism and the pathogenesis of insulin resistance. *Physiological reviews* **87**(2): 507-520.

Schoonjans K, Annicotte J-S, Huby T, Botrugno O, Fayard E, Ueda Y, Chapman J and Auwerx J (2002) Liver receptor homolog 1 controls the expression of the scavenger receptor class B type I. *EMBO Reports* **3**(12): 1181-1187.

Schrodinger, LLC (2010) The PyMOL Molecular Graphics System, Version 1.3r1.

Schrodinger L (2016) In the Ligand Interaction Diagram, what are the maximum lengths of hydrogen bonds, pi-pi stacking interactions, pi-cation interactions and metal

coordination interactions that are identified and displayed?, in *Schrodinger Knowledge Base*.

Schwaderer J, Gaiser AK, Phan TS, Delgado ME and Brunner T (2017) Liver receptor homolog-1 (NR5a2) regulates CD95/Fas ligand transcription and associated T-cell effector functions. *Cell Death Dis* **8**(4): e2745.

Sethi A, Eargle J, Black AA and Luthey-Schulten Z (2009a) Dynamical networks in tRNA:protein complexes. *Proceedings of the National Academy of Sciences of the United States of America* **106**(16): 6620-6625.

Sethi A, Eargle J, Black AA and Luthey-Schulten Z (2009b) Dynamical networks in tRNA:protein complexes. *Proceedings of the National Academy of Sciences of the United States of America* **106**(16): 6620-6625.

Shiau AK, Barstad, D., Loria, P.M., Cheng, L., Kushner, P.J., Agard, D.A., Greene, G.L. (1998) The Structural Basis of Estrogen Receptor/ Coactivator Recognition and the Antagonism of This Interaction by Tamoxifen. *Cell* **95**(7): 927-937.

Shin D, Osborne, T.F. (2008) Peroxisome Proliferator-activated Receptor-gamma Coactivator 1alpha Activation of CYP7A1 during Food Restriction and Diabetes Is Still Inhibited by Small Heterodimer Partner. *The Journal of biological chemistry* **283**(22): 15089-15096.

Solomon IH, Hager JM, Safi R, McDonnell DP, Redinbo MR and 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**(5): 1091-1102.

Song X, Kaimal R, Yan B and Deng R (2008) Liver receptor homolog 1 transcriptionally regulates human bile salt export pump expression. *J Lipid Res* **49**(5): 973-984.

Stec J (2010) Tandem Reaction Sequences on a Zirconocene Template, in *Chemistry*, PhD. Thesis, University of Southampton.

Stein S, Lemos, V. Xu, P., Demagny, H., Wang, X., Ryu, D., Jimenez, V., Bosch, F., Luscher, T.F., Oosterveer M.H., Schoonjans, K. (2016) Impaired SUMOylation of nuclear receptor LRH-1 promotes nonalcoholic fatty liver disease. *The Journal of clinical investigation* **127**(2): 583-592.

Stein S, Oosterveer MH, Mataki C, Xu P, Lemos V, Havinga R, Dittner C, Ryu D, Menzies KJ, Wang X, Perino A, Houten SM, Melchior F and Schoonjans K (2014) SUMOylation-dependent LRH-1/PROX1 interaction promotes atherosclerosis by decreasing hepatic reverse cholesterol transport. *Cell metabolism* **20**(4): 603-613.

Stein S and Schoonjans K (2015) Molecular basis for the regulation of the nuclear receptor LRH-1. *Current Opinion in Cell Biology* **33**: 26-34.

Thiruchelvam PT, Lai CF, Hua H, Thomas RS, Hurtado A, Hudson W, Bayly AR, Kyle FJ, Periyasamy M, Photiou A, Spivey AC, Ortlund EA, Whitby RJ, Carroll JS, Coombes RC, Buluwela L and Ali S (2011) The liver receptor homolog-1 regulates estrogen receptor expression in breast cancer cells. *Breast cancer research and treatment* **127**(2): 385-396.

Thomas E, Dixon, S. Whitby, R.J. (2006) A Rearrangement to a Zirconium-Alkenylidene in the Insertion of Dihalocarvvenoids and Acetylides into Ziconacycles. *Angew Chemie Int* **45**(43): 7070-7072.

van der Veen JN, Kennelly JP, Wan S, Vance JE, Vance DE and Jacobs RL (2017) The critical role of phosphatidylcholine and phosphatidylethanolamine metabolism in health and disease. *Biochim Biophys Acta Biomembr* **1859**(9 Pt B): 1558-1572.

Van Wart AT, Durrant J, Votapka L and Amaro RE (2014) Weighted Implementation of Suboptimal Paths (WISP): An Optimized Algorithm and Tool for Dynamical Network Analysis. *Journal of Chemical Theory and Computation* **10**(2): 511-517.

VanWart AT, Eargle J, Luthey-Schulten Z and Amaro RE (2012) Exploring Residue Component Contributions to Dynamical Network Models of Allostery. *Journal of Chemical Theory and Computation* **8**(8): 2949-2961.

Venteclef N, Jakobsson T, Ehrlund A, Damdimopoulos A, Mikkonen L, Ellis E, Nilsson LM, Parini P, Janne OA, Gustafsson JA, Steffensen KR and Treuter E (2010) GPS2-dependent corepressor/SUMO pathways govern anti-inflammatory actions of LRH-1 and LXRbeta in the hepatic acute phase response. *Genes & development* **24**(4): 381-395.

Venteclef N, Smith JC, Goodwin B and Delerive P (2006) Liver receptor homolog 1 is a negative regulator of the hepatic acute-phase response. *Molecular and cellular biology* **26**(18): 6799-6807.

Wagner M, Choi S, Panzitt K, Mamrosh JL, Lee JM, Zaufel A, Xiao R, Wooton-Kee R, Stahlman M, Newgard CB, Boren J and Moore DD (2016) Liver receptor homolog-1 is a critical determinant of methyl-pool metabolism. *Hepatology* **63**(1): 95-106.

Wagner RT, Xu X, Yi F, Merrill BJ and Cooney AJ (2010) Canonical Wnt/beta-catenin regulation of liver receptor homolog-1 mediates pluripotency gene expression. *Stem cells* **28**(10): 1794-1804.

Walker AK (2017) 1-Carbon Cycle Metabolites Methylate Their Way to Fatty Liver. *Trends Endocrinol Metab* **28**(1): 63-72.

Walker AK, Jacobs RL, Watts JL, Rottiers V, Jiang K, Finnegan DM, Shioda T, Hansen M, Yang F, Niebergall LJ, Vance DE, Tzoneva M, Hart AC and Naar AM (2011) A conserved SREBP-1/phosphatidylcholine feedback circuit regulates lipogenesis in metazoans. *Cell* **147**(4): 840-852.

Wang J, Wang W, Kollman PA and Case DA (2001) Antechamber: an accessory software package for molecular mechanical calculations. *J Am Chem Soc* **222**: U403.

Wang L, Zuercher WJ, Consler TG, Lambert MH, Miller AB, Orband-Miller LA, McKee DD, Willson TM and Nolte RT (2006) X-ray crystal structures of the estrogen-related receptor-gamma ligand binding domain in three functional states reveal the molecular basis of small molecule regulation. *The Journal of biological chemistry* **281**(49): 37773-37781.

Weikum ER, Liu X and Ortlund EA (2018) The nuclear receptor superfamily: A structural perspective. *Protein Sci*.

Weikum ER, Tuntland ML, Murphy MN and Ortlund EA (2016) A Structural Investigation into Oct4 Regulation by Orphan Nuclear Receptors, Germ Cell Nuclear Factor (GCNF), and Liver Receptor Homolog-1 (LRH-1). *Journal of Molecular Biology* **428**(24): 4981-4992.

Whitby RJ, Dixon S, Maloney PR, Delerive P, Goodwin B, Parks DJ and Wilson TM (2006) Identification of Small Molecule Agonists of the Orphan Nuclear

Receptros Liver Receoptor Homolog-1 and Steroidogenic Factor-1. *Journal of medicinal chemistry* **49**: 6652-6655.

Whitby RJ, Stec J, Blind RD, Dixon S, Leesnitzer LM, Orband-Miller LA, Williams SP, Willson TM, Xu R, Zuercher WJ, Cai F and Ingraham HA (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* **54**(7): 2266-2281.

Xiao L, Wang, Y., Xu, K., Hu, H., Xu, Z., Wu, D., You, W., Ng, C-F., Yu, S.. Chan, F.L. (2018) Nuclear Receptor LRH-1 Functions to Promote Castration-Resistant Growth of Prostate Cancer via Its Promotion of Introtumoral Androgen Biosynthesis. *Cancer research* **78**(9): 2205-2218.

Xu P, Oosterveer, M.H., Stein, S., Demagny, H., Ryu, D., Moullan, N., Comment, A., Auwerx, J., Schoonjans, K. (2016) LRH-1-dependent programming of mitochondrial glutamine processing drives liver cancer. *Genes & development* **30**: 1255-1260.

Yang FM, Pan CT, Tsai HM, Chiu TW, Wu ML and Hu MC (2009) Liver receptor homolog-1 localization in the nuclear body is regulated by sumoylation and cAMP signaling in rat granulosa cells. *FEBS J* **276**(2): 425-436.

Ye JC, G., Zaretskaya, I., Cutcutache, I., Rozen, S., Madden, T.L. (2012) Primer-BLAST: A tool to design target-specific primers for polymerase chain reaction. *BMC Bioinformatics* **13**(134).

Zhang Y, Yang Z, Whitby R and Wang L (2011) Regulation of miR-200c by nuclear receptors PPARalpha, LRH-1 and SHP. *Biochemical and biophysical research communications* **416**(1-2): 135-139.

Zhang Z, Chen H, Bai H and Lai L (2007) Molecular dynamics simulations on the oligomer-formation process of the GNNQQNY peptide from yeast prion protein Sup35. *Biophys J* **93**(5): 1484-1492.

Zhou J, Suzuki, T., Kovacic, A., Saito, R., Miki, Y., Ishida, T., Moriya, T., Simpson, E.R., Sasano, H., Clyne, C.D. (2005) Interactions between Prostaglandin E2, Liver Receptor Homologue-1, and Aromatase in Breast Cancer. *Cancer research* **65**(2): 657-663.