Distribution Agreement

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

Signature:

_____ _____

Emma Hope D'Agostino Date

STRUCTURE-FUNCTION STUDIES OF THE NR5A NUCLEAR RECEPTORS
By

Emma Hope D'Agostino
Doctor of Philosophy

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

Eric A. Ortlund, Ph.D.
Advisor

Lawrence Boise, Ph.D.
Committee Member

Graeme L. Conn, Ph.D.
Committee Member

John 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

STRUCTURE-FUNCTION STUDIES OF THE NR5A NUCLEAR RECEPTORS

By

Emma Hope D'Agostino
B.S., B.A., University of North Carolina at Chapel Hill, 2016

Advisor: Eric A. Ortlund, Ph.D

An abstract of
A dissertation submitted to the Faculty of the
James T. Laney School of Graduate Studies of Emory University
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy
in Graduate Division of Biological and Biomedical Science
Biochemistry, Cell, and Developmental Biology
2020

ABSTRACT

STRUCTURE-FUNCTION STUDIES OF THE NR5A NUCLEAR RECEPTORS

By Emma Hope D'Agostino

Nuclear receptors (NRs) comprise 48 ligand-regulated transcription factors. NRs have unique, high-affinity ligands, making them excellent pharmacological targets. Many NRs respond to lipids or lipid metabolites and have proven challenging to target due to the promiscuity and metabolic lability of lipids. This work focuses on two phospholipid (PL)-regulated NRs which together form the human NR5A subfamily: steroidogenic factor-1 (SF-1) and liver receptor homolog-1 (LRH-1). The NR5A receptors regulate development, metabolism, and steroidogenesis, and are putative therapeutic targets for obesity, diabetes, inflammatory bowel diseases, and several cancers. Small molecule screens for NR5As have yielded few results. Recombinantly expressed NR5As co-purify with bacterial PLs that are difficult to displace in compound screens; an easily accessible direct binding assay has not been previously described, complicating efforts to validate and improve lead compounds; and the largely hydrophobic binding pocket provides few residues to anchor a scaffold.

Here, we describe a novel fluorescence polarization competition assay which directly detects ligand binding and quantifies binding affinities. We demonstrate the development of agonists which combine elements of earlier LRH-1 agonists with PLs to drive improvements in efficacy, leading to the first LRH-1 agonist with *in vivo* efficacy in a murine model of colitis. We solve the first synthetic agonist-bound crystal structure of SF-1, giving insight into the mechanism of action of this receptor for future small molecule development. Finally, we show that SF-1, while previously reported to function exclusively as a monomer, can also dimerize *in vitro* and in cells and that the oligomerization state is ligand responsive. Together, these studies significantly progress our ability to study the NR5A receptors and our understanding of their ligand-mediated mechanisms of action.

STRUCTURE-FUNCTION STUDIES OF THE NR5A NUCLEAR RECEPTORS

By

Emma Hope D'Agostino
B.S., B.A., University of North Carolina at Chapel Hill, 2016

Advisor: Eric A. Ortlund, Ph.D.

A dissertation submitted to the Faculty of the
James T. Laney School of Graduate Studies at Emory University
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy
in Biochemistry, Cell, and Developmental Biology
2020

Acknowledgements

To my lab: The one thing I knew before I went to graduate school was that the most important decision I would make was choosing my lab and my mentor. I could not have made a better decision. Emily, I knew the first time we met that I wanted you to train me, and I thank you for taking me on in your final year and showing me how to juggle multiple projects in a hyperorganized fashion. Matt, you have always been available to bounce around ideas, answer the questions I am too embarrassed to ask anyone else, and make everyone laugh, and I'm glad that I wasn't the senior grad student without you for too long. Suzanne, thank you for teaching me so much about how to do science well and for allowing me to come on to your project – I'm really proud of the work we've done together. Cato, I could not have picked a better person to be my first mentee, and it has been a joy watching you grow into a mature scientist. I am so proud of you, and I know that you will go far and succeed in whatever direction you choose. Sam, I am so glad to have had a friend to go through the ups and downs of grad school with and stay after hours musing about our futures. Know that you belong here and wherever you ultimately decide to go. Filipp, thank you for keeping me from breaking just about every machine in the lab at one time or another and always patiently and expertly explaining science to us. Anamika, I have really enjoyed partnering with you on LRH-1 these last few years, and while I am sad I didn't get to follow in your cryoEM footsteps, I learned a lot from watching you pursue a completely unfamiliar set of methods and I have admired your tenacity on this incredibly challenging project. Xu, I have so enjoyed getting to know you more at the end of this journey, whether it's explaining the complex grammatical structures of the English language or the political structures of the American system, and I am glad to call you my friend. Jennifer Colucheese, I don't think I could have done this Ph.D. without you. You have always been there with a walk, a snack, a pun, a video, a good eye roll, solid advice – I am so thankful to have had you

literally by my side through our plant portal these last five years. Don't overwater Reginald, do not buy more Moon Cheese, and don't get too attached to anybody after me.

Eric, when I originally began looking at Emory it was because I wanted to be close to home. Matt mentioned that he had a former postdoc that could be a fit for a mentor, but I wasn't sold on nuclear receptors. Then Brian mentioned this incredible postdoc he'd worked with who was a professor at Emory and the nicest guy ever. I took a second look at the nuclear receptor research, and even though I think you ignored multiple emails from me while I was being recruited, I could tell from everyone I talked to that you would be a fantastic mentor. You have always put my interests first, from building a project that was catered to my future career, to never once looking even slightly annoyed at the many doctor's appointments and medical phone calls I have had over the years, to being excited at all of the non-academic careers I considered (ok, maybe slightly more excited the one time I said I might consider being a PI). Your door is always open, and you are always excited to talk about science, especially if I say the magic words "do you want to come look at a crystal?" You have also always, from the beginning of my time in the lab, fostered an environment of independence and let me make decisions about my projects, and I know that has equipped me well for the future. Thank you doesn't begin to cover the last five years, but: thank you.

Skylar, we say it probably weekly by now, but we wouldn't be here without each other. I don't know what decision of fate dropped us into the same cohort, but I am so glad it happened. From the frustrations of Foundations, to wobbling in our new labs, to running retreat and a few Training Grant symposia, to struggling through stats, to borrowing dogs, adopting cats, a medical emergency or two, and a couple of new diseases, it's been quite a ride. It is safe to say that I did not imagine "finishing my Ph.D. in the middle of a pandemic" would be a box I would ever be able to tick, and I would not be getting out of bed every day if I didn't know that I had our outdoor gatherings to get to. Thank you for always, no questions asked, being there, and for always making sure that we time

our meltdowns to not coincide. Know that you belong in science, in whatever career you want to have, and I am proud of how hard you have fought to get your Ph.D.

We do not simply appear in Ph.D. programs; somebody must have encouraged us along the way. For me, the training I received at the University of North Carolina at Chapel Hill was and continues to be instrumental in my path to my degree. Brian and Kelly, thank you for helping me figure out what I needed to do to get to the career I wanted, and for always welcoming me home with a home-cooked meal and some good backyard chicken watching. Thank you also for always enthusiastically encouraging whatever was best for me, regardless of what anyone else thought. It can be difficult to contemplate going against the "traditional route" of academia, and it means the world for my Carolina parents to be cheering me on. Matt, thank you for taking me on in my first lab experience. Thank you for treating me like more than an undergrad, for giving me independent projects and the opportunity to present at a conference, and for continuing to mentor me long past my time in your lab. I could not have asked for a more positive experience in research as an undergrad and it set me up to succeed at Emory. And finally, Rebecca, I do not think I would be a scientist without you. You have been a close friend and mentor to me for six years. You taught me how to hold a pipette, how to think about science and organize my projects, you took my texts when I started in Eric's lab and didn't want to ask a stupid question, and you have been an incredible source of support again during this pandemic. I know that you will find your dream position, and I am so excited for all the undergrads who will benefit from your mentorship going forward.

To my friends and family, thank you for being excited when the science worked and sad when the proteins weren't cooperating. Thank you for listening to me agonize over future job prospects and helping me think through priorities. Thank you to my barn family for my weekly outside time with the ponies and our regular "barn dinners" that gave me another Atlanta home. Emily, I especially appreciate your genuine interest in understanding what it is I do all day. Claire, thank you for starting

our Book Club and giving me another community – I look forward to another 20 years of our friendship. Bri, Hannah, Alyssa, and Calvin, the connections we formed at Carolina have carried us into our first post-grad years, and I am excited for many more together. Dad, I am so glad to have had so much time with you these last four years with our Thursday night dinners and weekend grilling. Momma, I'm glad these years did not hold as many health adventures as either of us expected, and that we got to relive our feis mom and daughter glory years instead (with a few health adventures because come on, I couldn't resist). I came to Emory in large part so that I could be near you, and I have been glad every day to get to spend every birthday, Thanksgiving, Christmas, Mother's Day, Father's Day, and any other day I wanted with you all.

Table of Contents

Abbreviations

6HP	hexahydropentalene
ACACB	acetyl-CoA carboxylase 2
ACC	adrenocortical carcinoma
ACTH	adrenocorticotropic hormone
AF-H	activation function helix
AFS	activation function surface
CA	carboxylic acid
CHAPS	3-[(3-cholamidopropyl)dimethylammonio]-1propanesulfonic acid
ChoP	phosphorylcholine
DAX-1/NR0B1	dosage-sensitive sex-reversal adrenal hypoplasia congenital critical region on the X chromosome gene 1/ nuclear receptor 0B1
DBD	DNA-binding domain
DCIA	7-diethylamino-3-((4'-(iodoacetyl)amino)phenyl)4- methylcoumarin
DLPC	1,2-dilauroyl-sn-glycero-3-phosphocholine
DLPE	1,2-dilauroyl-sn-glycero-3-phosphoethanolamine
DPPE	dipalmitoylphosphoethanolamine
DSS	dextran sodium sulfate
EMSA	electrophoretic mobility shift assay
FAM	fluoresceinamine
FASN	fatty acid synthase
FP	fluorescence polarization
Ftz-F1	Fushi Tarazu
FRET	fluorescence resonance energy transfer
FSH	follicle-stimulating hormone
FXR	farnesoid X receptor
HDX-MS	hydrogen deuterium exchange-mass spectrometry
IBD	inflammatory bowel diseases
IκBβ	NF-κB inhibitor beta
IPTG	isopropyl-1-thio-d-galactopyranoside
LBD	ligand-binding domain
LBP	ligand-binding pocket
LRH-1	liver receptor homolog-1
MARCoNI	microarray assay for real-time coregulator nuclear receptor interaction
NBD	nitrobenzoxadiazole
NCOR1	nuclear receptor corepressor 1
NGF1-B	nerve growth factor 1-B
NR	nuclear receptor
NR5A	nuclear receptor 5A
PC	phosphatidylcholine
PE	phosphatidylethanolamine
PEI	polyethylenimine
$PI(4,5)P_2$	phosphatidylinositol 4,5-bisphosphate
PL	phospholipid
PMSF	phenylmethylsulphonyl fluoride
PPI	protein-protein interaction

RE	relative efficacy
RXR	retinoid X receptor
SCD1	stearoyl-CoA desaturase-1
SEC	size exclusion chromatography
SF-1	steroidogenic factor-1
SHP/NR0B2	small heterodimer partner/ nuclear receptor 0B2
SREBP1	sterol regulatory element-binding protein 1
T1D	type 1 diabetes
T2D	type 2 diabetes
TCEP	*tris*(2-carboxyethyl) phosphine
TNF-α	tumor necrosis factor-alpha
UC	ulcerative colitis
VMH	ventromedial nucleus of the medial hypothalamus

CHAPTER 1: THE NR5A NUCLEAR RECEPTORS

Introduction

The nuclear receptor (NR) superfamily is one of the largest superfamilies of human receptors, comprising 48 members which control diverse functions including development, steroidogenesis, and metabolism. These transcription factors bind to their target DNA response elements (REs) to control gene expression in response to binding of small lipophilic ligands. The specificity of NRs for both their REs and ligands make them excellent drug targets. Drug discovery efforts have yielded FDA-approved therapies for 17 out of 48 NRs,[1] with the focus largely on the steroid and retinoid receptors. The untargeted receptors are either orphans, with as-yet unidentified ligands, or bind to lipids or lipid metabolites. Lipid-binding NRs are more difficult to target, and there has only been one therapeutic class developed which targets lipid-binding NRs thus far, the peroxisome proliferator-activated receptor subfamily (Table 1.1). Two phospholipid-regulated receptors which show promise as drug targets are steroidogenic factor-1 (SF-1; NR5A1) and liver receptor homolog-1 (LRH-1; NR5A2), which together compose the human NR5A subfamily.

NR5A Genes

SF-1 is located on chromosome 9 and is composed of seven exons.[2] There is one isoform of SF-1 in humans, although secondary isoforms are present in other species including rat and mouse.[2] LRH-1 is located on chromosome 1 and composed of eight exons.[3] There are three isoforms, and the primary isoform lacks the canonical A/B N-terminal region seen in most nuclear receptors (see below).

NR5A Structure

Like most NRs, the NR5As have a modular domain composition comprising an N-terminal DNA-binding domain (DBD) connected by a flexible hinge region to a C-terminal ligand-binding domain (LBD) (Figure 1.1).[3] SF-1, like the primary isoform of LRH-1, lacks the N-terminal A/B

region, termed the activation function-1 surface, which is involved in ligand-independent activation in other NRs. The NR5A DBDs contain two zinc finger motifs that flank a DNA-reading helix that positions in the DNA major groove. The NR5As have a unique addendum to the DBD, termed the Fushi Tarazu (Ftz-F1) region. This short region C-terminal to the DBD binds three base pairs of DNA 5' to the canonical six base pair NR RE allowing the NR5As to recognize nine base pairs of DNA as monomers, rather than binding repeating units of six base pairs as dimers like most other NRs.[4] The NR5As bind to the canonical response element YCAAGGYCR, where Y is any pyrimidine and R is any purine.[4] The LBD demonstrates the canonical NR fold, consisting of 12 alpha helices and two beta sheets between helices five and six, and contains a ligand-binding pocket and the activation function surface (AFS). The AFS comprises helices three, four, and the activation function helix (AF-H), and serves to bind coregulator proteins, shown as green fragments in Figure 1.1.

Expression

The NR5As exhibit distinct, largely nonoverlapping expression patterns, coexisting primarily in the ovary. SF-1 is expressed in the adrenal glands, the theca cells of the ovary, the Leydig and Sertoli cells of the testes, and the ventromedial nucleus of the hypothalamus (VMH).[5] LRH-1 is expressed in the liver, pancreas, colon, pre-adipocytes, the granulosa cells and corpora lutea of the ovary, and the testes.[3, 6] Both receptors are also expressed early in the developing embryo, with SF-1 in the urogenital ridge at day 9.5 and LRH-1 in multiple tissues as early as day 6.5.[7, 8]

Biological Roles/ Function

The primary role of SF-1 is as a master regulator of steroidogenesis in the adrenal glands and gonads. SF-1 regulates several aspects of the steroid synthesis process, including: the expression of every adrenal and gonadal steroidogenic gene; the uptake, transport, and *de novo* synthesis of cholesterol to supply steroidogenesis; and the expression of receptors for adrenocorticotropic

hormone (ACTH) and follicle-stimulating hormone (FSH) to stimulate adrenal and gonadal steroidogenesis.[9] SF-1 is also necessary for proper development of the adrenal glands, gonads, and hypothalamus.[7] Complete loss of SF-1 is lethal at birth due to the inability to synthesize glucocorticoids and mineralocorticoids, though exogenously-delivered steroids allow survival.[9] SF-1 mutations in humans lead to congenital adrenal hypoplasia, characterized by adrenal insufficiency and male to female sex reversal.[9] In the VMH, SF-1 regulates metabolism; the VMH is critical for maintaining energy homeostasis. The precise role of SF-1 is unclear in the VMH, but loss of SF-1 in this region leads to obesity due to increased food intake, decreased energy expenditure, and impaired glucose tolerance and insulin sensitivity.[10]

The primary role of LRH-1 is as a master metabolic regulator in the liver, pancreas, and colon. LRH-1 controls lipid, bile acid, and glucose homeostasis by regulating *de novo* lipogenesis, bile acid synthesis, glucose uptake, and reverse cholesterol transport into the liver.[3, 11] Like SF-1, LRH-1 controls steroidogenesis in the ovaries, as well as the breast preadipocytes. LRH-1 is required for ovulation and progesterone production in the ovaries through its regulation of the genes for the steroidogenic acute regulatory protein (*Star*), scavenger receptor B1(*Scarb1*), and cytochrome P450 side chain cleavage enzyme (*Cyp11a1*), and activates transcription of aromatase in breast preadipocytes, leading to estradiol synthesis.[12] Finally, LRH-1 controls pluripotency during development, directly inducing *Oct4* transcription, a key pluripotency regulator, and loss of LRH-1 is embryonic lethal.[8]

Together, these receptors have distinct but complementary roles in steroidogenesis, development, and metabolism. More work remains to fully understand the mechanism of some of these roles, such as the precise genomic targets of SF-1 in the VMH. The diverse roles of these receptors, particularly as regulators of metabolism and steroidogenesis, have made them attractive pharmacological targets.

Putative Ligands

The NR5As are adopted orphan receptors and have been shown through extensive study to bind various phospholipids (PL) (summarized in Figure 1.2). Initial biochemical and crystallographic studies identified co-purified bacterial PL in the binding pockets of these receptors, including phosphatidylglycerol and phosphatidylethanolamine, leading to the suggestion that PLs were the native ligand.[13-15] Both NR5As have been shown to bind directly through *in vitro* assays and X-ray crystallography to phosphatidylinositol species including PIP_2 and PIP_3, though activation in cells has not been shown.[13, 16, 17] LRH-1 and SF-1 bind directly to and are activated in cells by medium-chain, saturated phosphatidylcholine (PC) species, and SF-1 has also been crystallized with 1-palmitoyl-2-steroyl-sn-glycero-3phosphocholine (16:0-18:0 PC).[5, 18-20] LRH-1 is activated in mice by one of these species, 1,2-dilauroyl-sn-glycero-3-phosphocholine (DLPC). These PC species are dietary and present in only small amounts in mammalian cells; thus, they may not be the "true" native ligand. Studies to identify native activating ligands are lacking, and the specific endogenous PL ligands for these receptors have not yet been identified. However, based on the available biochemical and structural information, it is believed that the native ligands are PL species.

NR5As in Disease

The NR5As hold therapeutic promise in a number of disease areas, including several cancers, metabolic, and inflammatory diseases.

Cancer

SF-1 has been identified as a likely oncogene in adrenocortical carcinoma (ACC), an aggressive cancer of the adrenal cortex. ACC is a rare and deadly cancer, with 0.5 to 2 cases per million per year and 5-year survival rates ranging from 16-40%.[21] The 9q34 chromosomal region, which includes SF-1, is frequently amplified in adenomas, particularly pediatric adenomas.[21] The amplification and subsequent upregulation of SF-1 leads to increased proliferation in an ACC cell model.[22] Moreover,

SF-1 is an accurate diagnostic marker of ACC, and the level of expression of SF-1 protein in ACC tumors correlates with prognosis.[23]

LRH-1 has been identified as a target of interest in several cancers, primarily of the breast, colon, and pancreas. Perhaps the clearest role for LRH-1 is in the exacerbation of estrogen receptor positive (ER+) breast cancer. LRH-1 1) increases estrogen ligand availability to ER by increasing aromatase expression in breast preadipocytes adjoining the breast tumor;[24] 2) participates in a transcriptional feedforward loop with ER, as they are transcriptional targets of each other;[25, 26] and 3) promotes cleavage of E-cadherin, leading to increased migration and invasion of tumor cells.[27] In colon cancer, LRH-1 is involved in an entirely separate pathway, in which it cooperates with beta-catenin to drive tumor progression. These two transcription factors coactivate each other to initiate transcription of cyclins D1 and E1 and c-Myc, promoting proliferation.[28] Reducing LRH-1 levels in an LRH-1 haplosufficiency murine colorectal cancer model and an in-cell knockdown reduced tumor burden and cell proliferation, respectively.[28, 29] Finally, LRH-1's role in pancreatic cancer is somewhat unclear. A genome-wide association study identified the LRH-1 gene as a pancreatic cancer susceptibility locus, though the functional effects of the single nucleotide polymorphisms were not investigated.[30] Subsequent studies revealed LRH-1 to be upregulated in pancreatic ductal adenocarcinomas and pancreatic cancer cell lines, and silencing LRH-1 in these cancer cell lines reduced cell proliferation.[31] Somewhat paradoxically, loss of LRH-1 in a heterozygous mouse model sensitizes the pancreas to damage and creates a pre-inflammatory state similar to early pancreatitis, suggesting that LRH-1 protects against the development of pancreatic cancer.[32] Further work is needed to delineate the particular role of LRH-1 in the development and progression of pancreatic cancer.

Metabolic Disorders

The NR5A receptors are master regulators of metabolism, SF-1 through the VMH and LRH-1 through the liver. This level of control has made them targets of interest for the treatment of obesity and type II diabetes. Complete SF-1 knockout mice (rescued from steroid deficiency with exogenous corticosteroid delivery and adrenal gland transplants) display late-onset obesity, weighing nearly twice as much as their wildtype littermates by eight weeks of age due to decreased activity.[33] A VMH-specific SF-1 knockout mouse developed similar late-onset obesity due to both decreased activity and increased food intake.[10] The VMH SF-1 knockout mice also demonstrated impaired glucose tolerance and leptin and insulin sensitivity compared to their wildtype counterparts, implicating SF-1 in metabolic hormone regulation. These studies suggest that activating SF-1 in the VMH may protect against the development of obesity and type 2 diabetes (T2D); currently, there is no available agonist which could target SF-1 in the VMH. LRH-1 regulates glucose, cholesterol, and lipid metabolism in the liver, and controls several pathways which may be beneficial in the setting of obesity. LRH-1 regulates glucose through direct transcriptional regulation of glucokinase and increases glucose uptake into the cell through the GLUT4 transporter.[11, 34] It promotes reverse cholesterol transport through the Scarb1 transporter;[35] controls bile acid synthesis through a feedback loop with the nuclear receptors farsenoid X receptor (FXR) and small heterodimer partner (SHP);[36] and regulates *de novo* lipogenesis.[18] Combined, these effects may combat T2D and obesity. In an obese mouse model, activation of LRH-1 with a PL agonist improved glucose homeostasis and reduced *de novo* lipogenesis. Further work is needed to confirm the viability of these targets in metabolic disease.

Inflammation/ Autoimmune Disease

LRH-1 has recently drawn attention for its potential as a target for inflammatory/ autoimmune diseases including type I diabetes (T1D) and inflammatory bowel diseases (IBD). LRH-1 is both anti-inflammatory and pro-regeneration, and its activation in these disease settings may therefore be

beneficial. In T1D, activation of LRH-1 with a synthetic agonist reduced islet apoptosis in cells and significantly decreased and/ or reversed the development of diabetes in both chemical (streptozotocin) and autoimmune (RIP-B7.1 mice) murine models of T1D.[37] These effects were achieved via increased alpha to beta cell differentiation and promotion of an anti-inflammatory M2 macrophage phenotype. In IBD, LRH-1 combats inflammation in the colon by locally increasing cortisol production through upregulation of *CYP11A1* and *CYP11B1* transcription.[38] Additionally, LRH-1 promotes intestinal stem cell renewal through upregulation of Wnt/B-catenin signaling.[28] Overexpression of human LRH-1 in mice in an immune model of colitis improved disease severity and decreased inflammatory markers, and activation of LRH-1 in murine intestinal organoids with a synthetic agonist promoted steroidogenic and anti-inflammatory gene expression.[39, 40] These data provide an exciting foundation for future studies of LRH-1 in inflammatory disease.

Synthetic Modulation of NR5A Receptors

Given the breadth of potential therapeutic applications of NR5A antagonists and agonists for cancer, metabolic, and inflammatory diseases, there has been considerable interest in the development of synthetic modulators for this receptor subclass. Synthetic modulators are preferred over native PL ligands because PLs are rapidly metabolized, easily remodeled, and unlikely to be specific for a particular receptor. However, due to the overwhelmingly hydrophobic ligand-binding pockets (LBPs) of the NR5As, it has proved difficult to identify and progress synthetic modulators. The modulators which have been described to date are summarized in Figure 1.3.

Antagonists

Compounds to decrease NR5A activity are sought for the treatment of cancer. To date, five antagonist classes have been identified: three benzothiophene LRH-1 antagonists;[41] AC-45594, an SF-1 inverse agonist;[42] SID7965943 and SID7970631, SF-1-specific isoquinolinone antagonists;[43] SR1848, an LRH-1/ SF-1 dual antagonist;[44] and Cpd3, an LRH-1 specific antagonist.[45] While

SID7965943, SID7970631, and Cpd3 have been shown to bind directly to the NR5As, SR1848 utilizes an unknown mechanism to antagonize LRH-1 that does not involve binding in the LBP (the benzothiophenes and AC-45594 have not been extensively characterized).[44, 46] There is no structural information available for these antagonists and as such, it has not yet been possible to establish structure-activity relationships to improve the affinity or potency of these compounds. None have been tested in animal models, perhaps due to the relatively low (micromolar) affinity and potency of all but the isoquinolinone classes. It remains to be seen whether a more effective antagonist scaffold may exist for the NR5As.

Agonists

Agonists are sought to increase NR5A activity for the treatment of obesity, diabetes, and inflammatory disease. As with antagonists, discovering agonist scaffolds has proven challenging due to the hydrophobic LBPs of these receptors. To date, only one synthetic agonist scaffold has been reported, a hexahydropentalene scaffold with the original compound name of GSK8470, which is active towards both NR5As.[47, 48] One derivative of the original parent compound, termed BL001/ 5f, was used in the aforementioned T1D mouse models and successfully prevented and/ or reversed T1D.[37, 47] However, in general this compound class has required extensive modification for use as a synthetic agonist. The parent compound GSK8470 was acid labile and thus derivatized to the more stable RJW100.[48] RJW100 showed only modest in-cell activity and *in vitro* affinity for the NR5As. An extensive exploration of the RJW100 scaffold using biochemical, structural, and *in vivo* methods yielded the most potent LRH-1 agonists to date, with 1000-fold improved affinity and potency compared to RJW100.[40, 49] The best compound in this class shows activity in mice in a model of colitis (unpublished, Chapter 3).

It remains to be seen whether a similar highly effective compound can be designed to target SF-1. While there are now several X-ray crystal structures of LRH-1 in complex with small molecules,

until recently there were no structures solved of SF-1 in complex with any small molecules, which has hindered small molecule development for this receptor. The first structure of SF-1 in complex with a hexahydropentalene synthetic agonist is described in **Chapter 4** (unpublished).

Questions Addressed in This Work

The major question addressed in this work is how small molecules can be designed to modulate nuclear receptor activity and control gene expression. The use of X-ray crystallography to guide small molecule design has proven to be a powerful means towards generating highly potent compounds which strongly activate LRH-1 in particular. In **Chapter 2** we describe a novel fluorescence polarization competition binding assay which enabled, for the first time, the rapid quantification of small molecule binding to the NR5A receptors. This assay led to more efficient optimization of our chemical library and aided in the detection of high-affinity compounds in our pharmaceutical development program. We describe a subset of these high-affinity compounds, named the "phospholipid mimetics," in **Chapter 3**, which were designed to capture the best aspects of the hexahydropentalene agonists described above and PLs. This work contains the first small molecule which demonstrates *in vivo* activation of LRH-1 in the context of inflammatory bowel disease and lessens disease severity. In **Chapter 4** we report the first crystal structure of SF-1 bound to a small molecule agonist, 6N-10CA, developed by our lab. We reveal that 6N-10CA makes an extensive hydrogen bond network deep in the SF-1 binding pocket; further work is ongoing to determine the mechanism of action of this agonist. In **Chapter 5**, we detail the preliminary findings by our lab that SF-1 may not only function as a monomer, but also as a dimer. Finally, in **Chapter 6** we place this work within the NR field and propose ideas for future studies.

Figures and Tables

NRNC Symbol	Receptor	Native Ligand[50-52]	Approved Therapy[1]	Application
NR0B1	Dosage-sensitive sex reversal, adrenal hypoplasia critical region, on chromosome X, gene 1	Orphan	None	
NR0B2	Small heterodimer partner	Retinoids?	None	
NR1A1	Thyroid hormone receptor alpha	Thyroid hormone	Liothyronine, levothyroxine	Hormone replacement
NR1A2	Thyroid hormone receptor beta	Thyroid hormone	Liothyronine, levothyroxine	Hormone replacement
NR1B1	Retinoic acid receptor alpha	Retinoic acid, arotinoid acid	Adapalene; alitretinoin & tamibarotene; isotretinoin, tazarotene & acitretin	Dermatologic; antineoplastic; keratolytic
NR1B2	Retinoic acid receptor beta	Retinoic acid, arotinoid acid	Alitretinoin; adapalene; tazarotene	Antineoplastic; dermatologic; keratolytic
NR1B3	Retinoic acid receptor gamma	Retinoic acid, arotinoid acid	Alitretinoin; adapalene; tazarotene	Antineoplastic; dermatologic; keratolytic
NR1C1	Peroxisome proliferator-activated receptor alpha	Pirinixic acid, palmitic acid	Clofibrate; fenofibrate, gemfibrozil	Antichloesteremic; antilipemic
NR1C2	Peroxisome proliferator-activated receptor beta/delta	Fatty acids, prostaglandins	none	
NR1C3	Peroxisome proliferator-activated receptor gamma	Fatty acids, prostaglandins	Pioglitazone, rosiglitazone; treoprostinil	Hypoglycemic; antihypertensive
NR1D2	Rev-ErbA-alpha	Heme	None	
NR1D3	Rev-ErbA-beta	Heme	None	
NR1F1	RAR-related orphan-receptor-alpha	Cholesterol, ATRA	None	
NR1F2	RAR-related orphan-receptor-beta	Cholesterol, ATRA	None	
NR1F3	RAR-related orphan-receptor-gamma	Cholesterol, ATRA	None	
NR1H2	Liver X receptor-beta	Oxysterols	None	
NR1H3	Liver X receptor-alpha	Oxysterols	None	
NR1H4	Farnesoid X receptor	Bile acids	None	
NR1I1	Vitamin D receptor	Vitamin D	Calcitriol; dihydrotachysterol	Antiosteoporosis; anti-migraine
NR1I2	Pregnane X receptor	Xenobiotics		
NR1I3	Constitutive androstane receptor	Androstanol, phenobarbitol	None	
NR2A1	Hepatocyte nuclear receptor-4-alpha	Fatty acids?	None	
NR2A2	Hepatocyte nuclear receptor-4-gamma	Fatty acids?	None	
NR2B1	Retinoic X receptor alpha	Retinoids	Alitretinoin, bexarotene; adapalene	Antineoplastic; dermatologic
NR2B2	Retinoic X receptor beta	Retinoids	Alitretinoin, bexarotene; adapalene	Antineoplastic; dermatologic

NR2B3	Retinoic X receptor gamma	Retinoids	Alitretinoin, bexarotene; adapalene	Antineoplastic; dermatologic
NR2C1	Testicular receptor 2	Orphan	None	
NR2C2	Testicular receptor 4	Fatty acids, polyunsaturated fatty acids	None	
NR2E1	Tailless	Retinoids?	None	
NR2E3	Photoreceptor cell-specific nuclear receptor	Retinoids?	None	
NR2F1	Chicken ovalbumin upstream promoter-transcription factor I	Orphan	None	
NR2F2	Chicken ovalbumin upstream promoter-transcription factor II	Retinoids?	None	
NR2F6	Chicken ovalbumin upstream promoter-transcription factor III	Orphan	None	
NR3A1	Estrogen receptor alpha	Estrogens	Chlorotrianisene, estradiol, estrogens, dienestrol, diethylstilbestrol, raloxifene; clomifene; danazol, fulvestrant, tamoxifen; desogestrel, norgestrel, progesterone	Hormone replacement; fertility; antineoplastic; contraceptive
NR3A2	Estrogen receptor beta	Estrogens	Estradiol, raloxifene; tamoxifen	Hormone replacement; antineoplastic
NR3B1	Estrogen-related receptor-alpha	Orphan	None	
NR3B2	Estrogen-related receptor-beta	Orphan	None	
NR3B3	Estrogen-related receptor-gamma	Orphan	None	
NR3C1	Glucocorticoid Receptor	Cortisol	Betamethasone, budesonide, dexamethasone, difluprednante, flunisolide, flumethasone, hydrocortamate, amcinonide, methylprednisolone, prednisone	Anti-inflammatory, corticosteroids, glucocorticoids, anti-allergy, anti-asthma
NR3C2	Mineralocorticoid receptor	Aldosterone	Fludrocortisone; desoxycorticosterone pivalate; sprinolactone	Anti-inflammatory; hormone replacement; anti-hypertensive
NR3C3	Progesterone receptor	Progesterone	Drospirenone, etonogestrel, levonorgestrel, progesterone; dydrogesterone; nilutamid; ulipristsl	Contraceptive; antidysmennorheal; antineoplastic; abortificant

NR3C4	Androgen Receptor	testosterone	Bicalutamide, dromostanolone, fluoxymesterone, flutamide, nilutamide; cyproterone; nandrolone; oxandrolone; testestosterone	Antineoplastic; antihirsutism; antianemic; anabolic; hormone replacement
NR4A1	Nerve growth factor 1B	Orphan	None	
NR4A2	Nuclear receptor related 1	Orphan	None	
NR4A3	Neuron-derived orphan receptor 1	Orphan	None	
NR5A1	Steroidogenic factor 1	Phospholipids	None	
NR5A2	Liver receptor homolog 1	Phospholipids	None	
NR6A1	Germ cell nuclear factor	Orphan	None	

Table 1.1. Nuclear receptor family therapeutic landscape. Of the 48 human NRs, FDA-approved therapies are available for 17 receptors, 11 remain orphans for which native ligands have not been identified, and 20 have identified ligands but have not yet been targeted in the clinic.

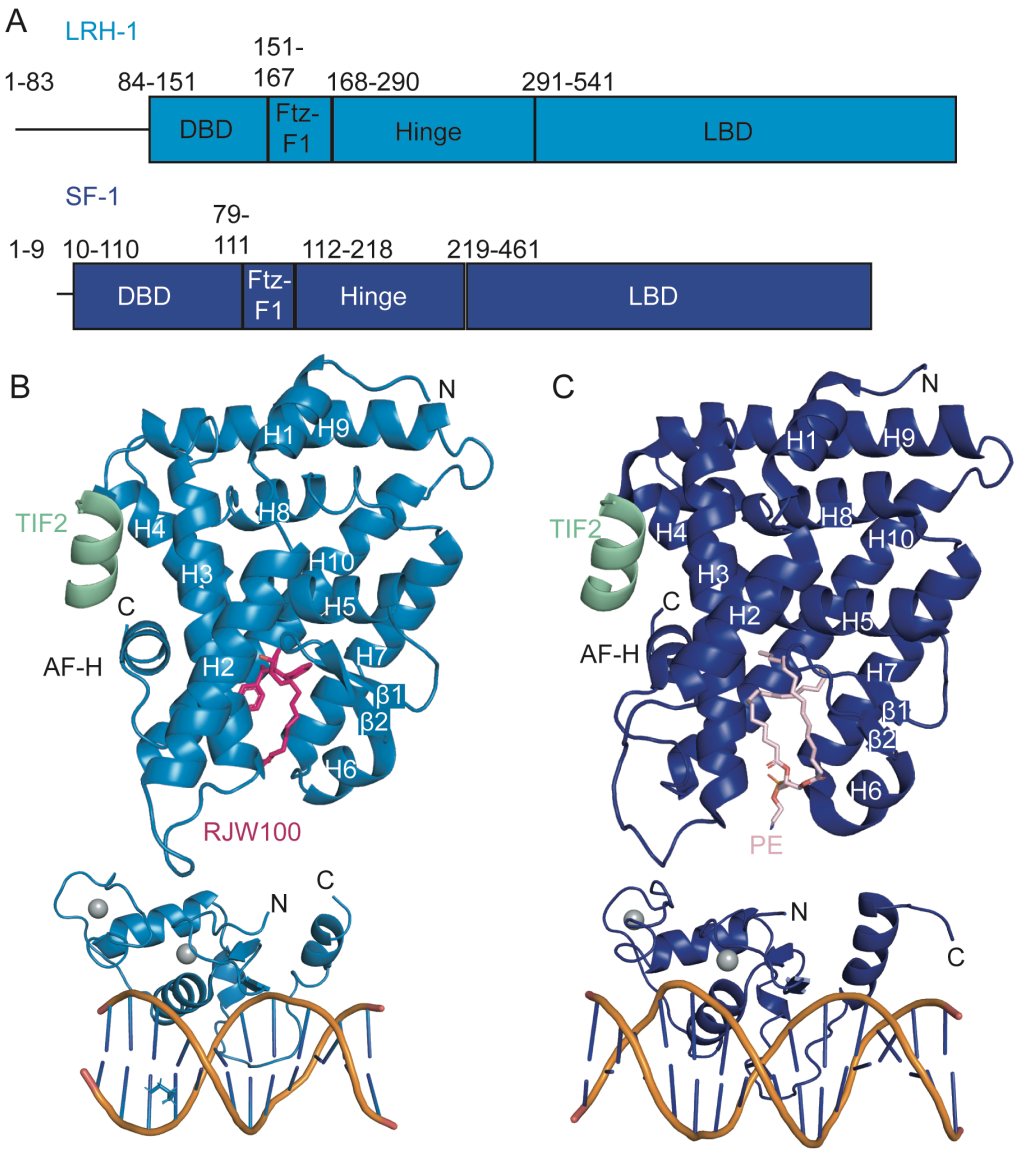


Figure 1.1. NR5A structure. (A) Domain architectures of LRH-1 (top, light blue) and SF-1 (bottom, dark blue). Each contains a DNA-binding domain (DBD), the Ftz-F1 box, a flexible hinge region, and a ligand-binding domain (LBD). LRH-1 is shown as isoform 2, which is the predominant isoform and does not contain the activation function surface at the N-terminus. (B) LRH-1 structure from ligand-bound LBD (PDB: 5L11), with ligand RJW100 shown in pink, and DBD (PDB: 5L0M) bound to *Oct4* DNA element. (C) SF-1 structure from phosphatidylethanolamine (PE)-bound LBD (PDB: 1ZDT) and inhibin-alpha-bound DBD (PDB: 2FF0). Full-length structures are unavailable for these receptors; thus, the hinge regions are not shown.

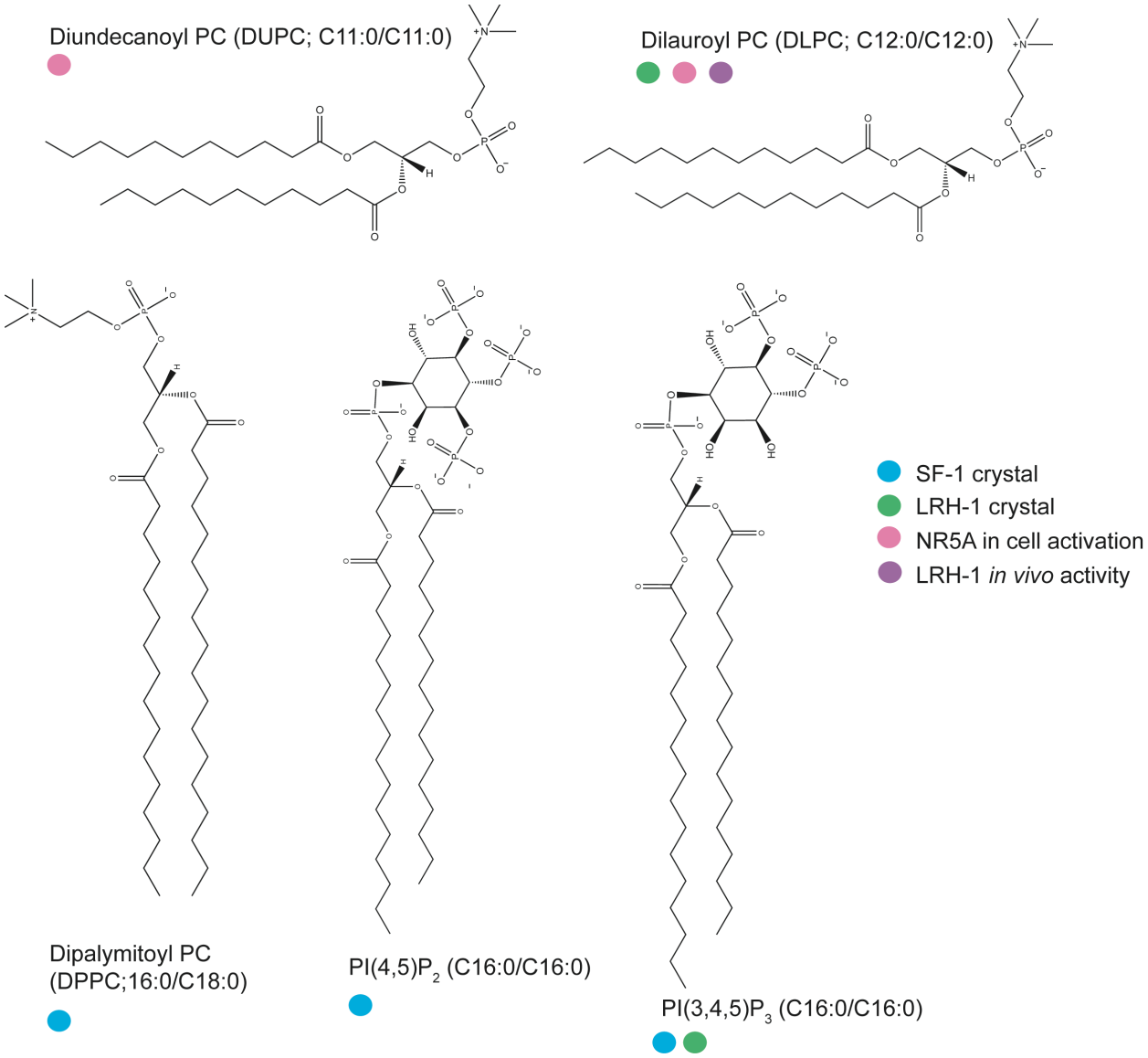


Figure 1.2. NR5A Mammalian Phospholipid Ligands. The NR5A receptors have been characterized with multiple phosphatidylcholine and phosphatidylinositol species *in vitro*, in cells, and *in vivo*. Shown are the PLs which have been characterized to date; colored dots indicate whether each species has been crystallized with each receptor (SF-1, blue; LRH-1, green), shown to activate the receptors in cells (pink), or shown to activate LRH-1 *in vivo* (purple).

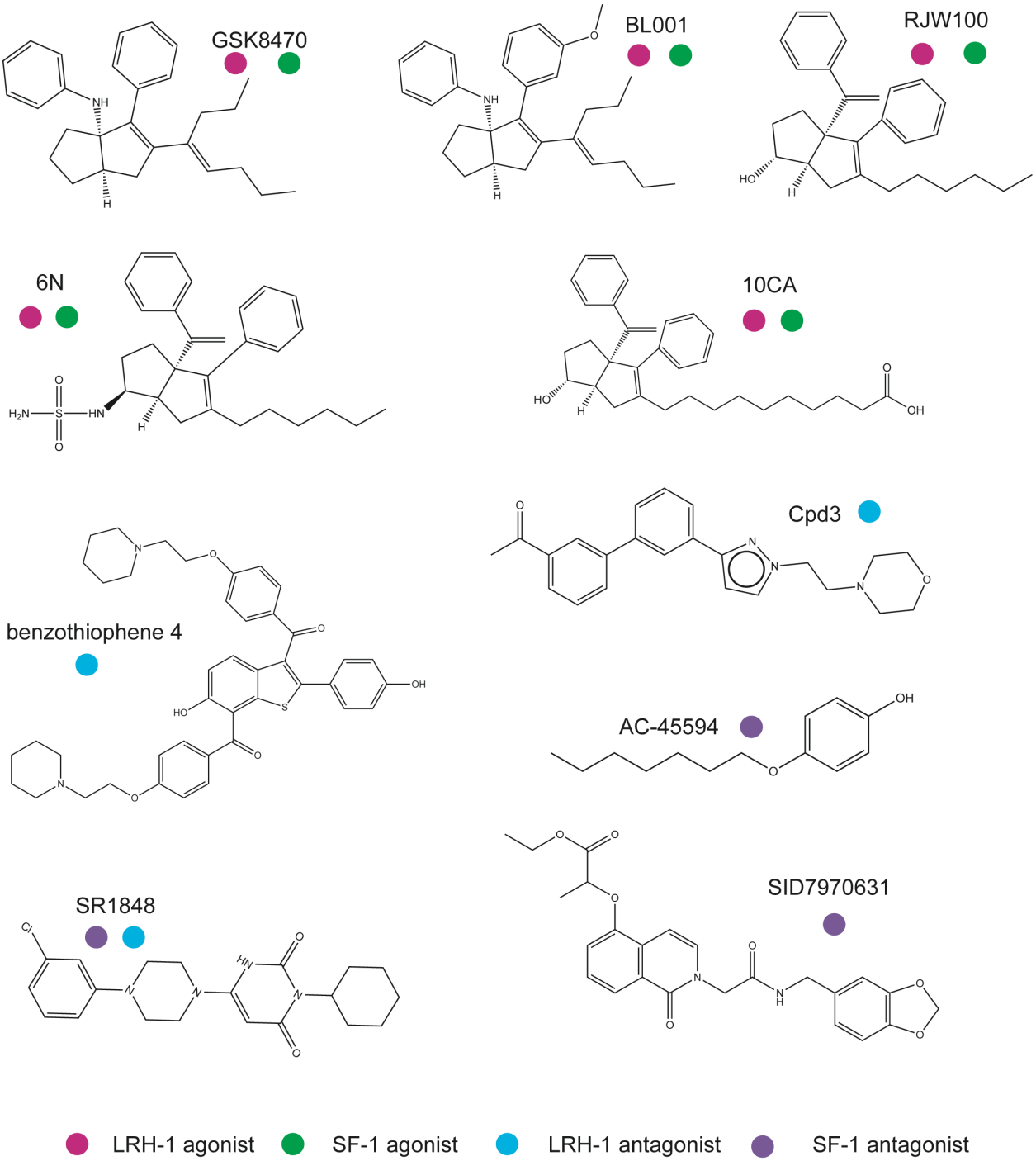


Figure 1.3. NR5A synthetic modulators. Shown are significant synthetic modulators which have been reported for the NR5A subfamily. *Top*, agonists, *bottom*, antagonists. Colored dots specify whether each compound has been validated for LRH-1, SF-1, or both.

References

1. Rask-Andersen M, Almen MS, Schioth HB. Trends in the exploitation of novel drug targets. Nat Rev Drug Discov. 2011;10(8):579-90. Epub 2011/08/02. doi: 10.1038/nrd3478. PubMed PMID: 21804595.
2. Nawata H, Yanase T, Oba K, Ichino I, Saito M, Goto K, Ikuyama S, Sakai H, Takayanagi R. Human Ad4BP/SF-1 and its related nuclear receptor. J Steroid Biochem Mol Biol. 1999;69(1-6):323-8. Epub 1999/07/27. doi: 10.1016/s0960-0760(99)00081-3. PubMed PMID: 10419009.
3. Fayard E, Auwerx J, Schoonjans K. LRH-1: an orphan nuclear receptor involved in development, metabolism and steroidogenesis. Trends Cell Biol. 2004;14(5):250-60. Epub 2004/05/08. doi: 10.1016/j.tcb.2004.03.008. PubMed PMID: 15130581.
4. Wilson TE, Fahrner TJ, Milbrandt J. The orphan receptors NGFI-B and steroidogenic factor 1 establish monomer binding as a third paradigm of nuclear receptor-DNA interaction. Mol Cell Biol. 1993;13(9):5794-804. Epub 1993/09/01. doi: 10.1128/mcb.13.9.5794. PubMed PMID: 8395013; PMCID: PMC360322.
5. Meinsohn MC, Smith OE, Bertolin K, Murphy BD. The Orphan Nuclear Receptors Steroidogenic Factor-1 and Liver Receptor Homolog-1: Structure, Regulation, and Essential Roles in Mammalian Reproduction. Physiol Rev. 2019;99(2):1249-79. Epub 2019/02/28. doi: 10.1152/physrev.00019.2018. PubMed PMID: 30810078.
6. Lazarus KA, Wijayakumara D, Chand AL, Simpson ER, Clyne CD. Therapeutic potential of Liver Receptor Homolog-1 modulators. J Steroid Biochem Mol Biol. 2012;130(3-5):138-46. Epub 2012/01/24. doi: 10.1016/j.jsbmb.2011.12.017. PubMed PMID: 22266285.
7. Parker KL, Schimmer BP. Steroidogenic factor 1: a key determinant of endocrine development and function. Endocr Rev. 1997;18(3):361-77. Epub 1997/06/01. doi: 10.1210/edrv.18.3.0301. PubMed PMID: 9183568.
8. Gu P, Goodwin B, Chung AC, Xu X, Wheeler DA, Price RR, Galardi C, Peng L, Latour AM, Koller BH, Gossen J, Kliewer SA, Cooney AJ. Orphan nuclear receptor LRH-1 is required to maintain Oct4 expression at the epiblast stage of embryonic development. Mol Cell Biol. 2005;25(9):3492-505. Epub 2005/04/16. doi: 10.1128/MCB.25.9.3492-3505.2005. PubMed PMID: 15831456; PMCID: PMC1084298.
9. Val P, Lefrancois-Martinez AM, Veyssiere G, Martinez A. SF-1 a key player in the development and differentiation of steroidogenic tissues. Nucl Recept. 2003;1(1):8. Epub 2003/11/05. doi: 10.1186/1478-1336-1-8. PubMed PMID: 14594453; PMCID: PMC240021.
10. Kinyua AW, Yang DJ, Chang I, Kim KW. Steroidogenic Factor 1 in the Ventromedial Nucleus of the Hypothalamus Regulates Age-Dependent Obesity. PLoS One. 2016;11(9):e0162352. Epub 2016/09/07. doi: 10.1371/journal.pone.0162352. PubMed PMID: 27598259; PMCID: PMC5012732.
11. Oosterveer MH, Mataki C, Yamamoto H, Harach T, Moullan N, van Dijk TH, Ayuso E, Bosch F, Postic C, Groen AK, Auwerx J, Schoonjans K. LRH-1-dependent glucose sensing determines intermediary metabolism in liver. J Clin Invest. 2012;122(8):2817-26. Epub 2012/07/10. doi: 10.1172/JCI62368. PubMed PMID: 22772466; PMCID: PMC3408738.
12. Duggavathi R, Volle DH, Mataki C, Antal MC, Messaddeq N, Auwerx J, Murphy BD, Schoonjans K. Liver receptor homolog 1 is essential for ovulation. Genes Dev. 2008;22(14):1871-6. Epub 2008/07/17. doi: 10.1101/gad.472008. PubMed PMID: 18628394; PMCID: PMC2492734.
13. 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, Ingraham HA. Structural analyses reveal phosphatidyl inositols as ligands

for the NR5 orphan receptors SF-1 and LRH-1. Cell. 2005;120(3):343-55. Epub 2005/02/15. doi: 10.1016/j.cell.2005.01.024. PubMed PMID: 15707893.
14. Wang W, Zhang C, Marimuthu A, Krupka HI, Tabrizizad M, Shelloe R, Mehra U, Eng K, Nguyen H, Settachatgul C, Powell B, Milburn MV, West BL. The crystal structures of human steroidogenic factor-1 and liver receptor homologue-1. Proc Natl Acad Sci U S A. 2005;102(21):7505-10. Epub 2005/05/18. doi: 10.1073/pnas.0409482102. PubMed PMID: 15897460; PMCID: PMC1140416.
15. Ortlund EA, Lee Y, Solomon IH, Hager JM, Safi R, Choi Y, Guan Z, Tripathy A, Raetz CR, McDonnell DP, Moore DD, Redinbo MR. Modulation of human nuclear receptor LRH-1 activity by phospholipids and SHP. Nat Struct Mol Biol. 2005;12(4):357-63. Epub 2005/02/22. doi: 10.1038/nsmb910. PubMed PMID: 15723037.
16. Sablin EP, Blind RD, Uthayaruban R, Chiu HJ, Deacon AM, Das D, Ingraham HA, Fletterick RJ. Structure of Liver Receptor Homolog-1 (NR5A2) with PIP3 hormone bound in the ligand binding pocket. J Struct Biol. 2015;192(3):342-8. Epub 2015/09/30. doi: 10.1016/j.jsb.2015.09.012. PubMed PMID: 26416531; PMCID: PMC4651778.
17. Blind RD, Sablin EP, Kuchenbecker KM, Chiu HJ, Deacon AM, Das D, Fletterick RJ, Ingraham HA. The signaling phospholipid PIP3 creates a new interaction surface on the nuclear receptor SF-1. Proc Natl Acad Sci U S A. 2014;111(42):15054-9. Epub 2014/10/08. doi: 10.1073/pnas.1416740111. PubMed PMID: 25288771; PMCID: PMC4210282.
18. Lee JM, Lee YK, Mamrosh JL, Busby SA, Griffin PR, Pathak MC, Ortlund EA, Moore DD. A nuclear-receptor-dependent phosphatidylcholine pathway with antidiabetic effects. Nature. 2011;474(7352):506-10. Epub 2011/05/27. doi: 10.1038/nature10111. PubMed PMID: 21614002; PMCID: PMC3150801.
19. Musille PM, Pathak M, Lauer JL, Hudson WH, Griffin PR, Ortlund EA. Antidiabetic phospholipid-nuclear receptor complex reveals the mechanism for phospholipid-driven gene regulation. Nat Struct Mol Biol. 2012;19(5):532-S2. Epub 2012/04/17. doi: 10.1038/nsmb.2279. PubMed PMID: 22504882; PMCID: PMC3960984.
20. Sablin EP, Blind RD, Krylova IN, Ingraham JG, Cai F, Williams JD, Fletterick RJ, Ingraham HA. Structure of SF-1 bound by different phospholipids: evidence for regulatory ligands. Mol Endocrinol. 2009;23(1):25-34. Epub 2008/11/08. doi: 10.1210/me.2007-0508. PubMed PMID: 18988706; PMCID: PMC2646595.
21. Creemers SG, Hofland LJ, Korpershoek E, Franssen GJ, van Kemenade FJ, de Herder WW, Feelders RA. Future directions in the diagnosis and medical treatment of adrenocortical carcinoma. Endocr Relat Cancer. 2016;23(1):R43-69. Epub 2015/10/18. doi: 10.1530/ERC-15-0452. PubMed PMID: 26475053.
22. Lalli E. Adrenocortical development and cancer: focus on SF-1. J Mol Endocrinol. 2010;44(6):301-7. Epub 2010/03/05. doi: 10.1677/JME-09-0143. PubMed PMID: 20200142.
23. Duregon E, Volante M, Giorcelli J, Terzolo M, Lalli E, Papotti M. Diagnostic and prognostic role of steroidogenic factor 1 in adrenocortical carcinoma: a validation study focusing on clinical and pathologic correlates. Hum Pathol. 2013;44(5):822-8. Epub 2012/11/20. doi: 10.1016/j.humpath.2012.07.025. PubMed PMID: 23158211.
24. Clyne CD, Speed CJ, Zhou J, Simpson ER. Liver receptor homologue-1 (LRH-1) regulates expression of aromatase in preadipocytes. J Biol Chem. 2002;277(23):20591-7. Epub 2002/04/03. doi: 10.1074/jbc.M201117200. PubMed PMID: 11927588.
25. Annicotte JS, Chavey C, Servant N, Teyssier J, Bardin A, Licznar A, Badia E, Pujol P, Vignon F, Maudelonde T, Lazennec G, Cavailles V, Fajas L. The nuclear receptor liver receptor homolog-1 is an estrogen receptor target gene. Oncogene. 2005;24(55):8167-75. Epub 2005/08/11. doi: 10.1038/sj.onc.1208950. PubMed PMID: 16091743; PMCID: PMC2259230.

26. 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, Ali S. The liver receptor homolog-1 regulates estrogen receptor expression in breast cancer cells. Breast Cancer Res Treat. 2011;127(2):385-96. Epub 2010/07/08. doi: 10.1007/s10549-010-0994-9. PubMed PMID: 20607599.
27. Chand AL, Herridge KA, Thompson EW, Clyne CD. The orphan nuclear receptor LRH-1 promotes breast cancer motility and invasion. Endocr Relat Cancer. 2010;17(4):965-75. Epub 2010/09/08. doi: 10.1677/ERC-10-0179. PubMed PMID: 20817789.
28. Schoonjans K, Dubuquoy L, Mebis J, Fayard E, Wendling O, Haby C, Geboes K, Auwerx J. Liver receptor homolog 1 contributes to intestinal tumor formation through effects on cell cycle and inflammation. Proc Natl Acad Sci U S A. 2005;102(6):2058-62. Epub 2005/02/03. doi: 10.1073/pnas.0409756102. PubMed PMID: 15684064; PMCID: PMC548586.
29. Bayrer JR, Mukkamala S, Sablin EP, Webb P, Fletterick RJ. Silencing LRH-1 in colon cancer cell lines impairs proliferation and alters gene expression programs. Proc Natl Acad Sci U S A. 2015;112(8):2467-72. Epub 2015/02/13. doi: 10.1073/pnas.1500978112. PubMed PMID: 25675535; PMCID: PMC4345603.
30. Petersen GM, Amundadottir L, Fuchs CS, Kraft P, Stolzenberg-Solomon RZ, Jacobs KB, Arslan AA, Bueno-de-Mesquita HB, Gallinger S, Gross M, Helzlsouer K, Holly EA, Jacobs EJ, Klein AP, LaCroix A, Li D, Mandelson MT, Olson SH, Risch HA, Zheng W, Albanes D, Bamlet WR, Berg CD, Boutron-Ruault MC, Buring JE, Bracci PM, Canzian F, Clipp S, Cotterchio M, de Andrade M, Duell EJ, Gaziano JM, Giovannucci EL, Goggins M, Hallmans G, Hankinson SE, Hassan M, Howard B, Hunter DJ, Hutchinson A, Jenab M, Kaaks R, Kooperberg C, Krogh V, Kurtz RC, Lynch SM, McWilliams RR, Mendelsohn JB, Michaud DS, Parikh H, Patel AV, Peeters PH, Rajkovic A, Riboli E, Rodriguez L, Seminara D, Shu XO, Thomas G, Tjonneland A, Tobias GS, Trichopoulos D, Van Den Eeden SK, Virtamo J, Wactawski-Wende J, Wang Z, Wolpin BM, Yu H, Yu K, Zeleniuch-Jacquotte A, Fraumeni JF, Jr., Hoover RN, Hartge P, Chanock SJ. A genome-wide association study identifies pancreatic cancer susceptibility loci on chromosomes 13q22.1, 1q32.1 and 5p15.33. Nat Genet. 2010;42(3):224-8. Epub 2010/01/27. doi: 10.1038/ng.522. PubMed PMID: 20101243; PMCID: PMC2853179.
31. Benod C, Vinogradova MV, Jouravel N, Kim GE, Fletterick RJ, Sablin EP. Nuclear receptor liver receptor homologue 1 (LRH-1) regulates pancreatic cancer cell growth and proliferation. Proc Natl Acad Sci U S A. 2011;108(41):16927-31. Epub 2011/09/29. doi: 10.1073/pnas.1112047108. PubMed PMID: 21949357; PMCID: PMC3193228.
32. Cobo I, Martinelli P, Flandez M, Bakiri L, Zhang M, Carrillo-de-Santa-Pau E, Jia J, Sanchez-Arevalo Lobo VJ, Megias D, Felipe I, Del Pozo N, Millan I, Thommesen L, Bruland T, Olson SH, Smith J, Schoonjans K, Bamlet WR, Petersen GM, Malats N, Amundadottir LT, Wagner EF, Real FX. Transcriptional regulation by NR5A2 links differentiation and inflammation in the pancreas. Nature. 2018;554(7693):533-7. Epub 2018/02/15. doi: 10.1038/nature25751. PubMed PMID: 29443959; PMCID: PMC6121728.
33. Majdic G, Young M, Gomez-Sanchez E, Anderson P, Szczepaniak LS, Dobbins RL, McGarry JD, Parker KL. Knockout mice lacking steroidogenic factor 1 are a novel genetic model of hypothalamic obesity. Endocrinology. 2002;143(2):607-14. Epub 2002/01/18. doi: 10.1210/endo.143.2.8652. PubMed PMID: 11796516.
34. Bolado-Carrancio A, Riancho JA, Sainz J, Rodriguez-Rey JC. Activation of nuclear receptor NR5A2 increases Glut4 expression and glucose metabolism in muscle cells. Biochem Biophys Res Commun. 2014;446(2):614-9. Epub 2014/03/19. doi: 10.1016/j.bbrc.2014.03.010. PubMed PMID: 24632207.

35. 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, Schoonjans K. SUMOylation-dependent LRH-1/PROX1 interaction promotes atherosclerosis by decreasing hepatic reverse cholesterol transport. Cell Metab. 2014;20(4):603-13. Epub 2014/09/02. doi: 10.1016/j.cmet.2014.07.023. PubMed PMID: 25176150.
36. Lu TT, Makishima M, Repa JJ, Schoonjans K, Kerr TA, Auwerx J, Mangelsdorf DJ. Molecular basis for feedback regulation of bile acid synthesis by nuclear receptors. Mol Cell. 2000;6(3):507-15. Epub 2000/10/13. doi: 10.1016/s1097-2765(00)00050-2. PubMed PMID: 11030331.
37. 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, Gauthier BR. LRH-1 agonism favours an immune-islet dialogue which protects against diabetes mellitus. Nat Commun. 2018;9(1):1488. Epub 2018/04/18. doi: 10.1038/s41467-018-03943-0. PubMed PMID: 29662071; PMCID: PMC5902555.
38. Coste A, Dubuquoy L, Barnouin R, Annicotte JS, Magnier B, Notti M, Corazza N, Antal MC, Metzger D, Desreumaux P, Brunner T, Auwerx J, Schoonjans K. LRH-1-mediated glucocorticoid synthesis in enterocytes protects against inflammatory bowel disease. Proc Natl Acad Sci U S A. 2007;104(32):13098-103. Epub 2007/08/03. doi: 10.1073/pnas.0702440104. PubMed PMID: 17670946; PMCID: PMC1941823.
39. Bayrer JR, Wang H, Nattiv R, Suzawa M, Escusa HS, Fletterick RJ, Klein OD, Moore DD, Ingraham HA. LRH-1 mitigates intestinal inflammatory disease by maintaining epithelial homeostasis and cell survival. Nat Commun. 2018;9(1):4055. Epub 2018/10/12. doi: 10.1038/s41467-018-06137-w. PubMed PMID: 30305617; PMCID: PMC6180039.
40. Mays SG, Flynn AR, Cornelison JL, Okafor CD, Wang H, Wang G, Huang X, Donaldson HN, Millings EJ, Polavarapu R, Moore DD, Calvert JW, Jui NT, Ortlund EA. Development of the First Low Nanomolar Liver Receptor Homolog-1 Agonist through Structure-guided Design. J Med Chem. 2019;62(24):11022-34. Epub 2019/08/17. doi: 10.1021/acs.jmedchem.9b00753. PubMed PMID: 31419141.
41. Rey J, Hu H, Kyle F, Lai CF, Buluwela L, Coombes RC, Ortlund EA, Ali S, Snyder JP, Barrett AG. Discovery of a new class of liver receptor homolog-1 (LRH-1) antagonists: virtual screening, synthesis and biological evaluation. ChemMedChem. 2012;7(11):1909-14. Epub 2012/09/11. doi: 10.1002/cmdc.201200307. PubMed PMID: 22961990.
42. Del Tredici AL, Andersen CB, Currier EA, Ohrmund SR, Fairbain LC, Lund BW, Nash N, Olsson R, Piu F. Identification of the first synthetic steroidogenic factor 1 inverse agonists: pharmacological modulation of steroidogenic enzymes. Mol Pharmacol. 2008;73(3):900-8. Epub 2007/12/07. doi: 10.1124/mol.107.040089. PubMed PMID: 18055761.
43. Madoux F, Li X, Chase P, Zastrow G, Cameron MD, Conkright JJ, Griffin PR, Thacher S, Hodder P. Potent, selective and cell penetrant inhibitors of SF-1 by functional ultra-high-throughput screening. Mol Pharmacol. 2008;73(6):1776-84. Epub 2008/03/13. doi: 10.1124/mol.108.045963. PubMed PMID: 18334597; PMCID: PMC3228235.
44. Corzo CA, Mari Y, Chang MR, Khan T, Kuruvilla D, Nuhant P, Kumar N, West GM, Duckett DR, Roush WR, Griffin PR. Antiproliferation activity of a small molecule repressor of liver receptor homolog 1. Mol Pharmacol. 2015;87(2):296-304. Epub 2014/12/05. doi: 10.1124/mol.114.095554. PubMed PMID: 25473120; PMCID: PMC4293447.
45. Benod C, Carlsson J, Uthayaruban R, Hwang P, Irwin JJ, Doak AK, Shoichet BK, Sablin EP, Fletterick RJ. Structure-based discovery of antagonists of nuclear receptor LRH-1. J Biol Chem.

2013;288(27):19830-44. Epub 2013/05/15. doi: 10.1074/jbc.M112.411686. PubMed PMID: 23667258; PMCID: PMC3707686.
46. D'Agostino EH, Flynn AR, Cornelison JL, Mays SG, Patel A, Jui NT, Ortlund EA. Development of a Versatile and Sensitive Direct Ligand Binding Assay for Human NR5A Nuclear Receptors. ACS Med Chem Lett. 2020;11(3):365-70. Epub 2020/03/19. doi: 10.1021/acsmedchemlett.9b00442. PubMed PMID: 32184971; PMCID: PMC7074214.
47. Whitby RJ, Dixon S, Maloney PR, Delerive P, Goodwin BJ, Parks DJ, Willson TM. Identification of small molecule agonists of the orphan nuclear receptors liver receptor homolog-1 and steroidogenic factor-1. J Med Chem. 2006;49(23):6652-5. Epub 2006/12/13. doi: 10.1021/jm060990k. PubMed PMID: 17154495.
48. Whitby RJ, Stec J, Blind RD, Dixon S, Leesnitzer LM, Orband-Miller LA, Williams SP, Willson TM, Xu R, Zuercher WJ, Cai F, Ingraham HA. Small molecule agonists of the orphan nuclear receptors steroidogenic factor-1 (SF-1, NR5A1) and liver receptor homologue-1 (LRH-1, NR5A2). J Med Chem. 2011;54(7):2266-81. Epub 2011/03/12. doi: 10.1021/jm1014296. PubMed PMID: 21391689; PMCID: PMC4151520.
49. Flynn AR, Mays SG, Ortlund EA, Jui NT. Development of Hybrid Phospholipid Mimics as Effective Agonists for Liver Receptor Homologue-1. ACS Med Chem Lett. 2018;9(10):1051-6. Epub 2018/10/23. doi: 10.1021/acsmedchemlett.8b00361. PubMed PMID: 30344916; PMCID: PMC6187417.
50. Zhi X, Zhou XE, Melcher K, Xu HE. Structures and regulation of non-X orphan nuclear receptors: A retinoid hypothesis. J Steroid Biochem Mol Biol. 2016;157:27-40. Epub 2015/07/15. doi: 10.1016/j.jsbmb.2015.06.012. PubMed PMID: 26159912.
51. Lin SJ, Yang DR, Yang G, Lin CY, Chang HC, Li G, Chang C. TR2 and TR4 Orphan Nuclear Receptors: An Overview. Curr Top Dev Biol. 2017;125:357-73. Epub 2017/05/22. doi: 10.1016/bs.ctdb.2017.02.002. PubMed PMID: 28527578.
52. Polvani S, Tarocchi M, Tempesti S, Galli A. Nuclear receptors and pathogenesis of pancreatic cancer. World J Gastroenterol. 2014;20(34):12062-81. Epub 2014/09/19. doi: 10.3748/wjg.v20.i34.12062. PubMed PMID: 25232244; PMCID: PMC4161795.

CHAPTER 2: DEVELOPMENT OF A VERSATILE AND SENSITIVE DIRECT LIGAND BINDING ASSAY FOR HUMAN NR5A NUCLEAR RECEPTORS[1a]

Emma H. D'Agostino,[*] Autumn R. Flynn,[#] Jeffery L. Cornelison,[#] Suzanne G. Mays,[*] Anamika Patel,[*] Nathan T. Jui,[#] and Eric A. Ortlund[*]

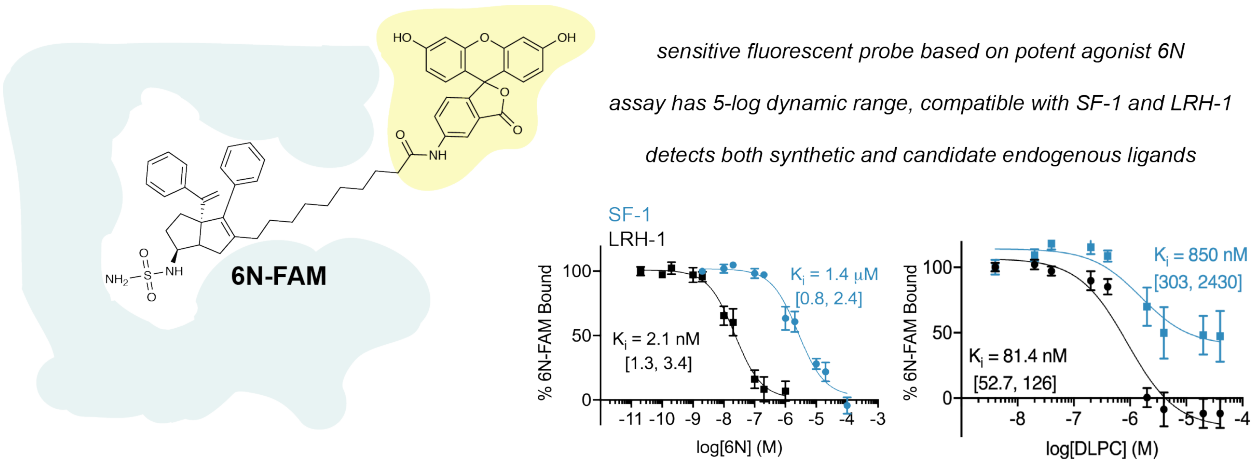
[*]Department of Biochemistry and [#]Department of Chemistry, Emory University, Atlanta, Georgia 30322, United States

This manuscript describes a novel fluorescence polarization binding assay for the human NR5A nuclear receptors, liver receptor homolog-1 (LRH-1) and steroidogenic factor-1 (SF-1). Although we had been developing NR5A agonists for several years, prior to the development of this binding assay we had no way to measure binding affinities to guide our drug development efforts. We also could not easily screen potential endogenous ligands for these receptors, as their endogenous ligand(s) is not known, though they are presumed to be phospholipids. The development of this assay has greatly facilitated our own drug development efforts and has great potential to accelerate the NR5A therapeutic development field. This work was accepted for publication in ACS Medicinal Chemistry Letters for a special Women in Chemistry issue in November of 2019.

---

[1]This chapter adapted from the previously published work D'Agostino EH, Cornelison JL, Mays SG, Flynn AR, Patel A, Jui NT, Ortlund EA. Development of a Robust Direct Binding Assay for Phospholipid-Sensing Human NR5A Nuclear Receptors. *ACS Med Chem Lett.* 2019 Nov 21; 11(3):365-370.

[a]E.H.D., A.R.F., J.L.C., S.G.M., N.T.J., and E.A.O. participated in research design. E.H.D., A.R.F., J.L.C., S.G.M., and A.P. conducted experiments. E.H.D., A.R.F., J.L.C., and S.G.M. performed data analysis. E.H.D., A.R.F., S.G.M., N.T.J., and E.A.O. contributed to the writing of the manuscript.



Abstract

As regulators of steroidogenesis, development, and metabolism, the nuclear receptor 5A (NR5A) subfamily members steroidogenic factor 1 (SF-1) and liver receptor homologue 1 (LRH-1) are important pharmacological targets for cancers and metabolic diseases. Evaluation of small molecule modulators and candidate endogenous ligands for these orphan receptors has been hindered by the lack of accessible, robust direct-binding assays. Here, we leverage the potency of our new NR5A agonist (6N) to create a high-affinity probe for fluorescence polarization competition assays by conjugating 6N to fluorescein (FAM). The 6N-FAM probe tightly binds the NR5A receptors and detects direct binding of synthetic and phospholipid ligands. For 25 LRH-1 agonists, affinity predicts potency in cellular activation assays, demonstrating the potential for this assay in drug discovery. Moreover, phospholipids dilauroylphosphatidylcholine and phosphatidylinositol(4,5)phosphate bind with high affinity, demonstrating this assay is robust for evaluation of candidate endogenous ligands for human NR5A receptors.

Introduction

The human nuclear receptor (NR) superfamily comprises 48 ligand-regulated transcription factors that regulate diverse biological processes including metabolism, inflammation, immune response, development, and steroidogenesis. NRs show exquisite specificity for their endogenous ligands and respond by driving specific transcriptional changes. Their powerful control of gene expression makes them attractive pharmacological targets, and genetic gain and loss of function studies have revealed tremendous potential for this receptor class. However, only 17 NRs have been successfully targeted in the clinic.[1] Of the remaining NRs, many respond to abundant lipids and lipid metabolites, and elucidating native ligands and synthetic modulators has been challenging.[2,3]

Two lipid-sensing NRs with promising therapeutic potential are steroidogenic factor-1 (SF-1; NR5A1; Figure 1A) and liver receptor homologue-1 (LRH-1; NR5A2; Figure 2.1B), the two human NR5A subfamily members. SF-1 regulates steroidogenesis in the ovaries and adrenal glands[4] and energy homeostasis in the ventromedial hypothalamus.[5] LRH-1 regulates steroidogenesis[6–9] in the ovaries, breast preadipocytes, and intestinal epithelium and glucose,[10] cholesterol,[11-12] and bile acid homeostasis in the liver, intestine, and pancreas. SF-1 and LRH-1 are critical for development: SF-1 is necessary for endocrine organ development and differentiation,[13] and LRH-1 is required for the maintenance of stem cell pluripotency.[14] Both NR5As also drive cancer progression, with SF-1 involved in adrenocortical carcinoma and LRH-1 in cancers of the breast, colon, pancreas, and prostate.[15–19] These diverse roles make the NR5As attractive pharmaceutical targets.

Despite the therapeutic promise of the NR5As, development of synthetic modulators has been challenging. Phospholipids (PLs) are the putative native ligands for the NR5As, which bind multiple phosphatidylcholine and phosphatidylinositol species.[20–22] The hydrophobicity of these native ligands and their corresponding ligand-binding pockets creates two challenges in designing ligand-binding assays and screens.

First, NR5As favor ligands with low aqueous solubility, hindering ligand binding detection. Second, recombinant proteins copurify with phospholipids, further confounding ligand binding detection. Though direct binding assays have been reported for the NR5A receptors, they are not amenable to rapid compound screening.[21,23−25] Screens have largely relied on indirect methods such as coregulator recruitment and have only identified a handful of small molecule modulators.[23,25−28] The ability to measure direct binding in the ligand-binding pocket would greatly facilitate synthetic ligand screening and development. Thus, we sought to develop a direct binding assay for efficient quantification of binding affinities of a small compound library. Fluorescence polarization (FP) is a direct, equilibrium binding assay commonly used with NRs. It is solution based, allowing molecules to retain their native state, uses minimal material, and allows parallel evaluation of several compounds using plate-based fluorescence detectors. We recently developed an NR5A agonist, 6N, with low nanomolar potency which facilitated the synthesis of a fluorescent probe for use in FP.[29] Here, we report an FP competition assay using a novel fluorescent probe synthesized by conjugating 6N to a fluoresceinamine (FAM) moiety. This assay detects binding of synthetic ligands from multiple classes and of potential endogenous phospholipid ligands with a dynamic range from single-digit nanomolar to midmicromolar. Affinities of a small set of synthetic agonists correlate with potencies in cellular LRH-1 activation assays, demonstrating the potential for this assay to predict in-cell activity prior to undertaking more expensive and time-intensive methods for characterization of candidate NR5A modulators.

Results

Probe Design

We designed a novel fluorescent probe based on our recent discovery of 6N, which has low nanomolar EC_{50} in luciferase reporter assays.[29] High-affinity compounds are important for FP-based competition binding assays because probe affinity limits detection of K_i for competing ligands.[30] We hypothesized

that this potent agonist would bind the NR5A receptors with high affinity and serve as a scaffold for an FP probe.

The 6N agonist was rationally designed based on our crystallographic studies with the hexahydropentalene NR5A agonist, RJW100.[31,32] Substitution of a sulfamide for the RJW100 hydroxyl group enhanced polar interactions in the LRH-1 binding pocket and improved potency 100-fold over RJW100 in cellular activation assays. Guided by the crystal structure of LRH-1-6N, we extended the 6N hexyl "tail" and installed a fluoresceinamine (FAM) moiety. The linker length was sufficient to position the FAM substituent outside the pocket without interfering with desired deep-pocket contacts anchoring the probe (Figure 1C−D). Tail modifications on the hexahydropentalene scaffold are easily incorporated, and NR5A receptors can accommodate a variety of modifications.[33] Synthesis of the designed probe involved elaboration of diol **1**, the synthesis of which was reported previously.[33] Ley-Griffith oxidation afforded the corresponding ketoacid. Esterification gave rise to **2** (in 72% yield over two steps), and diastereoselective reductive amination to **3** provided the endo amine necessary for installation of the sulfamide, which drives potency of the probe. Sulfamide assembly[29] and global deprotection gave **5** which was coupled with fluoresceinamine to furnish the probe 6N-FAM (**6**) (Scheme 2.1).

Assay Development

We first determined the affinities of SF-1 and LRH-1 for 6N- FAM. Purified SF-1 or LRH-1 ligand-binding domain was titrated against several constant 6N-FAM concentrations to determine optimal conditions. We chose 10 nM 6N-FAM as it maximized signal and sensitivity in competition experiments (below). The K_d of the probe using these conditions was 1.0 nM for LRH-1 (95% confidence interval: [0.8, 1.3]), and 12.3 nM [9.0, 16.7] for SF-1 (Figure 2.2A, S2.1).

To validate 6N-FAM in a competition assay, we measured the K_i values of unlabeled 6N (Figure 2.2B, S2.1). The unlabeled probe should completely outcompete the labeled probe with a similar inhibition

constant (K_i) to the forward binding constant (K_d).[30] Optimized reaction conditions are described in full detail in the Supporting Information. For both LRH-1 and SF-1, unlabeled 6N dose-dependently decreased milli- polarization values and completely outcompeted the probe (Figure 2.2B, S2.1). For LRH-1, 6N bound with a K_i of 2.1 nM (95% CI: [1.3, 3.4]), in agreement with the forward binding K_d. Affinities of 6N-FAM and 6N were similar when apo-LRH-1 was used instead of DLPC-exchanged protein (Figure 2.2C−D, S2.1). Thus, DLPC does not significantly impact affinity measurements, eliminating the need to strip and refold the protein. Surprisingly, the affinity of unlabeled 6N for SF-1 was much lower than the K_d of the probe, perhaps indicating that the FAM linker makes additional interactions with SF-1 versus LRH-1.

High-Affinity Probe Increases Sensitivity For Detecting Mammalian Phospholipid Binding

Phospholipid binding assays are challenging to develop, as lipids prefer micellular environments and aggregate in solution. We have previously reported a liposome-based fluorescence resonance energy transfer (FRET)-based assay for LRH-1.[24] This assay utilizes donor-quencher vesicles harboring nitrobenzoxadiazole (NBD)-labeled 1,2-dilauroyl-sn-glycero-3- phosphoethanolamine (DLPE) and 7-diethylamino-3-((4'- iodoacetyl)amino)phenyl)-4-methylcouramin (DCIA)-labeled LRH-1 and requires the nonspecific lipid chaperone β-cyclodextrin to enhance lipid exchange. Though this assay measures binding of a variety of lipids, its lower range of detection is 1 μM due to the relatively low affinity of DCIA- LRH-1 for NBD-DLPE. Thus, we sought to determine whether the FP competition assay, which has a low nanomolar limit of detection, could be used to measure phospholipid binding and evaluate candidate endogenous NR5A ligands.

DLPC binds both LRH-1 and SF-1 and is of pharmacological interest due to its ability to suppress lipogenesis and improve insulin resistance in obese mice.[20] We detected DLPC K_i values of 850 nM [303, 2430] and 81.4 nM [52.7, 126] for SF-1 and LRH-1, respectively (Figures 2.3A and S2.1). This is 20-fold greater than the 1.9 μM affinity we measured using the FRET assay with LRH-1. Thus, the

FP assay increases our dynamic range 1,000-fold compared to the FRET assay and expands our ability to evaluate potential endogenous ligands for the NR5As.

We also measured binding to phosphatidylinositol 4,5- bisphosphate (PI(4,5)P_2), as multiple phosphatidylinositol species have been crystallized with NR5As (Figures 2.3B and S2.1).[21,22] Phosphatidylinositols bind NR5As with high affinity in an electrophoretic mobility shift (EMSA)-based assay.[21,22] Affinities for PIP(4,5)P_2 binding of 418 nM [265, 646] and 64.3 nM [39.8, 103] for SF-1 and LRH-1, respectively, were similar to those obtained in the EMSA assay for SF-1 binding to PIP(4,5)P_2 (~250 nM) and LRH-1 binding to PI(3,4,5)P_3 (120 ± 9 nM).

The ligand exchange detected with these PLs indicates that NR5As may follow a canonical model of nuclear receptor activation. Previous reports have proposed that NR5As bind ligand upon folding, are constitutively ligand-bound and active, and do not exist in apo form.[34,35] These data indicate that NR5A receptors are dynamic and capable of heterotypic ligand exchange.

Affinity Correlates with Biological Activity and Receptor Stability for Synthetic Agonists

We have previously used thermal shift assays to detect ligand binding to LRH-1 and assess the effects of ligands on global protein stability. For a recently reported subset of these agonists, we have shown that 50% unfolding temperature (T_m) values for LRH-1−ligand complexes strongly correlate with EC_{50} values for LRH-1 activity in cellular luciferase reporter assays. However, many compounds that bind the NR5As do not induce a strong thermal shift response. We sought to determine whether the FP competition assay could be used to predict compound activity in cells by measuring K_i values for this set of compounds based on the RJW100 scaffold. These compounds, referred to as the R^1 series, include 6N and contain modifications at the 1-position hydroxyl (Figure 2.4A). We found that LRH-1 K_i values for the R^1 series correlate with EC_{50} values from luciferase reporter assays (Figure 2.4B, Pearson $r = 0.67$, $p = 0.0087$), indicating that the FP assay can be a screening tool to predict in-cell activity for LRH-1, offering exciting potential for future use in compound development. The LRH-1

K_i values also correlate with T_m values for the R^1 compounds (Figure 4C, Pearson r = −0.66, p = 0.0027; Figures S2.1−3). We have not measured in-cell activation of SF-1 by the entire set of R^1 compounds; thus, further investigation is needed to determine whether K_i and EC_{50} correlate for SF-1. However, there is no correlation between the K_i and T_m values for the R^1 series with SF-1 (Figure 2.4D, Pearson r = −0.084, p = 0.73; Figures S1−2, 4), perhaps due to distinct effects on SF-1 conformation. The R^1 compounds were designed based on LRH-1 structural studies and generally exhibit poor affinity (>1 μM) for SF-1. It remains to be seen whether binding affinities will predict stabilization or in-cell activity of SF-1 for higher affinity compounds.

FP Competition Assay Accurately Quantifies Binding of Synthetic Modulators

To further validate our FP competition assay, we compared K_i values to previously reported K_d values (Figures 2.5 and S2.1). The FP affinity of RJW100 for SF-1 was similar to the value determined by EMSA (EMSA: 1200 ± 270 nM; FP: 3.3 μM [1.9, 5.7]; Figure 2.5A).[21] Our FP values were in agreement with affinities calculated by equilibrium surface plasmon resonance (SPR) for the Cpd3 antagonist for LRH-1 (SPR: 1.5 ± 0.3 μM; FP: 2.4 μM [0.9, 5.2]; Figure 2.5B).[23,25] Interestingly, the FP assay showed a higher affinity for the PME9 agonist to LRH-1 than SPR (FP, 7.0 μM [3.8, 13.0]; SPR, 62.9 μM; Figure 2.5C), perhaps due to the time difference between the two assays. SPR was conducted with 60-s contact times, whereas the FP assay requires overnight equilibration to achieve maximum affinity.

We next examined the SF-1 isoquinolinone antagonist SID7969543; no binding data is available for this compound, but its IC_{50} is 30 nM.[36] While SID7969543 displaced the 6N-FAM at high concentrations, we were unable to calculate affinity (Figure 2.5D). Finally, we tested binding of the LRH-1 antagonist SR1848, for which the authors could not detect binding.[28] Our FP assay also did not detect binding (Figure 2.5E), suggesting a novel mechanism of LRH-1 inhibition and also underscoring the difficulty of synthetic modulator development without a direct binding assay to verify compound binding in the ligand-binding pocket.

Discussion

The NR5A receptors are promising therapeutic targets for metabolic diseases and several cancers, but the hydrophobicity of their binding pockets and preferred ligands has made compound screening and development exceptionally challenging. We present an FP competition assay to quantify direct ligand binding to NR5A receptors with a 5-log dynamic range. Fluorescence polarization is a simple, inexpensive assay commonly used to quantify ligand binding for NRs (e.g., the PolarScreen FP competition assay is available for seven NRs, Thermo Fisher Scientific, Waltham, MA). It is solution-based, retaining native protein conformation, and equilibrium-based, permitting for measurement of binding regardless of ligand exchange kinetics. We have optimized the assay for 384-well plates, allowing measurement of several compounds in parallel. The assay format ensures that binding will only be detected if a competitor binds in the ligand-binding pocket. This is particularly important for NR5As given that previous screens have largely relied on indirect or virtual screening methods.[23,27,28,31] The flexibility in buffer components provided by FP is critical given the general insolubility of NR5A ligands. We have successfully used 6.7% v/v DMSO and ethanol to increase competitor solubility, allowing competitor ligand concentrations up to 200 μM. We have shown that K_i values correlate with in-cell potencies for a series of related NR5A agonists, indicating that this in vitro assay can predict biological activity. In additional to small molecules, the assay detects binding of candidate endogenous phospholipid ligands, which are still under investigation for this subfamily. This assay will be invaluable in continued drug design efforts for these attractive pharmacological targets

Methods

Protein Expression and Purification.

Human LRH-1 LBD (residues 299-541) in the pLIC-His vector was transformed in *E. coli* strain BL21(pLysS) for expression. Cultures (6 L in Liquid Broth, LB) were grown in the presence of ampicillin and chloramphenicol at 37 °C to an OD_{600} of 0.6. Protein expression was induced with 1

mM isopropyl-1-thio-D-galactopyranoside (IPTG) and grown for 4 hours at 30 °C. Cell pellets were resuspended in Buffer A (20 mM Tris-HCl (pH 7.4), 150 mM NaCl, 5% glycerol, 25 mM imidazole), DNAse, lysozyme, and phenylmethylsulfonyl fluoride (PMSF). Resuspended cells were sonicated and clarified by centrifugation at 16,000xg for 45 minutes in a Sorvall RC 6+. Protein was purified from the lysate by nickel affinity chromatography (HisTrap FF; GE Healthcare, Little Chalfront, UK): lysate was flowed over the column, washed with Buffer A, and eluted with Buffer B (20 mM Tris-HCl (pH 7.4), 150 mM NaCl, 5% glycerol, 500 mM imidazole). Protein was incubated with DLPC (4-fold molar excess) overnight at 4 °C, repurified by size exclusion into assay buffer (150 mM NaCl, 20 mM Tris-HCl (pH 7.4), 5% glycerol) concentrated to approximately 3 mg/mL,and stored at -80 °C until use.

SF-1 LBD (residues 218-461) in the pLIC-His vector was transformed in *E. coli* strain BL21(pLysS) for expression. Cultures (6L LB) were grown in the presence of ampicillin and chloramphenicol at 37 °C to an OD_{600} of 0.6. Expression was induced with 0.5 mM IPTG and cultures were grown overnight at 18 °C. Protein was purified by nickel affinity chromatography as described for LRH-1 (Buffer A: 20 mM Tris-HCl (pH 7.4), 5% glycerol, 500 mM NaCl, 25 mM imidazole, 0.5 mM TCEP; Buffer B: 20 mM Tris-HCl (pH 7.4), 5% glycerol, 500 mM NaCl, 500 mM imidazole, 0.5 mM TCEP), followed by overnight DLPC exchange and size exclusion chromatography into assay buffer. Pure SF-1 protein was concentrated to approximately 3 mg/mL and stored at -80 °C until use.

Generation of apo LRH-1. To extract lipids from the LRH-1 LBD, 4.5 mL of purified protein (15 mg) was treated with 18.75 mL of chloroform-methanol solution (1:2 v/v) and vortexed briefly. An additional 2.5 mL chloroform:water solution (1:1 v/v) was added and the mixture was vortexed again. The stripped and unfolded protein was pelleted by centrifugation at 1000 rpm for 10 minutes. The resulting protein pellet was dissolved into 0.5 mL of buffer containing 50 mM Tris (pH 8.0), 6 M guanidine hydrochloride and 2 mM DTT. Protein was refolded by fast dilution at 4 °C into 50 mL of

buffer containing 20 mM Tris (pH 8.5), 1.7 M urea, 4% glycerol and 2 mM DTT. The final urea concentration was adjusted to 2 M, and protein was concentrated to ~ 15 mL, followed by overnight dialysis against assay buffer (see below) containing 2 mM DTT at 4 °C. Refolded protein was purified by size exclusion chromatography to remove aggregates and remaining unfolded protein.

Fluorescence Polarization. All assays were conducted in black, polystyrene, non-binding surface 384-well plates (Corning Inc., Corning, NY) with 30 μL volumes in assay buffer. Binding affinity for 6N-FAM was determined using 10 nM 6N-FAM and protein concentrations ranging from 1^{-10}–5^{-5} M (SF-1) or 1^{-11}–5^{-6} M (LRH-1). Plates were incubated overnight at 4 °C and centrifuged at 2,000xg for 2 minutes before polarization measurement. Polarization was monitored on a Neo plate reader (Biotek, Winooski, VT) at an excitation/emission wavelength of 485/528 nm. Nine technical replicates were conducted over three experiments and compiled binding data were baseline-corrected to wells with no protein and fit with a one-site binding curve in GraphPad Prism version 7 (GraphPad, Inc., La Jolla, CA).

Competition assays were performed in accordance with development guidelines.[33] For LRH-1, 10 nM 6N-FAM (10 times the affinity of LRH-1 for 6N-FAM, necessary to obtain adequate signal) and 5 nM LRH-1 (80% of the forward binding B_{max}) were used. For SF-1, 10 nM 6N-FAM (0.8 times the affinity of SF-1 for 6N-FAM) and 25 nM SF-1 (60% of the forward binding B_{max}) were used. Competitor ligand concentration ranged from 2^{-11}-2^{-4} M, and competitor ligand volume was kept constant to maintain constant DMSO in each well (6.7% v/v). Eight technical replicates were performed over two experiments, and GraphPad Prism version 7 was used to analyze compiled data using a one-site, fit K_i curve, with normalization to 6N competition.

For assays with lipids, lipids were solubilized in chloroform and transferred to a clean glass tube. Lipids were dried via evaporation to produce multilamellar sheets. These were resuspended in ethanol and sonicated (twice x 30 seconds) to produce small vesicles for use in FP assays.

Differential Scanning Fluorimetry. Purified protein, pre-exchanged with DLPC (0.2 mg/mL), was combined with agonists overnight at 4 °C in assay buffer. SYPRO orange dye was added to the complexes the next day, at a final dilution of 1:1000. Complexes were heated at a rate of 0.5 °C/ minute on a StepOne Plus thermocycler, using the ROX filter for fluorescence detection. The melting temperature (T_m, 50% unfolding) was calculated using the Bolzman equation (GraphPad Prism, V7). Assays were conducted with nine technical replicates over three experiments.

Figures

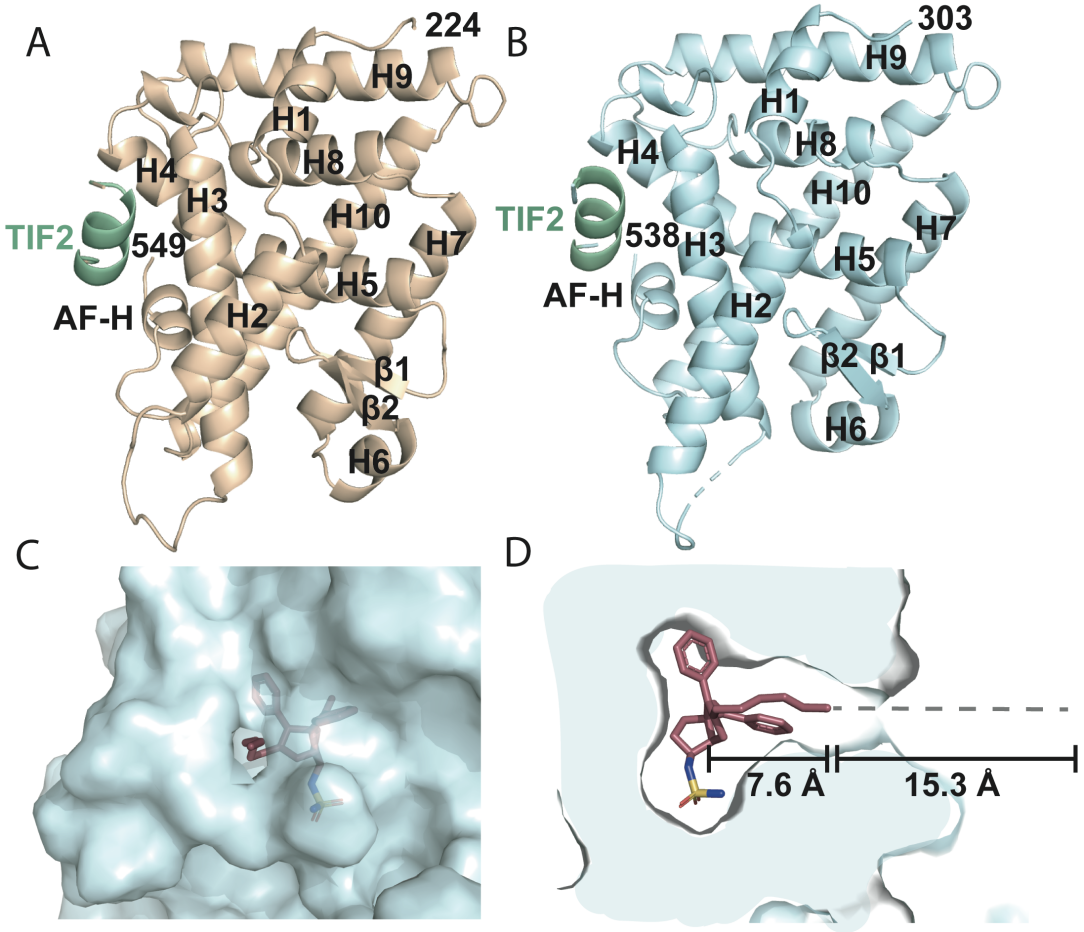
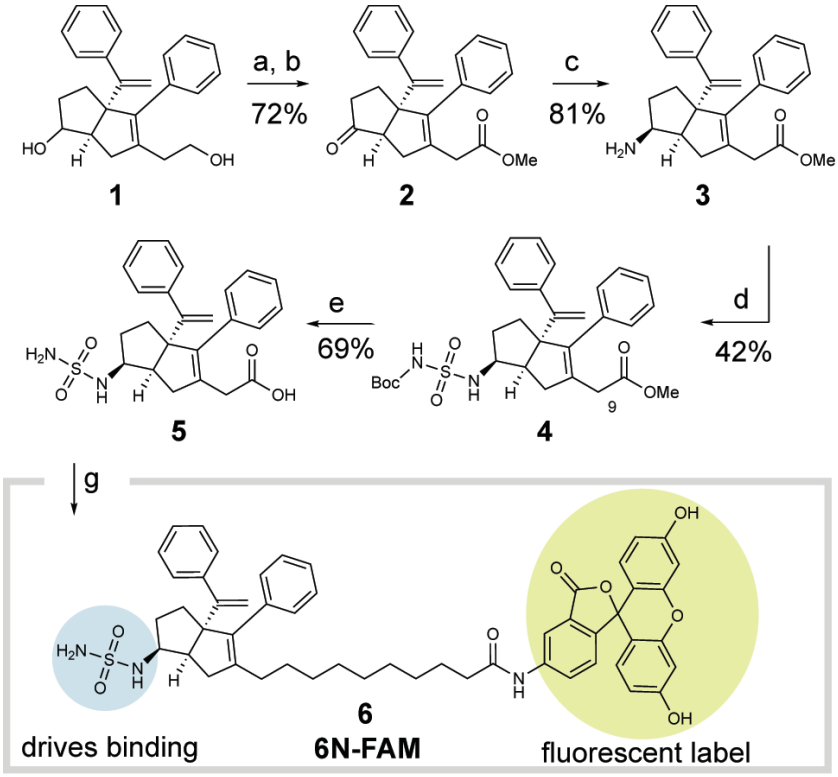


Figure 2.1. Structure-guided design of NR5A probe. Structures of the ligand-binding domains of (A) SF-1 (PDB: 1ZDT) and (B) LRH-1 (PDB: 6OQY). (C) There is a clear exit tunnel from the ligand- binding

pocket (LRH-1 shown) which can accommodate the 6N- FAM linker, and (D) the linker (red) and FAM molecule (dashed line) provide sufficient length to exit the pocket mouth and leave the FAM moiety solvent-exposed for FP detection.



Scheme 2.1. Chemical synthesis of 6N-FAM (6)[a]

[a]Reagents and conditions: (a) Tetrapropylammonium perrhuthenate, N-methyl morpholine oxide, H_2O, MeCN, 23 °C, 16 h; (b) MeOH, conc. aq. HCl, 23 °C, 16 h; (c) NH_3 (7N in MeOH), titanium(IV) isopropoxide, 23 °C, 6h. Sodium borohydride, 16 h; (d) Chlorosulfonylisocyanate, ^{t}BuOH, DCM, 0 to 23 °C, 45 min, then TEA, 0 to 23 °C, 3 h; (e) 1,4-dioxane: conc. aq. HCl (3:1 v/v), 40 °C, 14 h; (f) EDCI, fluoreseinamine isomer 1, DMF, 23 °C, 5 h.

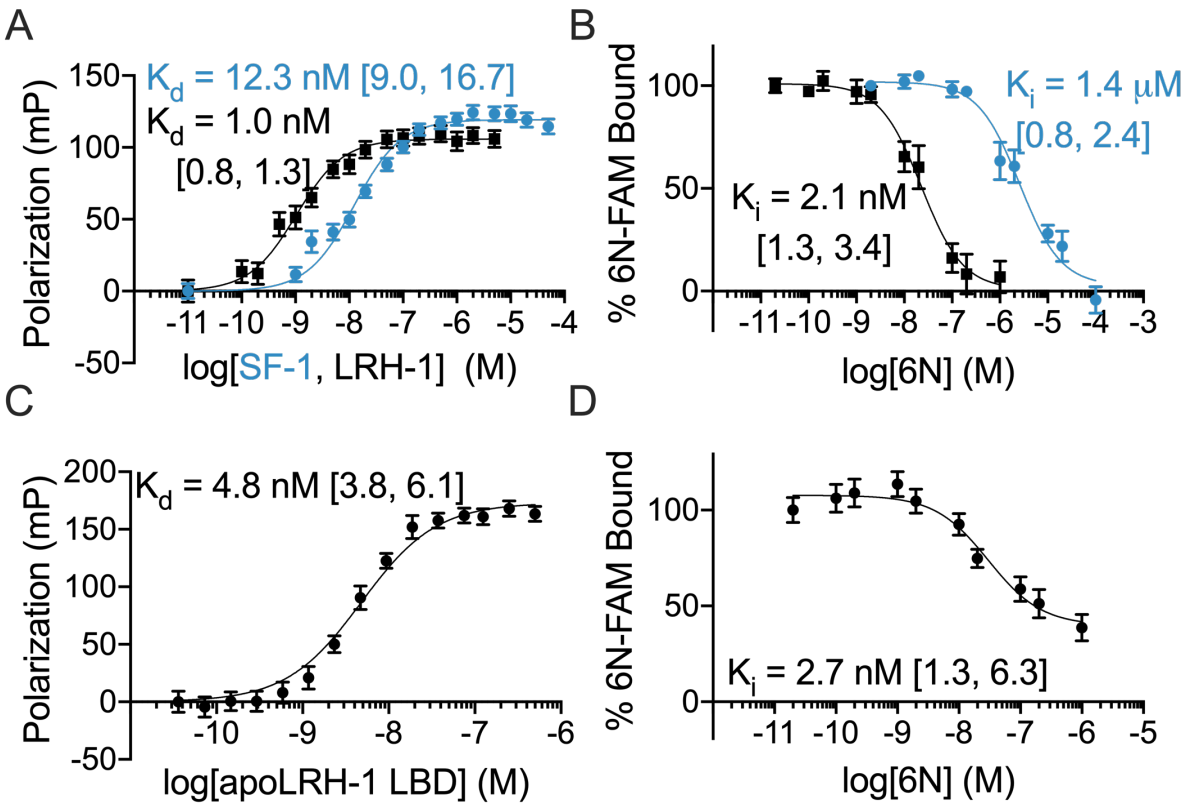


Figure 2.2. Validation of fluorescence polarization. (A) Binding of 6N- FAM to SF-1 or LRH-1 (n = 9). Insets indicate K_d values (95% CI in square brackets). (B) Competitive displacement of the 6N-FAM probe using unlabeled 6N (n = 8). 6N completely displaced 6N-FAM from both SF-1 and LRH-1. Insets indicate the K_i (95% CI in square brackets). (C) 6N-FAM and (D) Unlabeled 6N binds apo LRH-1 with comparable affinity to DLPC-exchanged protein (C, n = 2; D, n = 8). Blue lines, SF-1; black lines, LRH-1; error bars are SEM. Competition experiments used 10 nM 6N-FAM, 5 nM LRH-1, 25 nM SF-1.

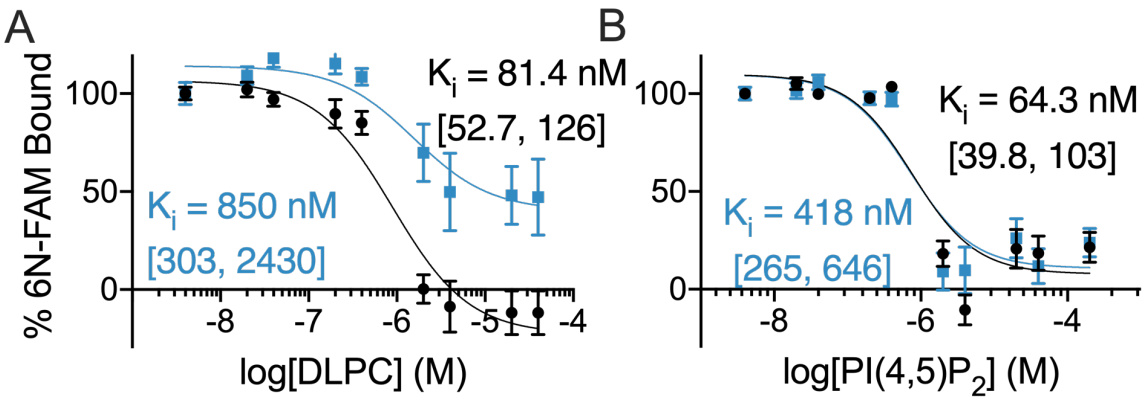


Figure 2.3. FP assay detects lipid binding. Both NR5As bind (A) dilauroylphosphatidylcholine (DLPC) and (B) PI(4,5)P_2 (n = 8). Blue lines, SF-1; black lines, LRH-1; 95% CI is in square brackets; error bars are SEM. Experiments used 10 nM 6N-FAM, 5 nM LRH-1, 25 nM SF-1.

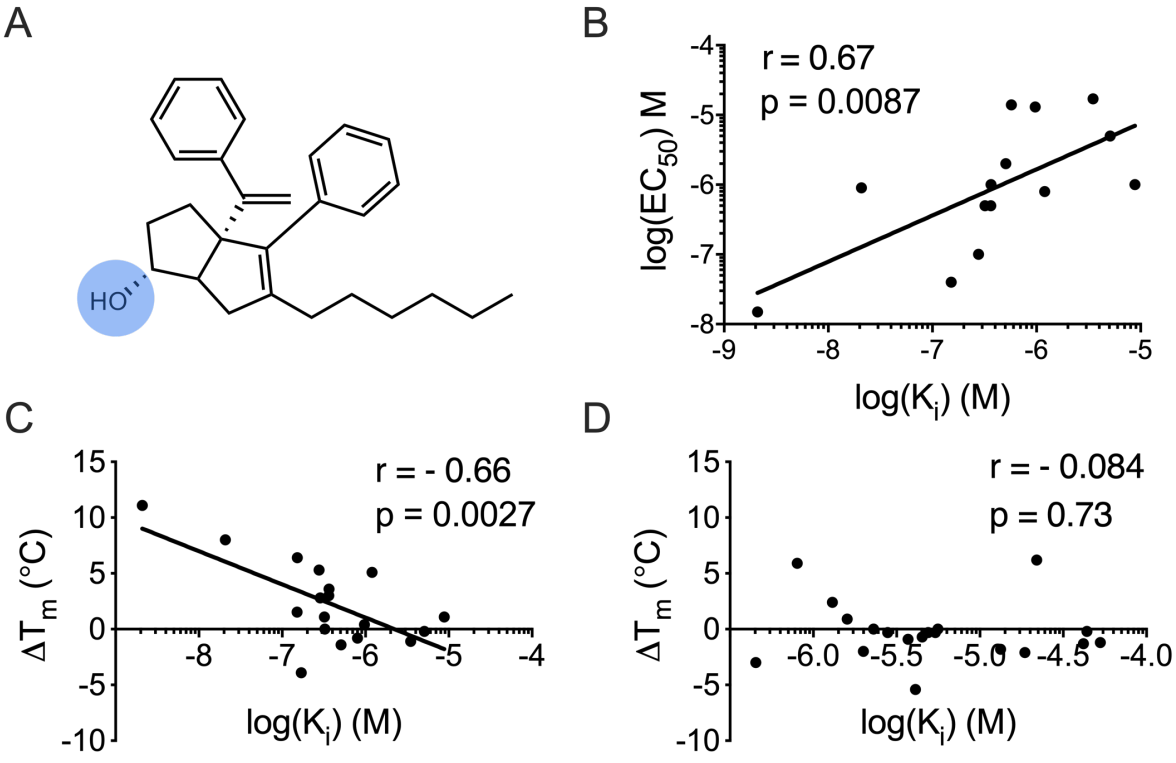


Figure 2.4. Binding affinity correlates with in-cell activity and receptor stability for LRH-1, but not SF-1. (A) Modifications were made to the 1-position hydroxyl ("R^1") of RJW100, shown in blue, to generate 24

derivatives (reported in). (B) Affinity of R^1 compounds correlates with in-cell potency for LRH-1 in a luciferase reporter assay. Affinity of R^1 compounds also correlated with their effect on receptor stability for LRH-1 (C), but not for SF-1 (D).

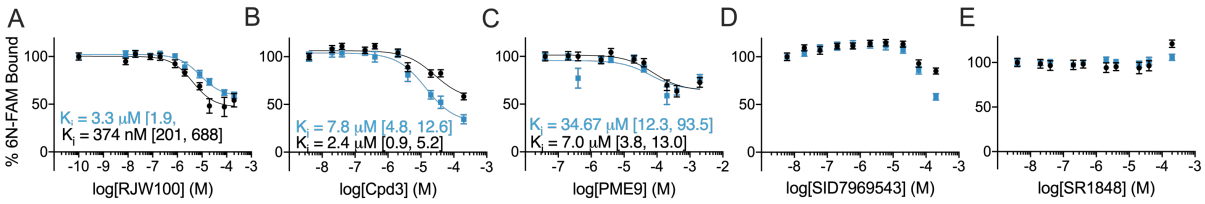
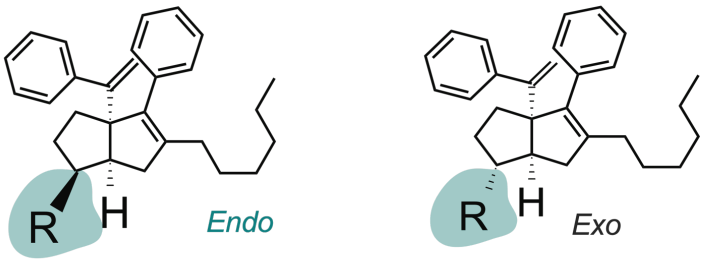


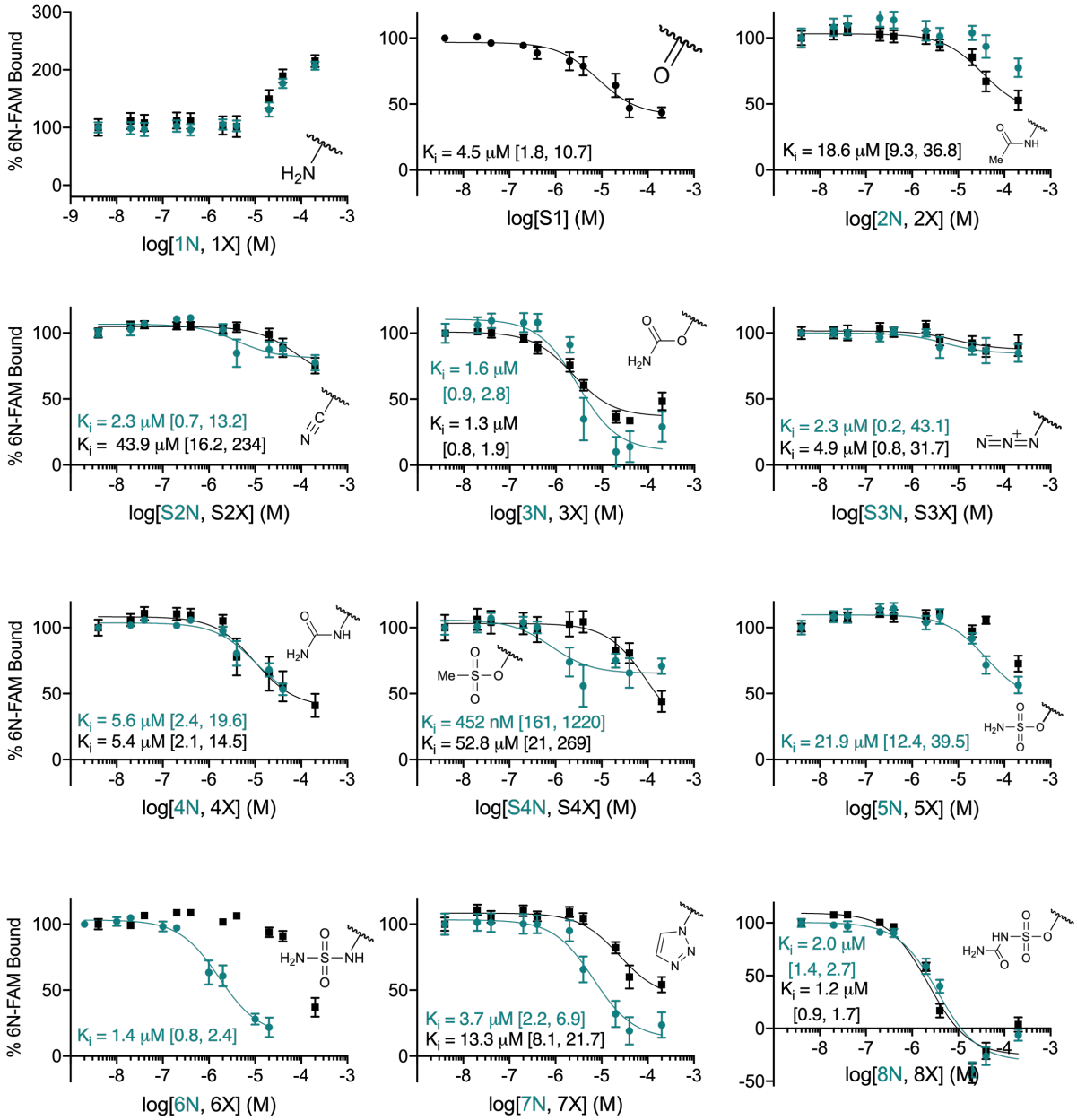
Figure 2.5. FP measurements for synthetic ligands. Both NR5As bind RJW100 (A), Cpd3 (B), and PME9 (C). (D) Binding affinity for SF-1 antagonist SID7969543 cannot be calculated, although probe displacement is detected at high doses. (D) Binding is undetectable for SR1848. Blue lines, SF-1; black lines, LRH-1; error bars are SEM (n = 8). Experiments used 10 nM 6N-FAM, 5 nM LRH-1, 25 nM SF-1.

Compound	SF-1 K_i [95% confidence interval]	LRH-1 K_i [95% confidence interval]
6N-FAM (K_d)	12.3 nM [9.0, 16.7]	1.0 nM [0.8, 1.3]; Apo, 4.8 nM [3.8, 6.1]
6N	1.4 μM [0.8, 2.4]	2.1 nM [1.3, 3.4]; Apo, 2.7 nM [1.3, 6.3]
DLPC	850 nM [303, 2430]	81.4 nM [52.7, 126]
$PI(4,5)P_2$	9.6 μM [4.3, 21.5]	756 nM [382, 1470]
1N	cnc	cnc
1X	cnc	cnc
S1	4.5 μM [1.8, 10.7]	801 nM [434, 1460]
2N	cnc	973 nM [453, 2060]
2X	18.6 μM [9.3, 36.8	3.5 μM [1.5, 8.5]
S2N	2.3 μM [0.7, 13.2]	326 nM [181, 580]
S2X	43.9 μM [16.1, 234]	cnc
3N	1.6 μM [0.9, 2.8]	287 nM [174, 468]
3X	1.3 μM [0.8, 1.9]	151 nM [83.0, 263]
S3N	2.8 μM [0.2, 43.1]	cnc

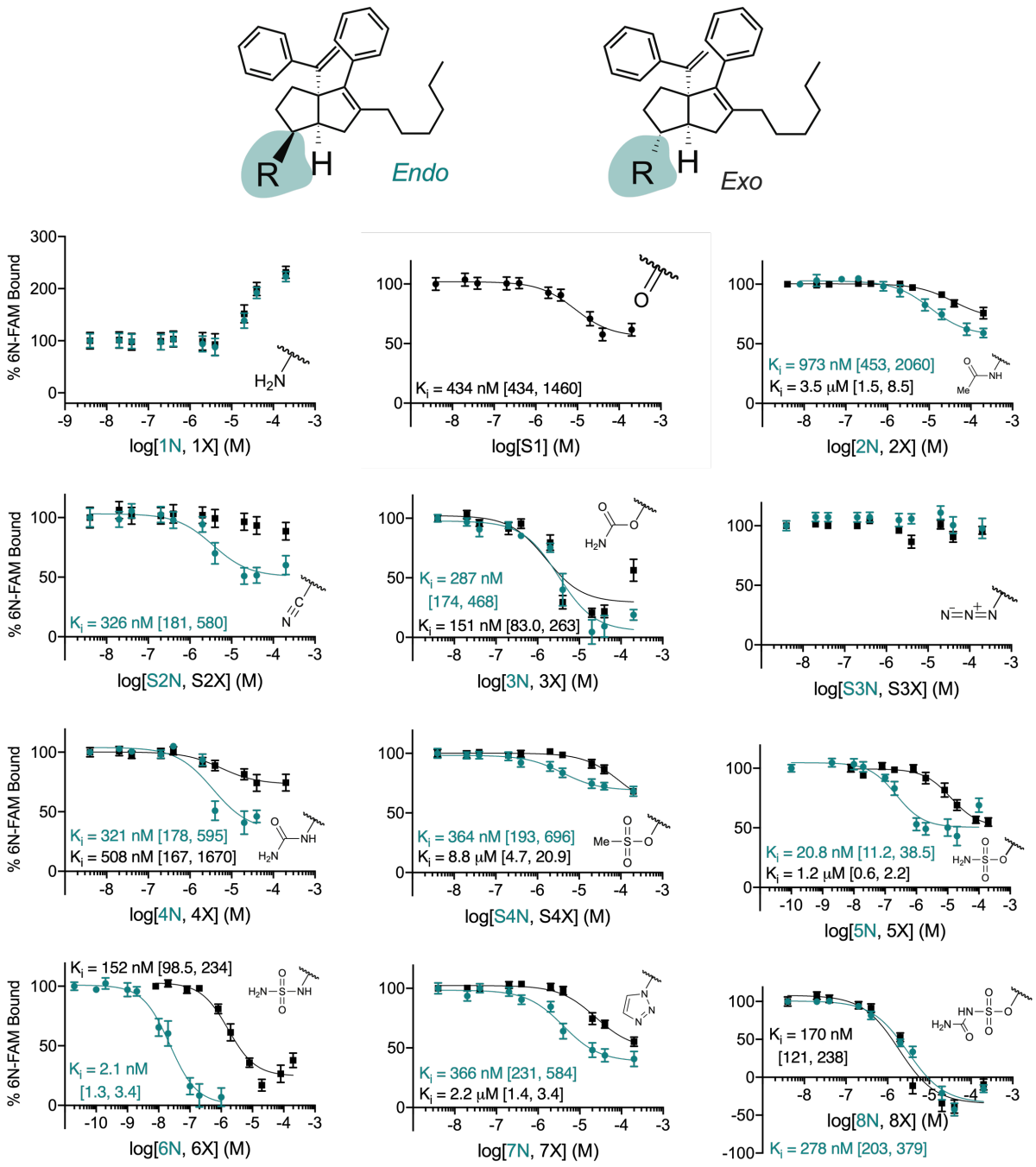
S3X	4.9 μM, 0.8, 31.7]	cnc
4N	5.6 μM [2.4, 19.6]	321 nM [178, 595]
4X	5.4 μM [2.1, 14.5]	508 nM [167, 1670]
5N	21.9 μM [12.4, 39.5]	20.8 nM [11.2, 38.5]
5X	cnc	1.2 μM [0.6, 2.2]
6N	1.4 μM [0.8, 2.4]	2.1 nM [1.3, 3.4]
6X	cnc	152 nM [98.5, 234]
7N	3.7 μM [2.2, 6.9]	366 nM [231, 584]
7X	13.3 μM [8.1, 21.7]	2.2 μM [1.4, 3.4]
8N	2.0 μM [1.4, 2.7]	278 nM [203, 379]
8X	1.2 μM [0.9, 1.7]	170 nM [121, 238]
RJW100	3.3 μM [1.9, 5.7]	316 nM [179, 555]
Cpd3	7.8 μM [4.8, 12.6]	2.4 μM [0.9, 5.2]
PME9	34.6 μM [12.3, 93.5]	7.0 μM [3.8, 13.0]
SID7969543	cnc	cnc
SR1848	cnc	cnc

Supplementary Figure S2.1. Summary K_i table. K_d and K_i values are presented in the order in which they appear in the manuscript. Experiments were analyzed in Graphpad Prism, v7 using a one-site fit K_i curve (n=8). 95% confidence are reported for each K_i value; cnc (could not calculate) indicates that confidence intervals or K_i values could not be calculated.


R
H
Endo
R
H
Exo

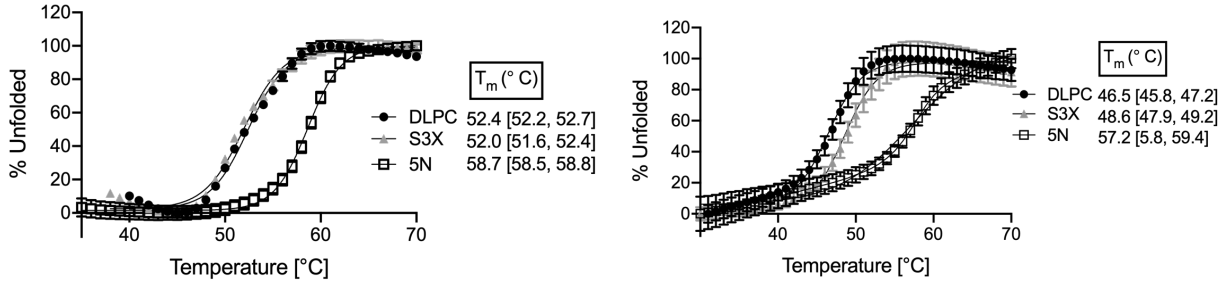

% 6N-FAM Bound
log[1N, 1X] (M)
log[S1] (M)
Ki = 4.5 μM [1.8, 10.7]
log[2N, 2X] (M)
Ki = 18.6 μM [9.3, 36.8]
log[S2N, S2X] (M)
Ki = 2.3 μM [0.7, 13.2]
Ki = 43.9 μM [16.2, 234]
log[3N, 3X] (M)
Ki = 1.6 μM
[0.9, 2.8]
Ki = 1.3 μM
[0.8, 1.9]
log[S3N, S3X] (M)
Ki = 2.3 μM [0.2, 43.1]
Ki = 4.9 μM [0.8, 31.7]
log[4N, 4X] (M)
Ki = 5.6 μM [2.4, 19.6]
Ki = 5.4 μM [2.1, 14.5]
log[S4N, S4X] (M)
Ki = 452 nM [161, 1220]
Ki = 52.8 μM [21, 269]
log[5N, 5X] (M)
Ki = 21.9 μM [12.4, 39.5]
log[6N, 6X] (M)
Ki = 1.4 μM [0.8, 2.4]
log[7N, 7X] (M)
Ki = 3.7 μM [2.2, 6.9]
Ki = 13.3 μM [8.1, 21.7]
log[8N, 8X] (M)
Ki = 2.0 μM
[1.4, 2.7]
Ki = 1.2 μM
[0.9, 1.7]

Supplementary Figure 2.2. SF-1 binding to R1 compounds. Curves for SF-1 binding to R^1 compounds are shown, with *endo* stereoisomers of the R^1 substituent in teal and *exo* stereoisomers in black. Experiments analyzed in Graphpad Prism, v7 using a one-site fit K_i curve (n=8). 95% confidence are reported for each K_i value; curves are not shown if confidence intervals or K_i values could not be calculated. Error bars are shown as SEM.



Supplementary Figure S2.3. LRH-1 binding to R[1] compounds. Curves for LRH-1 binding to R[1] compounds are shown, with *endo* stereoisomers of the R[1] substituent in teal and *exo* stereoisomers in black. Experiments were analyzed in Graphpad Prism, v7 using a one-site fit K_i curve (n=8). 95% confidence

are reported for each K_i value; curves are not shown if confidence intervals or K_i values could not be calculated. Error bars are shown as SEM.



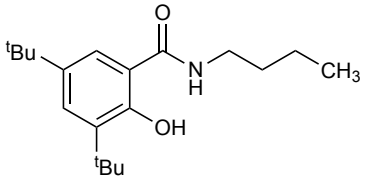
Supplementary Figure S2.4. Representative thermal shift curves. Left, SF-1; right, LRH-1. Representative melting curves for DLPC and two synthetic agonists are shown. S3X stabilizes LRH-1, but not SF-1, whereas 5N stabilizes both receptors. Experiments were analyzed in Graphpad Prism, v7 using the Boltzman equation (n=9). 95% confidence intervals are shown for each T_m value; error bars are shown as SEM.

Supplementary Figure S2.5. Detailed Chemical Syntheses

Chemical synthesis

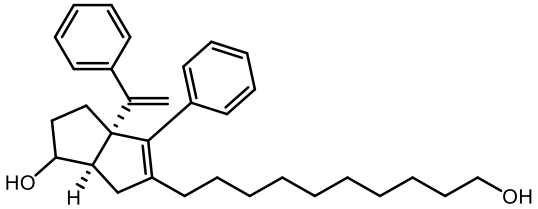
All reactions were carried out in oven-dried glassware, equipped with a stir bar and under a nitrogen atmosphere with dry solvents under anhydrous conditions, unless otherwise noted. Solvents used in anhydrous reactions were purified by passing over activated alumina and storing under argon. Yields refer to chromatographically and spectroscopically (^{1}H NMR) homogenous materials, unless otherwise stated. Reagents were purchased at the highest commercial quality and used without further purification, unless otherwise stated. *n*-Butyllithium (*n*-BuLi) was used as a 2.5 M solution in hexanes (Aldrich), was stored at 4 °C and titrated prior to use. Organic solutions were concentrated under reduced pressure on a rotary evaporator using a water bath. Chromatographic purification of products was accomplished using forced-flow chromatography on 230-400 mesh silica gel. Preparative thin-layer chromatography (PTLC) separations were carried out on 1000μm SiliCycle silica gel F-254 plates. Thin-layer chromatography (TLC) was performed on 250μm SiliCycle silica gel F-254 plates. Visualization of the developed chromatogram was performed by fluorescence quenching or by staining using $KMnO_4$, p-anisaldehyde, or ninhydrin stains. ^{1}H and ^{13}C NMR spectra were obtained from the Emory University NMR facility and recorded on a INOVA 600 (600 MHz), INOVA 500 (500 MHz), INOVA 400 (400 MHz), VNMR 400 (400 MHz), or Mercury 300 (300 MHz), and are internally referenced to residual protio solvent signals. Data for ^{1}H NMR are reported as follows: chemical shift (ppm), multiplicity (s = singlet, d = doublet, t = triplet, q = quartet, m = multiplet, dd = doublet of doublets, dt = doublet of triplets, ddd= doublet of doublet of doublets, dtd= doublet of triplet of doublets, b = broad, etc.), coupling constant (Hz), integration, and assignment, when applicable. Data for decoupled ^{13}C NMR are reported in terms of chemical shift and multiplicity when applicable. Liquid chromatography mass spectrometry (LC-MS) was performed on an Agilent 6120 mass spectrometer with an Agilent 1220 Infinity liquid chromatography inlet. Preparative high pressure liquid chromatography (Prep-HPLC) was performed on an Agilent 1200 Infinity Series

chromatograph using an Agilent Prep-C18 30 x 250 mm 10 μm column. HPLC analyses were performed using the following conditions.
Method A: A linear gradient using water and 0.1 % formic acid (FA) (Solvent A) and MeCN and 0.1% FA (Solvent B); t = 0 min, 75% B, t = 4 min, 99% B (held for 1 min), then 50% B for 1 min, was employed on Agilent Zorbax SB-C18 1.8 micron, 2.1 mm x 50 mm column (flow S2 rate 0.8 mL/min). The UV detection was set to 254 nm. The LC column was maintained at ambient temperature.
Method B: A linear gradient using water and 0.1 % formic acid (FA) (Solvent A) and MeCN and 0.1% FA (Solvent B); t = 0 min, 50% B, t = 4 min, 99% B (held for 1 min), then 50% B for 1 min, was employed on Agilent Zorbax SB-C18 1.8 micron, 2.1 mm x 50 mm column (flow S2 rate 0.8 mL/min). The UV detection was set to 254 nm. The LC column was maintained at ambient temperature.
Method C: An isocratic method using 60% MeCN, 40% water, and 0.1 % FA was employed on an Agilent Zorbax SB-C18 1.8 micron, 2.1 mm x 50 mm column (flow rate 0.8 mL/min). The UV detection was set to 254 nm. The LC column was maintained at ambient temperature.



PME9 3,5-di-*tert*-butyl-*N*-butyl-2-hydroxybenzamide (**PME9**)
To a round bottom flask charged with stir bar was added 3,5-di-*tert*-butyl-2-hydroxybenzoic acid (311 mg, 1.2 mmol, 1.0 equiv) in THF. 4-nitrophenylchloroformate (284 mg, 1.4 mmol, 1.1 equiv) was added, followed by triethylamine (200 μL, 1.4, 1.1 equiv). The reaction was stirred at room temperature for 1 h. The volatiles were concentrated and the crude residue was partitioned between ethyl acetate and water and extracted three times with ethyl acetate. The combined organic layers were washed with $NaHCO_3$ and brine, dried with $MgSO_4$, filtered, and concentrated. The crude activated ester was then dissolved in THF and treated with n-butylamine (140 μL, 1.4mmol, 1.1 equiv) and triethylamine (210 μL, 1.5mmol, 1.2 equiv). The reaction was stirred overnight. After reaction completion, the solution was concentrated and subjected to silica gel chromatography in 10% EtOAc/hexanes to afford the title compound as a colorless solid (162 mg, 43% yield over 2 steps). Spectral data were consistent with literature values from: de Jesus Cortez F, Suzawa M, Irvy S, Bruning JM, Sablin E, Jacobson MP, et al. Disulfide-Trapping Identifies a New, Effective Chemical Probe for Activating the Nuclear Receptor Human LRH-1 (NR5A2). *PLoS ONE,* **2016**, 11(*7*): e0159316
^{1}H NMR (600 MHz, $CDCl_3$) δ 12.73 (s, 1H), 7.44 (d, *J* = 2.3 Hz, 1H), 7.10 (d, *J* = 2.3 Hz, 1H), 6.24 (s, 1H), 3.58 – 3.34 (m, 2H), 1.66 – 1.57 (m, 2H), 1.45 – 1.35 (m, 11H), 1.29 (s, 9H), 0.94 (t, *J* = 7.5 Hz, 3H).

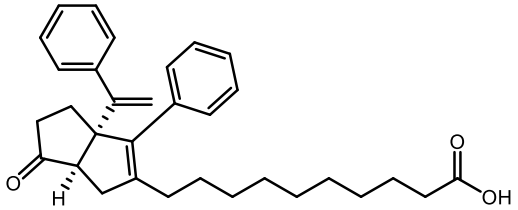
HPLC Method A, **LRMS** (ESI, APCI) *m/z*: calc'd for $C_{19}H_{32}NO_2$ $(M+H)^+$ 306.2, found 305.9.
Purity established by HPLC Method A: >99%.



5-(10-hydroxydecyl)-4-phenyl-3a-(1-phenylvinyl)-1,2,3,3a,6,6a-hexahydropentalen-1-ol (1): A slight modification of the procedure of Flynn et al. was used. Prior to use in the reaction, all reagents were dried by azeotropic removal of water using benzene. A dry round bottom flask containing bis(cyclopentadienyl)zirconium(IV) dichloride (1.403 g, 4.8 mmol, 1.2 equiv) under nitrogen, was dissolved in anhydrous, degassed tetrahydrofuran (THF, 5 mL/mmol enyne) and cooled to -78 °C. The resulting solution was treated with *n*-BuLi (3.84 mL, 9.6 mmol, 2.4 equiv.) and the light yellow solution was stirred for 50 minutes. A solution of **tert-butyldimethyl((7-phenylhept-1-en-6-yn-3-yl)oxy)silane** (1.202 g, 4.0 mmol, 1.0 equiv) in anhydrous, degassed THF (5 mL/mmol) was added. The resulting salmon-colored mixture was stirred at -78 °C for 45 minutes, the cooling bath removed, and the reaction mixture was allowed to warm to ambient temperature with stirring (2.5 hours total). The reaction mixture was then cooled to -78 °C for 15 minutes and **tert-butyl((10,10-dibromodecyl)oxy)diphenylsilane** (2.492 g, 4.4 mmol, 1.1 equiv) was added as a solution in anhydrous THF (5 mL/mmol) followed by freshly prepared lithium diisopropylamide (LDA, 4.4 mL, 4.4 mmol, 1.0 M, 1.1 equiv.). After 30 minutes, a freshly prepared solution of lithium phenylacetylide (14.4 mmol, 3.6 equiv.) in anhydrous THF (2 mL/mmol) was added dropwise and the resulting rust-colored solution was stirred at -78 °C for 1 hour. The reaction was quenched with methanol and saturated aqueous sodium bicarbonate and allowed to warm to room temperature, affording a light yellow slurry that stirred overnight. The slurry was then poured onto water and extracted with ethyl acetate four times. The combined organic layers were washed with brine, dried with Na_2SO_4, filtered, and concentrated *in vacuo* to afford a crude mixture. The resulting crude mixture was dissolved in 200 mL of 1:1 DCM:MeOH in a round bottom flask then 0.5 mL of concentrated HCl added. The resulting solution was stirred at room temperature for 2.5 hours before concentrating *in vacuo* and subjecting to silica gel chromatography (5-50% EtOAc/hexanes eluent) to afford the title compound as a yellow oil and 1.7:1 mixture of diastereomers used in the next step without separation. (1.47 g, 80% over 2 steps).

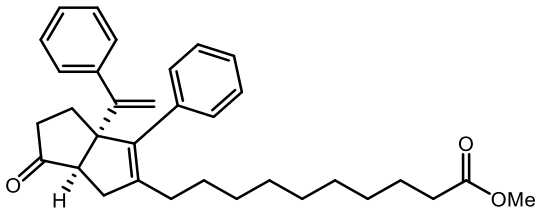
Exo diastereomer: **^{1}H NMR** (600 MHz, $CDCl_3$) δ 7.37 – 7.23 (m, 8H), 7.20 (t, *J* = 8.0 Hz, 2H), 5.07 (d, *J* = 1.4 Hz, 1H), 4.99 (d, *J* = 1.4 Hz, 1H), 3.96 – 3.93 (m, 1H), 3.64 (t, *J* = 6.6 Hz, 2H), 2.36 (dd, *J* = 16.9, 9.3 Hz, 1H), 2.29 (dd, *J* = 9.3, 1.8 Hz, 1H), 2.13 – 1.98 (m, 4H), 1.75 – 1.63 (m, 2H), 1.56 (p, *J* = 6.8 Hz, 3H), 1.43 – 1.17 (m, 14H).

Endo diastereomer: **^{1}H NMR** (600 MHz, $CDCl_3$) δ 7.37 – 7.23 (m, 8H), 7.20 (t, *J* = 7.5 Hz, 2H), 5.07 (d, *J* = 1.4 Hz, 1H), 4.93 (d, *J* = 1.4 Hz, 1H), 4.18 (td, *J* = 8.9, 5.6, 1H), 3.64 (t, *J* = 6.6 Hz, 2H), 2.62 (dd, *J* = 17.5, 2.1 Hz, 1H), 2.48 (td, *J* = 8.7, 2.0 Hz, 1H), 2.13 – 1.98 (m, 4H), 1.84 (dq, *J* = 10.0, 4.9 Hz, 1H), 1.75 – 1.63 (m, 2H), 1.56 (p, *J* = 6.8 Hz, 2H), 1.43 – 1.17 (m, 14H).

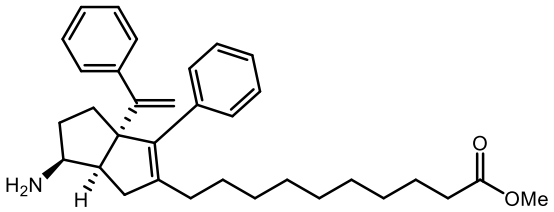


10-(-6-oxo-3-phenyl-3a-(1-phenylvinyl)-1,3a,4,5,6,6a-hexahydropentalen-2-yl)decanoic acid: To a solution of **5-(10-hydroxydecyl)-4-phenyl-3a-(1-phenylvinyl)-1,2,3,3a,6,6a-hexahydropentalen-1-ol (1)** in acetonitrile (592 mg, 1.3 mmol, 0.1 M) was added tetrapropylammonium perruthenate (45.3 mg, 0.13 mmol, 0.1 equiv.), *N*-methylmorpholine *N*-oxide (2.29 g, 12.9 mmol, 10 equiv.), and water (0.24 mL, 12.9 mmol, 10 equiv.) and stirred at room temperature overnight. The reaction solution was then filtered through a pad of silica with 99:1 EtOAc:AcOH to collect the title compound as a yellow oil (608 mg, quant.).

^{1}H NMR (500 MHz, $CDCl_3$) δ 7.40 – 7.19 (m, 10H), 5.22 (d, *J* = 1.5 Hz, 1H), 5.11 (d, *J* = 1.4 Hz, 1H), 2.46 (d, *J* = 7.8 Hz, 1H), 2.36 – 2.26 (m, 4H), 2.16 – 1.95 (m, 5H), 1.91 (dd, *J* = 16.5, 7.8 Hz, 1H), 1.61 (p, *J* = 7.5 Hz, 2H), 1.46 – 0.97 (m, 12H). Carboxylic acid proton (-COO**H**) not observed.
^{13}C NMR (126 MHz, $CDCl_3$) δ 223.0, 179.9, 153.3, 145.0, 142.6, 137.5, 136.8, 129.1, 128.4, 128.2, 127.7, 127.2, 127.1, 115.4, 65.6, 55.7, 38.9, 37.6, 34.2, 30.1, 29.8, 29.4, 29.3, 29.1, 28.5, 27.7, 24.8.
HPLC method A, **LRMS** (ESI, APCI) *m/z*: calc'd for $C_{32}H_{39}O_3$ $(M+H)^+$ 471.3, found 470.8.

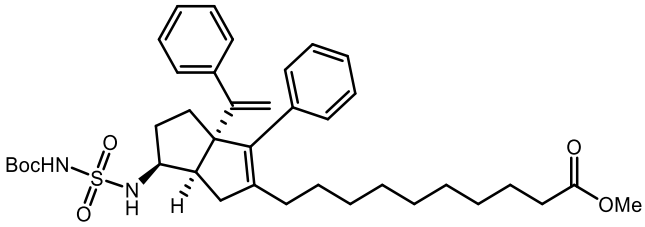


methyl 10-(6-oxo-3-phenyl-3a-(1-phenylvinyl)-1,3a,4,5,6,6a-hexahydropentalen-2-yl)decanoate (2): To a solution of **10-(-6-oxo-3-phenyl-3a-(1-phenylvinyl)-1,3a,4,5,6,6a-hexahydropentalen-2-yl)decanoic acid** in methanol (945 mg, 2 mmol, 0.1 M) was added 5 drops of concentrated HCl and stirred at room temperature overnight. Reaction solution was then concentrated *in vacuo* and filtered through a pad of silica to collect the title compound as a yellow oil (930 mg, 96%).
^{1}H NMR (500 MHz, $CDCl_3$) δ 7.40 – 7.19 (m, 10H), 5.22 (d, *J* = 1.3 Hz, 1H), 5.11 (d, *J* = 1.3 Hz, 1H), 3.66 (s, 3H), 2.46 (d, *J* = 7.7 Hz, 1H), 2.34 – 2.25 (m, 4H), 2.16 – 1.95 (m, 5H), 1.91 (dd, *J* = 16.5, 7.8 Hz, 1H), 1.60 (p, *J* = 7.5 Hz, 2H), 1.33 – 1.10 (m, 12H).
^{13}C NMR (126 MHz, $CDCl_3$) δ 222.7, 174.4, 153.3, 144.9, 142.6, 137.5, 136.7, 129.0, 128.3, 128.2, 127.7, 127.1, 127.1, 115.3, 65.5, 55.6, 51.5, 38.8, 37.6, 34.2, 30.0, 29.7, 29.4, 29.33, 29.27, 29.2, 28.4, 27.7, 25.0.
HPLC method A, **LRMS** (ESI, APCI) *m/z*: calc'd for $C_{33}H_{41}O_3$ $(M+H)^+$ 485.3, found 484.9.

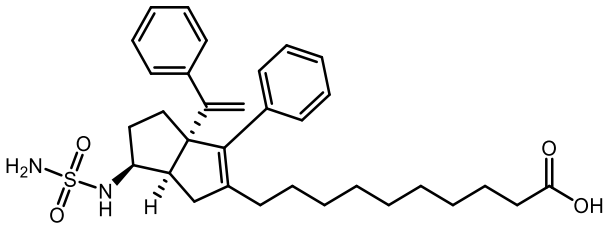


methyl 10-(6-amino-3-phenyl-3a-(1-phenylvinyl)-1,3a,4,5,6,6a-hexahydropentalen-2-yl)decanoate (3): To a screw top test tube charged with a stir bar was added **methyl 10-(6-oxo-3-phenyl-3a-(1-phenylvinyl)-1,3a,4,5,6,6a-hexahydropentalen-2-yl)decanoate (2)** (350 mg, 0.72 mmol, 1.0 equiv.) and ethanol (3 mL) and sealed. Ammonia (7 M in methanol, 0.52 mL, 3.61 mmol, 5.0 equiv.) then titanium(IV) isopropoxide (0.33 mL, 1.08 mmol, 1.5 equiv.) were added via syringe and stirred at room temperature for 6 hours. The test tube cap was then removed and sodium borohydride (82 mg, 2.16 mmol, 3 equiv.) added portion-wise. The resulting solution was stirred at room temperature overnight before being quenched with EtOAc, saturated aqueous potassium sodium tartrate, and 2 M aqueous sodium hydroxide. The resulting slurry was then sonicated in the reaction tube for 10 minutes before adding to a separatory funnel. The aqueous layer was then drained and remaining EtOAc washed with 2 x 20 mL of aqueous potassium sodium tartrate and 2M sodium hydroxide then 20 mL water and 20 mL brine. The remaining organic layer was then dried over Na_2SO_4, filtered, and concentrated *in vacuo* to give the title compound as a yellow oil (283 mg, 81%).
^{1}H NMR (500 MHz, $CDCl_3$) δ 7.35 – 7.22 (m, 8H), 7.20 (t, *J* = 2.0 Hz, 1H), 7.18 (t, *J* = 1.5 Hz, 1H), 5.07 (d, *J* = 1.4 Hz, 1H), 4.93 (d, *J* = 1.5 Hz, 1H), 3.66 (s, 3H), 3.31 (td, *J* = 8.7, 5.8 Hz, 1H), 2.49 –

2.40 (m, 2H), 2.29 (t, J = 7.6 Hz, 2H), 2.12 – 2.00 (m, 4H), 1.85 – 1.78 (m, 1H), 1.72 – 1.67 (m, 2H), 1.60 (p, J = 7.5 Hz, 2H), 1.42 – 1.15 (m, 12H). Amine protons (-N**H**$_2$) not observed.
^{13}C NMR (126 MHz, $CDCl_3$) δ 174.5, 155.1, 144.3, 143.0, 139.5, 137.2, 129.9, 127.9, 127.8, 127.7, 126.8, 126.7, 115.3, 69.6, 55.3, 51.6, 34.4, 34.3, 33.3, 30.0, 29.9, 29.52, 29.50, 29.4, 29.3, 28.1, 25.1.
HPLC method A, **LRMS** (ESI, APCI) m/z: calc'd for $C_{33}H_{44}NO_2$ $(M+H)^+$ 486.3, found 485.8

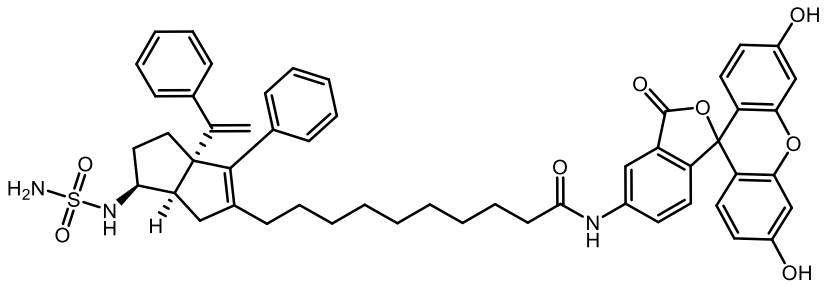


methyl 10-(6-((N-(tert-butoxycarbonyl)sulfamoyl)amino)-3-phenyl-3a-(1-phenylvinyl)-1,3a,4,5,6,6a-hexahydropentalen-2-yl)decanoate (4): To a solution of *tert*-butyl alcohol (47 mg, 0.64 mmol, 1.1 equiv.) in anhydrous DCM (0.6 mL) in an oven-dried flask under nitrogen at 0 °C was added neat chlorosulfonylisocyanate (0.050 mL, 0.58 mmol, 1.0 equiv.) and stirred for 45 minutes, warming to room temperature in that time. The resulting solution was then added via syringe to a solution of **methyl 10-(6-amino-3-phenyl-3a-(1-phenylvinyl)-1,3a,4,5,6,6a-hexahydropentalen-2-yl)decanoate (3)** (283 mg, 0.58 mmol, 1.0 equiv.) and triethylamine (0.12 mL, 0.87 mmol, 1.5 equiv.) in anhydrous DCM (0.6 mL) under nitrogen in an oven-dried flask at 0 °C. The reaction was then stirred and warmed to room temperature over 3 hours before diluting with DCM and washing with 2 x 10 mL 0.5 M aqueous HCl, 10 mL water and 10 mL brine. The organic layer was then dried over Na_2SO_4, filtered, and concentrated *in vacuo* to give crude material. This material was subjected to silica gel chromatography (10-40% EtOAc/hexanes) to collect material taken crude to the next step.

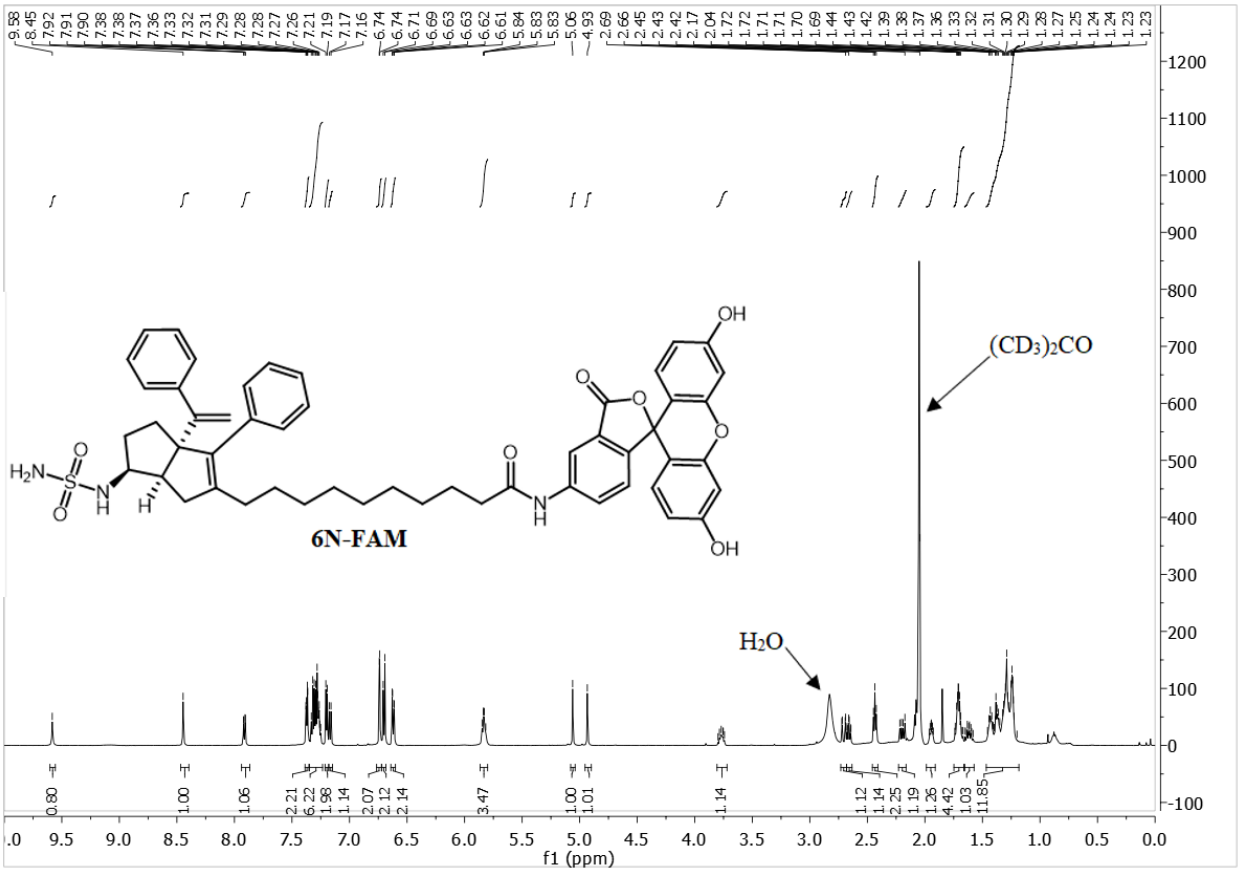


10-(3-phenyl-3a-(1-phenylvinyl)-6-(sulfamoylamino)-1,3a,4,5,6,6a-hexahydropentalen-2-yl)decanoic acid (5): A 20 mL scintillation vial was charged with a stir bar and **methyl 10-(6-((N-(tert-butoxycarbonyl)sulfamoyl)amino)-3-phenyl-3a-(1-phenylvinyl)-1,3a,4,5,6,6a-hexahydropentalen-2-yl)decanoate (4)** (160 mg, 0.24 mmol) and cooled to 0 °C. A 3:1 v/v solution of dioxane and concentrated HCl (2 mL) was then added and allowed to warm to room temperature and stirred for 24 hours before heating to 40 °C for 14 hours. The reaction solution was then diluted with EtOAc and washed with 3 x 5 mL 0.5 M aqueous HCl, 5 ml water, and 5 mL brine. The organic layer was then dried over Na_2SO_4, filtered, and concentrated *in vacuo* to give the title compound as a brown oil (94 mg, 29% over 2 steps).
^{1}H NMR (500 MHz, $CDCl_3$) δ 7.35 – 7.22 (m, 8H), 7.19 (t, J = 2.0 Hz, 1H), 7.17 (t, J = 1.6 Hz, 1H), 5.10 (d, J = 1.3 Hz, 1H), 4.95 (d, J = 1.4 Hz, 1H), 4.84 (d, J = 7.8 Hz, 1H), 4.72 (s, 2H), 3.78 (dtd, J = 11.2, 8.3, 6.0 Hz, 1H), 2.62 (td, J = 8.9, 2.1 Hz, 1H), 2.42 (dd, J = 17.7, 2.1 Hz, 1H), 2.34 (t, J = 7.4 Hz, 2H), 2.16 (dd, J = 17.6, 9.0 Hz, 1H), 2.11 – 2.00 (m, 2H), 1.99 – 1.92 (m, 1H), 1.75 – 1.68 (m,

2H), 1.62 (p, *J* = 7.4 Hz, 2H), 1.53 – 1.42 (m, 1H), 1.43 – 1.18 (m, 12H). Carboxylic acid proton (-COO**H**) not observed.
^{13}C NMR (126 MHz, $CDCl_3$) δ 179.3, 154.3, 143.8, 143.01 139.3, 136.8, 129.8, 128.0, 127.8, 127.0, 126.9, 115.7, 68.9, 57.2, 47.5, 35.6, 34.0, 32.6, 31.8, 29.9, 29.7, 29.2, 29.1, 29.0, 28.9, 27.9, 24.6.
HPLC method B, **LRMS** (ESI, APCI) *m/z*: calc'd for $C_{32}H_{43}N_2O_4S$ $(M+H)^+$ 551.3, found 551.8.



N-(3',6'-dihydroxy-3-oxo-3H-spiro[isobenzofuran-1,9'-xanthen]-5-yl)-10-(3-phenyl-3a-(1-phenylvinyl)-6-(sulfamoylamino)-1,3a,4,5,6,6a-hexahydropentalen-2-yl)decanamide (6N-FAM): A 1 dram vial was charged with a stirbar, **10-(3-phenyl-3a-(1-phenylvinyl)-6-(sulfamoylamino)-1,3a,4,5,6,6a-hexahydropentalen-2-yl)decanoic acid (5)** (28.6 mg, 0.05 mmol, 1.0 equiv.), fluoresceinamine isomer 1 (21 mg, 0.06 mmol, 1.2 equiv.), EDCI (11.5 mg, 0.06 mmol, 1.2 equiv.) and DMF (0.65 mL). The resulting solution was stirred at room temperature for 16 hours before diluting with MeCN and subjecting to preparative HPLC to collect the title compound. (9.9 mg, 22%).
^{1}H NMR (600 MHz, $(CD_3)_2CO$) δ 9.58 (s,1H), 8.45 (s, 1H), 7.91 (dd, J = 8.3, 1.9 Hz, 1H), 7.39 – 7.35 (m, 2H), 7.35 – 7.23 (m, 6H), 7.20 (d, J = 7.5 Hz, 2H), 7.17 (d, J = 8.3 Hz, 1H), 6.74 (d, J = 2.3 Hz, 2H), 6.70 (d, J = 8.7 Hz, 2H), 6.62 (dd, J = 8.7, 2.4 Hz, 2H), 5.86 – 5.80 (m, 3H), 5.06 (s, 1H), 4.93 (s, 1H), 3.81 – 3.72 (m, 1H), 2.71 (dd, J = 17.6, 2.2 Hz, 1H), 2.66 (td, J = 9.0, 2.2 Hz, 1H), 2.43 (t, J = 7.4 Hz, 2H), 2.20 (dd, J = 17.5, 9.1 Hz, 1H), 1.95 (m, 1H), 1.75 – 1.66 (m, 4H), 1.62 (qd, J = 11.1, 7.0 Hz, 1H), 1.47 – 1.18 (m, 12H). Phenolic protons (Ar-O**H**) not observed.
^{13}C NMR (151 MHz, $(CD_3)_2CO$) 172.8, 169.6, 160.6, 156.3, 153.6, 145.1, 144.4, 142.2, 139.9, 138.2, 130.8, 130.39, 130.35, 130.31, 129.0, 128.8, 128.7, 128.6, 127.8, 127.7, 125.3, 115.6, 113.5, 112.0, 103.5, 103.4, 69.9, 58.0, 48.6, 37.9, 36.6, 33.7.
HPLC method C, **LRMS** (ESI, APCI) *m/z*: calc'd for $C_{52}H_{52}N_3O_8S$ $(M-H)^-$ 878.4, found 878.1 $(M-H)^-$.
Purity established by HPLC Method C: 96%



Chemical Sources

SR1848 and Cpd3 were gifted from Patrick Griffin (Scripps University). SID7969543 was purchased from Tocris. DLPC was purchased from Avanti Polar Lipids. PI(4,5)P_2 was purchased from Cayman Chemical.

References

1. Rask-Andersen, M.; Almen, M. S.; Schioth, H. B. Trends in the exploitation of novel drug targets. Nat. Rev. Drug Discovery 2011, 10 (8), 579−590.

2. Zhi, X.; Zhou, X. E.; Melcher, K.; Xu, H. E. Structures and regulation of non-X orphan nuclear receptors: A retinoid hypothesis. J. Steroid Biochem. Mol. Biol. 2016, 157, 27−40.

3. Kelly, M. E.; Mohan, H. M.; Baird, A. W.; Ryan, E. J.; Winter, D. C. Orphan Nuclear Receptors in Colorectal Cancer. Pathol. Oncol. Res. 2018, 24 (4), 815−819.

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

5. Majdic, G.; Young, M.; Gomez-Sanchez, E.; Anderson, P.; Szczepaniak, L. S.; Dobbins, R. L.; McGarry, J. D.; Parker, K. L. Knockout mice lacking steroidogenic factor 1 are a novel genetic model of hypothalamic obesity. Endocrinology 2002, 143 (2), 607− 614.

6. Peng, N.; Kim, J. W.; Rainey, W. E.; Carr, B. R.; Attia, G. R. The role of the orphan nuclear receptor, liver receptor homologue-1, in the regulation of human corpus luteum 3beta-hydroxysteroid dehydrogenase type II. J. Clin. Endocrinol. Metab. 2003, 88 (12), 6020−6028.

7. Kim, J. W.; Peng, N.; Rainey, W. E.; Carr, B. R.; Attia, G. R. Liver receptor homolog-1 regulates the expression of steroidogenic acute regulatory protein in human granulosa cells. J. Clin. Endocrinol. Metab. 2004, 89 (6), 3042−3047.

8. Clyne, C. D.; Speed, C. J.; Zhou, J.; Simpson, E. R. Liver receptor homologue-1 (LRH-1) regulates expression of aromatase in preadipocytes. J. Biol. Chem. 2002, 277 (23), 20591−20597.

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

10. Oosterveer, M. H.; Mataki, C.; Yamamoto, H.; Harach, T.; Moullan, N.; van Dijk, T. H.; Ayuso, E.; Bosch, F.; Postic, C.; Groen, A. K.; Auwerx, J.; Schoonjans, K. LRH-1-dependent glucose sensing determines intermediary metabolism in liver. J. Clin. Invest. 2012, 122 (8), 2817−2826.

11. Stein, S.; Oosterveer, M. H.; Mataki, C.; Xu, P.; Lemos, V.; Havinga, R.; Dittner, C.; Ryu, D.; Menzies, K. J.; Wang, X.; Perino, A.; Houten, S. M.; Melchior, F.; Schoonjans, K. SUMOylation-dependent LRH-1/PROX1 interaction promotes atherosclerosis by decreasing hepatic reverse cholesterol transport. Cell Metab. 2014, 20 (4), 603− 613.

12. Lu, T.; Makishima, M.; JJ, R.; Schoonjans, K.; Kerr, T.; Auwerx, J.; DJ, M. Molecular Basis for Feedback Regulation of Bile Acid Synthesis by Nuclear Receptors. Mol. Cell 2000, 6, 507−515.

13. Shinoda, K.; Lei, H.; Yoshii, H.; Nomura, M.; Nagano, M.; Shiba, H.; Sasaki, H.; Osawa, Y.; Ninomiya, Y.; Niwa, O.; et al. Developmental defects of the ventromedial hypothalamic nucleus and pituitary gonadotroph in the Ftz-F1 disrupted mice. Dev. Dyn. 1995, 204 (1), 22−29.

14. Gu, P.; Goodwin, B.; Chung, A. C.; Xu, X.; Wheeler, D. A.; Price, R. R.; Galardi, C.; Peng, L.; Latour, A. M.; Koller, B. H.; Gossen, J.; Kliewer, S. A.; Cooney, A. J. Orphan nuclear receptor LRH-1 is required to maintain Oct4 expression at the epiblast stage of embryonic development. Molecular and cellular biology 2005, 25 (9), 3492−3505.

15. Doghman, M.; Karpova, T.; Rodrigues, G. A.; Arhatte, M.; De Moura, J.; Cavalli, L. R.; Virolle, V.; Barbry, P.; Zambetti, G. P.; Figueiredo, B. C.; Heckert, L. L.; Lalli, E. Increased steroidogenic factor-1 dosage triggers adrenocortical cell proliferation and cancer. Mol. Endocrinol. 2007, 21 (12), 2968−2987.

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

17. Chand, A. L.; Herridge, K. A.; Thompson, E. W.; Clyne, C. D. The orphan nuclear receptor LRH-1 promotes breast cancer motility and invasion. Endocr.-Relat. Cancer 2010, 17 (4), 965−975.

18. Bayrer, J. R.; Mukkamala, S.; Sablin, E. P.; Webb, P.; Fletterick, R. J. Silencing LRH-1 in colon cancer cell lines impairs proliferation and alters gene expression programs. Proc. Natl. Acad. Sci. U. S. A. 2015, 112 (8), 2467−2472.

19. Lin, Q.; Aihara, A.; Chung, W.; Li, Y.; Huang, Z.; Chen, X.; Weng, S.; Carlson, R. I.; Wands, J. R.; Dong, X. LRH1 as a driving factor in pancreatic cancer growth. Cancer Lett. 2014, 345 (1), 85−90.

20. Lee, J. M.; Lee, Y. K.; Mamrosh, J. L.; Busby, S. A.; Griffin, P. R.; Pathak, M. C.; Ortlund, E. A.; Moore, D. D. A nuclear-receptor- dependent phosphatidylcholine pathway with antidiabetic effects. Nature 2011, 474 (7352), 506−510.

21. Blind, R. D.; Sablin, E. P.; Kuchenbecker, K. M.; Chiu, H. J.; Deacon, A. M.; Das, D.; Fletterick, R. J.; Ingraham, H. A. The signaling phospholipid PIP3 creates a new interaction surface on the nuclear receptor SF-1. Proc. Natl. Acad. Sci. U. S. A. 2014, 111 (42), 15054−15059.

22. Sablin, E. P.; Blind, R. D.; Uthayaruban, R.; Chiu, H. J.; Deacon, A. M.; Das, D.; Ingraham, H. A.; Fletterick, R. J. Structure of Liver Receptor Homolog-1 (NR5A2) with PIP hormone bound in the ligand binding pocket. J. Struct. Biol. 2015, 192 (3), 342−348.

23. Benod, C.; Carlsson, J.; Uthayaruban, R.; Hwang, P.; Irwin, J. J.; Doak, A. K.; Shoichet, B. K.; Sablin, E. P.; Fletterick, R. J. Structure- based discovery of antagonists of nuclear receptor LRH-1. J. Biol. Chem. 2013, 288 (27), 19830−19844.

24. Musille, P. M.; Kossmann, B. R.; Kohn, J. A.; Ivanov, I.; Ortlund, E. A. Unexpected Allosteric Network Contributes to LRH-1 Coregulator Selectivity. J. Biol. Chem. 2016, 291 (3), 1411−1426.

25. de Jesus Cortez, F.; Suzawa, M.; Irvy, S.; Bruning, J. M.; Sablin, E.; Jacobson, M. P.; Fletterick, R. J.; Ingraham, H. A.; England, P. M. Disulfide-Trapping Identifies a New, Effective Chemical Probe for Activating the Nuclear Receptor Human LRH-1 (NR5A2). PLoS One 2016, 11 (7), No. e0159316.

26. Whitby, R. J.; Dixon, S.; Maloney, P. R.; Delerive, P.; Goodwin, B.; Parks, D. J.; Wilson, T. M. Identification of Small Molecule Agonists of the Orphan Nuclear Receptros Liver Receoptor Homolog-1 and Steroidogenic Factor-1. J. Med. Chem. 2006, 49, 6652−6655.

27. Rey, J.; Hu, H.; Kyle, F.; Lai, C. F.; Buluwela, L.; Coombes, R. C.; Ortlund, E. A.; Ali, S.; Snyder, J. P.; Barrett, A. G. Discovery of a new class of liver receptor homolog-1 (LRH-1) antagonists: virtual screening, synthesis and biological evaluation. ChemMedChem 2012, 7 (11), 1909−1914.

28. Corzo, C. A.; Mari, Y.; Chang, M. R.; Khan, T.; Kuruvilla, D.; Nuhant, P.; Kumar, N.; West, G. M.; Duckett, D. R.; Roush, W. R.; Griffin, P. R. Antiproliferation activity of a small molecule repressor of liver receptor homolog 1. Mol. Pharmacol. 2015, 87 (2), 296−304.

29. Mays, S. G.; Flynn, A. R.; Cornelison, J. L.; Okafor, C. D.; Wang, H.; Wang, G.; Huang, X.; Donaldson, H. N.; Millings, E. J.; Polavarapu, R.; Moore, D. D.; Calvert, J. W.; Jui, N. T.; Ortlund, E. A. Development of the First Low Nanomolar Liver Receptor Homolog-1 Agonist through Structure-guided Design. J. Med. Chem. 2019, DOI: 10.1021/acs.jmedchem.9b00753.

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

31. Whitby, R. J.; Stec, J.; Blind, R. D.; Dixon, S.; Leesnitzer, L. M.; Orband-Miller, L. A.; Williams, S. P.; Willson, T. M.; Xu, R.; Zuercher, W. J.; Cai, F.; Ingraham, H. A. Small molecule agonists of the orphan nuclear receptors steroidogenic factor-1 (SF-1, NR5A1) and liver receptor homologue-1 (LRH-1, NR5A2). J. Med. Chem. 2011, 54 (7), 2266−2281.

32. Mays, S. G.; Okafor, C. D.; Whitby, R. J.; Goswami, D.; Stec, J.; Flynn, A. R.; Dugan, M. C.; Jui, N. T.; Griffin, P. R.; Ortlund, E. A. Crystal Structures of the Nuclear Receptor, Liver Receptor Homolog 1, Bound to Synthetic Agonists. J. Biol. Chem. 2016, 291 (49), 25281−25291.

33. Flynn, A. R.; Mays, S. G.; Ortlund, E. A.; Jui, N. T. Development of Hybrid Phospholipid Mimics as Effective Agonists for Liver Receptor Homologue-1. ACS Med. Chem. Lett. 2018, 9 (10), 1051−1056.

34. Fayard, E.; Auwerx, J.; Schoonjans, K. LRH-1: an orphan nuclear receptor involved in development, metabolism and steroido- genesis. Trends Cell Biol. 2004, 14 (5), 250−260.

35. Sablin, E. P.; Krylova, I. N.; Fletterick, R. J.; Ingraham, H. A. Structural basis for ligand-independent activation of the orphan nuclear receptor LRH-1. Mol. Cell 2003, 11 (6), 1575−1585.

36. Madoux, F.; Li, X.; Chase, P.; Zastrow, G.; Cameron, M. D.; Conkright, J. J.; Griffin, P. R.; Thacher, S.; Hodder, P. Potent, selective and cell penetrant inhibitors of SF-1 by functional ultra-high-throughput screening. Mol. Pharmacol. 2008, 73 (6), 1776−1784.

CHAPTER 3: TAPPING INTO A PHOSPHOLIPID-LRH-1 SIGNALING AXIS YIELDS A POWERFUL ANTI-INFLAMMATORY SIGNALING AGENT WITH *IN VIVO* ACTIVITY AGAINST COLITIS[1a]

Suzanne G. Mays[#*], Emma H. D'Agostino[#*], Autumn R. Flynn[&], Xiangcheng Wang[%], Guohui Wang[%], Xu Liu[#], Elizabeth J. Millings[#,$], C. Denise Okafor[#], Anamika Patel[#], M. Lee Cato[#], Jeffery L. Cornelison[&], Dianna Melchers[@], Rene Houtman[@], David D. Moore[%], John W. Calvert[$], Nathan T. Jui[&], and Eric A. Ortlund[#¤]

Departments of [#]Biochemistry, [&]Chemistry, and [$]Surgery, Emory University, Atlanta, GA 30322; [%]Department of Molecular and Cell Biology, Baylor College of Medicine, Houston, TX, [@]PamGene International B.V., 5211 HH Den Bosch, the Netherlands

*These authors contributed equally ¤Co-corresponding authors

This manuscript describes a series of LRH-1 small molecule agonists designed to mimic native phospholipid activation. Our previous agonists had a distinct binding mode from that of phospholipids, and in this work we demonstrated successful modification of our agonist scaffold to attain phospholipid-like binding and activation using extensive structural, biochemical, in cell, and *in vivo* studies. Our lead agonist from this effort exhibited striking *in vivo* efficacy in an ulcerative colitis mouse model, activating LRH-1 to decrease disease severity. A version of this manuscript has been deposited in the BioRxiv server, and it has been submitted for publication in Nature Chemical Biology.

---

[1]This chapter adapted from the BioRxiv preprint: Mays, S.G., D'Agostino, E.H., Flynn, A.R., Huang, X., Wang, G., Millings, E.J., Okafor, C.D., Patel, A., Cato, M.L., Cornelison, J.L., Melchers, D., Houtman, R., Moore, D.D., Calvert, J.W., Jui, N.T., Ortlund, E.A. Tapping into a phospholipid-LRH-1 axis yields a powerful anti-inflammatory agent with *in vivo* activity against colitis. *BioRXIV,* 2020, Sept 2.

[a]S.G.M., E.H.D., A.R.F., E.J.M., C.D.O., A.P., R.H., D.D.M., J.W.C., N.T.J., and E.A.O. participated in research design. S.G.M., E.H.D., A.R.F., X.H., G.W., X.L., E.J.M., C.D.O., A.P., M.L.C., J.L.C., and D.M. conducted experiments. S.G.M., E.H.D., A.R.F., X.H., G.W., X.L., E.J.M., C.D.O., A.P., M.L.C., J.L.C., D.M., and R.H. performed data analysis. S.G.M, E.H.D., and E.A.O. contributed to the writing of the manuscript. All authors have given approval to the final version of the manuscript. E.H.D. conducted all wildtype binding assays, generated substitutions in bacterial vector, conducted Y520 binding assays, solved both crystal structures, and contributed to the manuscript.

Abstract

As ligands for nuclear hormone receptors (NRs), phosphatidylcholines are powerful signaling molecules. Here, we demonstrate that mimicking phosphatidylcholine-NR interactions is a robust strategy for improving agonists of liver receptor homolog-1 (LRH-1), a promising therapeutic target for diabetes and colitis. Conventional LRH-1 modulators only occupy part of the binding pocket, leaving vacant a region important for phosphatidylcholine binding and allostery. Therefore, we constructed a set of hybrid molecules that incorporate elements of natural phosphatidylcholines into the scaffold of a synthetic LRH-1 agonist. The phosphatidylcholine-mimicking group increases LRH-1 binding affinity and transcriptional activity via formation of productive interactions with residues that coordinate the phosphatidylcholine headgroup. In organoid and *in vivo* models of colitis, the best new agonist upregulates LRH-1-controlled anti-inflammatory genes and significantly improves colonic histopathology and disease-related weight loss. This is the first evidence of *in vivo* efficacy for an LRH-1 modulator in colitis, a leap forward in agonist development.

Introduction

Phospholipids (PLs) comprise a diverse family of amphipathic lipids characterized by a phosphate-containing headgroup and two fatty acyl tails of varied lengths and saturation states. PLs play many roles in physiology, including forming lipid bilayers, signaling for apoptosis, and activating G-protein coupled receptors.[1-3] PLs are labile and are synthesized and catabolized in response to stimuli such as fed or fasting states, stress, and circadian rhythms,[4-6] yielding metabolites that perform specialized functions.

Phosphatidylcholines (PCs) are a subtype of PL that are the most abundant component of eukaryotic cellular membranes and are vital for maintaining membrane integrity and curvature.[7] In addition to these structural roles, PCs are important cell signaling mediators. Catabolism of PCs

produces bioactive molecules such as arachidonic acid and diacylglycerol.[8] PC catabolism also generates labile methyl groups used for DNA methylation and for synthesis of molecules such as nucleotides and amino acids.[9] Catabolism of PCs is not required for bioactivity: certain intact PCs exert broad effects on metabolic homeostasis by acting as ligands for nuclear hormone receptors (NRs).[4,10-15] As NR ligands, PCs represent a novel, understudied class of hormone.

Particularly strong evidence supports a role for PCs as ligands of liver receptor homolog-1 (LRH-1). LRH-1 is an orphan NR that regulates cholesterol homeostasis, metabolism, proliferation, and intestinal inflammation and has garnered attention as a novel therapeutic target for diseases such as diabetes, cancer, and inflammatory bowel diseases.[16-22] LRH-1 binds a wide range of PLs *in vitro,* and medium-chained, saturated PCs activate it exogenously.[16,23] Moreover, several models have shown a relationship between endogenous PC levels and LRH-1 activity. Diet-induced depletion of PCs in mice induces an "antagonist" pattern of gene expression in the liver similar to LRH-1 knockout mice.[24] Similarly, diet-induced depletion of PCs in *C. elegans* inhibits the worm LRH-1 ortholog, causing fat accumulation that is rescued by PC agonists.[25] In human hepatocytes, LRH-1 senses PCs generated via methyl metabolism to control beta-oxidation of fatty acids and mitochondrial biogenesis.[26] Finally, oral administration of dilauroylphosphatidylcholine (DLPC, PC 12:0/12:0) induces potent anti-diabetic effects in mouse models of insulin resistance. The anti-diabetic effects are LRH-1-dependent and occur via repression of *de novo* lipogenesis in the liver.[16]

The sensitivity of LRH-1 to PC levels suggests a regulatory circuit connecting PC availability to LRH-1-controlled gene expression. However, this circuit has been difficult to elucidate. Natural PCs are insoluble and rapidly metabolized, and unlikely to be LRH-1-selective, which makes them difficult to work with in the laboratory. These poor pharmacological properties also limit the use of PCs as therapeutics. We therefore designed a set of "PL-mimics" by fusing PL headgroups (or phosphate bioisosteres) to the synthetic LRH-1 agonist RJW100 (Figure 1a-b).[27] These hybrid

molecules contain the hexahydropentalene (6HP) core from RJW100, appended with a polar moiety that mimics a PC headgroup at position R^4 (Figure 3.1a-b).[27] Based on structural analyses, we hypothesized that the modified R^4 groups would make similar interactions to PCs in the LRH-1 binding pocket and stimulate receptor activity. Indeed, the agonists activate LRH-1 robustly *in vitro* and promote strong patterns of coregulator protein recruitment. We present new crystal structures showing that the agonists are anchored deeply in the pocket via the 6HP core and that the added polar groups make PL-like contacts at the mouth of the pocket as designed. The best of this class is active *in vivo,* suppressing lipogenic genes in the liver and inducing intestinal anti-inflammatory genes in separate mouse studies. Anti-inflammatory effects in the gut profoundly reduce inflammation and weight loss in a murine model of colitis. This work demonstrates that strategic targeting of the PC binding site improves activity, elucidates the mechanism of action of these compounds, and provides the first *in vivo* evidence of efficacy in colitis for a small molecule modulator of LRH-1.

Results

Structure-guided design of PL-mimics.

Crystal structures of LRH-1 bound to PLs depict an unusual ligand binding mode.[12-15,28] Whereas most NR ligands are fully engulfed within the binding pocket, PLs bind LRH-1 with a portion of the headgroup protruding into the solvent (Figure 3.1a). Lipid phosphates coordinate polar residues near the pocket mouth, while the acyl chains line the pocket interior, making numerous hydrophobic contacts (Figure 3.1a). We previously focused on strengthening deep pocket interactions made by RJW100, resulting in the discovery of the potent agonists 6N and 6Na.[29,30] We used an alternative approach to design the PL-mimics, seeking to promote interactions near the mouth of the pocket. Notably, the alkyl "tail" of RJW100 (position R^4) overlaps with the PL fatty acyl chains and follows a trajectory toward the pocket mouth (Figure 3.1a).[27] Based on this

observation, we synthesized sets of compounds modified at R^4 with alkyl tails of varied lengths at that terminate in polar, PL-mimicking groups (Figure 3.1b).[27]

Modifications improve binding affinity and activity.

The addition of a phosphorylcholine (ChoP) or carboxylic acid (CA) to position R^4 on the RJW100 scaffold dramatically increases LRH-1 transcriptional activity, depending on the length of the alkyl linker connecting the bicyclic core to the modified group.[27] Compounds containing linkers of 9-10 carbons are the strongest activators for both classes, increasing E_{max} nearly two-fold over RJW100 (Figure 3.1c). The optimal linker is 18-20 Å from the bicyclic core to the terminal polar group for the ChoPs and 15-17 Å for the CAs (Figure 3.1c, Table S3.1).[27] Compounds with short alkyl linkers (4-5 carbons) do not significantly activate the receptor above baseline at doses up to 30 μM,[27] and diverse modifications elsewhere on the RJW100 scaffold[29] affect potency (EC_{50}) but rarely increase E_{max} (Figure 3.1d). The PL-mimicking groups also improve binding affinity in a linker length-dependent manner (Figure 3.1e-f, S3.1-S3.3). ChoPs and CAs with longer linkers (9-11 carbons) have the highest affinities, with K_i values between 2-11 nM (Figure 3.1e-f, Table S3.1). To test whether the improved affinity and activity are caused by the longer alkyl linker or the terminal polar group, we synthesized and tested a set of diols (terminally hydroxy-modified at R^4) with varying linker lengths (Figure 3.1b). The diols bind and activate LRH-1 similarly to RJW100 regardless of linker length, suggesting that the charged polar group is the driver of the improved properties (Figure 3.1c, 1g, S3.3-S3.4). Together, these data suggest that PL-mimicking groups form productive interactions with LRH-1, greatly increasing transcriptional output.

Structural basis for improved binding and activity.

To understand the mechanism of action of the PL-mimics, we determined X-ray crystal structures of the LRH-1 LBD bound to either 10CA or 9ChoP at resolutions of 2.3 and 2.55 Å, respectively (Figure 3.2a-b, Table S3.2). Both structures depict the ligand bound at a single site in

the ligand binding pocket (Figure 3.2a-b). The peptide from the TIF2 coregulator, added to stabilize the complex, is bound at the activation function surface (AFS), comprised of portions of helices 3, 4, and the activation function helix (AF-H, Figure 3.2a-b).

The electron density surrounding the ligands unambiguously indicates their positions in the ligand binding pocket, including the extended alkyl linkers and terminal polar groups (Figure 3.2c). The agonists' 6HP cores, styrene substituents and hydroxyl groups adopt nearly identical positions in the deep pocket to RJW100 (Figure S3.5). Remarkably, the alkyl tails of both 10CA and 9ChoP extend toward the mouth of the pocket, with the terminal polar groups close to the DLPC phosphate (seen by superposition with PDB 4DOS,[12] Figure 3.2d). The CA and ChoP groups make very similar contacts near the mouth of the pocket as DLPC, including hydrogen bonds with the sidechains of residues Y516 and K520 and with the α-amino group of G421 (Figure 3.2e-f). The agonists also make a water-mediated hydrogen bond with the α-amino group of L424 via the CA and phosphate groups (also made by DLPC), strengthening their associations with the pocket mouth. Molecular dynamics simulations with the 10CA structure predict that the interactions with Y516 and K520 persist for 50.2% and 37.6% of a 1 μs simulation, respectively. This is similar to DLPC, which interacts for 49.4% of the time with Y516 and 33.2% with K520. The interaction with the backbone amide of G421 is less stable for 10CA than DLPC (27.0% of the simulation *versus* 48.1% for DLPC).

We were particularly interested in how the PL-mimics affect receptor conformation near the mouth of the pocket, since flexibility there is important for allosteric signaling by DLPC.[23] Specifically, the helix 6/ β-sheet region is used to sense ligands and communicate with the coregulator binding surface to recruit appropriate coregulator proteins.[23] PL binding expands the pocket mouth relative to apo LRH-1 via a ~3 Å displacement of helix 6 by the sn1 acyl chain.[31,32] 10CA and 9ChoP expand the pocket mouth through a different mechanism. They do not shift helix

6; instead, they displace the C-terminus of helix 10 by 2.0-2.3 Å relative to apo LRH-1 (from superposition with PDB 4PLD)[23] and 3.5-3.9 Å relative to the LRH-1-DLPC structure, Figure 3.2g). To form hydrogen bonds with Y516 and K520, the R^4 polar groups of the PL-mimics move toward the C-terminus of helix 10 relative to the DLPC phosphate.

To interrogate the function of the PL-like interactions made in the structures, we mutated key LRH-1 residues and measured the impact on binding affinity and transcriptional activity. Introduction of either a Y516A or K520A mutation decreases DLPC's binding affinity by 40- and 24-fold, respectively, *versus* WT LRH-1 (Figure 3.2h, S3.6). We also observe statistically significant reductions in binding affinities for longer-tailed CAs and ChoPs for both mutants, with the exception of 9ChoP for the K520A mutation (Figure 3.2h, S3.6). Compounds with the 6HP scaffold that do not contact these residues (RJW100 and 6N)[32,33] are insensitive or less sensitive to these mutations (Figure 3.2h, S3.7). In luciferase reporter assays, mutations to Y516A or K520A do not affect basal activity of LRH-1; however, the Y516A mutation reduces activation by all four agonists by 24-33% relative to WT protein (Figure 3.2i). Compounds 9CA and 10CA are very poor activators of the K520A mutant, reducing its activation by around 60% *versus* WT LRH-1, while 9ChoP and 10ChoP activate it equally well as WT protein. Together, these results demonstrate that one or more of the interactions near the mouth of the pocket contribute to the increased affinity and efficacy conferred by the extended polar group at R^4.

Effects on LRH-1 conformation and coregulator recruitment.

To understand how the PL-mimics switch LRH-1 into the active state, we investigated how activating *versus* non-activating CAs affect LRH-1 conformation in solution using hydrogen-deuterium exchange mass spectrometry (HDX-MS). We used four CAs for HDX-MS, two that bind LRH-1 but are inactive (4CA and 5CA) and the two most active agonists in this series (9CA and 10CA). Relative to the inactive compounds, 9CA and 10CA stabilize the LRH-1 activation

function surface (AFS), a dynamic region in the LBD that serves as a binding site for coregulator proteins (Figure 3.3a, Figure S3.7). Ligand-induced conformational changes to the AFS tune coregulator associations and modulate transcriptional activity. Compounds 9CA and 10CA stabilize portions of the AFS, including helix 3 and the loop preceding the activation function helix (pre-AF-H loop) relative to the shorter-tailed CAs (Figure 3.3a-c). 10CA also stabilizes the AF-H and part of helix 4 relative to 4CA (Figure 3.3a-b).

The stabilization of the AFS observed with HDX-MS suggests that 9CA and 10CA may affect coregulator associations. To test this, we conducted coregulator profiling using the Microarray Assay for Real-time Coregulator-Nuclear Receptor Interactions (MARCoNI), a microarray that quantifies associations of 64 coregulators from a library of 154 peptides containing NR interaction motifs.[34] We first investigated how long- *versus* short-tailed agonists affect coregulator binding to the LRH-1 LBD, using the same CAs as the HDX-MS experiments. Since LRH-1 copurifies with PL from *E. coli*, thought to act as weak agonists,[12-14] we generated apo LRH-1 LBD by stripping the ligands and refolding the receptor[12] (Figure S3.9). Apo LRH-1 LBD preferentially binds peptides from known LRH-1 corepressors (*i.e.* nuclear receptor corepressor 1 (NCOR1), nuclear receptor subfamily 0 group B member 1 (NR0B1, also called DAX-1), and nuclear receptor subfamily 0 group B member 2 (NR0B2, also called SHP) (Figure S3.9). The strongest interactor is NF-κB inhibitor beta (IκBβ, Figure S3.9), a direct repressor of the retinoid X receptor[35] that has not been previously reported to bind LRH-1. Both 9CA and 10CA disrupt many of the strongest coregulator interactions with apo-LRH-1 (*i.e.* IκBβ, NCOR1, and transcriptional cofactor of c-*fos* (TREF1)), while strongly recruiting a specific complement of coregulators (*e.g.* ligand dependent nuclear receptor corepressor (LCOR), nuclear receptor coactivator A1 (NCOA1) and nuclear receptor coactivator A3 (NCOA3), Figure 3.4a). In contrast, 4CA and 5CA have much weaker effects on coregulator associations (Figure 3.4a). These results suggest that the longer-tailed

agonists promote conformational changes to the LBD that affect coregulator recognition of the AFS, consistent with observations from HDX-MS.

In addition to demarking differences between active and inactive compounds, the coregulator recruitment pattern of 9CA and 10CA is also distinct from agonists in different classes. This is seen through comparison to MARCoNI results with 6N and 6N-aniline (6Na), which share the RJW100 core scaffold but are modified elsewhere on the molecule.[33,36] Unsupervised hierarchical clustering of the data shows a striking separation by ligand class: the most structurally similar ligands cluster together (Figure 3.4b). The close relationship between 9CA and 10CA in the clustering analysis appears to be due to common effects on coregulator dissociations. This distinguishes them from 4CA, which exhibits an otherwise similar (albeit weaker) pattern of coregulator associations to the longer-tailed compounds (Figure 3.4b). Coregulator recruitment by 6N and 6Na is weak relative to the longer-tailed CAs, but it also displays a different pattern: for example, 9CA and 10CA promote dissociation of 16 coregulator peptides in common, while 6N and 6Na affect a distinct set of five coregulator peptides (Figure 3.4c).

Many of the observations from the experiment with isolated LBD are recapitulated in a MARCoNI study using FL-LRH-1. This is not a direct correlate to the LBD experiment, since apo-FL protein was unstable, and FL-LRH-1 in this experiment was bound to copurifying *E. coli* PL. Interestingly, the recombinant FL-LRH-1 exhibits a nearly identical coregulator binding pattern to untreated apo-LBD (Figure S3.9). The FL protein also responds very similarly to 10CA, strongly recruiting peptides from LCOR, NCOA1, and SHP and strongly dissociating IκBβ, NCoR1, and TREF (Figure S3.9). A total of 85% of the peptides affected by at least 1.5-fold over baseline by 10CA in the LBD experiment are also altered by more than 1.5-fold with FL-LRH-1 (Figure 3.4d). As in the LBD experiment, the effects of 6N are relatively weak and mainly involve coregulator dissociation, including IκBβ and several canonical LRH-1 corepressors (*e.g.* NCOR1, PROX1, SHP,

and DAX-1, Figure S3.9). Differences between 10CA and 6N are more apparent with FL-LRH-1 than LBD: 22 peptides are recruited by one compound and displaced by the other (Figure 3.4e). These striking differences in patterns of coregulator recruitment by 10CA *versus* 6N are indicative of distinct ligand-induced conformational changes at the LRH-1 AFS, demonstrating that receptor allostery is tunable through ligand modifications. These studies also reveal the strong effects on coregulator recruitment driven by the strategic addition of a PL-mimicking group.

10CA reduces expression of lipogenic genes in the liver.

LRH-1 is highly expressed in the liver, where it regulates bile acid biosynthesis, ER-stress resolution, and *de novo* lipogenesis.[16,17,37] We tested whether 10CA, one of the most active of the new compounds in luciferase assays, could activate endogenous LRH-1 in Huh7 hepatocytes. A 24-hour treatment of 10CA dose-dependently upregulates known LRH-1 transcriptional targets SHP (*NR0B2*) and polo-like kinase 3 (*PLK3*). 10CA also downregulates targets involved in *de novo* lipogenesis, including sterol regulatory element-binding protein 1 (*SREBPF1,* also called SREBP1-c), acetyl-CoA carboxylase 2 (*ACACB*), fatty acid synthase (*FASN*), and stearoyl-CoA desaturase-1 (*SCD1*) (Figure 3.5a). Interestingly, downregulation of these lipogenic genes in the liver is associated with LRH-1-dependent antidiabetic effects by DLPC.[16]

To determine whether 10CA can activate LRH-1 in the liver *in vivo*, we overexpressed human LRH-1 (hLRH-1) in C57BL/6J mice using AAV8 adenovirus. The overexpression strategy was chosen because 10CA and analogs are poor activators of mouse LRH-1 in luciferase reporter assays (data not shown). Two weeks after viral infection, mice were treated with 10CA at 0.1, 1, or 10 mg/kg in a series of five intraperitoneal injections spanning three days (Figure 3.5b). Overexpression of hLRH-1 was verified by qRT-PCR (Figure 3.5c), and agonist-induced gene expression changes in the liver were assessed by Nanostring.[38] The most prominent effect of 10CA treatment is the downregulation of lipogenic genes (Figure 3.5d-e). The affected genes were the

same as those downregulated in Huh7 cells (*FASN, SREBPF1, ACACB,* and *SCD1*). The downregulation was dependent on hLRH-1 expression, since they were unchanged in mice not expressing hLRH-1 (Figure S3.10).

Efficacy of 10CA in organoid and murine models of colitis.

LRH-1 is a novel therapeutic target for inflammatory bowel diseases such as ulcerative colitis (UC), a chronic disease characterized by inflammation of the colon and decreased integrity of the intestinal epithelium.[39] LRH-1 regulates local glucocorticoid production in the intestine that promotes cell survival, which is important for maintenance of the intestinal epithelial barrier.[40-43] Loss of LRH-1 function in mice decreases glucocorticoid production in the intestinal epithelium and decreases expression of *Cyp11a1* and *Cyp11b1*, genes that regulate corticosterone production.[40] These defects are rescued by overexpression of hLRH-1.[43] We therefore investigated efficacy of 10CA in UC. We began with an *in vitro* model, using organoids derived from intestinal crypts of humanized LRH-1 mice (enteroids). In these enteroids, hLRH-1 is expressed in the context of conditional deletion of endogenous LRH-1 in the intestinal epithelium,[43] and they mimic ulcerative colitis (UC) when stimulated with tumor necrosis factor alpha (TNF-α) to induce inflammation.[29,43,44] Overnight treatment with 10CA induces expression of *Cyp11a1* and *Cyp11b1*, increases expression of the anti-inflammatory cytokine *Il-10*, and decreases expression of pro-inflammatory cytokines *Il-1B* and *TNF-α* (Figure 3.6a-c). These changes in gene expression require the expression of hLRH-1, since they are not observed in epithelium-specific LRH-1-deficient enteroids (Figure S3.11). Therefore, 10CA strongly upregulates LRH-1-controlled steroidogenesis in enteroids with concomitant anti-inflammatory gene expression changes.

To test efficacy of 10CA *in vivo,* we utilized a dextran sulfate sodium (DSS)-induced murine model of UC, using epithelium-specific humanized LRH-1 mice. Administration of DSS rapidly damages the intestinal epithelium, resulting in intestinal inflammation and compromised barrier

function.[45] Mice received seven days of DSS treatment, followed by five days of daily intraperitoneal injections of 10CA or vehicle. DSS administration caused rapid weight loss in both groups beginning around day 5, typical for this disease model.[46] Vehicle-treated animals lost an average of 8.6% initial body weight by the end of the study (Figure 3.6d). In contrast, mice treated with 10CA began re-gaining weight on day 9 and were within 1% of baseline at the end of the study (Figure 3.6d). 10CA-treated mice had markedly improved scores for disease activity, which quantify severity of symptoms such as diarrhea and blood in the stool (Figure 3.6e), and scores for histology and gross morphology, which quantify inflammation, tissue damage, and crypt loss in the colon (Figure 3.6e). Representative colonic tissue sections illustrate the striking differences between 10CA-treated and control mice, with examples of crypt loss highlighted in yellow (Figure 3.6f-g). This is the first LRH-1 small molecule agonist to show *in vivo* efficacy in a model of colitis, demonstrating the potential of targeting LRH-1 for treatment of this chronic disease.

Discussion

The discovery that PCs act as nuclear receptor ligands has illuminated the potential for novel signaling axes connecting PC metabolism with gene expression. Compelling evidence from gain and loss of function studies connects PC availability to LRH-1-controlled transcriptional programs involving lipid and methyl metabolism. A close relationship between PC generation via the one-carbon pathway and the response of LRH-1 to dietary PCs has led to the intriguing hypothesis that LRH-1 senses PCs as a proxy for nutrient availability. Likewise, PCs have anti-inflammatory effects in the intestinal mucosa that are not fully explained by their physical role in this tissue.[47-50] The agonists described herein provide tools to probe a potential LRH-1-PC regulatory circuit, which is challenging to elucidate with natural PCs due to their lability and promiscuity.[51]

Crystallographic data reveals very different binding modes of natural PL and the synthetic LRH-1 modulator RJW100, with the former making polar interactions near the mouth of the pocket

and the latter being anchored in the deep pocket.[31,32] The hybrid PL-mimics were designed with the rationale that appending a polar moiety to an extended R^4 alkyl tail would allow the agonists to interact with both regions of the pocket. With two new crystal structures, we demonstrate that this strategy was successful: the 6HP cores nearly perfectly superpose with RJW100, while the polar groups in the extended tails form hydrogen bonds with several residues engaged by the DLPC phosphate group (Figure 3.2). Disrupting these hydrogen bonds via mutagenesis reduces binding affinity and decreases activity of the longer-tailed agonists (Figure 3.2). Compounds with shorter tails or lacking a charged group at the R^4 terminus are not more active than RJW100, further demonstrating that the PL-mimicking interactions are responsible for the improved binding and activity profiles (Figure 3.1). HDX-MS experiments show that the longer-tailed CAs stabilize the LRH-1 AFS relative to shorter-tailed CAs (Figure 3.3). The AFS stabilization is associated with much stronger effects on coregulator recruitment than seen with shorter-tailed CAs, including dissociation of many factors bound by apo-LRH-1 (Figure 3.4). Together, these data suggest a model in which specific interactions made by the PL-mimicking group promote conformational changes at the AFS that activate the receptor by modulating coregulator associations. Notably, the allostery observed for the PL-mimics is different from that of DLPC, which relies upon flexibility and motion of the region near helix 3.6.[23] The PL-mimics do not displace helix 6 in the crystal structures (Figure 2), and the flexibility of helix 6 is unchanged in HDX-MS experiments with 9CA and 10CA *versus* shorter-tailed CAs (Figure 3.3). The agonists therefore mimic PC interactions but act more as hybrids between PCs and synthetic agonists with respect to effects on receptor conformation and allostery. Nevertheless, the PL-mimics recapitulate many gene expression changes made by exogenous DLPC, including downregulation of lipogenic genes in hepatocytes and mouse liver (Figure 3.5). Downregulation of the same set of lipogenic genes and consequent suppression of *de novo* lipogenesis is thought to be the main mechanism through which DLPC exerts

anti-diabetic effects, including reducing liver fat accumulation and improving insulin resistance.[16] This suggests a potential for the new agonists in diabetes and metabolic disorders associated with obesity, which is an ongoing area of research in our laboratories.

LRH-1 is as an important regulator of intestinal cell renewal and local inflammation in the gut.[41,43] One of the most exciting findings in these studies is the strong efficacy in UC *in vivo*. While gain and loss of function studies have demonstrated the potential of targeting LRH-1 in colitis,[43] and the agonist 6N activates steroidogenic genes in enteroids,[33] this is the first demonstration of a synthetic molecule to exhibit *in vivo* efficacy. In DSS-induced UC, LRH-1 activation by 10CA markedly decreases disease markers, inhibits colonic crypt loss, and ameliorates disease-associated weight loss (Figure 3.6). Studies in enteroids link this efficacy to anti-inflammatory gene expression patterns (Figure 3.6). Targeting LRH-1 in UC is a particularly intriguing strategy, as its activation suppresses inflammation locally in the gut.[41,52] UC is commonly treated with corticosteroids, anti-TNFα therapies, or other systemic immunosuppressants that often cause adverse effects associated with immunosuppression in tissues other than the colon[53]. Thus, LRH-1 agonists may offer a more precise approach to UC treatment, reducing the potential for on-target, adverse effects in other tissues.

Together, this work has demonstrated an exciting potential for incorporating PL-mimicking interactions into LRH-1 agonists. The PL-mimics are valuable tools to study LRH-1 activation by ligands and to broaden our understanding of PL-regulated transcriptional programs. They also have potential as therapeutics for diseases associated with aberrant LRH-1 activity, such as diabetes and UC.

Methods

Chemical synthesis. We have previously described the synthesis of the phosphorylcholine derivatives (4ChoP-11ChoP, also called 12a-12h) and carboxylic acid derivatives (4CA-11CA, also called 13a-

13h) modified at the R^4 position of the RJW100 scaffold.[27] Methods for synthesis of diols (4OH-11OH) are described in detail in the supplemental materials. Synthetic methods for RJW100 and RJW100 analogs modified at positions R^1, R^2, or R^3 (compounds 1N-8N; 1X-8X; 9-23, used to generate Figure 2B) have also been published previously.[29]

Cell culture. Hela and Huh7 cells were purchased from Atlantic Type Culture Collection. Hela cells were cultured in MEMα medium supplemented with 10% charcoal-stripped fetal bovine serum. Huh7 cells were cultured in DMEM/F12 medium (Gibco) supplemented with 10% fetal bovine serum. Cells were maintained under standard culture conditions.

Reporter gene assays. Hela cells were seeded into 96-well culture plates at 7500 cells/ well and transfected with reporter plasmids and either pCI empty vector or full-length LRH-1-pCI. Reporter plasmids consisted of (1) the pGL3 basic vector containing a portion of the SHP promoter containing the LRH-1 response element cloned upstream of firefly luciferase and (2) a constitutively active vector encoding *Renilla* luciferase used for normalization. Transfections utilized the FugeneHD transfection reagent at a ratio of 5 μl Fugene: 2 μg DNA. Cells were treated with agonists 24 hours after transfection at concentrations indicated in the figure legends. 24 hours after treatment, luminescence was quantified using the DualGlo kit from Promega on a BioTek Neo plate reader. Values for EC_{50} and E_{max} were calculated by fitting data to a three-parameter dose-response curve in GraphPad Prism, v7.

Calculation of Relative Efficacy (RE). This value was used to compare data from luciferase reporter assays that were conducted on different dates by normalizing values to data from the compound RJW100 assessed in parallel experiments. RE 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 $(Max_{cpd} - Min_{cpd}) / (Max_{RJW100} - Min_{RJW100})$, 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.

Protein purification

LRH-1 LBD. LRH-1 LBD (residues 299-541) in the pMSC7 vector was transformed in *E. coli* strain BL21(pLysS) for expression. Cultures were grown at 37 °C in liquid broth to an OD_{600} of 0.6 prior to induction of protein expression with 1 mM isopropyl-1-thio-D-galactopyranoside (IPTG). Cultures were grown for 4 hours at 30 °C following induction. Protein was purified by nickel affinity chromatography with Buffer A of 20 mM Tris HCl (pH 7.4), 150 mM NaCl, 5% glycerol, and 25 mM imidazole and Buffer B of 150 mM Tris HCl (pH 7.4), 150 mM NaCl, 5% glycerol, and 500 mM imidazole (HisTrap FF; GE Healthcare, Little Chalfront, UK). Protein was incubated with DLPC (4-fold molar excess) overnight at 4 °C, repurified by size exclusion into assay buffer (150 mM NaCl, 20 mM Tris-HCl (pH 7.4), and 5% glycerol, concentrated to approximately 3 mg/mL, and stored at -80 °C until use. The same purification strategy was utilized for protein containing the point mutations described in the text.

Full-length LRH-1 purification. The gene encompassing the full-length LRH-1 (FL-LRH-1; residues 1-495, isoform 1) was codon-optimized for *E. coli,* sub-cloned into a modified PET28b vector with a His-SUMO tag and transformed into *E. coli* strain BL21 (DE3) cells. Cells were grown in LB media at 37 °C to an OD_{600} of 0.5. Protein expression was induced with 0.2 mM IPTG and grown overnight at 16 °C. Cells were harvested by centrifugation and resuspended into lysis buffer containing 20 mM Tris (pH 7.5), 500 mM NaCl, 5% glycerol, 25 mM imidazole, 0.5 mM *tris*(2-carboxyethyl) phosphine (TCEP), and 0.1 mM phenylmethylsulphonyl fluoride (PMSF). Cells were lysed by sonication for 8 minutes and clarified by centrifugation. Protein was purified by nickel affinity chromatography with elution buffer of 20 mM Tris HCl (pH 7.5), 500 mM NaCl, 5% glycerol, 0.5 mM TCEP and 125 mM imidazole (HisTrap FF; GE Healthcare, Little Chalfront, UK).

The His-Sumo tag was removed using tobacco etch virus (TEV) protease at 4 °C with dialysis against buffer containing 20 mM Tris (pH 7.5), 500 mM NaCl, 5% glycerol and 0.5 mM TCEP. Ion-exchange chromatography was used to isolate cleaved protein and remove DNA with tandem Hitrap Q-SP columns (GE Healthcare). Protein was flowed through the Q column onto the SP column, from which it was eluted as a single peak using a linear gradient of NaCl from 0.35 M to 1 M. Finally, the protein was further purified to homogeneity by size exclusion chromatography. Functionality of the purified protein was determined by testing its ability to bind DNA containing an LRH-1 response element, and its ability to bind ligand (see "*Ligand binding assay*" for details). DNA binding was measured using fluorescence polarization with a 6-carboxyfluorescein (FAM)-labeled double-stranded DNA probe containing the LRH-1 binding element from the CYP7A1 promoter. FL-LRH-1 was titrated against 10 nM FAM-CYP7A1 in buffer containing 20 mM Tris (pH 7.5), 150 mM NaCl, 5% glycerol and 0.5 mM TCEP, with a final volume of 50 μL, and incubated at room temperature for 30 min. Polarization was monitored on a Synergy 4 microplate reader (Biotek, Winooski, VT) at an excitation/emission wavelength of 485/528 nm. Curves were fit using a one-site specific binding equation with baseline-subtraction of protein alone in GraphPad Prism, v7. One biological replicate was conducted with two technical replicates.

Generation of apo LRH-1. To extract lipids from the LRH-1 LBD, 1 mL of purified protein (3 mg) was treated with 3.75 mL of chloroform-methanol solution (1:2 v/v) and vortexed briefly. An additional 2.5 mL chloroform:water solution (1:1 v/v) was added and the mixture was vortexed again. The stripped and unfolded protein was pelleted by centrifugation at 1000 rpm for 10 minutes. The resulting protein pellet was dissolved into 0.5 mL of buffer containing 50 mM Tris (pH 8.0), 6 M guanidine hydrochloride and 2 mM DTT. Protein was refolded by fast dilution at 4 °C into 25 mL of buffer containing 20 mM Tris (pH 8.5), 1.7 M urea, 4% glycerol and 2 mM DTT. The final urea concentration was adjusted to 2 M, and the protein was concentrated to ~ 1.5 mL, followed by

overnight dialysis against PBS containing 2 mM DTT at 4 °C. Refolded protein was purified by size exclusion chromatography to remove aggregates and remaining unfolded protein, and ligand binding ability was verified by fluorescence polarization as described below.

Mutagenesis. Mutations were introduced to LRH-1 in the pMSC7 and pCI vectors using the Quikchange Lightning Site-Directed Mutagenesis Kit as directed by the manufacturer (Agilent). Constructs were sequenced prior to use in binding or activity assays.

Ligand binding assay. All assays were conducted in black, polystyrene, non-binding surface 384-well plates (Corning Inc., Corning, NY) with 30 μL volumes in assay buffer (150 mM NaCl, 20 mM Tris-HCl pH 7.4, 5% glycerol). Binding affinity of the fluorescein amidite (FAM)-labeled probe (6N-FAM) for WT and mutant LRH-1 was determined by titrating purified protein in the presence of 10 nM 6N-FAM, as previously described.[54] The affinity of the probe for WT and mutant LRH-1 LBD from these initial experiments was used to determine probe and protein concentrations for competition experiments with unlabeled agonists. Competition assays were conducted with 10 nM 6N-FAM (10, 0.6, and 1.1 times the K_d of wildtype, K520A, and Y516A LRH-1, respectively, for 6N-FAM). For protein concentration, 5, 10, and 15 nM wildtype, K520A, or Y516A LRH-1 were used, representing 50-80% of the B_{max} from the 6N-FAM binding curves for each protein variant. Assay conditions were validated using unlabeled compound 6N. Competition experiments used unlabeled agonists at a concentration range of 200 pM-200 μM. Plates were incubated overnight at 4 °C and centrifuged at 2,000 x g for 2 minutes before polarization measurement. Polarization was monitored on a Neo plate reader (Biotek, Winooski, VT) at an excitation/emission wavelength of 485/528 nm. The experiment was conducted twice in quadruplicate, and GraphPad Prism version 7 was used to analyze compiled data. Data were first baseline subtracted with the lowest concentration of competitor as the baseline. To compare the ability of each competitor to displace the 6N-FAM probe from LRH-1, data were then normalized to the unlabeled 6N competition curve, as 6N

completely competes the 6N-FAM probe, such that a baseline of 0 represents complete competition of the 6N-FAM probe for a given agonist. Finally, data were fit using a one-site, fit K_i curve in GraphPad Prism. This equation fits the K_i of the unlabelled competing ligand using the EC_{50} of the competition curve, the concentration of the fluorescent probe, and the K_d of the probe for each receptor, as follows:

Equation 1: Calculating the EC_{50} of the competing ligand.

Y=Bottom + (Top-Bottom)/(1+10^(X-LogEC_{50}))

Equation 2: Calculating K_i from EC_{50}, probe concentration, and probe K_d.

logEC_{50}=log(10^logK_i*(1+[Probe, nM]/Probe Kd, nM))

Mutants were normalized to their respective unlabeled 6N competition curves, which had different shifts in polarization than wildtype. To compare K_i values for wildtype versus K520A and Y516A mutants, two-way ANOVA was conducted followed by Dunnett's test for multiple comparisons (GraphPad Prism v7).

Hydrogen-deuterium exchange mass spectrometry (HDX-MS). Following affinity purification and removal of the His-tag as described above, LRH-1 LBD protein was repurified by size exclusion chromatography into an assay buffer of phosphate-buffered saline, pH 7.5, plus 5% glycerol. Protein aliquots (2 mg/ml) were incubated with carboxylic acid derivatives 4CA, 5CA, 9CA, or 10CA at 5-fold molar excess, overnight at 4 °C. Samples were diluted 1:7 (v/v) in labeling buffer to initialize exchange reactions (same as the protein buffer but containing 99.9 % D_2O instead of H_2O). The reactions were quenched after 0, 10, 100, 1000 and 10000 seconds by adding equal volume of precooled quenching buffer (100 mM phosphate, 0.5 M tris(2-carboxyethyl) phosphine, 0.8% formic acid, and 2% acetonitrile, pH 2.5). Reactions for each sample at each time point were performed in triplicate. Quenched samples were passed through an Enzymate BEH pepsin column (Waters Corp, Milford, MA). Fragmented peptides were separated by an C18 UPLC column and analyzed by a Q-

Tof Premier mass spectrometer. ProteinLynx Global SERVER™ (PLGS) was used to identify peptides by database searching of LRH-1 LBD sequence. The HDX-MS data were further processed using DynamX (v3.0) and the differential HDX between ligand-bound states was calculated by comparing the relative fractional uptake for each residue at given time.

In vitro NR-coregulator recruitment by MARCoNI

Experiment with LRH-1 LBD: Assay mixes consisted of 50 nM purified His-LRH-1 LBD (either in the apo state or pre-loaded with compounds), 25 nM ALEXA488-conjugated penta-His antibody (Qiagen # 35310), 50 μM DTT, 20mM Tris (pH 7.4), 250mM NaCl and 0.5 mM TCEP.

Experiment with FL-LRH-1: Assay mixes of 50 nM His-SUMO-hLRH-1-FL (bound to copurifying PL from *E. coli* or preloaded with compound during purification), 25 nM ALEXA488-conjugated penta-His antibody (Qiagen # 35310), 50 μM DTT, and 10 μM freshly added compound (or 2% DMSO) were made in 20mM Tris(pH 7.4), 250mM NaCl and 0.5 mM TCEP and stored on ice.

LRH-1 in the assay mixes was analyzed by applying Microarray Assay for Real-time Coregulator Nuclear Receptor Interaction (MARCoNI), using PamChip #88101 with 154 unique coregulator sequences, as described previously.[55] In short, each condition was tested using three technical replicates (arrays), and LRH-1 binding to each coregulator motif was quantified using BioNavigator software (PamGene). The compound-induced log-fold change (LFC) of LRH-1 binding to each coregulator peptide and statistical significance (Student's t-test) were calculated and visualized using R software (R Core Team, 2017). Compound and interaction (dis)similarity were calculated by Hierarchical Clustering on Euclidean Distance and Ward's agglomeration. Upset plots were generated using the R package UpSetR v1.4.0.[56]

Crystallography. 6X-His-LRH-1 LBD was expressed and purified by nickel affinity chromatography as described above. The His tag was cleaved overnight at 4 °C using TEV protease with dialysis against

buffer containing 150 mM NaCl, 5% glycerol, and 20 mM Tris HCl (pH 7.4). Cleaved protein was isolated using nickel affinity chromatography. To generate protein-ligand complexes, purified protein was incubated overnight at 4 °C with 4-5-fold molar excess of ligand. Complexes were purified by size exclusion chromatography into crystallization buffer consisting of 150 mM NaCl, 100 mM ammonium acetate (pH 7.4), 1 mM DTT, and 1 mM EDTA, 2 mM 3-[(3-cholamidopropyl)dimethylammonio]-1propanesulfonic acid (CHAPS) and then incubated with 4-fold molar excess of Tif2 peptide for two hours at room temperature (Tif2 peptide sequence was H_3N-KENALLRYLLDK-CO_2^-). Crystallization conditions for each complex were as follows: LRH-1-10CA-Tif2: The LRH-1-10CA-TIF2 complex was concentrated to 5 mg/mL and screened using the Classics screen (Qiagen) and a Phoenix Liquid Handler (Art Robbins Instruments) in 96-well sitting drop plates. Crystals were generated at room temperature in 0.4 M N/K tartrate. LRH-1-9ChoP-Tif2: Crystals were generated by soaking into LRH-1-10CA-Tif2 crystals. First, we generated LRH-1-10CA-Tif2 crystals in larger drops by microseeding, using the LRH-1-10CA-Tif2 crystals described above as seed stocks. Fresh LRH-1-10CA-Tif2 complex was concentrated to 5.2 mg/mL, and the seed stock was added to the protein at a dilution of 1:100. The complex was then crystallized by sitting drop vapor diffusion at 4 °C, in drops containing 1 µl protein/seed stock plus 2 µl crystallant. The crystallant was 9-19% tert-butanol, 0-6 % glycerol, 0.1 M tri-sodium citrate pH 4.6. The 9ChoP ligand was soaked into the crystal drops at 1 µM (~1% DMSO) for two days, and 1 µM 9ChoP was also added to the cryoprotectant.

Structure Determination. Crystals were flash-frozen in liquid nitrogen using a cryoprotectant of 30% glycerol in crystallant. Data were collected remotely from Argonne National Laboratory (South East Regional Collaborative Access Team, Lemont, IL) using the 22ID beamline. Data were processed and scaled using HKL2000[57] and phased by molecular replacement using Phaser-MR in Phenix[58] with PDB 5L11 as the search model. For the 9ChoP structure, the 10CA structure was used as the

search model. Coot[59] and Phenix.refine[58] were used for model building and refinement, respectively. Figures were constructed using Pymol.[60] Ligplot+[61] was used to identify residues interacting with the ligands. Measurements to determine displacement of helix 10 were between C-alphas of residue N523 for each protein.

Molecular dynamics simulations.

Model construction:

For model construction, two crystal structures of LRH-1 LBD in complex with Tif2 were used. These were (1) PDB 7JYD (10CA ligand) and (2) PDB 4DOS (DLPC). The structures were modified at the N and C termini so that they included 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 side chains.

Simulations:

All complexes were solvated in an octahedral box of TIP3P water with a 10 Å buffer around the complex. Na^+ and Cl^- ions were added to neutralize the complex and achieve physiological conditions (150 mM sodium chloride). All systems were generated with the Xleap module of AmberTools,[62] using the ff14SB force field.[63] Parameters for all ligands were obtained using the Antechamber module[64] of AmberTools. All minimizations and simulations were performed with Amber16.[62] Using a minimization protocol of 5000 steps of steepest descent followed by 5000 steps of conjugate gradient, systems were minimized in an iterative manner: i) with 500 kcal/mol.Å^{2}restraints on all atoms; ii) with 500 kcal/ mol.Å^2 restraints on restraints retained on ligand and peptide atoms; iii) with restraints removed from all atoms. Following minimization, 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. Equilibration was performed using a 12-ns MD run 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 equilibration. All restraints were removed, and production runs were performed for 1 μs in the NPT ensemble. A 2-fs time step was used for all simulations. All bonds between heavy atoms and hydrogens were fixed with the SHAKE algorithm.[65] A 10-Å cutoff was used for nonbonded interactions. Long range electrostatics were evaluated using Particle Mesh Ewald (PME). For analysis, frames were obtained at 20 ps intervals from each simulation. The CPPTRAJ module[66] of AmberTools was used for structural averaging of MD trajectories and distance calculations.

Agonist treatment and RNA extraction in Huh7 cells. Huh7 cells were plated in 6-well plates at 5x10^5 cells/well. After 24 hours, cells were treated with vehicle (0.1% DMSO in media) or indicated concentrations of 10CA. Cells were collected for RNA after 24 hours of treatment. RNA extraction was performed using the RNAeasy Mini RNA extraction kit (Qiagen).

Animals. Experimental protocols for the AAV8 overexpression study were approved by the Institute for Animal Care and Use Committee at T3 Laboratories and Emory University 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. C57BL/6J mice (male; 10-12 weeks of age) were purchased from Jackson Labs (Bar Harbor, ME, USA). Mice were housed at room temperature in a 12-hr light/12-hr dark vivarium, with *ad libitum* access to food and water.

For the experiments involving enteroids and DSS, the study protocol was approved by the Animal Care and Use Committee of the Baylor College of Medicine and was in accordance with the Guide for the Care and Use of Laboratory Animals [DHHS publication no. (NIH) 85-23, revised 1985, Office of Science and Health Reports, DRR/NIH, Bethesda, MD 20205]. Animals were housed and bred in a specific pathogen free (SPF) facility. An inducible intestinal epithelia cell (IEC) knockout line was created by crossing animals harboring CreERT2 under control of the villin

promoter with LRH-1 floxed (LRH-$1^{f/f}$)animals and bred to homozygosity (LRH-1 KO, LRH-$1^{f/f}$;Villin-Cre+). The humanized LRH-1 line (LRH-$1^{f/f}$;hLRH-$1^{\Delta\Delta}$;Villin-Cre+) was generated with Rosa26-LoxP-STOP-LoxP-hLRH-1 animals, crossed into our LRH-1 KO line, which conditionally overexpresses hLRH-1 in the intestinal epithelium with the knockout of endogenous mLRH-1.

Viral overexpression and drug treatment. Adeno-associated virus serotype 8 (AAV8) expressing N-terminal Avi-tagged, human LRH-1 was obtained from Vector Biolabs (Malvern, PA, USA). C57BL/6J mice were administered $1x10^{12}$ genome copies (GC)/mL via femoral vein injection. All experiments were conducted two weeks after virus treatment. Mice were administered 0.1, 1.0, or 10 mg/ kg of 10CA or vehicle (5% ethanol in sunflower oil) via intraperitoneal injection five times over a 3-day period. Mice were treated each morning and evening for two days and on the morning of day 3. Mice were euthanized on day 3, approximately four hours after the final drug treatment. Mouse liver tissue was collected and flash-frozen in liquid nitrogen at the time of euthanasia. Tissue samples were subsequently stored at -80 °C. To extract RNA, liver samples were cryo-pulverized in liquid nitrogen, then homogenized in TRI® Reagent (Invitrogen) and RNA was extracted using the TRI® Reagent protocol followed by miRNAeasy Mini RNA extraction kit (Qiagen). A portion of the RNA was retained for NanoString analysis, and a portion was subjected to qRT-PCR to verify overexpression of hLRH-1. Reverse transcription was performed using the High Capacity cDNA Reverse Transcription Kit (Applied Biosystems). qPCR was performed using the *Power* SYBR™ Green PCR Master Mix (Applied Biosystems) and run on a StepOnePlus™ Real-Time PCR instrument (Applied Biosystems). Data were analyzed via the $\Delta\Delta C_t$ method, with *Tata binding protein* (*Tbp*) serving as the reference gene. Primers for hLRH-1 detection by qPCR were CTTTGTCCCGTGTGTGGAGAT (forward) and GTCGGCCCTTACAGCTTCTA (reverse) and for Tbp were TGCACAGGAGCCAAGAGTGAA (forward) and CACATCACAGCTCCCCACCA (reverse).

NanoString Gene Expression Analysis. RNA was extracted from liver tissue as detailed above. RNA concentration was determined from OD values on a Nanodrop 1000. Sample profiles were assessed on a 2100 Bioanalyzer Instrument (Agilent) using the RNA 6000 Nano (25ng/ul – 1,000ng/ul) or Pico assay (0.5ng/ul – 20ng/ul). All RNA Integrity (RIN) scores were >9.0. 25ng of total RNA was hybridized with biotin labeled capture probes and fluorescently labeled reporter probes from NanoString Technologies, Inc. for 18 hours at 65 °C. Following hybridization, samples were injected into a NanoString SPRINT cartridge and loaded onto the nCounter SPRINT Profiler instrument where excess capture probe and reporter probe were removed, and hybridized mRNAs were immobilized for imaging. Following image acquisition, mRNA counts were extracted from raw RCC files using nSolver™ data analysis software (ver.4.0, NanoString Technologies, Inc.) and exported as .csv files for further analysis. One-way ANOVA followed by pairwise t-tests were used to determine statistical significance.

Humanized LRH-1 Mouse Intestinal Enteroid Culture. Intestinal crypt culture (enteroids) were derived from LRH-1 KO (LRH-$1^{f/f}$;Villin-Cre+), and hLRH-1 (LRH-$1^{f/f}$;hLRH-$1^{\Delta\Delta}$;Villin-Cre+) male mice (6−8 weeks old). Briefly, the small intestine was isolated and flushed with ice-cold PBS, opened longitudinally, then cut into 1−2 mm pieces. Intestinal fragments were incubated in an EDTA-containing solution (4 mM) at 4 °C for 60 min on a tube rocker. The intestinal fragment suspension was fractionated by vertical shaking manually and crypt-containing fractions passed through a 70 μm cell strainer for plating in Matrigel. The crypt-Matrigel suspension was allowed to polymerize at 37 °C for 15 min. Intestinal organoids were grown in base culture media (Advanced DMEM/F12 media, HEPES, GlutaMax, penicillin, and streptomycin) supplemented with growth factors (EGF, Noggin, R-spondin, R&D Systems), B27 (Life Technologies), N2 (Life Technologies), and N-acetyl cysteine (NAC, Sigma). Intestinal enteroids were passaged every 3 days. Established LRH-1h enteroids were treated with mouse TNF-α overnight to provoke inflammatory changes, then treated

with vehicle (DMSO) or compound 10CA (1 μM) overnight. Following the treatment, enteroid tissues were harvested for real-time PCR.

RNA Isolation and RT-qPCR. Intestinal enteroids were washed in ice cold PBS and suspended in Trizol solution (Sigma). RNA was isolated with RNeasy spin columns (Qiagen). DNAse-treated total RNA was used to generate cDNA using Superscript II (Quanta). Sybr green-based qPCR (Kapa Biosystems) was performed on a Roche LightCycler 480 II with primers as shown below. The $\Delta\Delta C_t$ method was used for calculating gene expression fold changes using Rplp0 (ribosomal protein, large, P0, known as 36B4) as the reference. Primer sequences are shown in Table 1 below.

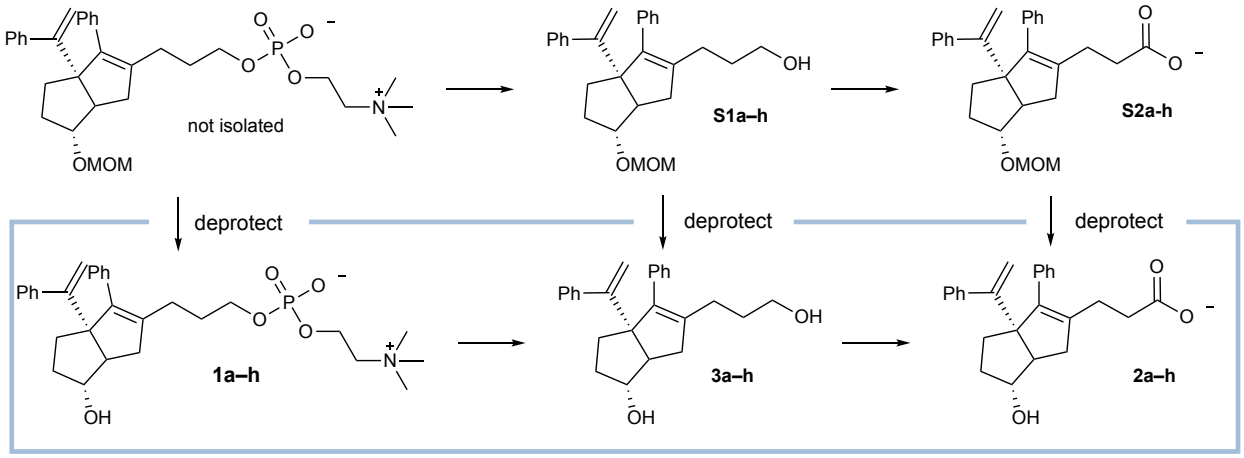
General Synthetic Information

All reactions were carried out in oven-dried glassware, equipped with a stir bar and under a nitrogen atmosphere with dry solvents under anhydrous conditions, unless otherwise noted. Solvents used in anhydrous reactions were purified by passing over activated alumina and storing under argon. Yields refer to chromatographically and spectroscopically (^{1}H NMR) homogenous materials, unless otherwise stated. Reagents were purchased at the highest commercial quality and used without further purification, unless otherwise stated. n-Butyllithium (n-BuLi) was used as a 1.6 M or a 2.5 M solution in hexanes (Aldrich), was stored at 4°C and titrated prior to use. Organic solutions were concentrated under reduced pressure on a rotary evaporator using a water bath. Chromatographic purification of products was accomplished using forced-flow chromatography on 230-400 mesh silica gel. Preparative thin-layer chromatography (PTLC) separations were carried out on 1000μm SiliCycle silica gel F-254 plates. Thin-layer chromatography (TLC) was performed on 250μm SiliCycle silica gel F-254 plates. Visualization of the developed chromatogram was performed by fluorescence quenching or by staining using $KMnO_4$, p-anisaldehyde, or ninhydrin stains.

^{1}H and ^{13}C NMR spectra were obtained from the Emory University NMR facility and recorded on a Bruker Avance III HD 600 equipped with cryo-probe (600 MHz), INOVA 600 (600 MHz),

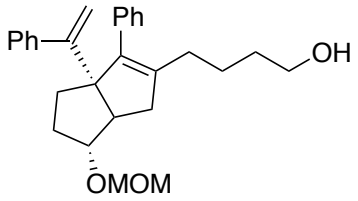
INOVA 500 (500 MHz), INOVA 400 (400 MHz), VNMR 400 (400 MHz), or Mercury 300 (300 MHz), and are internally referenced to residual protio solvent signals. Data for ^{1}H NMR are reported as follows: chemical shift (ppm), multiplicity (s = singlet, d = doublet, t = triplet, q = quartet, m = multiplet, dd = doublet of doublets, dt = doublet of triplets, ddd= doublet of doublet of doublets, dtd= doublet of triplet of doublets, b = broad, etc.), coupling constant (Hz), integration, and assignment, when applicable. Data for decoupled ^{13}C NMR are reported in terms of chemical shift and multiplicity when applicable. IR spectra were recorded on a Thermo Fisher Diamond-ATR and reported in terms of frequency of absorption (cm $^{-1}$). High Resolution mass spectra were obtained from the Emory University Mass Spectral facility. Gas Chromatography Mass Spectrometry (GC-MS) was performed on an Agilent 5977A mass spectrometer with an Agilent 7890A gas chromatography inlet. Liquid Chromatography Mass Spectrometry (LC-MS) was performed on an Agilent 6120 mass spectrometer with an Agilent 1220 Infinity liquid chromatography inlet. Preparative High-Pressure Liquid chromatography (Prep-HPLC) was performed on an Agilent 1200 Infinity Series chromatograph using an Agilent Prep-C18 30 x 250 mm 10 μm column, or an Agilent Prep-C18 21.2 x 100 mm, 5 μm column.

Synthesis of PC Mimics 1–3

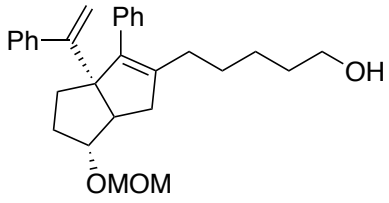


Compounds **S1a-S1h, S2a–h, 1a–h, and 2a–h** were synthesized and purified as described previously.[1]

Hybrid Precursors: Compounds S1a–S1h



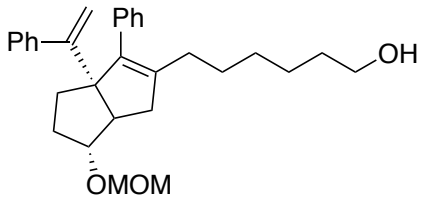
4-(6-*exo*-(methoxymethoxy)-3-phenyl-3a-(1-phenylvinyl)-1,3a,4,5,6,6a-hexahydropentalen-2-yl)butan-1-ol (**S1a**): **^{1}H NMR** (400 MHz, $CDCl_3$) δ 7.35 – 7.11 (m, 10H), 5.03 (d, *J* = 1.5 Hz, 1H), 4.99 (d, *J* = 1.5 Hz, 1H), 4.60 – 4.52 (m, 2H), 3.81 – 3.71 (m, 1H), 3.54 (t, *J* = 6.2 Hz, 2H), 3.29 (s, 3H), 2.47 – 2.35 (m, 1H), 2.31 – 2.26 (m, 1H), 2.09 – 1.99 (m, 4H), 1.78 – 1.54 (m, 4H), 1.52 – 1.35 (m, 3H). For the *endo* diastereomer (characteristic signals): **^{1}H NMR** (400 MHz, $CDCl_3$) δ 5.06 (d, *J* = 1.3 Hz, 1H), 4.84 (d, *J* = 1.4 Hz, 1H), 4.00 (td, *J* = 9.8, 6.1 Hz, 1H), 2.64 (dd, *J* = 17.1, 2.4 Hz, 1H), 2.53 (td, *J* = 9.0, 2.2 Hz, 1H). **^{13}C NMR** (126 MHz, $CDCl_3$) δ 154.4, 144.0, 140.8, 139.7, 137.3, 129.6, 127.8, 127.70, 127.65, 126.6, 114.9, 94.7, 86.7, 69.1, 62.6, 55.2, 52.8, 40.4, 32.7, 32.4, 31.5, 29.4, 24.0. **LRMS** (ESI, APCI) *m/z*: calc'd for $C_{27}H_{31}O_2$ $[M\text{-}OCH_3]^+$ 387.2, found 386.9.



[1] Flynn, A.R., Mays, S.G., Ortlund, E.A., and Jui, N.T.; *ACS Med. Chem. Lett.* **2018**, *9*(10), 1051–1056.

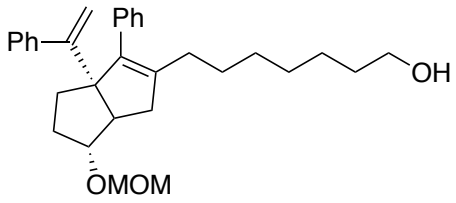
5-(6-*exo*-(methoxymethoxy)-3-phenyl-3a-(1-phenylvinyl)-1,3a,4,5,6,6a-hexahydropentalen-2-yl)pentan-1-ol (**S1b**): For the *exo* diastereomer: **^{1}H NMR** (600 MHz, $CDCl_3$) δ 7.41 – 7.10 (m, 10H), 5.01 (d, *J* = 1.4 Hz, 1H), 4.97 (d, *J* = 1.4 Hz, 1H), 4.60 – 4.51 (m, 2H), 3.76 (s, 1H), 3.55 (t, *J* = 6.5 Hz, 2H), 3.27 (s, 3H), 2.39 (d, *J* = 8.6 Hz, 1H), 2.29 (dd, *J* = 18.2, 9.6 Hz, 1H), 2.10 – 1.88 (m, 4H), 1.75 – 1.50 (m, 2H), 1.50 – 1.42 (m, 2H), 1.41 – 1.22 (m, 5H).

For the *endo* diastereomer (characteristic signals): **^{1}H NMR** (600 MHz, $CDCl_3$) δ 5.07 (d, *J* = 1.3 Hz, 1H), 4.84 (d, *J* = 1.4 Hz, 1H), 4.00 (td, *J* = 10.3, 5.8 Hz, 1H), 2.63 (dd, *J* = 17.3, 2.3 Hz, 1H), 2.52 (td, *J* = 9.2, 2.1 Hz, 1H). **LRMS** (ESI, APCI) *m*/*z*: calc'd for $C_{28}H_{33}O_2$ $[M\text{-}OCH_3]^+$ 401.2, found 401.2.



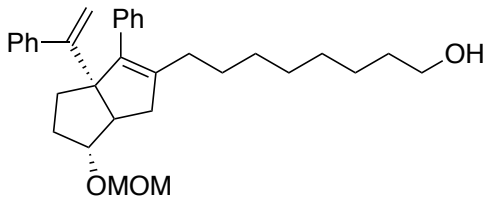
6-(6-*exo*-(methoxymethoxy)-3-phenyl-3a-(1-phenylvinyl)-1,3a,4,5,6,6a-hexahydropentalen-2-yl)hexan-1-ol (**S1c**): For the *exo* diastereomer: **^{1}H NMR** (600 MHz, $CDCl_3$) δ 7.36 – 7.11 (m, 10H), 5.03 (s, 1H), 4.98 (s, 1H), 4.61 – 4.51 (m, 2H), 3.77 (s, 1H), 3.57 (t, *J* = 6.7 Hz, 2H), 3.29 (s, 3H), 2.40 (d, *J* = 9.1, 1.8 Hz, 1H), 2.30 (dd, *J* = 16.9, 9.3 Hz, 1H), 2.10 – 1.96 (m, 4H), 1.79 – 1.53 (m, 2H), 1.54 – 1.45 (m, 2H), 1.43 – 1.23 (m, 7H). For the *endo* diastereomer (characteristic signals): **^{1}H NMR** (600 MHz, $CDCl_3$) δ 5.08 (d, *J* = 1.4 Hz, 1H), 4.85 (d, *J* = 1.4 Hz, 1H), 4.01 (td, *J* = 9.8, 5.6 Hz, 1H), 2.65 (dd, *J* = 17.3, 2.2 Hz, 1H), 2.54 (td, *J* = 9.1, 2.3 Hz, 1H).

^{13}C NMR (126 MHz, $CDCl_3$) δ 154.5, 144.1, 141.1, 139.4, 137.4, 134.8, 129.6, 127.8, 127.63, 126.59, 114.9, 94.7, 86.7, 69.1, 62.8, 55.1, 52.8, 40.5, 32.6, 32.41, 31.42, 29.6, 29.4, 27.8, 25.5. **LRMS** (ESI, APCI) *m*/*z*: calc'd for $C_{29}H_{35}O_2$ $[M\text{-}OCH_3]^+$ 415.3, found 415.3.



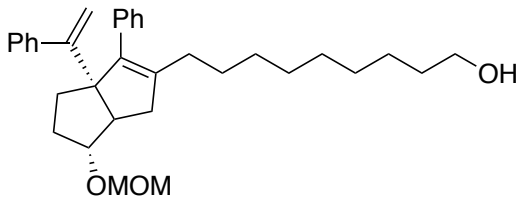
7-(6-*exo*-(methoxymethoxy)-3-phenyl-3a-(1-phenylvinyl)-1,3a,4,5,6,6a-hexahydropentalen-2-yl)heptan-1-ol (**S1d**): For the *exo* diastereomer: **^{1}H NMR** (600 MHz, $CDCl_3$) δ 7.46 – 7.02 (m, 10H), 5.05 (s, 1H), 5.00 (s, 1H), 4.62 – 4.56 (m, 2H), 3.79 (s, 1H), 3.66 – 3.60 (m, 2H), 3.30 (s, 3H), 2.41 (d, *J* = 9.4 Hz, 1H), 2.32 (dd, *J* = 17.0, 9.2 Hz, 1H), 2.08 – 1.92 (m, 4H), 1.78 – 1.67 (m, 2H), 1.59 – 1.49 (m, 2H), 1.44 – 1.19 (m, 9H). For the *endo* diastereomer (characteristic signals): **^{1}H NMR** (600 MHz, $CDCl_3$) δ 5.08 (d, *J* = 1.3 Hz, 1H), 4.85 (d, *J* = 1.2 Hz, 1H), 4.00 (td, *J* = 9.6, 5.6 Hz, 1H), 2.65 (dd, *J* = 17.4, 2.1 Hz, 1H), 2.53 (td, *J* = 8.9, 2.2 Hz, 1H).

^{13}C NMR (126 MHz, $CDCl_3$) δ 154.5, 144.1, 141.2, 139.3, 137.5, 129.6, 127.8, 127.63, 126.57, 114.9, 94.7, 86.7, 69.1, 63.0, 55.1, 52.8, 40.5, 32.7, 32.4, 31.4, 29.64, 29.58, 29.2, 27.7, 25.6. **LRMS** (ESI, APCI) *m/z*: calc'd for $C_{30}H_{37}O_2$ $[M\text{-}OCH_3]^+$ 429.3, found 428.8.



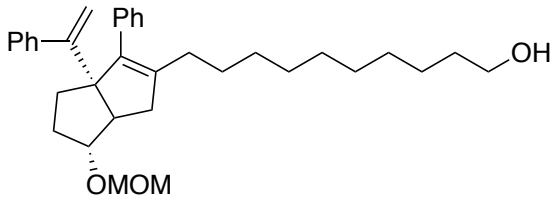
8-(6-*exo*-(methoxymethoxy)-3-phenyl-3a-(1-phenylvinyl)-1,3a,4,5,6,6a-hexahydropentalen-2-yl)octan-1-ol (S1e): For the *exo* diastereomer: **^{1}H NMR** (500 MHz, $CDCl_3$) δ 7.41 – 7.18 (m, 10H), 5.05 (d, *J* = 1.5 Hz, 1H), 5.01 (d, *J* = 1.5 Hz, 1H), 4.63 – 4.53 (m, 2H), 3.79 (s, 1H), 3.62 (t, *J* = 6.7 Hz, 2H), 3.31 (s, 3H), 2.42 (d, *J* = 9.4, 1.8 Hz, 1H), 2.32 (dd, *J* = 16.9, 9.1 Hz, 1H), 2.11 – 1.97 (m, 4H), 1.83 – 1.54 (m, 4H), 1.59 – 1.50 (m, 2H), 1.38 – 1.17 (m, 9H). For the *endo* diastereomer (characteristic signals): **^{1}H NMR** (500 MHz, $CDCl_3$) δ 5.08 (d, *J* = 1.3 Hz, 1H), 4.85 (d, *J* = 1.3 Hz,

1H), 4.01 (td, *J* = 9.6, 5.8 Hz, 1H), 2.65 (dd, *J* = 17.4, 2.1 Hz, 1H), 2.53 (td, *J* = 8.9, 2.2 Hz, 1H). **^{13}C NMR** (126 MHz, $CDCl_3$) δ 154.5, 144.1, 141.3, 139.3, 137.5, 129.6, 127.8, 127.6, 126.59, 126.55, 114.8, 94.7, 86.7, 69.1, 63.0, 55.1, 52.8, 40.5, 32.8, 32.4, 31.4, 29.7, 29.6, 29.33, 29.27, 27.8, 25.7. **LRMS** (ESI, APCI) *m/z*: calc'd for $C_{31}H_{39}O_2$ $[M\text{-}OCH_3]^+$ 443.3, found 442.9.

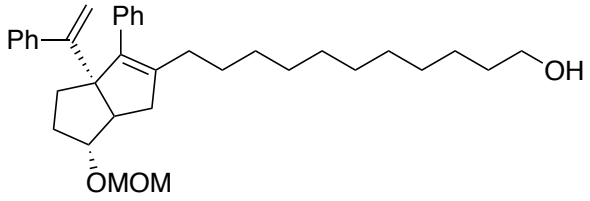


9-(6-*exo*-(methoxymethoxy)-3-phenyl-3a-(1-phenylvinyl)-1,3a,4,5,6,6a-hexahydropentalen-2-yl)nonan-1-ol (**S1f**): For the *exo* diastereomer: **^{1}H NMR** (500 MHz, $CDCl_3$) δ 7.39 – 7.15 (m, 10H), 5.06 (d, *J* = 1.4 Hz, 1H), 5.01 (d, *J* = 1.5 Hz, 1H), 4.62 – 4.56 (m, 2H), 3.80 (s, 1H), 3.62 (t, *J* = 6.7 Hz, 2H), 3.32 (s, 3H), 2.43 (d, *J* = 9.1, 1.7 Hz, 1H), 2.33 (dd, *J* = 16.9, 9.2 Hz, 1H), 2.11 – 1.99 (m, 4H), 1.79 – 1.60 (m, 4H), 1.60 – 1.51 (m, 2H), 1.43 – 1.17 (m, 11H).

For the *endo* diastereomer (characteristic signals): **^{1}H NMR** (500 MHz, $CDCl_3$) δ 5.09 (d, *J* = 1.4 Hz, 1H), 4.86 (d, *J* = 1.4 Hz, 1H), 4.01 (td, *J* = 9.6, 5.8 Hz, 1H), 2.66 (dd, *J* = 17.3, 2.4 Hz, 1H), 2.54 (td, *J* = 9.0, 2.3 Hz, 1H). **^{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 $C_{32}H_{41}O_2$ $[M\text{-}CH_3O]^+$ 457.3, found 457.8.



10-(6-*exo*-(methoxymethoxy)-3-phenyl-3a-(1-phenylvinyl)-1,3a,4,5,6,6a-hexahydropentalen-2-yl)decan-1-ol (**S1g**): For the *exo* diastereomer: **^{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). For the *endo* diastereomer (characteristic signals): **^{1}H NMR** (500 MHz, $CDCl_3$) δ 5.10 (d, *J* = 1.3 Hz, 1H), 4.87 (d, *J* = 1.4 Hz, 1H), 4.03 (td, *J* = 9.7, 5.6 Hz, 1H), 2.68 (dd, *J* = 17.3, 2.3 Hz, 1H), 2.56 (td, *J* = 9.0, 2.3 Hz, 1H). **^{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.

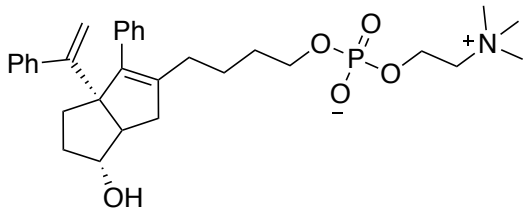


11-(6-*exo*-(methoxymethoxy)-3-phenyl-3a-(1-phenylvinyl)-1,3a,4,5,6,6a-hexahydropentalen-2-yl)undecan-1-ol (**S1h**): For the *exo* diastereomer: **^{1}H NMR** (500 MHz, $CDCl_3$) δ 7.42 – 7.16 (m, 10H), 5.08 (d, *J* = 1.5 Hz, 1H), 5.03 (d, *J* = 1.5 Hz, 1H), 4.65 – 4.56 (m, 3H), 3.82 (s, 1H), 3.64 (t, *J* = 6.7 Hz, 2H), 3.33 (s, 3H), 2.45 (d, *J* = 9.2, 1.7 Hz, 1H), 2.36 (dd, *J* = 16.9, 9.2 Hz, 1H), 2.13 – 2.00 (m, 4H), 1.89 – 1.61 (m, 4H), 1.62 – 1.51 (m, 2H), 1.41 – 1.19 (m, 15H).

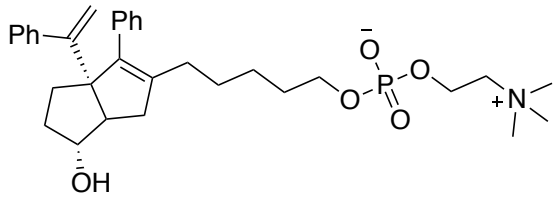
For the *endo* diastereomer (characteristic signals): **^{1}H NMR** (500 MHz, $CDCl_3$) δ 5.11 (d, *J* = 1.4 Hz, 1H), 4.88 (d, *J* = 1.4 Hz, 1H), 4.03 (td, *J* = 9.8, 5.8 Hz, 1H), 2.68 (dd, *J* = 17.3, 2.3 Hz, 1H), 2.57 (td, *J* = 9.0, 2.3 Hz, 1H). **^{13}C NMR** (126 MHz, $CDCl_3$) δ 154.5, 144.1, 141.4, 139.2, 137.5, 129.6, 127.8, 127.7, 126.64, 126.58, 114.9, 94.7, 86.8, 69.1, 62.9, 55.2, 52.7, 40.6, 32.8, 32.5, 31.5, 29.7, 29.7, 29.61,

29.58, 29.54, 29.46, 29.4, 27.9, 25.8. **LRMS** (ESI, APCI) *m/z*: calc'd for $C_{34}H_{45}O_2$ $[M\text{-}CH_3O]^+$ 485.3, found 484.9.

Phosphorylcholines: Compounds 1a–h.

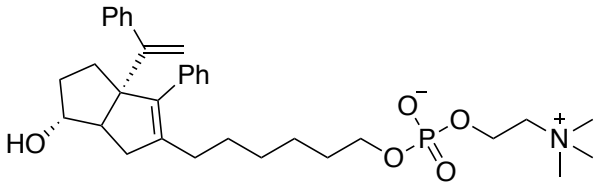


4-(6-*exo*-hydroxy-3-phenyl-3a-(1-phenylvinyl)-1,3a,4,5,6,6a-hexahydropentalen-2-yl)butyl (2-(trimethylammonio)ethyl) phosphate (**1a**): **^{1}H NMR** (500 MHz, $CDCl_3$) δ 7.28 – 7.13 (m, 10H), 5.02 (s, 1H), 4.99 (s, 1H), 4.19 (s, 2H), 3.81 (s, 1H), 3.76 (s, 2H), 3.64 (s, 2H), 3.21 (s, 9H), 2.26 – 2.05 (m, 3H), 2.01 – 1.87 (m, 1H), 1.68 – 1.36 (m, 9H). **^{31}P NMR (**121 MHz, $CDCl_3$) δ -0.76. **LRMS** (ESI, APCI) *m/z*: calc'd for $C_{31}H_{43}O_5NP$ $[M+H]^+$: 540.3, found 540.3.

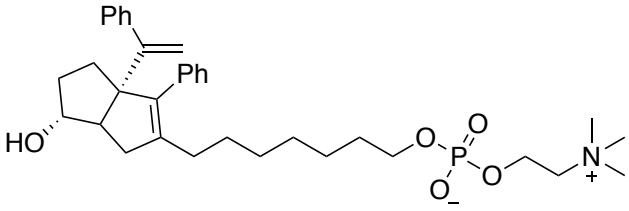


5-(6-*exo*-hydroxy-3-phenyl-3a-(1-phenylvinyl)-1,3a,4,5,6,6a-hexahydropentalen-2-yl)pentyl (2-(trimethylammonio)ethyl) phosphate (1b): ^{1}H NMR (600 MHz, $CDCl_3$) δ 7.31 – 7.24 (m, 4H), 7.25 – 7.14 (m, 6H), 4.99 (s, 2H), 4.19 (s, 2H), 3.86 (s, 1H), 3.75 (s, 2H), 3.63 (s, 2H), 3.20 (s, 9H), 2.25 – 2.18 (m, 1H), 2.16 – 2.04 (m, 2H), 1.95 – 1.87 (m, 1H), 1.70 – 1.60 (m, 1H), 1.60 – 1.52

(m, 4H), 1.53 – 1.43 (m, 1H), 1.42 – 1.30 (m, 3H), 1.31 – 1.18 (m, 2H). **^{31}P NMR** (121 MHz, $CDCl_3$) δ -0.82. **LRMS** (ESI, APCI) *m/z*: calc'd for $C_{32}H_{45}O_5NP$ $[M+H]^+$: 554.3, found 554.2.

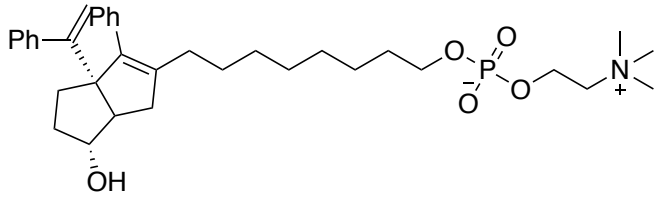


6-(6-*exo*-hydroxy-3-phenyl-3a-(1-phenylvinyl)-1,3a,4,5,6,6a-hexahydropentalen-2-yl)hexyl (2-(trimethylammonio)ethyl) phosphate (**1c**): **^{1}H NMR** (500 MHz, $CDCl_3$) δ 7.41 – 7.26 (m, 4H), 7.28 – 7.12 (m, 6H), 5.03 (s, 1H), 5.01 (s, 1H), 4.24 (s, 2H), 3.88 (s, 1H), 3.80 – 3.67 (m, 4H), 3.28 (s, 9H), 2.28 – 2.05 (m, 3H), 2.01 – 1.80 (m, 1H), 1.72 – 1.49 (m, 3H), 1.40 – 0.82 (m, 10H). **^{31}P NMR** (121 MHz, Chloroform-*d*) δ -0.59. **LRMS** (ESI, APCI) *m/z*: calc'd for $C_{33}H_{47}O_5NP$ $[M+H]^+$ 568.3, found 568.2.

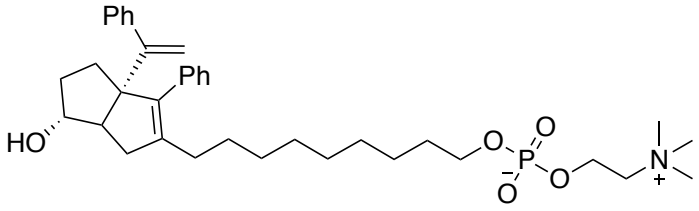


7-(6-*exo*-hydroxy-3-phenyl-3a-(1-phenylvinyl)-1,3a,4,5,6,6a-hexahydropentalen-2-yl)heptyl (2-(trimethylammonio)ethyl) phosphate (**1d**): **^{1}H NMR** (500 MHz, $CDCl_3$) δ 7.37 – 7.11 (m, 10H), 5.02 (s, 1H), 4.99 (s, 1H), 4.21 (s, 2H), 3.87 (s, 1H), 3.77 (s, 2H), 3.67 (s, 1H), 3.25 (s, 9H), 2.29 – 2.16 (m, 1H), 2.15 – 2.02 (m, 2H), 2.00 – 1.90 (m, 1H), 1.72 – 1.57 (m, 3H), 1.57 – 1.48 (m, 3H), 1.37 – 1.14 (m, 9H). **^{13}C NMR** (300 MHz, $CDCl_3$) δ 154.6, 144.1, 140.9, 139.6, 137.4, 129.6,

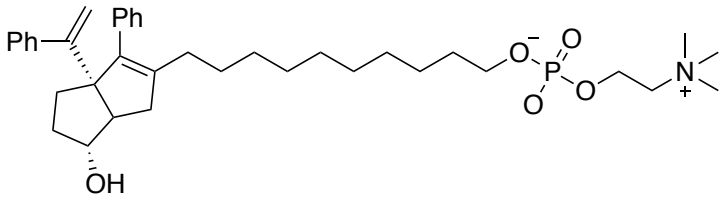
127.8, 127.7, 126.64, 114.7, 81.5, 69.1, 66.1, 65.7, 59.2, 55.7, 54.3, 40.0, 34.2, 32.0, 30.8, 29.5, 29.3, 29.2, 27.6, 25.6. **^{31}P NMR** (300 MHz, $CDCl_3$) δ -0.51. **HRMS** (ESI) *m/z*: calc'd for $C_{34}H_{49}O_5NP$ $[M+H]^+$: 582.3343, found 582.3338.



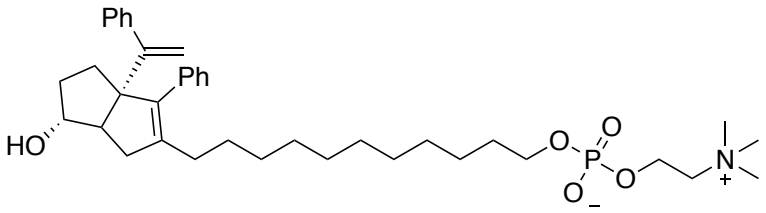
8-(6-*exo*-hydroxy-3-phenyl-3a-(1-phenylvinyl)-1,3a,4,5,6,6a-hexahydropentalen-2-yl)octyl (2-(trimethylammonio)ethyl) phosphate (1e): **^{1}H NMR** (600 MHz, $CDCl_3$) δ 7.33 –7.15 (m, 10H), 5.00 (s, 1H), 4.96 (s, 1H), 4.22 (s, 2H), 3.89 (s, 1H), 3.81 – 3.75 (m, 2H), 3.74 – 3.66 (m, 2H), 3.27 (s, 9H), 2.28 – 2.17 (m, 2H), 2.10 – 2.04 (m, 1H), 1.96 – 1.87 (m, 1H), 1.71 – 1.58 (m, 3H), 1.57 – 1.48 (m, 2H), 1.36 – 1.13 (m, 12H). **^{13}C NMR** (151 MHz, $CDCl_3$) δ 154.6, 144.1, 141.0, 139.5, 137.4, 129.6, 127.8, 127.6, 126.61, 126.55, 114.8, 81.7, 69.2, 66.3, 59.1, 55.6, 54.4, 53.4, 40.1, 37.1, 34.3, 32.0, 30.8, 29.4, 29.2, 28.9, 28.8, 27.5, 25.6, 22.6. **^{31}P NMR** (300 MHz, $CDCl_3$) δ -0.43. **LRMS** (ESI, APCI) *m/z*: calc'd for $C_{35}H_{51}O_5NP$ $[M+H]^+$ 596.3, found 596.3. **HRMS** (ESI) *m/z*: calc'd for $C_{35}H_{51}O_5NP$ $[M+H]^+$: 596.3499, found 596.3494.



9-(6-*exo*-hydroxy-3-phenyl-3a-(1-phenylvinyl)-1,3a,4,5,6,6a-hexahydropentalen-2-yl)nonyl (2-(trimethylammonio)ethyl) phosphate (**1f**): **^{1}H NMR** (600 MHz, $CDCl_3$) δ 7.32 – 7.14 (m, 10H), 5.01 (d, *J* = 7.6 Hz, 1H), 4.95 (d, *J* = 7.9 Hz, 1H), 4.20 (s, 2H), 3.87 (s, 1H), 3.75 (s, 3H), 3.67 (s, 3H), 3.25 (s, 9H), 2.27 – 2.18 (m, 2H), 2.06 – 1.98 (m, 2H), 1.94 (p, *J* = 7.0 Hz, 1H), 1.70 – 1.58 (m, 4H), 1.58 – 1.48 (m, 3H), 1.36 – 1.10 (m, 12H). **^{13}C NMR** (126 MHz, $CDCl_3$) δ 154.7, 144.2, 141.1, 139.4, 137.4, 129.7, 127.8, 127.7, 127.6, 126.63, 126.57, 114.9, 81.7, 69.3, 66.2, 65.9, 59.3, 55.7, 54.4, 40.2, 34.2, 32.1, 30.9, 29.5, 29.3, 29.2, 27.7, 27.6, 25.8. **^{31}P NMR** (121 MHz, $CDCl_3$) δ -0.73. **LRMS** (ESI, APCI) *m/z*: calc'd for $C_{36}H_{53}O_5NP$ $[M+H]^+$ 610.4, found 609.8. **HRMS** (ESI) *m/z*: calc'd for $C_{36}H_{53}O_5NP$ $[M+H]^+$: 610.3656, found 610.3655.

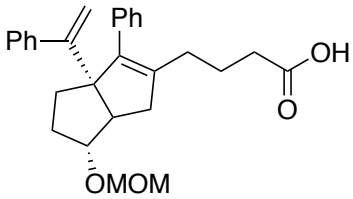


10-(6-*exo*-hydroxy-3-phenyl-3a-(1-phenylvinyl)-1,3a,4,5,6,6a-hexahydropentalen-2-yl)decyl (2-(trimethylammonio)ethyl) phosphate (**1g**): **^{1}H NMR** (600 MHz, $CDCl_3$) δ 7.34 – 7.14 (m, 10H), 5.03 (d, *J* = 4.9 Hz, 1H), 4.95 (d, *J* = 5.0 Hz, 1H), 4.24 (s, 2H), 3.90 (s, 1H), 3.79 (q, *J* = 6.4 Hz, 2H), 3.74 (s, 2H), 3.28 (s, 9H), 2.34 – 2.23 (m, 2H), 2.10 – 1.92 (m, 3H), 1.64 (q, *J* = 10.2, 6.5 Hz, 2H), 1.56 (t, *J* = 7.3 Hz, 2H), 1.37 – 1.10 (m, 16H). **^{13}C NMR** (126 MHz, $CDCl_3$) δ 154.7, 144.2, 141.1, 139.4, 137.4, 129.7, 127.73, 127.70, 127.6, 126.63, 126.56, 81.7, 69.3, 66.2, 65.9, 59.3, 55.6, 54.3, 40.2, 34.1, 32.2, 30.91, 30.86, 29.6, 29.3, 29.22, 29.17, 27.6, 25.8. **^{31}P NMR** (121 MHz, $CDCl_3$) δ -0.75. **LRMS** (ESI, APCI) *m/z*: calc'd for $C_{37}H_{55}O_5NP$ $[M+H]^+$: 624.4, found 624.3.



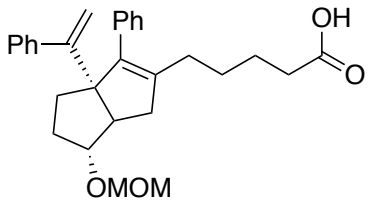
11-(6-*exo*-hydroxy-3-phenyl-3a-(1-phenylvinyl)-1,3a,4,5,6,6a-hexahydropentalen-2-yl)undecyl (2-(trimethylammonio)ethyl) phosphate (**1h**): **^{1}H NMR** (600 MHz, $CDCl_3$) δ 7.35 – 7.15 (m, 12H), 5.03 (s, 1H), 4.96 (s, 1H), 4.26 (s, 2H), 3.91 (s, 1H), 3.85 – 3.77 (m, 2H), 3.75 (s, 2H), 3.29 (s, 9H), 2.32 (dd, *J* = 16.7, 9.4 Hz, 1H), 2.26 (d, *J* = 9.6 Hz, 1H), 2.10 – 1.94 (m, 4H), 1.71 – 1.61 (m, 3H), 1.61 – 1.53 (m, 2H), 1.38 – 1.14 (m, 16H). **^{13}C NMR** (126 MHz, $CDCl_3$) δ 154.7, 144.2, 141.1, 139.3, 137.4, 129.7, 114.9, 81.7, 69.3, 66.2, 65.9, 59.3, 55.7, 54.4, 40.3, 34.0, 32.2, 31.6, 30.91, 30.86, 29.6, 29.5, 29.43, 29.40, 29.36, 29.3, 29.2, 27.7, 25.8, 22.7. **^{31}P NMR** (121 MHz, $CDCl_3$) δ -0.90.m**LRMS** (ESI, APCI) *m/z*: calc'd for $C_{38}H_{57}O_5NP$ $[M+H]^+$ 638.4, found 637.8 **HRMS** (ESI) *m/z*: calc'd for $C_{38}H_{57}O_5NP$ $[M+H]^+$: 638.3969, found 638.3974.

Carboxylate Precursors: Compounds S2a–h.

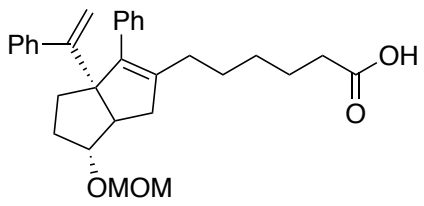


4-(6-*exo*-(methoxymethoxy)-3-phenyl-3a-(1-phenylvinyl)-1,3a,4,5,6,6a-hexahydropentalen-2-yl)butanoic acid (S2a): For the *exo* diastereomer: **^{1}H NMR** (500 MHz, $CDCl_3$) δ 7.47 – 7.05 (m, 10H), 5.07 (s, 1H), 5.01 (s, 1H), 4.66 – 4.52 (m, 2H), 3.80 (s, 1H), 3.31 (s, 3H), 2.45 (d, *J* = 8.9 Hz, 1H), 2.37 – 2.20 (m, 2H), 2.13 – 1.98 (m, 4H), 1.81 – 1.57 (m, 3H), 1.36 – 1.18 (m, 3H). For the *endo* diastereomer (characteristic signals): ^{1}H NMR (500 MHz, $CDCl_3$) δ 5.08 (s, 1H), 4.88 (s, 1H), 4.06 – 3.98 (m, 1H), 2.67 (d, *J* = 17.8 Hz, 1H). **^{13}C NMR** (126 MHz, cdcl$_3$) δ 179.0, 154.3, 143.9, 140.6,

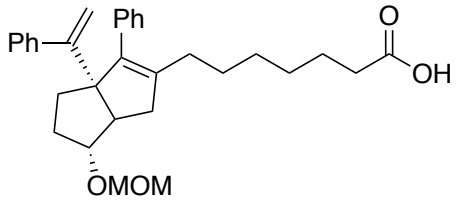
139.7, 137.1, 129.5, 127.80, 127.75, 127.7, 126.8, 115.0, 94.7, 86.6, 69.1, 55.2, 52.8, 40.3, 33.6, 32.4, 31.5, 29.1, 22.9. **LRMS** (ESI, APCI) *m/z*: calc'd for $C_{27}H_{29}O_3$ $[M\text{-}OCH_3]^+$ 401.2, found 400.8.



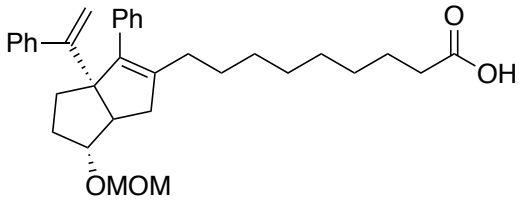
5-(6-*exo*-(methoxymethoxy)-3-phenyl-3a-(1-phenylvinyl)-1,3a,4,5,6,6a-hexahydropentalen-2-yl)pentanoic acid (**S2b):** For the *exo* diastereomer: **^{1}H NMR** (500 MHz, $CDCl_3$) δ 7.37 – 7.14 (m, 10H), 5.06 (d, *J* = 1.4 Hz, 1H), 5.02 (d, *J* = 1.4 Hz, 1H), 4.63 – 4.57 (m, 2H), 3.80 (s, 1H), 3.32 (s, 3H), 2.43 (d, *J* = 9.2 Hz, 1H), 2.35 – 2.26 (m, 3H), 2.10 – 2.00 (m, 4H), 1.79 – 1.53 (m, 4H), 1.46 – 1.20 (m, 3H). For the *endo* diastereomer (characteristic signals): **^{1}H NMR** (500 MHz, $CDCl_3$) δ 5.07 (d, *J* = 1.3 Hz, 1H), 4.85 (d, *J* = 1.3 Hz, 1H), 3.99 (dt, *J* = 9.6, 5.8 Hz, 1H), 2.63 (dd, *J* = 17.2, 1.8 Hz, 1H), 2.54 (dd, *J* = 9.5, 2.2 Hz, 1H). **^{13}C NMR** (151 MHz, $CDCl_3$) δ 179.5, 154.3, 144.0, 140.4, 139.9, 137.2, 129.5, 127.8, 127.7, 127.6, 126.7, 126.6, 114.9, 86.6, 69.1, 55.1, 52.7, 40.3, 32.4, 31.4, 29.9, 29.3, 27.2, 24.6. **LRMS** (ESI, APCI) *m/z*: calc'd for $C_{29}H_{33}O_4$ $[M\text{-}H]^-$ 445.2, found 445.1. Calc'd for $C_{28}H_{31}O_3$ $[M\text{-}OCH_3]^+$ 415.2, found 415.2.



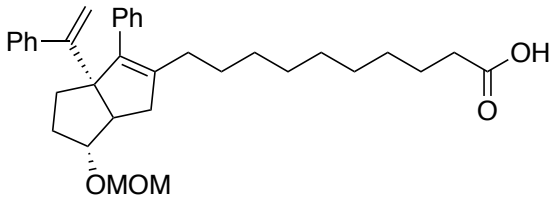
6-(6-*exo*-(methoxymethoxy)-3-phenyl-3a-(1-phenylvinyl)-1,3a,4,5,6,6a-hexahydropentalen-2-yl)hexanoic acid (**S2c**)**:** For the *exo* diastereomer: **^{1}H NMR** (500 MHz, $CDCl_3$) δ 7.36 – 7.11 (m, 10H), 5.05 (d, *J* = 1.4 Hz, 1H), 5.02 (d, *J* = 1.5 Hz, 1H), 3.80 (s, 1H), 3.33 (s, 3H), 2.41 (d, *J* = 8.7 Hz, 1H), 2.38 – 2.29 (m, 2H), 2.24 (dd, *J* = 17.0, 8.9 Hz, 1H), 2.12 – 1.98 (m, 4H), 1.79 – 1.49 (m, 7H), 1.47 – 1.18 (m, 4H). For the *endo* diastereomer: **^{1}H NMR** (600 MHz, $CDCl_3$) δ 5.06 (d, *J* = 1.3 Hz, 1H), 4.84 (d, *J* = 1.3 Hz, 1H), 4.00 (td, *J* = 9.8, 5.8 Hz, 1H), 2.63 (dd, *J* = 17.4, 2.2 Hz, 1H), 2.53 (td, *J* = 9.0, 2.2 Hz, 1H). **^{13}C NMR** (151 MHz, $CDCl_3$) δ 179.5, 144.1, 140.7, 139.5, 137.2, 129.6, 127.72, 127.69, 127.66, 126.66, 126.65, 115.0, 82.0, 69.3, 60.4, 53.4, 40.1, 34.0, 32.0, 31.6, 29.4, 29.0, 27.4, 27.0, 22.6, 21.1. **LRMS** (ESI, APCI) *m*/*z*: calc'd for $C_{30}H_{35}O_4$ $[M-H]^-$ 459.3, found 459.3. Calc'd for $C_{29}H_{33}O_3$ $[M-OCH_3]^+$ 429.2, found 429.3.



7-(6-*exo*-(methoxymethoxy)-3-phenyl-3a-(1-phenylvinyl)-1,3a,4,5,6,6a-hexahydropentalen-2-yl)heptanoic acid (**S2d**)**:** **^{1}H NMR** (500 MHz, $CDCl_3$) δ 7.60 – 7.02 (m, 10H), 5.05 (d, *J* = 1.5 Hz, 1H), 5.00 (d, *J* = 1.5 Hz, 1H), 4.66 – 4.52 (m, 2H), 3.79 (s, 1H), 3.31 (s, 3H), 2.42 (d, *J* = 9.3, 1.7 Hz, 1H), 2.40 – 2.27 (m, 3H), 2.10 – 1.97 (m, 4H), 1.79 – 1.50 (m, 4H), 1.52 – 1.18 (m, 7H). For the *endo* diastereomer (characteristic signals): **^{1}H NMR** (500 MHz, Chloroform-*d*) δ 5.08 (d, *J* = 1.4 Hz, 1H), 4.85 (d, *J* = 1.4 Hz, 1H), 4.01 (td, *J* = 9.1, 5.8 Hz, 1H), 2.64 (d, *J* = 17.5 Hz, 1H), 2.54 (t, *J* = 9.1 Hz, 1H). **^{13}C NMR** (126 MHz, $CDCl_3$) δ 179.5, 154.5, 144.1, 141.1, 139.4, 137.4, 134.6, 129.6, 127.8, 127.7, 126.6, 114.9, 94.7, 86.7, 69.1, 55.1, 52.8, 40.5, 33.9, 32.4, 31.4, 29.6, 29.2, 28.8, 27.6, 24.5. **LRMS** (ESI, APCI) *m*/*z*: calc'd for $C_{31}H_{37}O_4$ $[M-H]^-$ 473.3, found 473.4.

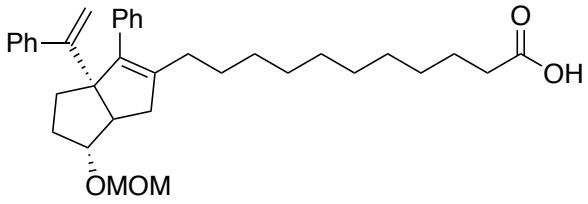


9-(6-*exo*-(methoxymethoxy)-3-phenyl-3a-(1-phenylvinyl)-1,3a,4,5,6,6a-hexahydropentalen-2-yl)nonanoic acid (**S2f**): For the *exo* diastereomer: **^{1}H NMR** (500 MHz, $CDCl_3$) δ 7.39 – 7.18 (m, 10H), 5.05 (d, *J* = 1.4 Hz, 1H), 5.00 (d, *J* = 1.5 Hz, 1H), 4.65 – 4.57 (m, 2H), 3.80 (d, *J* = 2.1 Hz, 1H), 3.32 (d, *J* = 0.7 Hz, 3H), 2.42 (d, 1H), 2.37 – 2.29 (m, 3H), 2.10 – 1.98 (m, 4H), 1.69 – 1.58 (m, 5H), 1.44 – 1.08 (m, 11H). For the *endo* diastereomer (characteristic signals): **^{1}H NMR** (500 MHz, $CDCl_3$) δ 5.09 (d, *J* = 1.3 Hz, 1H), 4.85 (d, *J* = 1.4 Hz, 1H), 4.02 (td, *J* = 9.8, 5.4 Hz, 1H), 2.65 (dd, *J* = 17.7, 2.1 Hz, 1H), 2.54 (td, *J* = 9.4, 2.1 Hz, 1H). **^{13}C NMR** (126 MHz, $CDCl_3$) δ 179.2, 154.5, 144.1, 141.3, 139.3, 137.4, 129.6, 127.8, 127.6, 126.61, 126.57, 114.9, 94.6, 86.8, 69.1, 55.1, 52.8, 40.5, 33.9, 32.4, 31.4, 29.7, 29.5, 29.2, 29.1, 29.0, 27.7, 24.7. **LRMS** (ESI, APCI) *m/z*: calc'd for $C_{33}H_{41}O_4$ $[M-H]^-$ 501.3, found 501.4. Calc'd for $C_{32}H_{39}O_3$ $[M-CH_3O]^+$ 471.3, found 470.8.



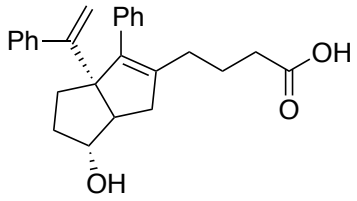
10-(6-*exo*-(methoxymethoxy)-3-phenyl-3a-(1-phenylvinyl)-1,3a,4,5,6,6a-hexahydropentalen-2-yl)decanoic acid (S2g): For the *exo* diastereomer: **^{1}H NMR** (500 MHz, $CDCl_3$) δ 7.38 – 7.18 (m, 10H), 5.05 (d, *J* = 1.4 Hz, 1H), 5.01 (d, *J* = 1.5 Hz, 1H), 4.61 – 4.59 (m, 2H), 3.80 (s, 1H), 3.32 (s, 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). For the *endo* diastereomer (characteristic signals): **^{1}H NMR** (500 MHz, $CDCl_3$) δ

5.09 (d, *J* = 1.3 Hz, 1H), 4.86 (d, *J* = 1.3 Hz, 1H), 4.02 (td, *J* = 9.7, 5.5 Hz, 1H), 2.65 (dd, *J* = 17.7, 2.2 Hz, 1H), 2.54 (td, *J* = 9.3, 2.2 Hz, 1H). **^{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.

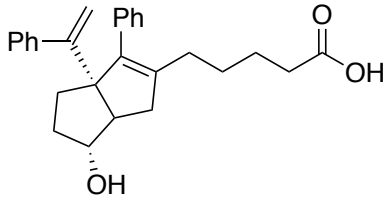


11-(6-*exo*-(methoxymethoxy)-3-phenyl-3a-(1-phenylvinyl)-1,3a,4,5,6,6a-hexahydropentalen-2-yl)undecanoic acid (**S2h**): For the *exo* diastereomer: **^{1}H NMR** (500 MHz, $CDCl_3$) δ 7.39 – 7.18 (m, 10H), 5.05 (d, *J* = 1.2 Hz, 1H), 5.00 (d, *J* = 1.2 Hz, 1H), 4.64 – 4.58 (m, 2H), 3.80 (s, 1H), 3.32 (s, 3H), 2.42 (d, *J* = 8.8 Hz, 1H), 2.38 – 2.30 (m, 3H), 2.11 – 1.98 (m, 4H), 1.72 – 1.59 (m, 5H), 1.44 – 1.17 (m, 16H). For the *endo* diastereomer (characteristic signals): **^{1}H NMR** (500 MHz, Chloroform-*d*) δ 5.09 (d, *J* = 1.2 Hz, 1H), 4.86 (d, *J* = 1.2 Hz, 1H), 4.02 (td, *J* = 9.4, 5.7 Hz, 1H), 2.66 (dd, *J* = 17.6, 2.2 Hz, 1H), 2.54 (td, *J* = 9.9, 2.1 Hz, 1H). **^{13}C NMR** (126 MHz, $CDCl_3$) δ 179.2, 154.5, 144.1, 141.4, 139.2, 137.5, 129.6, 127.8, 127.63, 126.60, 126.55, 114.9, 94.7, 86.7, 69.1, 55.1, 52.7, 40.5, 33.9, 32.4, 31.4, 29.7, 29.6, 29.38, 29.36, 29.3, 29.2, 29.0, 27.8, 24.7. **LRMS** (ESI, APCI) *m/z*: calc'd for $C_{35}H_{46}O_4$ $[M-H]^-$ 529.3, found 529.5. Calc'd for $C_{34}H_{43}O_3$ $[M-CH_3O]^+$ 499.3, found 498.9.

Carboxylates: Compounds 2a–2h

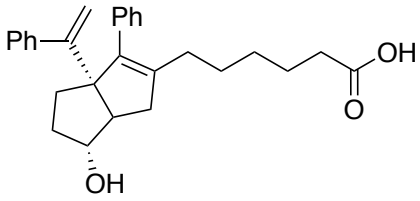


4-(6-*exo*-hydroxy-3-phenyl-3a-(1-phenylvinyl)-1,3a,4,5,6,6a-hexahydropentalen-2-yl)butanoic acid (**2a**): **^{1}H NMR** (500 MHz, $CDCl_3$) δ 7.38 – 7.16 (m, 10H), 5.08 (d, *J* = 1.4 Hz, 1H), 5.00 (d, *J* = 1.4 Hz, 1H), 3.97 (s, 1H), 2.42 – 2.36 (m, 1H), 2.37 – 2.31 (m, 2H), 2.30 – 2.23 (m, 2H), 2.14 – 2.06 (m, 4H), 1.75 – 1.65 (m, 5H). **^{13}C NMR** (126 MHz, $CDCl_3$) δ 178.7, 154.4, 144.0, 140.4, 139.6, 137.0, 129.6, 127.8, 127.7, 126.81, 126.75, 115.2, 82.0, 69.4, 55.7, 40.1, 34.0, 33.8, 32.0, 29.1, 22.9. **LRMS** (ESI, APCI) *m/z*: calc'd for $C_{26}H_{27}O_3$ $[M-H]^-$ 387.2, found 387.1. **HRMS** calc'd for $C_{26}H_{27}O_3$ $[M-H]^-$: 387.1966, found 387.1969.

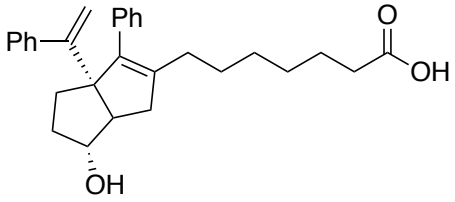


5-(6-hydroxy-3-phenyl-3a-(1-phenylvinyl)-1,3a,4,5,6,6a-hexahydropentalen-2-yl)pentanoic acid (2b): For the *exo* diastereomer: **^{1}H NMR** (600 MHz, $CDCl_3$) δ 7.56 – 6.96 (m, 10H), 5.05 (d, *J* = 1.4 Hz, 1H), 4.97 (d, *J* = 1.4 Hz, 1H), 3.93 (s, 1H), 2.38 – 2.23 (m, 4H), 2.13 – 1.97 (m, 5H), 1.74 – 1.62 (m, 3H), 1.53 (p, *J* = 7.4 Hz, 2H), 1.44 – 1.32 (m, 2H).

For the *endo* diastereomer (characteristic signals): ^{1}H NMR (600 MHz, $CDCl_3$) δ 5.07 (d, *J* = 1.6 Hz, 1H), 4.94 (d, *J* = 1.4 Hz, 1H), 4.20 (td, *J* = 9.0, 5.7 Hz, 1H), 2.64 (dd, *J* = 17.3, 2.0 Hz, 1H), 2.50 (td, *J* = 8.6, 1.9 Hz, 1H). **^{13}C NMR** (126 MHz, $CDCl_3$) δ 177.8, 154.5, 144.1, 140.3, 139.8, 137.2, 129.7, 127.8, 127.7, 126.8, 126.7, 115.1, 82.0, 69.4, 55.8, 40.1, 34.0, 33.5, 32.1, 29.3, 27.2, 24.6. **LRMS** (ESI, APCI) *m/z*: calc'd for $C_{27}H_{30}O_3$ $[M-H]^-$ 401.2, found 401.5. Calc'd for $C_{27}H_{29}O_2$ $[M-OH]^+$ 385.2, found 385.2 **HRMS** calc'd for $C_{27}H_{31}O_3$ $[M+H]^+$: 403.2268, found 403.2266.



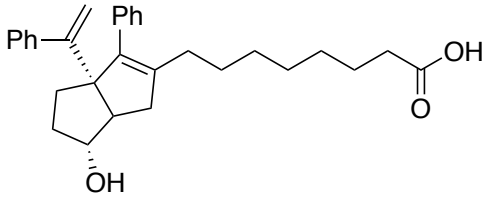
6-(6-*exo*-hydroxy-3-phenyl-3a-(1-phenylvinyl)-1,3a,4,5,6,6a-hexahydropentalen-2-yl)hexanoic acid (**2c**): **^{1}H NMR** (600 MHz, $CDCl_3$) δ 7.49 – 7.05 (m, 10H), 5.05 (d, *J* = 1.5 Hz, 1H), 4.97 (d, *J* = 1.4 Hz, 1H), 3.93 (s, 1H), 2.36 – 2.21 (m, 4H), 2.10 – 1.97 (m, 5H), 1.73 – 1.61 (m, 3H), 1.59 – 1.51 (m, 2H), 1.34 (p, *J* = 7.6 Hz, 2H), 1.28 – 1.20 (m, 2H). **^{13}C NMR** (126 MHz, $cdcl_3$) δ 179.4, 154.5, 144.1, 140.7, 139.5, 137.3, 129.6, 127.8, 127.70, 127.68, 126.7, 115.0, 82.1, 69.3, 55.7, 40.2, 34.0, 33.8, 32.0, 29.5, 29.0, 27.4, 24.5. **LRMS** (ESI, APCI) *m/z*: calc'd for $C_{28}H_{31}O_3$ $[M-H]^-$ 415.2, found 414.9. Calc'd for $C_{28}H_{31}O_2$ $[M-OH]^+$ 399.2, found 399.2 **HRMS** calc'd for $C_{28}H_{32}O_3Na$ $[M+Na]^+$: 439.2235, found 439.2237.



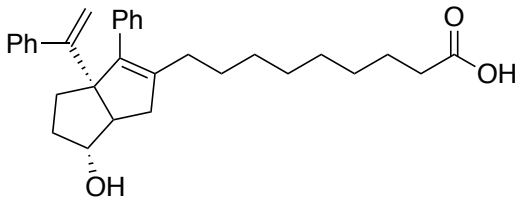
7-(6-*exo*-hydroxy-3-phenyl-3a-(1-phenylvinyl)-1,3a,4,5,6,6a-hexahydropentalen-2-yl)heptanoic acid (**2d**): **^{1}H NMR** (600 MHz, $CDCl_3$) δ 7.37 – 7.14 (m, 9H), 5.05 (d, *J* = 1.5 Hz, 1H), 4.97 (d, *J* = 1.4 Hz, 1H), 3.93 (s, 1H), 2.38 – 2.25 (m, 4H), 2.11 – 1.97 (m, 4H), 1.76 – 1.47 (m, 4H), 1.33 (p, *J* = 7.6 Hz, 2H), 1.29 – 1.14 (m, 6H). **^{13}C NMR** (126 MHz, $CDCl_3$) δ 179.9, 154.5, 144.1, 140.9, 139.3, 137.3, 129.7, 127.74, 127.70, 127.66, 126.7, 115.0, 82.1, 69.3, 55.6, 40.2, 34.0,

33.9, 32.0, 29.6, 29.2, 28.8, 27.6, 24.6. **LRMS** (ESI, APCI) *m/z*: calc'd for $C_{29}H_{33}O_3$ $[M-H]^-$ 429.3, found 429.3. Calc'd for $C_{29}H_{33}O_2$ $[M-OH]^+$: 413.2, found 413.28.

HRMS calc'd for $C_{29}H_{33}O_3$ $[M-H]^-$: 429.2435, found 429.2378.

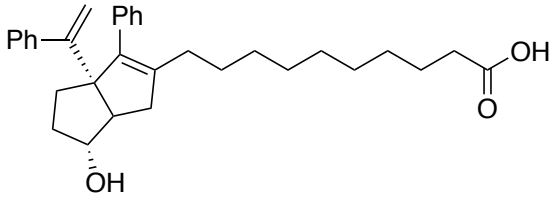


8-(6-*exo*-hydroxy-3-phenyl-3a-(1-phenylvinyl)-1,3a,4,5,6,6a-hexahydropentalen-2-yl)octanoic acid (**2e**): **^{1}H NMR** (600 MHz, $CDCl_3$) δ 7.81 – 7.13 (m, 10H), 5.05 (s, 1H), 4.97 (s, 1H), 3.93 (s, 1H), 2.37 – 2.22 (m, 5H), 2.09 – 1.94 (m, 5H), 1.79 – 1.58 (m, 6H), 1.38 – 1.16 (m, 5H). **^{13}C NMR** (126 MHz, $CDCl_3$) δ 180.0, 154.5, 144.1, 141.0, 139.2, 137.3, 129.7, 127.74, 127.68, 127.6, 126.7, 115.0, 82.1, 69.3, 55.6, 40.2, 34.0, 33.8, 32.0, 29.4, 28.94, 28.88, 27.7, 24.6, 21.0. **LRMS** (ESI, APCI) *m/z*: calc'd for $C_{30}H_{35}O_3$ $[M-H]^-$ 443.3, found 443.2. Calc'd for $C_{30}H_{35}O_2$ $[M-OH]^+$ 427.3, found 427.4. **HRMS** calc'd for $C_{30}H_{35}O_3$ $[M-H]^-$: 443.2592, found 443.2593.

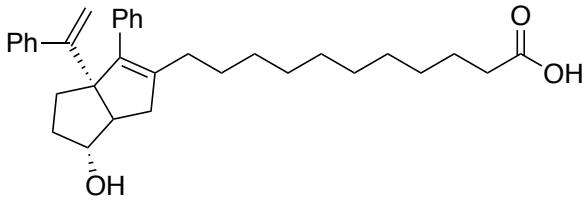


9-(6-*exo*-hydroxy-3-phenyl-3a-(1-phenylvinyl)-1,3a,4,5,6,6a-hexahydropentalen-2-yl)nonanoic acid (**2f**): **^{1}H NMR** (500 MHz, $CDCl_3$) δ 7.41 – 7.15 (m, 10H), 5.07 (d, *J* = 1.4 Hz,

1H), 4.99 (d, J = 1.4 Hz, 1H), 3.95 (s, 1H), 2.39 – 2.27 (m, 5H), 2.12 – 1.97 (m, 5H), 1.76 – 1.56 (m, 6H), 1.40 – 1.15 (m, 7H). **^{13}C NMR** (126 MHz, $CDCl_3$) δ 179.5, 154.6, 144.1, 141.1, 139.2, 137.4, 129.7, 127.74, 127.71, 127.6, 126.7, 126.6, 115.0, 82.1, 69.3, 55.7, 40.2, 34.0, 32.1, 29.6, 29.5, 29.2, 29.1, 29.0, 24.7. **LRMS** (ESI, APCI) m/z: calc'd for $C_{31}H_{37}O_3$ [M-H]$^-$ 457.3, found 457.3. Calc'd for $C_{31}H_{37}O_2$ [M-OH]$^+$ 441.3, found 440.8. **HRMS** (ESI) m/z: calc'd for $C_{31}H_{37}O_3$ [M-H]$^-$: 457.2748, found 457.2749.



10-(6-*exo*-hydroxy-3-phenyl-3a-(1-phenylvinyl)-1,3a,4,5,6,6a-hexahydropentalen-2-yl)decanoic acid (6HP-CA 2g): **^{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, 34.0, 33.9, 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-OH]$^+$: 455.3, found 454.8. **HRMS** (ESI) m/z: calc'd for $C_{32}H_{39}O_3$ [M-H]$^-$: 471.2882, found 471.2905. **FT-IR** (neat): 3360(b), 1708 (s) cm^{-1}.

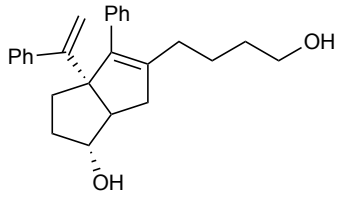


11-(6-*exo*-hydroxy-3-phenyl-3a-(1-phenylvinyl)-1,3a,4,5,6,6a-hexahydropentalen-2-yl)undecanoic acid (**2h**): **^{1}H NMR** (500 MHz, $CDCl_3$) δ 7.41 – 7.12 (m, 10H), 5.07 (d, *J* = 1.4 Hz, 1H), 4.99 (d, *J* = 1.5 Hz, 1H), 3.99 – 3.92 (s, 1H), 2.38 – 2.26 (m, 5H), 2.11 – 1.98 (m, 5H), 1.75 – 1.58 (m, 6H), 1.36 – 1.17 (m, 11H). **^{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, 40.2, 34.0, 33.9, 32.1, 29.7, 29.6, 29.4, 29.3, 29.2, 29.0, 27.8, 24.7. **LRMS** (ESI, APCI) *m/z*: calc'd for $C_{33}H_{41}O_3$ $[M-H]^-$ 485.3, found 485.3. Calc'd for $C_{33}H_{41}O_2$ $[M-OH]^+$: 469.3, found 468.9. **HRMS** (ESI) *m/z*: calc'd for $C_{33}H_{41}O_3$ $[M-H]^-$: 485.3061, found 485.3065.

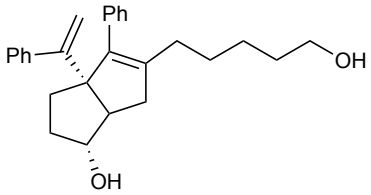
Diols: Compounds 3a–3h

General Procedure for Diol MOM Deprotection

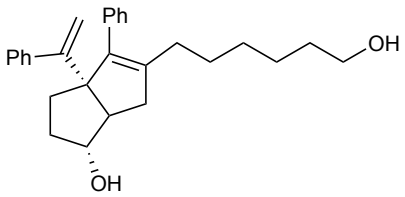
To a solution of **(S1a-h)** in MeCN was added concentrated HCl in excess (5-20 equiv.). The solution was stirred for 5 minutes or until the reaction was completed by TLC/LCMS. The resulting solution as concentrated *in vacuo* and subjected to preparatory HPLC to isolate the desired product as the major *exo* diastereomer.



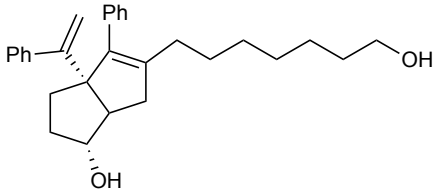
***Exo*-5-(4-hydroxybutyl)-4-phenyl-3a-(1-phenylvinyl)-1,2,3,3a,6,6a-hexahydropentalen-1-ol (3a): S1a** (18.3 mg, 0.04 mmol) was reacted and purified according to the general procedure to give the title compound as a clear, colorless oil (13.0 mg, 79% yield). **^{1}H NMR** (600 MHz, Chloroform-*d*) δ 7.33 – 7.24 (m, 6H), 7.23 (s, 2H), 7.20 – 7.16 (m, 2H), 5.05 (d, *J* = 1.4 Hz, 1H), 4.97 (d, *J* = 1.5 Hz, 1H), 3.93 (s, 1H), 3.55 (t, *J* = 6.2 Hz, 2H), 2.36 (dd, *J* = 16.9, 9.3 Hz, 1H), 2.29 (d, *J* = 6.9 Hz, 1H), 2.10 – 2.02 (m, 4H), 1.74 – 1.61 (m, 3H), 1.50 – 1.37 (m, 4H). **^{13}C NMR** (500 MHz, Chloroform-*d*) δ 154.6, 144.1, 140.8, 139.4, 137.3, 129.6, 127.7, 127.7, 127.6, 126.6, 116.6, 114.8, 81.8, 69.3, 62.4, 55.7, 40.1, 33.9, 32.7, 32.0, 29.4, 24.0, 22.5, 10.6. **HRMS** calcd for $C_{26}H_{31}O_2$ $[M+H]^+$: 375.23186, found 375.23145.



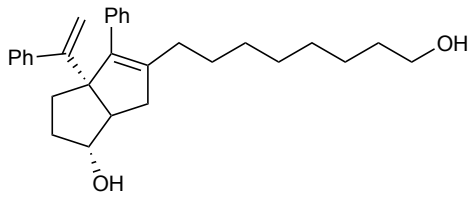
***Exo*-5-(5-hydroxypentyl)-4-phenyl-3a-(1-phenylvinyl)-1,2,3,3a,6,6a-hexahydropentalen-1-ol (3b): S1b** (21.2 mg, 0.05 mmol) was reacted and purified according to the general procedure to give the title compound as a clear, colorless oil (19.0 mg, >99% yield). **^{1}H NMR** (500 MHz, Chloroform-*d*) δ 7.43 – 7.28 (m, 7H), 7.28 – 7.19 (m, 3H), 5.09 (d, *J* = 1.4 Hz, 1H), 5.01 (d, *J* = 1.4 Hz, 1H), 3.97 (s, 1H), 3.61 (t, *J* = 6.6 Hz, 2H), 2.38 (dd, *J* = 16.8, 9.4 Hz, 1H), 2.31 (d, *J* = 9.3, 1.5 Hz, 1H), 2.16 – 2.03 (m, 4H), 1.77 – 1.65 (m, 3H), 1.56 – 1.48 (m, 3H), 1.43 – 1.25 (m, 3H). **^{13}C NMR** (500 MHz, Chloroform-*d*) δ 154.5, 144.1, 140.8, 139.3, 137.3, 131.5, 129.6, 128.2, 127.7, 127.6, 126.7, 126.6, 115.0, 82.0, 69.3, 62.9, 61.9, 55.8, 40.3, 34.0, 32.5, 32.1, 31.3, 29.7, 27.6, 25.8, 16.0. **HRMS** calc'd for $C_{27}H_{33}O_2$ $[M+H]^+$: 389.24751, found 398.24762.



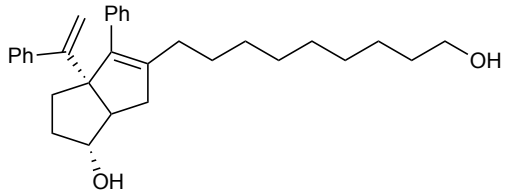
***Exo*-5-(6-hydroxyhexyl)-4-phenyl-3a-(1-phenylvinyl)-1,2,3,3a,6,6a-hexahydropentalen-1-ol (3c): S1c (**9.5 mg, 0.02 mmol) was reacted and purified according to the general procedure to give the title compound as a clear, colorless oil (6.8 mg, 79% yield). **^{1}H NMR** (600 MHz, Chloroform-*d*) δ 7.36 – 7.25 (m, 6H), 7.24 – 7.15 (m, 4H), 5.05 (s, 1H), 4.97 (s, 1H), 3.93 (s, 1H), 3.58 (t, *J* = 6.6 Hz, 2H), 2.35 (dd, *J* = 16.9, 9.4 Hz, 1H), 2.28 (d, *J* = 9.4 Hz, 1H), 2.11 – 1.97 (m, 4H), 1.75 – 1.62 (m, 3H), 1.53 – 1.46 (m, 2H), 1.38 – 1.15 (m, 6H). **^{13}C NMR** (600 MHz, Chloroform-*d*) δ 154.5, 144.1, 140.9, 139.2, 137.3, 129.7, 127.7, 127.6, 126.7, 126.6, 115.0, 110.0, 82.0, 69.3, 63.0, 55.8, 40.3, 34.0, 32.6, 32.1, 29.6, 29.4, 27.7, 25.5. **HRMS** calc'd for $C_{28}H_{35}O_2$ $[M+H]^+$: 403.26316, found 403.26338.



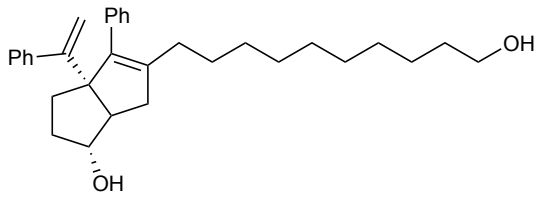
***Exo*-5-(7-hydroxyheptyl)-4-phenyl-3a-(1-phenylvinyl)-1,2,3,3a,6,6a-hexahydropentalen-1-ol (3d): S1d** (11.6 mg, 0.03 mmol) was reacted and purified according to the general procedure to give the title compound as a clear, colorless oil (8.5 mg, 81% yield). **^{1}H NMR** (600 MHz, Chloroform-*d*) δ 7.35 – 7.25 (m, 5H), 7.24 – 7.15 (m, 5H), 5.05 (d, *J* = 1.4 Hz, 1H), 4.97 (d, *J* = 1.4 Hz, 1H), 3.93 (s, 1H), 3.60 (t, *J* = 6.7 Hz, 2H), 2.35 (dd, *J* = 16.9, 9.3 Hz, 1H), 2.28 (d, *J* = 9.4 Hz, 1H), 2.09 – 1.98 (m, 5H), 1.73 – 1.61 (m, 3H), 1.36 – 1.19 (m, 10H). **^{13}C NMR** (600 MHz, Chloroform-*d*) δ 154.5, 144.1, 141.0, 139.2, 137.3, 129.7, 127.7, 127.6, 126.64, 126.59, 115.0, 82.0, 69.3, 63.0, 58.5, 55.8, 50.9, 40.2, 34.0, 32.7, 32.1, 29.63, 29.58, 29.2, 27.7, 25.6, 18.4. **HRMS** calc'd for $C_{29}H_{36}O_2Cl$ $[M+Cl]^-$: 451.24093, found 451.24179.



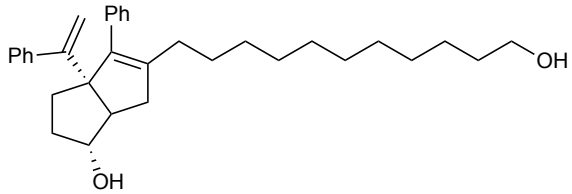
***Exo*-5-(8-hydroxyoctyl)-4-phenyl-3a-(1-phenylvinyl)-1,2,3,3a,6,6a-hexahydropentalen-1-ol (3e): S1e** (4.6 mg, 0.01 mmol) was reacted and purified according to the general procedure to give the title compound as a clear, colorless oil (2.3 mg, 55% yield) **^{1}H NMR** (600 MHz, Chloroform-*d*) δ 7.35 – 7.25 (m, 6H), 7.24 – 7.16 (m, 4H), 5.05 (d, *J* = 1.4 Hz, 1H), 4.97 (d, *J* = 1.4 Hz, 1H), 3.93 (s, 1H), 3.61 (t, *J* = 6.7 Hz, 2H), 2.35 (dd, *J* = 16.9, 9.3 Hz, 1H), 2.28 (d, *J* = 9.2 Hz, 1H), 2.10 – 2.01 (m, 2H), 2.00 – 1.97 (m, 2H), 1.74 – 1.61 (m, 3H), 1.35 – 1.16 (m, 13H). **HRMS** calc'd for $C_{30}H_{38}O_2Cl$ $[M+Cl]^-$: 465.25658, found 465.25703.



***Exo*-5-(9-hydroxynonyl)-4-phenyl-3a-(1-phenylvinyl)-1,2,3,3a,6,6a-hexahydropentalen-1-ol (3f): S1f** (9.4 mg, 0.02 mmol) was reacted and purified according to the general procedure to give the title compound as a clear, colorless oil (8.2 mg, 96% yield.) **^{1}H NMR** (600 MHz, Chloroform-*d*) δ 7.39 – 7.26 (m, 5H), 7.24 – 7.14 (m, 5H), 5.05 (d, *J* = 1.4 Hz, 1H), 4.97 (d, *J* = 1.4 Hz, 1H), 3.93 (s, 1H), 3.62 (t, *J* = 9.4 Hz, 2H), 2.34 (dd, *J* = 16.7, 9.4 Hz, 1H), 2.27 (d, *J* = 9.3 Hz, 1H), 2.12 – 2.03 (m, 2H), 2.01 – 1.96 (m, 1H), 1.74 – 1.58 (m, 3H), 1.34 – 1.17 (m, 18H). **^{13}C NMR** (600 MHz, Chloroform-*d*) δ 154.6, 144.2, 141.1, 139.1, 137.4, 129.7, 127.71, 127.70, 127.59, 126.64, 126.57, 115.0, 82.1, 69.3, 63.1, 55.8, 40.2, 34.0, 32.8, 32.1, 29.7, 29.6, 29.5, 29.4, 29.3, 27.8, 25.7. **HRMS** calc'd for $C_{31}H_{40}O_2Cl$ $[M+Cl]^-$: 479.27223, found 479.27260.



***Exo*-5-(10-hydroxydecyl)-4-phenyl-3a-(1-phenylvinyl)-1,2,3,3a,6,6a-hexahydropentalen-1-ol (3g): S1g** (81.4 mg, 0.16 mmol) was reacted and purified according to the general procedure to give the title compound as a clear, colorless oil (62.0 mg, 83% yield). **^{1}H NMR** (600 MHz, Chloroform-*d*) δ 7.35 – 7.25 (m, 5H), 7.24 – 7.16 (m, 5H), 5.05 (d, *J* = 1.4 Hz, 1H), 4.97 (d, *J* = 1.4 Hz, 1H), 3.93 (s, 1H), 3.61 (t, *J* = 6.7 Hz, 2H), 2.34 (dd, *J* = 16.9, 9.4 Hz, 1H), 2.27 (d, *J* = 9.2 Hz, 1H), 2.11 – 2.00 (m, 3H), 1.73 – 1.62 (m, 3H), 1.54 (dq, *J* = 8.2, 6.7 Hz, 2H), 1.34 – 1.17 (m, 17H). **LRMS [APCI]** calc'd for $C_{32}H_{41}O_2$ $[M-H]^-$: 457.3, found 457.2.



***Exo*-5-(11-hydroxyundecyl)-4-phenyl-3a-(1-phenylvinyl)-1,2,3,3a,6,6a-hexahydropentalen-1-ol (3h): S1h** (7.7 mg, 0.015 mmol) was reacted and purified according to the general procedure to give the title compound as a clear, colorless oil (5.4 mg, 78% yield). **^{1}H NMR** (600 MHz, Chloroform-*d*) δ 7.34 – 7.25 (m, 4H), 7.24 – 7.16 (m, 5H), 5.05 (d, *J* = 1.4 Hz, 1H), 4.96 (d, *J* = 1.3 Hz, 1H), 3.93 (s, 1H), 3.61 (t, *J* = 6.7 Hz, 2H), 2.33 (dd, *J* = 17.2, 9.8 Hz, 1H), 2.26 (d, *J* = 9.3 Hz, 1H), 2.09 – 1.99 (m, 1H), 2.02 – 1.99 (m, 5H), 1.72 – 1.61 (m, 3H), 1.57 – 1.51 (m, 2H), 1.40 – 1.11 (m, 10H). **LRMS [APCI]** calc'd for $C_{33}H_{43}O_2$ $[M-H]^-$: 471.3, found 471.0.

Table 1. List of primers for qRT-PCR with enteroids.

Gene	Forward Primer (5' → 3')	Reverse Primer (5' → 3')
m/hLRH-1	GTGTCTCAAT TTAAAATGGT GAATTACTCC TATGATGAAG	AAT AAGTTTGGGC CAATGTACAA GAGAGACAGG
Cyp11a1	GCTGGAAGGTGTAGCTCAGG	CACTGGTGTGGAACATCTGG
Cyp11b1	primers purchased from QuantiTect, Qiagen (NM_001033229, catalog # QT01198575)	
IL-10	GCCTTATCGGAAATGATCCAGT	GCTCCACTGCCTTGCTCTTATT
IL-1β	primers purchased from QuantiTect, Qiagen (NM_008361, catalog # QT01048355)	
TNFα	CCAGAAAAGACACCATGAGCAC	GGGCCATAGAACTGATGAGAGG
Rplp0(36B4)	GAAACTGCTGCCTCACATCCG	GCTGGCACAGTGACCTCACAC

Dextran sulfate sodium (DSS) induced colitis model: hLRH-1 mice (8 weeks old) were administered 2.5% DSS w/v in drinking water for 5 days, followed by regular water for 7 days. The 10CA (1mg/kg) or vehicle (0.9% normal saline) were administered by daily intraperitoneal injection beginning on D6 for seven days, during which body weights were monitored daily. At the end of treatment, disease activity scores were assessed according to Table S3. For colon histological analysis, the colon was divided into three segments (proximal third, middle third, and distal third). Each segment was embedded in paraffin, sectioned at 5 μm, and stained with hematoxylin and eosin. Histological analysis was performed in the Cellular and Molecular Morphology Core of the Digestive Disease Center at Baylor College of Medicine. The sections were blindly scored using a standard histologic colitis score. Three independent parameters were measured: severity of inflammation (0–3: none, slight, moderate, severe), depth of injury (0–3: none, mucosal, mucosal and submucosal, transmural),

and crypt damage (0–4: none, basal one-third damaged, basal two-thirds damaged, only surface epithelium intact, entire crypt and epithelium lost). The score of each parameter was multiplied by a factor reflecting the percentage of tissue involvement (×1, 0–25%; ×2, 26–50%; ×3, 51–75%; ×4, 76–100%) averaged per colon.

Acknowledgements

The authors thank the HDX-MS core in School of Medicine, Emory University for their technical assistance in data collection and analysis. This study was supported in part by the Emory Integrated Genomics Core (EIGC), which is subsidized by the Emory University School of Medicine and is one of the Emory Integrated Core Facilities. The authors are grateful to the beamline staff at Argonne National Laboratory, South East Regional Collaborative Team, for support during remote collection of crystal diffraction data. This work was supported in part by the National Institutes of Health under the following awards: T32GM008602 (S.G.M.), F31DK111171 (S.G.M.), T32GM008367-27 (E.H.D.), R01DK095750 (E.A.O.), R01DK114213 (E.A.O., N.T.J., J.W.C). The work was also supported by an Emory Catalyst Award (E.A.O., N.T.J), and USDA ARS 3092-5-001-057 (D.D.M.). E.H.D was supported by the National Science Foundation Graduate Research Fellowship.

Competing Interests Statement

The authors declare no competing financial interests.

Figures

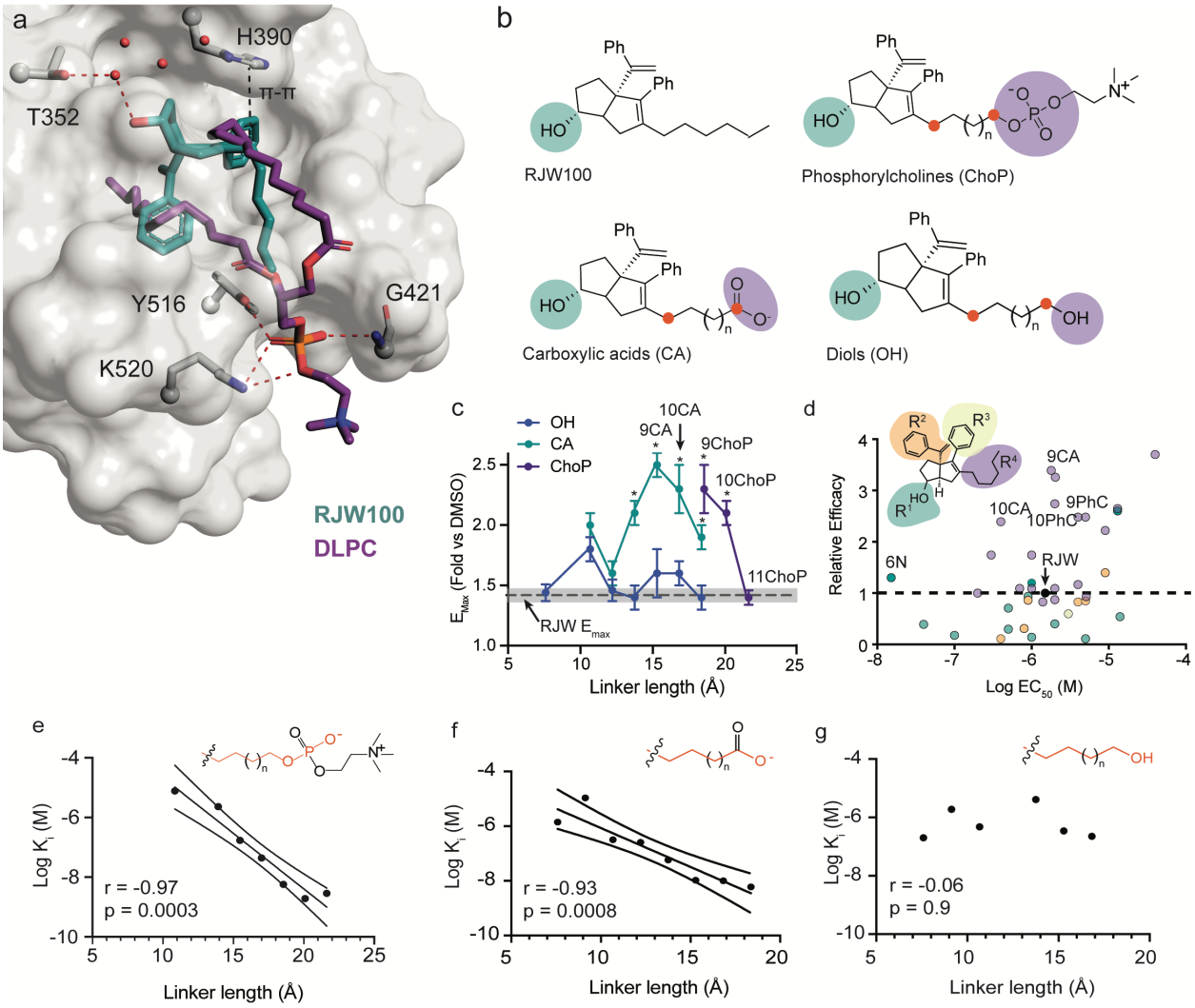


Figure 3.1. Phospholipid mimetic design, binding affinity, and activity. (A) LRH-1 ligand binding pocket (LBP) from PDB 5L11, showing the binding modes of synthetic agonist RJW100 (teal sticks), and phospholipid agonist DLPC (PDB 4DOS, purple sticks). Key interactions made by each agonist are highlighted. (B) Agonist design strategy. Phosphorylcholines or carboxylic isosteres were conjugated to the RJW100 core via alkyl linkers of 4-11 carbons.[27] Diols with alkyl linkers of 4-11 carbons but lacking a charged group at the R^4 terminus were synthesized for comparison with PL mimics. (C) Plot of E_{max} values from luciferase reporter assays as a function of linker length. Each point represents the mean ± SEM from 2-3 experiments. Compounds that did not activate sufficiently for

E_{max} calculation are omitted. *Dotted line* is the mean E_{max} of RJW100; grey shading indicates SEM. *, $p < 0.05$ *versus* RJW100 by two-way ANOVA followed by Sidak's multiple comparisons test. (D) Activity profiles of the PL-mimics *versus* closely-related LRH-1 agonists. *Inset,* RJW100 chemical structure indicating the modification sites of compounds depicted in the plot. Relative efficacy is E_{max} relative to RJW100, calculated as described in Methods. (G-I) Binding affinity (K_i) plotted as a function of linker length. *Lower insets*, Pearson correlation coefficients (r) and p-values. *Upper insets*, illustration of the part of R^4 measured to calculate linker length (orange) .

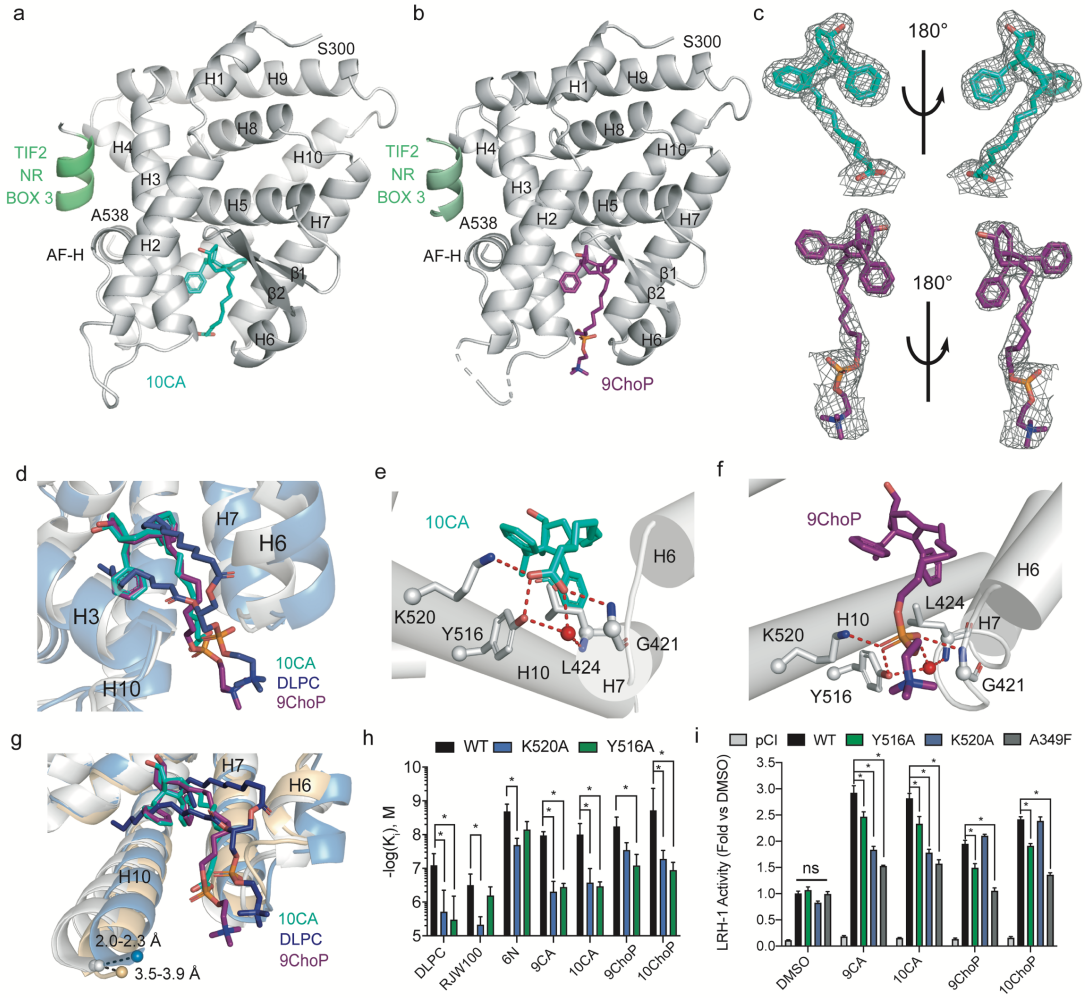


Figure 3.2. Phospholipid mimetics make PL-like interactions with LRH-1 ligand-binding pocket. Crystal structures of (*A*) 10CA (PDB: 7JYD) and (*B*) 9ChoP (PDB: 7JYE) bound to LRH-1 LBD and a

fragment of the Tif2 coregulator. (C) Omit maps showing electron density surrounding 10CA (*top*) and 9ChoP (*bottom*). Maps are F_o-F_c, contoured at 2σ. (D) Superposition of three ligands from LRH-1 structures: 10CA (cyan sticks), 9ChoP (purple sticks), or DLPC (dark blue sticks, from PBD 4DOS[31]). 10CA and DLPC protein backbones are shown as grey and light blue cartoons, respectively. (E-F) Close-up of the interactions made by the terminal polar groups of 10CA (*E*) and 9ChoP (*F)* at the LRH-1 binding pocket mouth. Hydrogen bonds are red dotted lines; water molecule is a red sphere. (G) Superposition of LRH-1 structures in the apo state (PBD 4PLD, pale orange), LRH-1-DLPC (blue), LRH-1-10CA (grey), and LRH-1-9ChoP (grey). DLPC displaces helix 6 relative to apo-LRH-1, whereas 10CA and 9ChoP displace helix 10. (H) Effects of K520A or Y516A mutations on binding affinity. Each bar represents the mean ± 95% CI from 2 experiments in quadruplicate. *, $p < 0.05$ by two-way ANOVA followed by Dunnett's test for multiple comparisons. (I) Luciferase reporter assays assess PL-mimetics ability to activate LRH-1 with K520A or Y516A mutations. The pocket-occluding A349F mutation was a negative control. Cells were treated with 30 μM of each compound for 24 hours prior to measurement of luciferase signal. *, $p < 0.05$ *versus* WT LRH-1 treated with each agonist; (two-way ANOVA followed by Dunnett's test for multiple comparisons).

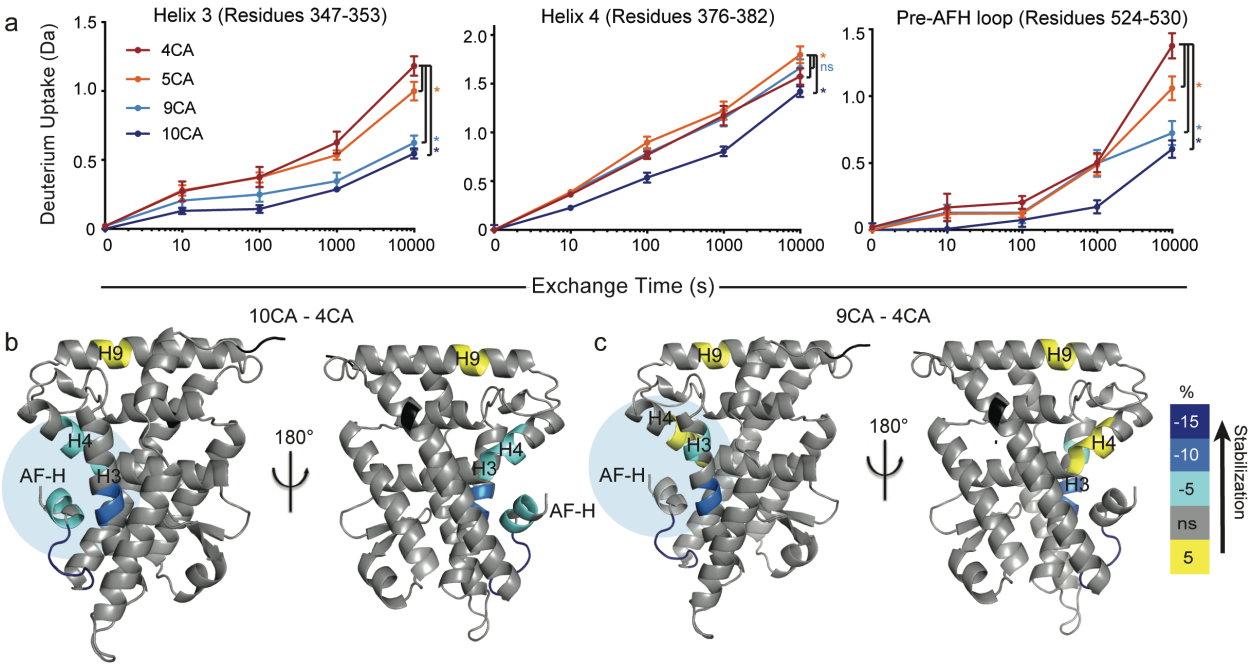


Figure 3.3. Longer-tailed PL mimetics stabilize the AFS in HDX-MS. (A) Deuterium uptake in regions within the activation function surface (AFS) of the LRH-1 LBD is more rapid and occurs to a greater extent for 4CA and 5CA compared to 9CA and 10CA. Each point represents the mean +/- SD for three replicates. Reduced deuterium exchange indicates greater stability. *, $p < 0.05$ (B-C) Differences in deuterium uptake (mean of all timepoints) are mapped onto the LRH-1 LBD for 10CA (*C*) and 9CA (*D*) *versus* 4C. The blue shaded circle indicates the location of the AFS. The scale bar indicates the colors associated with greater stabilization by 10CA or 9CA (teal and blue colors) or relative destabilization (yellow). Black indicates regions that were not mapped.

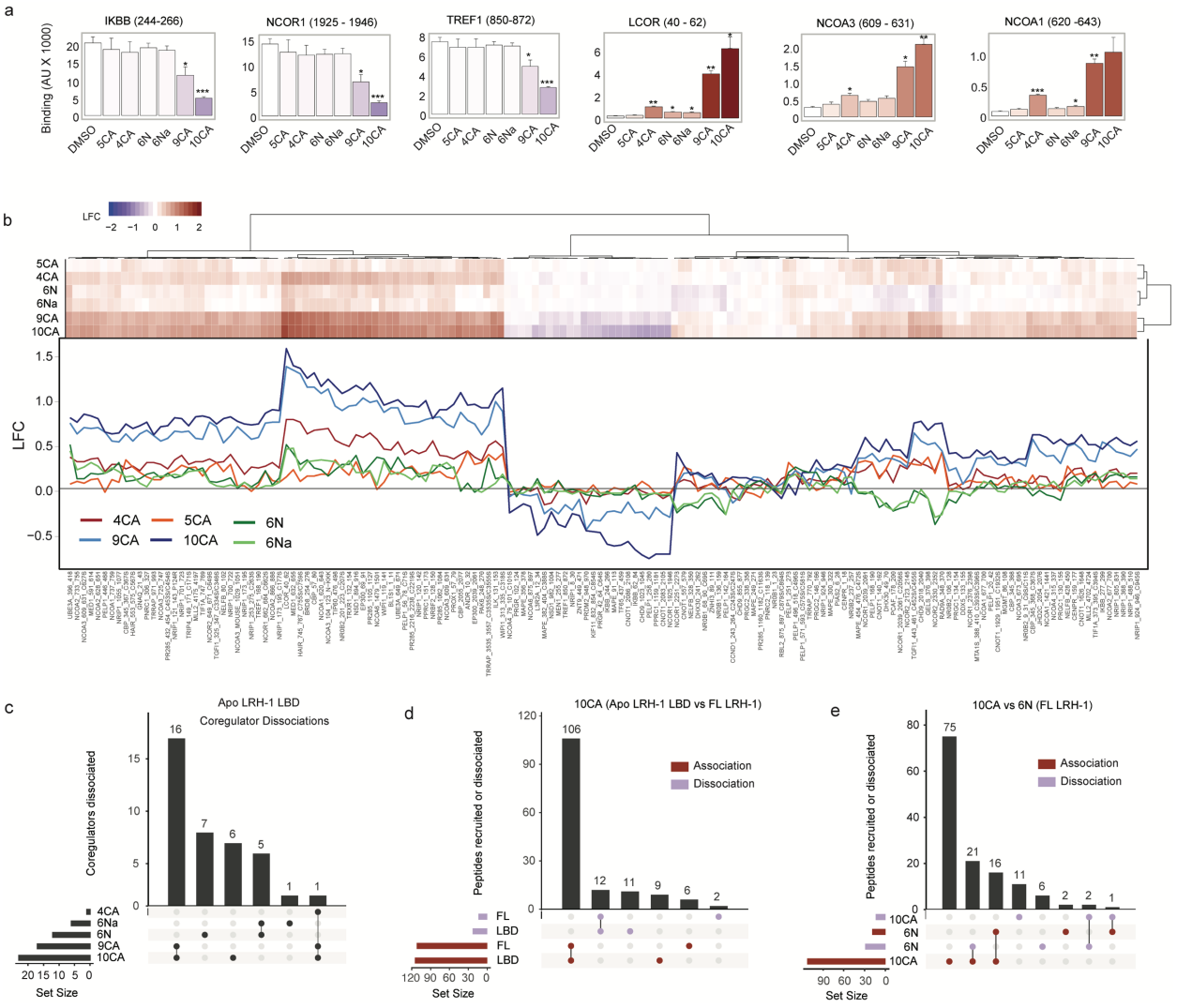


Figure 3.4. Coregulator recruitment profiling by MARCoNI. (A) Representative bar plots from the experiment, showing some of the most strongly affected coregulators by 9CA and 10CA. Each bar represents mean ± SEM from three replicates. Graph titles indicate the coregulator names, with the residue numbers of the coregulator peptide in parentheses. Statistically significant changes relative to DMSO control were identified by Student's t-test, post-hoc FDR, *, $p<0.05$, **, $p<0.01$ or ***, $p<0.001$. (B) Heat map and corresponding raw traces showing log fold change (LFC) in coregulator peptide binding relative to apo LRH-1-LBD. (C-E). Upset plots highlighting overlaps in coregulator binding events affected by at least 1.5-fold by various agonists or between experiments. Horizontal bars represent the number of events in each set, and vertical bars indicate numbers of

events that overlap in between the groups that have filled-in circles below the vertical bars. For example, in panel *(C),* the first vertical bar shows that a set of 16 coregulator peptides dissociate from LRH-1 in the presence of both 9CA and 10CA that are not affected by 6N and 6Na. (D) Upset plot comparing overlap in coregulator recruitment events by 10CA bound to full-length LRH-1 (FL) or LRH-1 ligand binding domain (LBD). Red bars indicate coregulators that associated to the 10CA-LRH-1 complex and purple bars indicate coregulator dissociation. (E) Upset plot highlighting an opposite recruitment pattern by compounds 10CA and 6N bound to FL-LRH-1.

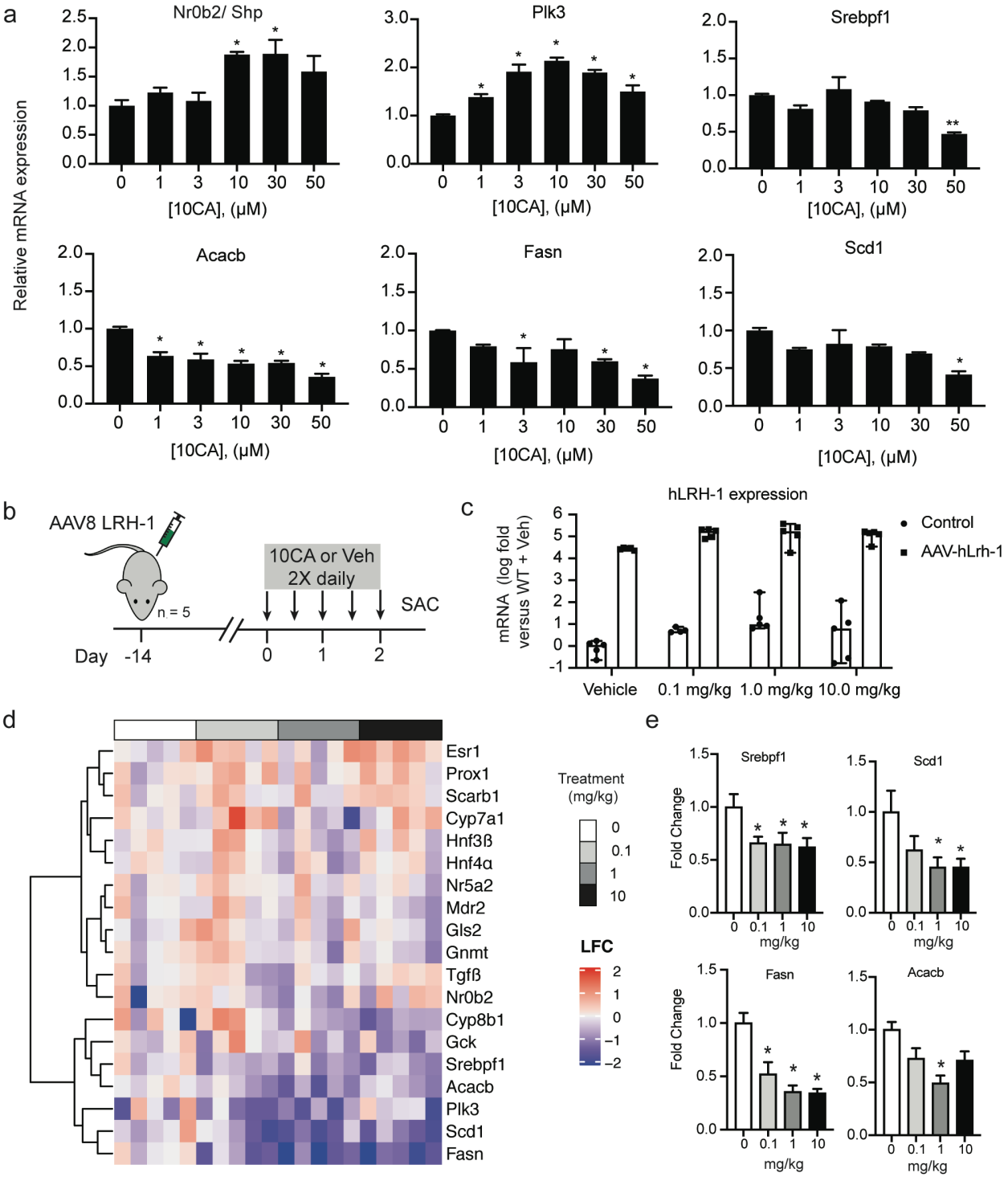

a
Relative mRNA expression
Nr0b2/ Shp
Plk3
Srebpf1
Acacb
Fasn
Scd1
[10CA], (μM)
0 1 3 10 30 50
b
AAV8 LRH-1
n = 5
10CA or Veh
2X daily
SAC
Day -14 0 1 2
c
hLRH-1 expression
mRNA (log fold versus WT + Veh)
Control
AAV-hLrh-1
Vehicle
0.1 mg/kg
1.0 mg/kg
10.0 mg/kg
d
Esr1
Prox1
Scarb1
Cyp7a1
Hnf3ß
Hnf4α
Nr5a2
Mdr2
Gls2
Gnmt
Tgfß
Nr0b2
Cyp8b1
Gck
Srebpf1
Acacb
Plk3
Scd1
Fasn
Treatment (mg/kg)
0
0.1
1
10
LFC
2
1
0
−1
−2
e
Srebpf1
Scd1
Fasn
Acacb
Fold Change
mg/kg

Figure 3.5. 10CA activates LRH-1 in hepatocytes and in the liver. (A) Treatment with 10CA dose-dependently upregulates expression of *SHP* and *PLK3* and downregulates LRH-1 targets involved in *de novo* lipogenesis in Huh7 hepatocytes (*SREBPF1*, *ACACB*, *FASN*, and *SCD1*). Cells were treated for 24 hours with doses of 10CA indicated in the x-axes. (B) Schematic of the experimental design for the mouse study. Mice were injected with AAV8-hLRH-1 intravenously to induce human LRH-1 expression, then after two weeks were treated with five doses of 10CA over three days. (C) Liver tissue from mice injected with hLRH-1-AAV8 was isolated and analyzed by qRT-PCR to measure levels of human LRH-1. Data are quantified relative to TBP. (D) Heatmap of Nanostring results in mice expressing hLRH-1 and treated with vehicle, 0.1, 1.0, or 10.0 mg/kg 10CA. (E) Individual bar graphs from Nanostring showing downregulation of lipogenic genes. *, $p < 0.05$.

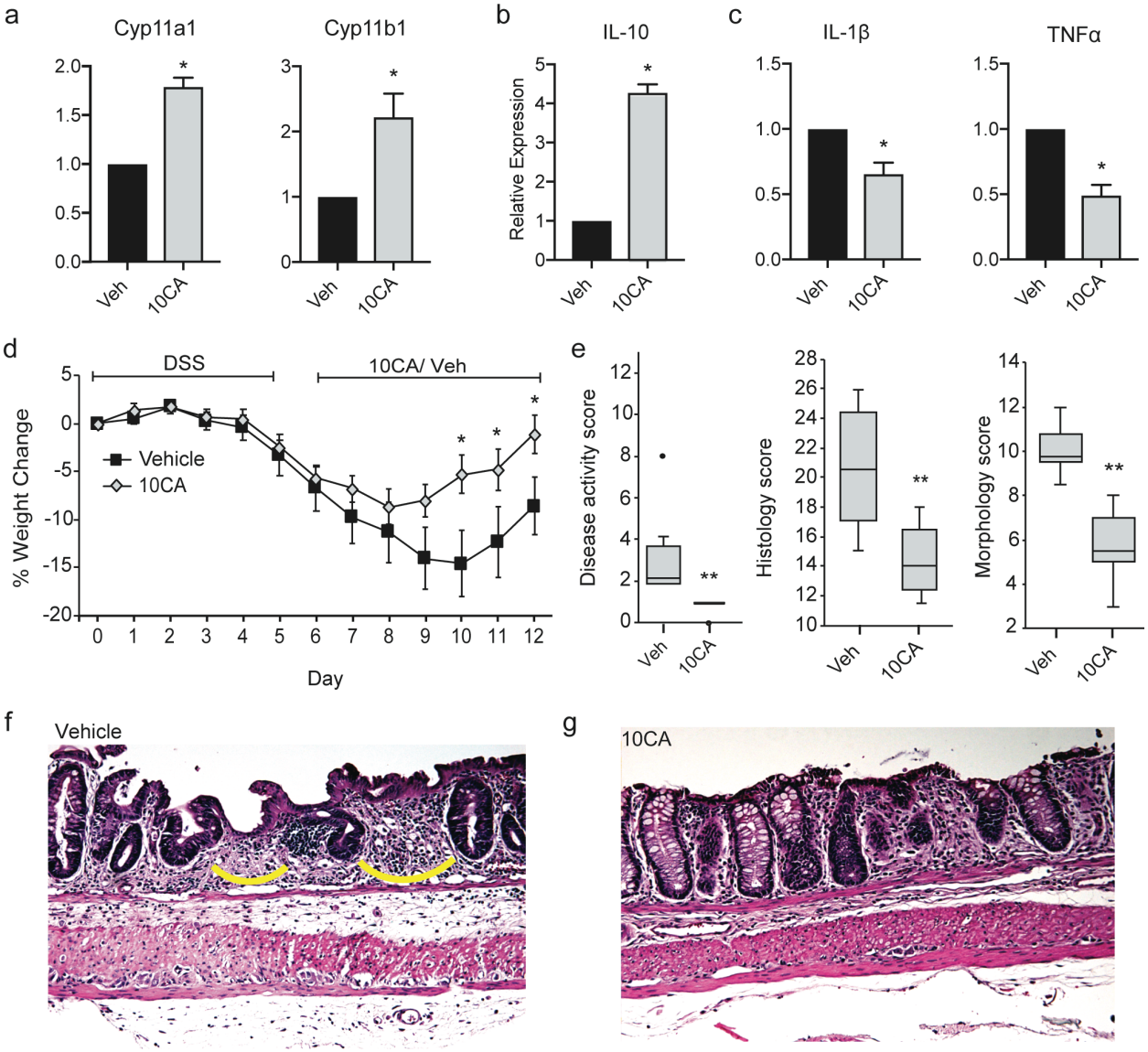


Figure 3.6. Efficacy of 10CA in organoid and in vivo models of colitis. (A-C) qRT-PCR measuring gene expression changes in mouse enteroids (n = 3) expressing human LRH-1 and treated with either DMSO (Veh) or 10CA at 1 μM for 24 hours. 10CA induces expression of steroidogenic genes *Cyp11a1* and *Cyp11b1 (A),* increases expression of the anti-inflammatory cytokine IL-10 (*B*) and suppresses expression of the pro-inflammatory cytokines IL-1b and TNFα (*C*). (D) Plot of weight changes of mice in the DSS experiment. Each point represents the mean ± SEM from 5 mice per group. Horizontal bars above the figure indicate periods of daily DSS or 10CA or Vehicle (Veh)

administration. (E) Scores for disease activity scores, histology, and morphology were calculated as described in the methods section and Table S3. Boxes indicate the median and 1st and 3rd quartiles and whiskers indicate the range . (F-G) Representative colon sections stained with H&E show reduced intestinal damage with 10CA treatment (*G*) compared to vehicle (*F*). Regions highlighted in yellow in panel (*F*) show examples of crypt loss.*, $p < 0.05$, **, $p < 0.01$ relative to vehicle-treated organoids or mice.

Authors will release the atomic coordinates and experimental data upon article publication. PDB IDs have been provided in figure legends and are as follows: LRH-1-10CA, 7JYD; LRH-1-9ChoP, 7JYE.

Table S3.1. Summary of linker lengths and key biological parameters for LRH-1 agonists.

Phosphorylcholines

Linker length (# Carbons)	Linker length (Å)	log K_i (M) Mean, [95% CI]	EC_{50} (μM)* Mean +/- SEM	E_{max}* (Mean Fold vs DMSO +/- SEM)
4	10.9	cnc	cnc	cnc
5	12.4	cnc	cnc	cnc
6	13.9	-5.6 [-6.0, -5.2]	cnc	cnc
7	15.5	-6.8 [-7.1, -6.4]	cnc	cnc
8	17.0	-7.4 [-7.7, -7.0]	>30	cnc
9	18.6	-8.2 [-8.5, -7.9]	7 +/- 2	2.3 +/- 0.2
10	20.1	-8.7 [-9.4, -8.1]	5 +/- 2	2.1 +/-0.1
11	21.6	-8.5 [-8.8, -8.3]	5.1 +/- 0.5	1.4 +/- 0.06

Carboxylic acids

Linker length (# Carbons)	Linker length (Å)	log K_i (M) Mean, [95% CI]	EC_{50} (μM)* Mean +/- SEM	E_{max}* (Mean Fold vs DMSO +/- SEM)
4	7.6	-5.8 [-6.0, -5.7]	cnc	cnc
5	9.1	-5.0 [-5.3, -4.4]	cnc	cnc
6	10.7	-6.5 [-6.8, -6.2]	9 +/- 4	2 +/- 0.1
7	12.2	-6.6 [-6.8, -6.4]	4 +/- 3	1.6 +/- 0.1

8	13.8	-7.2 [-7.7, -6.7]	4 +/- 3	2.1 +/- 0.1
9	15.3	-8.0 [-8.1, -7.9]	1.8 +/- 0.7	2.5 +/- 0.1
10	16.8	-8.0 [-8.3, -7.7]	0.4 +/- 0.2	2.3 +/- 0.2
11	18.4	-8.2 [-8.5, -8.0]	0.3 +/- 0.2	1.9 +/- 0.1

Diols

Linker length (# Carbons)	Linker length (Å)	log K_i (M) Mean, [95% CI]	EC_{50} (μM) Mean +/- SEM	E_{max} (Mean Fold vs DMSO +/- SEM)
4	7.6	-6.7 [-7.4, -6.0]	0.4 +/-0.5	1.44 +/- 0.07
5	9.1	-5.7 [-6.0, -5.5]	cnc	cnc
6	10.7	-6.3 [-6.6, -6.1]	1.0 +/- 0.8	1.8 +/- 0.1
7	12.2	cnc	0.2 +/- 0.3	1.46 +/-0.09
8	13.8	-5.4 [-5.6, -5.2]	0.7 +/- 1	1.4 +/- 0.1
9	15.3	-6.5 [-6.8, -6.1]	1 +/- 2	1.6 +/-0.2
10	16.8	-6.6 [-7.1, -6.2]	0.1+/-0.2	1.6 +/- 0.1
11	18.4	cnc	1 +/-1	1.4 +/- 0.1

* Values previously reported (Flynn *et al*, 2018)

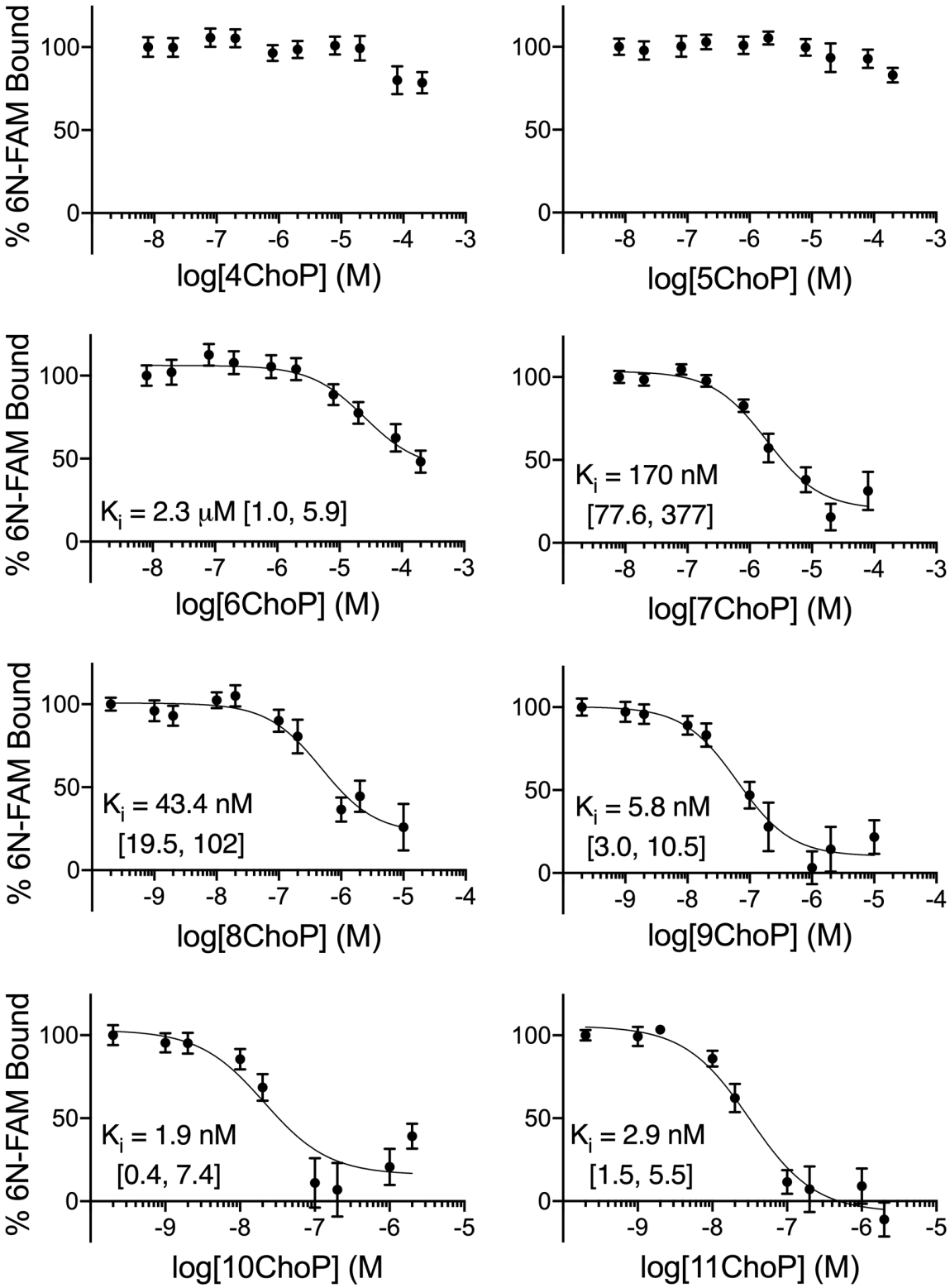

% 6N-FAM Bound
log[4ChoP] (M)
log[5ChoP] (M)
K_i = 2.3 μM [1.0, 5.9]
log[6ChoP] (M)
K_i = 170 nM
[77.6, 377]
log[7ChoP] (M)
K_i = 43.4 nM
[19.5, 102]
log[8ChoP] (M)
K_i = 5.8 nM
[3.0, 10.5]
log[9ChoP] (M)
K_i = 1.9 nM
[0.4, 7.4]
log[10ChoP] (M
K_i = 2.9 nM
[1.5, 5.5]
log[11ChoP] (M)

Figure S3.1. Fluorescence polarization competition: phosphorylcholines. Curves were first baseline subtracted with the lowest concentration of competitor agonist as baseline, then normalized to the unlabeled 6N curve to determine percent 6N-FAM bound as described in the methods. Insets are 95% confidence intervals; error shown is SEM. Assays were conducted in duplicate, each with four technical replicates.

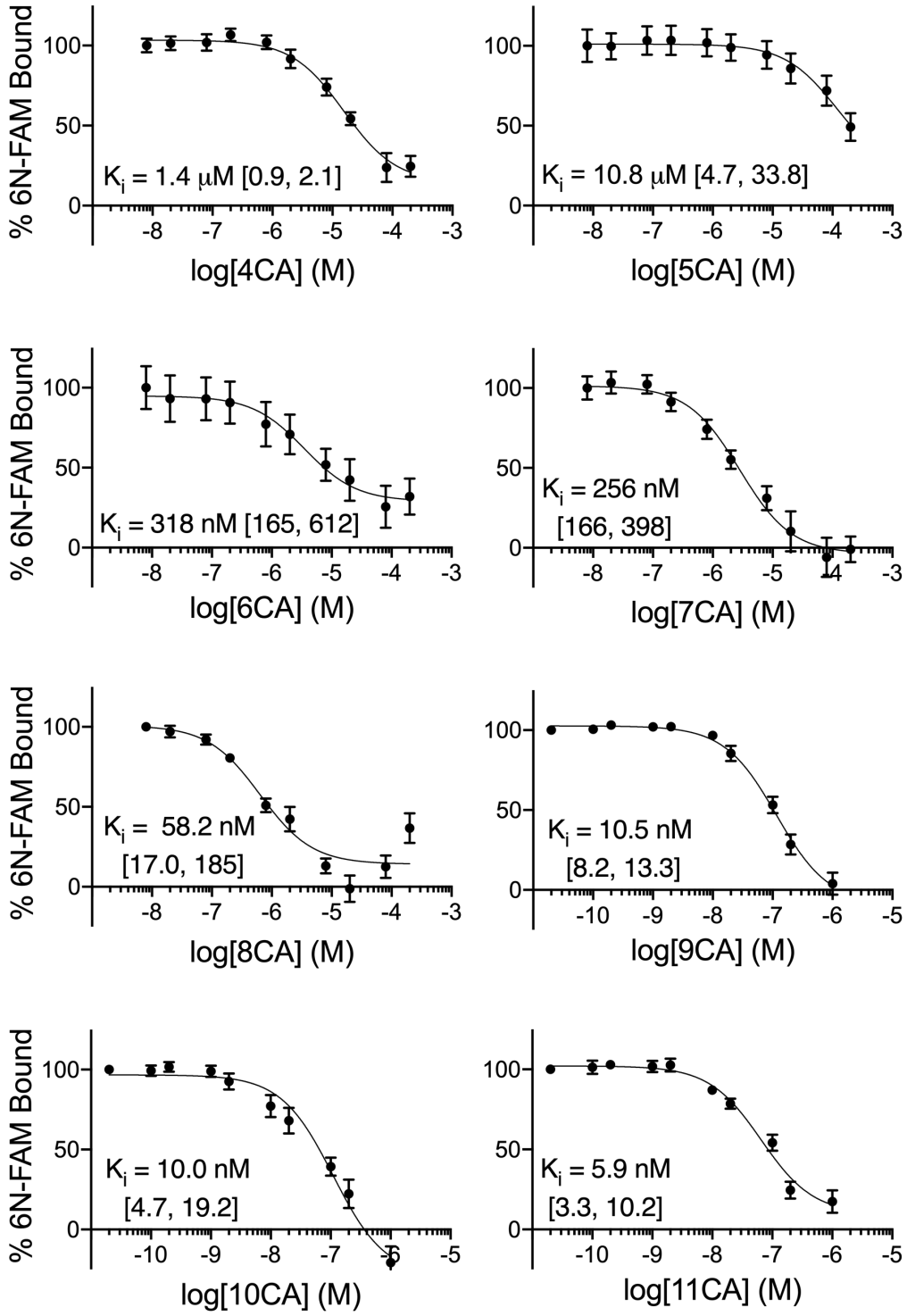


Figure S3.2. Fluorescence polarization competition: carboxylic acids. Curves were first baseline subtracted with the lowest concentration of competitor agonist as baseline, then normalized to the unlabeled 6N curve to determine percent 6N-FAM bound as described in the methods. Insets are 95% confidence

intervals; error shown is SEM. Assays were conducted in duplicate, each with four technical replicates.

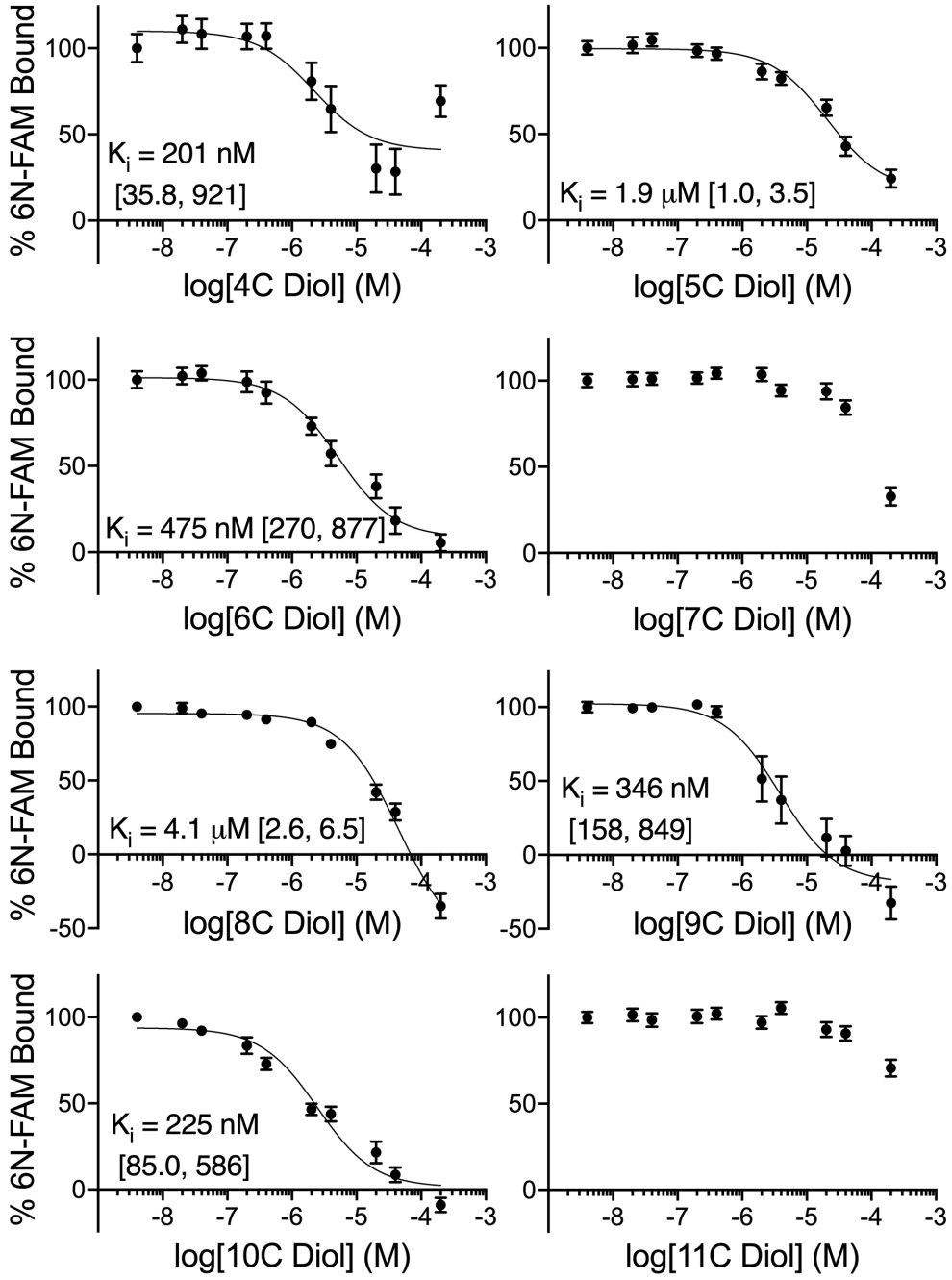


Figure S3.3. Fluorescence polarization competition: diols. Curves were first baseline subtracted with the lowest concentration of competitor agonist as baseline, then normalized to the unlabeled 6N curve to determine percent 6N-FAM bound as described in the methods. Insets are 95% confidence

intervals; error shown is SEM. Assays were conducted in duplicate, each with four technical replicates.

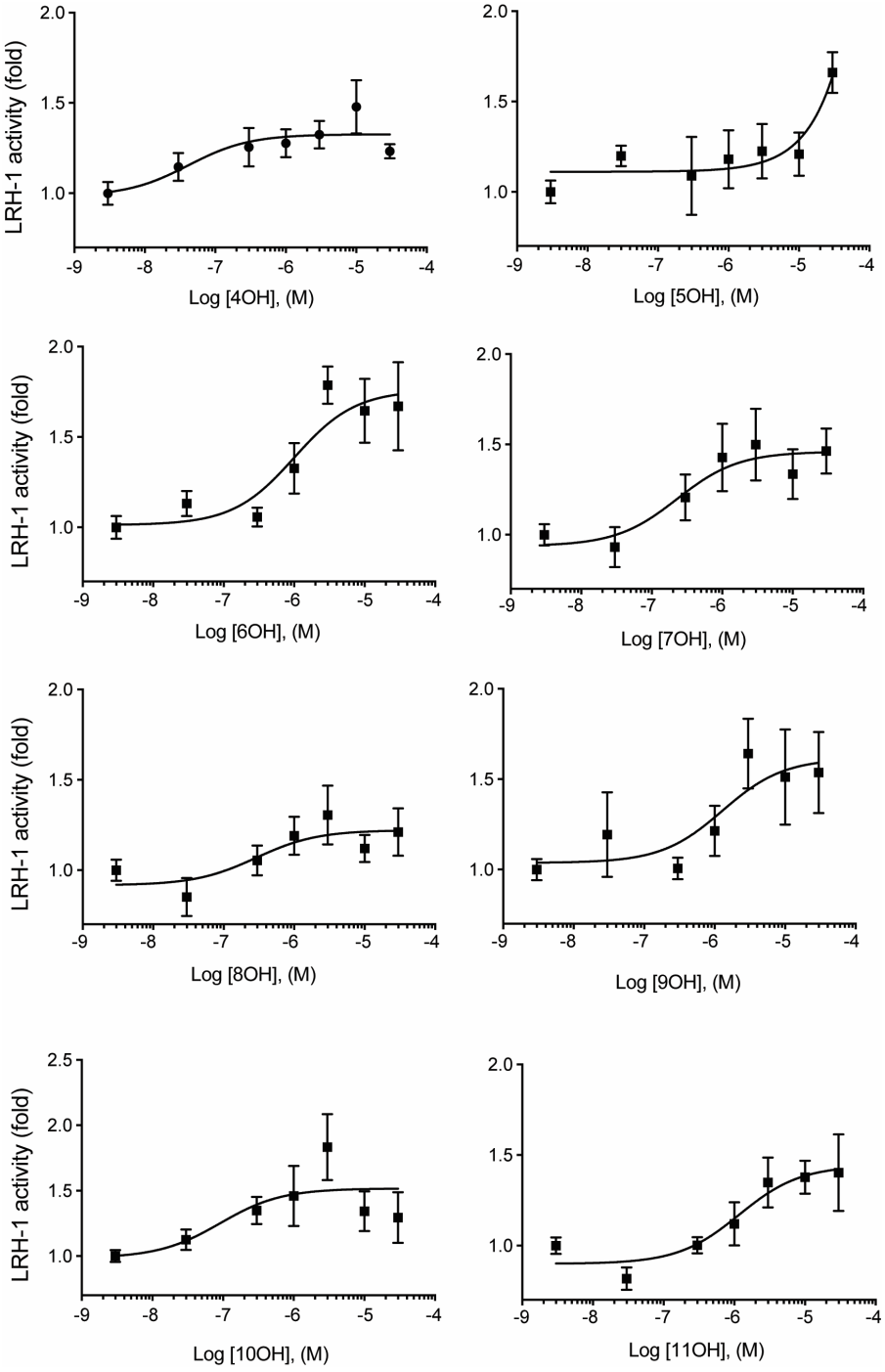


Figure S3.4. Luciferase reporter assays with diols. Hela cells were treated with the indicated concentration of each compound. Each point represents the mean ± SEM from two experiments conducted in triplicate.

Table S3.2: X-ray data collection and refinement statistics.

Data collection	LRH-1 + 10CA + TIF2	LRH-1 + 9ChoP + TIF2
Space group	$P3_221$	$P3_221$
Cell dimensions		
a, *b*, *c* (Å)	88.4, 88.4, 105.7	89.2, 89.2, 106.5
α, β, γ (□)	90, 90, 120	90, 90, 90
Resolution (Å)	38.27 - 2.30 (2.38 - 2.30)	34.21 – 2.55 (2.64-2.55)
R_{pim}	0.045 (0.305)	0.074 (0.542)
$I / \sigma I$	22.7 (1.79)	13.8 (1.31)
$CC_{1/2}$	97.0 (77.6)	98.7 (57.1)
Completeness (%)	99.8 (98.7)	99.9 (100.0)
Redundancy	15.8 (15.3)	13.8 (13.8)
Refinement		
Resolution (Å)	2.30	2.55
No. reflections	21639	16426
R_{work} / R_{free} (%)	21.3/23.3	17.6 / 20.1
No. atoms		
Protein	2037	2061
Water	38	18
Twin Law	n/a	-h, -k, l
B-factors		
Protein	66.3	63.7
Ligand	52.6	65.0
Water	60.2	49.7
R.m.s. deviations		
Bond lengths (Å)	0.002	0.001
Bond angles (°)	0.48	0.33
Ramachandran favored (%)	97.2	97.6
Ramachandran outliers (%)	0	0
PDB accession code	7JYD	7JYE

Values in parentheses indicate highest resolution shell.

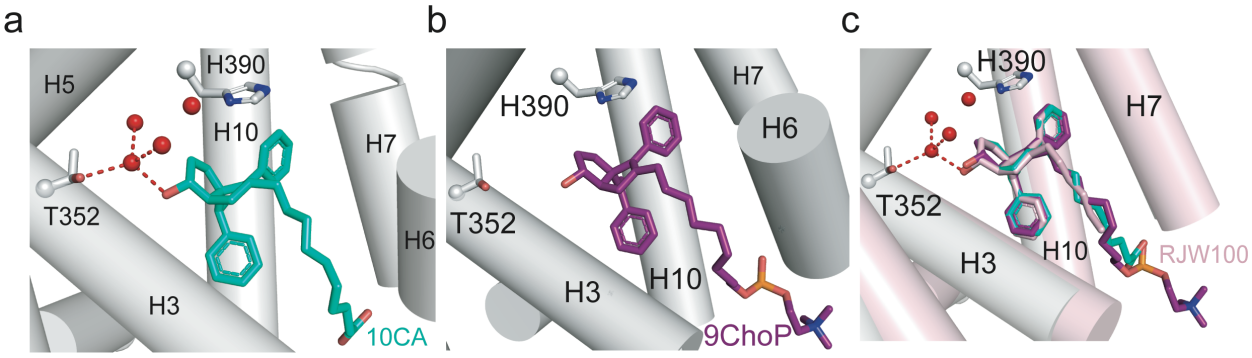


Figure S3.5. 6HP cores of 10CA and 9ChoP adopt the same position as RJW100 6HP core. (A) 10CA (PDB: 7JYD) maintains the deep pocket water network with T352 seen with RJW100 (PDB: 5L11). (B) 9ChoP (PDB: 7JYE) maintains the core position of the parent compound RJW100. Although there was not electron density to support the expected water network, this was not surprising given the high B-factors in the structure. (C) Superposition of 10CA (teal), 9ChoP (purple), and RJW100 (pink) ligands shows all three cores overlay in a nearly identical position. 10CA and 9ChoP backbone shown as 10CA alone in grey for simplicity; RJW100 shown as pink.

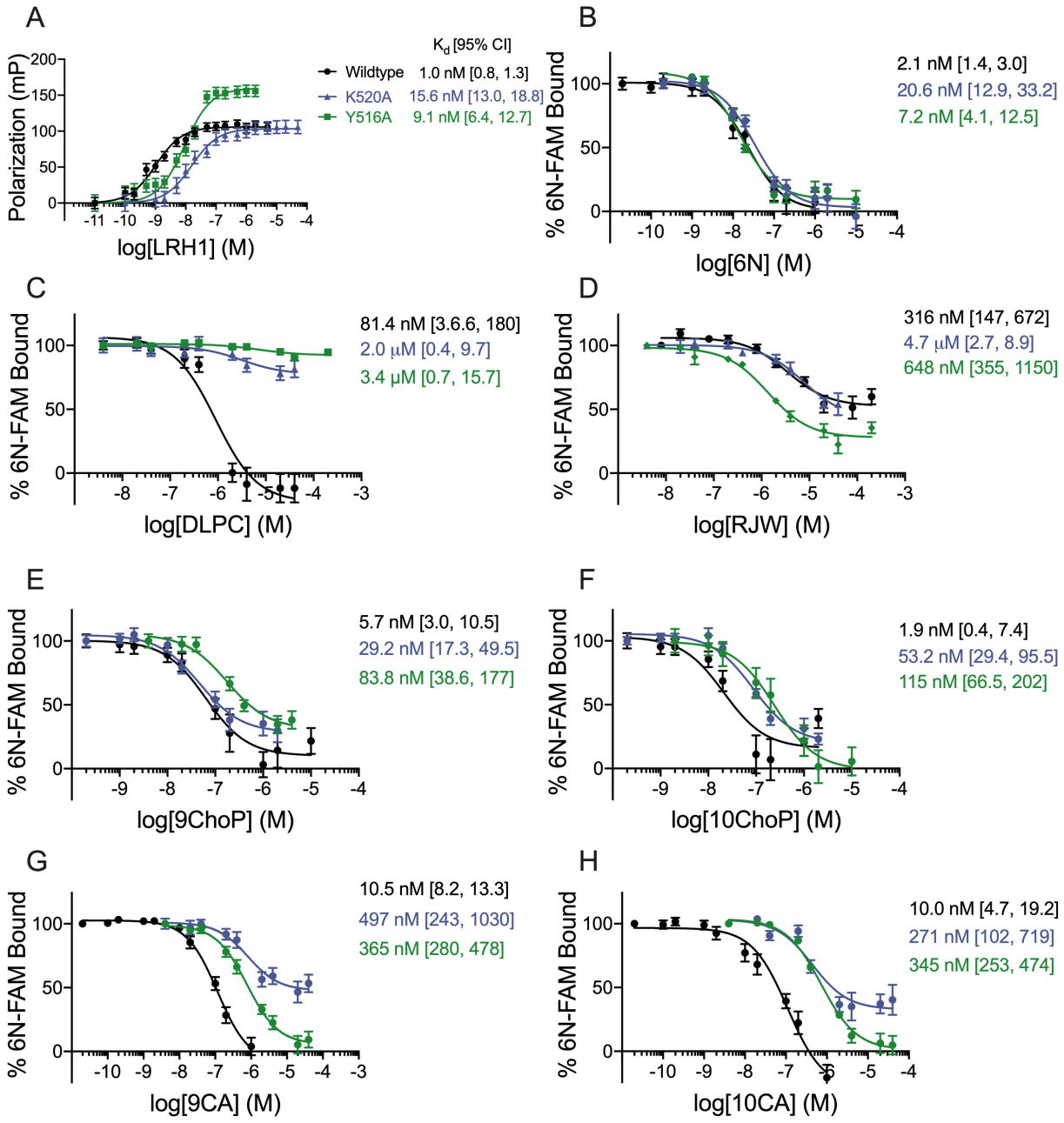


Figure S3.6. Fluorescence polarization: pocket mouth mutants. (A) Forward binding curve for the 6N-FAM probe. For B-H, curves were first baseline subtracted with the lowest concentration of competitor agonist as baseline, then normalized to the unlabeled 6N curve (for wildtype, K520A, or Y516A as appropriate) to determine percent 6N-FAM bound as described in the methods. Insets are 95% confidence intervals; error shown is SEM. Forward binding assays were conducted in triplicate, each

with three technical replicates; competition assays were conducted in duplicate, each with four technical replicates.

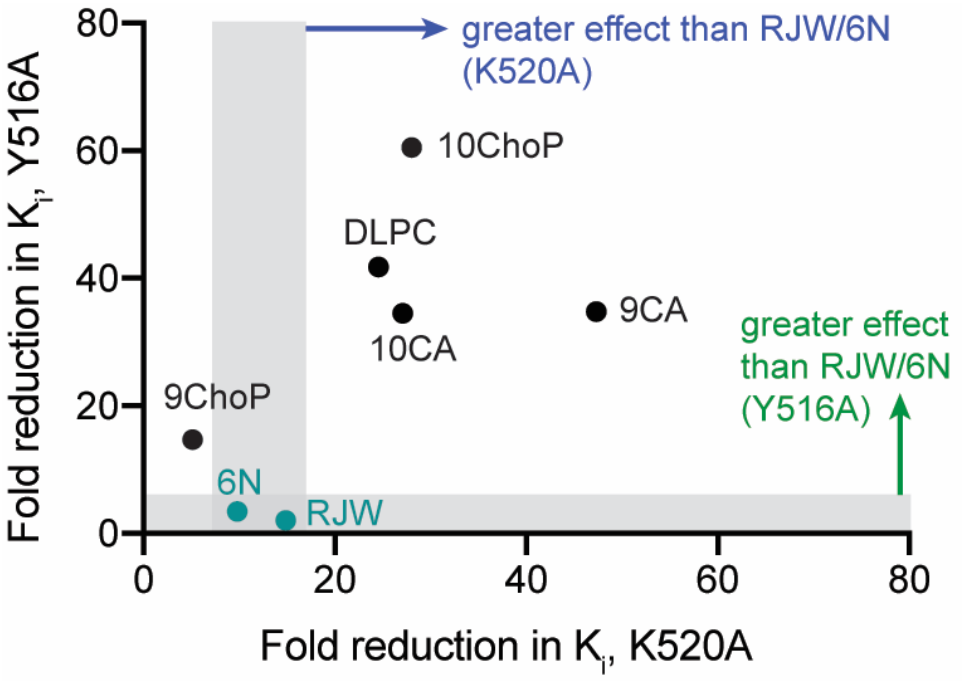


Figure S3.7. Fold reduction in K_i for Y516A versus K520A mutation versus WT K_i. Grey shading indicates the effects of the mutations RJW100 and 6N, which do not contact the mutated residues. DLPC and longer-tailed PL-mimics are more sensitive to these mutations, with the exception of 9ChoP, which is not significantly affected by the K520A mutation.

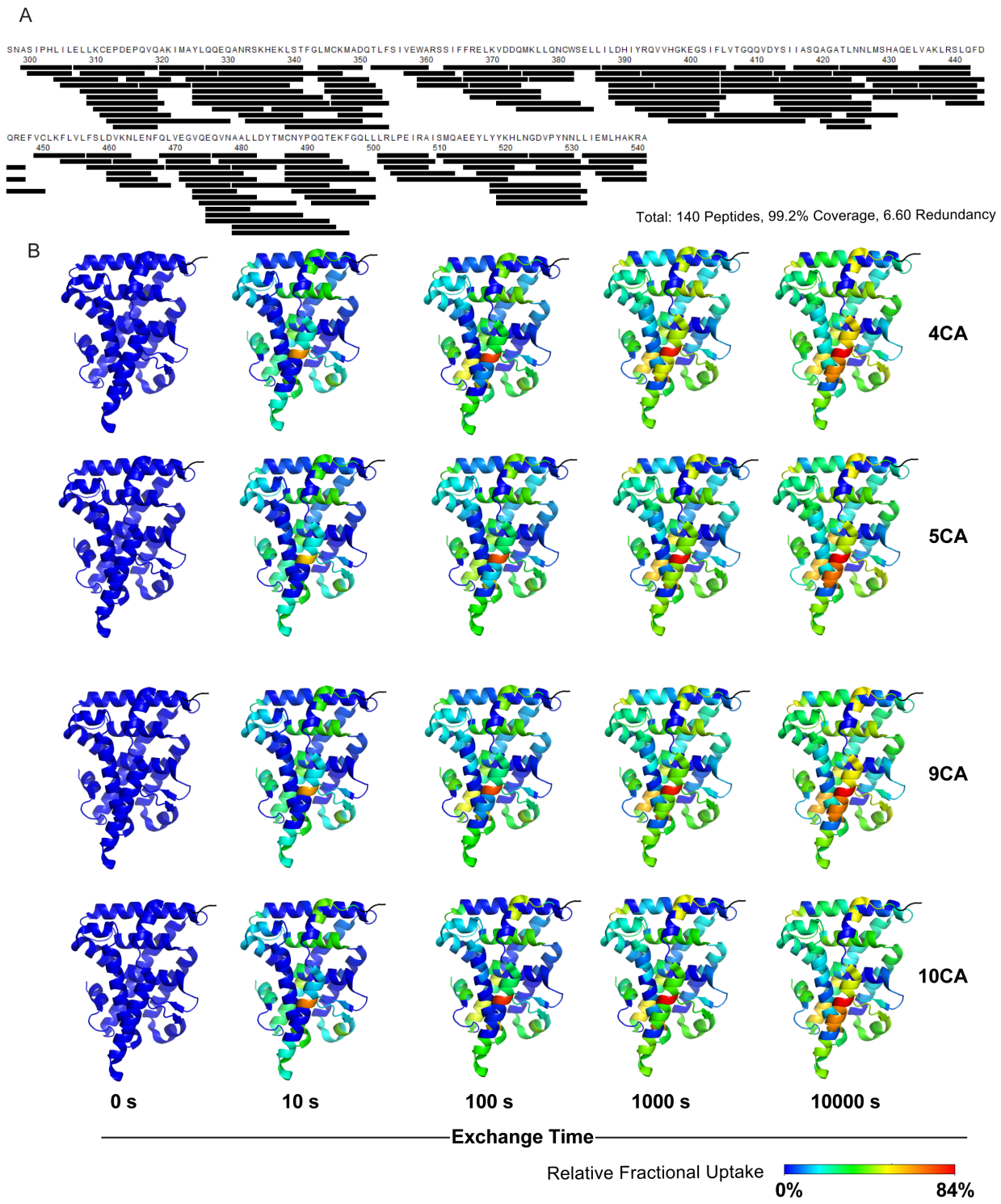
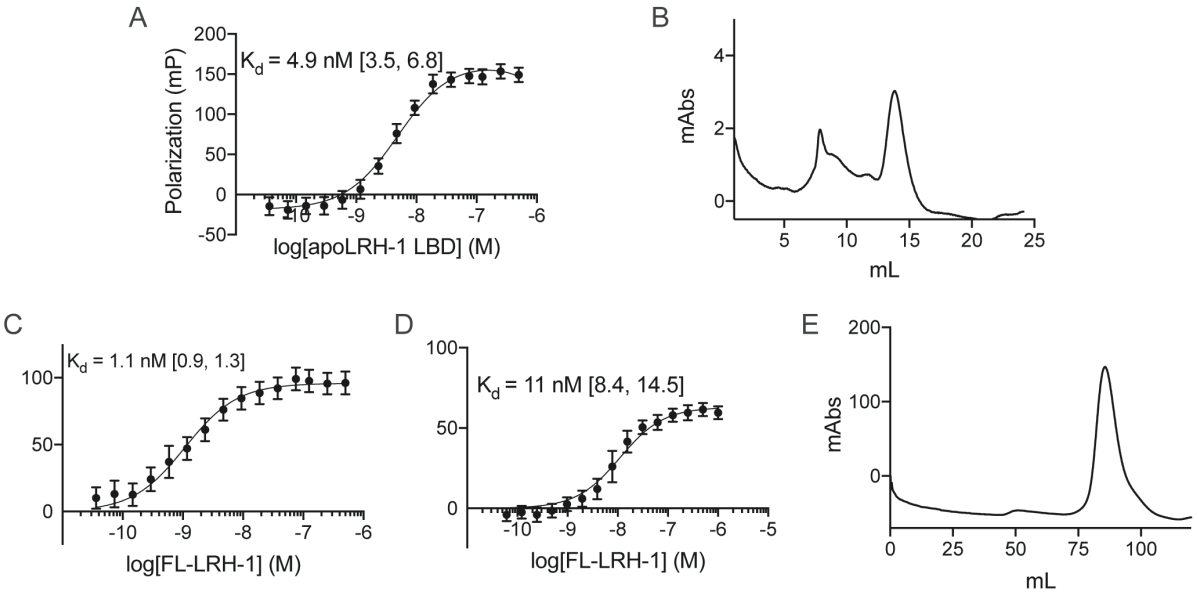
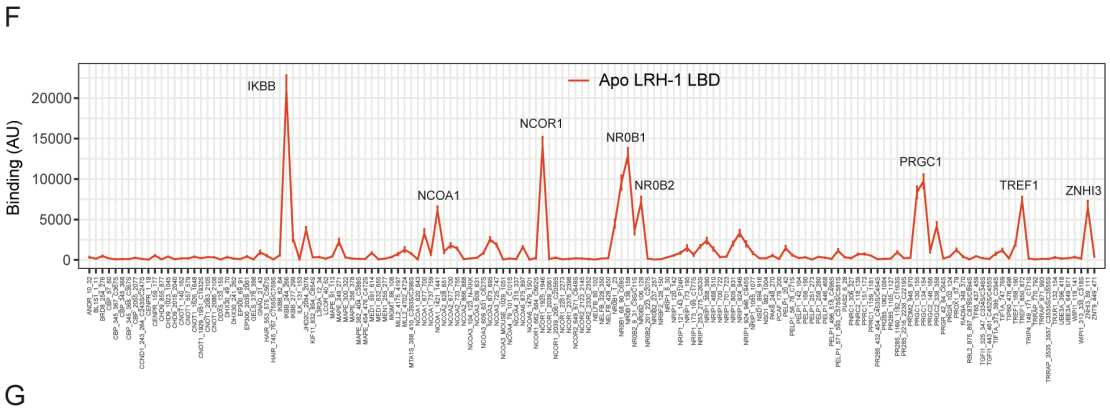
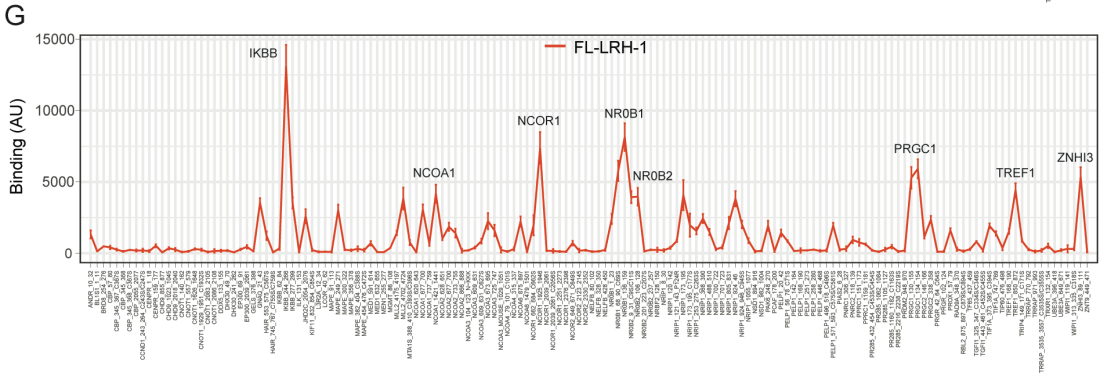


Figure S3.8. Peptide coverage and deuterium uptake in HDX-MS experiments. (A) Peptide coverage. (B) Deuterium uptake over time for each HDX experiment, mapped onto PBD 4DOS.


A
Kd = 4.9 nM [3.5, 6.8]
Polarization (mP)
log[apoLRH-1 LBD] (M)
B
mAbs
mL
C
Kd = 1.1 nM [0.9, 1.3]
log[FL-LRH-1] (M)
D
Kd = 11 nM [8.4, 14.5]
log[FL-LRH-1] (M)
E
mAbs
mL


F
Apo LRH-1 LBD
Binding (AU)
IKBB
NCOA1
NCOR1
NR0B1
NR0B2
PRGC1
TREF1
ZNHI3


G
FL-LRH-1
Binding (AU)
IKBB
NCOA1
NCOR1
NR0B1
NR0B2
PRGC1
TREF1
ZNHI3

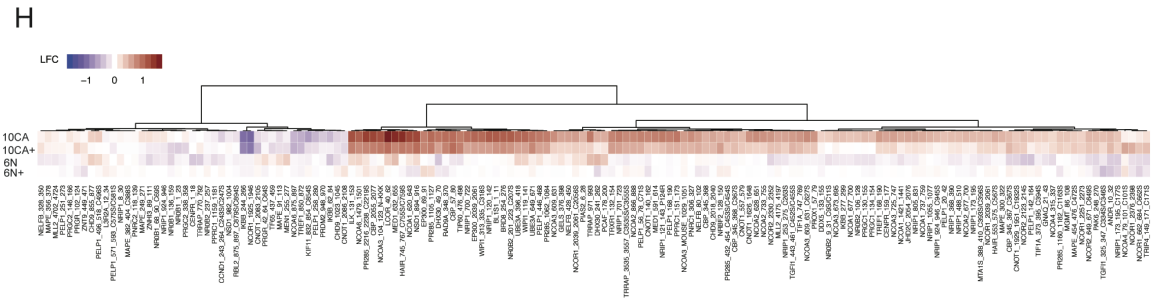

H
LFC
10CA
10CA+
6N
6N+

Figure S3.9. Protein purification and MARCoNI. (A) Binding curve of the 6N-FAM probe to apo LRH-1 LBD. Each point represents the mean ± SEM from one experiment conducted in duplicate. *Inset* indicates the K_d, with the 95% confidence interval in square brackets. (B) Profile from size exclusion chromatography showing the pure protein eluting at ~14 mL. (C) Binding curve of the 6N-FAM probe to FL-LRH-1. Each point represents the mean ± SEM from one experiment conducted in duplicate. *Inset* indicates the K_d, with the 95% confidence interval in square brackets. (D) Fluorescence polarization binding curve showing the association between FL-LRH-1 and the LRH-1 response element on the CYP7A1 promoter. *Inset* indicates the K_d, with the 95% confidence interval in square brackets. (E) Size exclusion chromatogram for FL-LRH-1, with clear elution peak at ~85 mL. F. Coregulator recruitment by apo LRH-1 LBD, showing recruitment of coregulator peptides, including previously identified LRH-1-interacting motifs from NR0B1, NR0B2, and NCOR1, as well as the novel interactor, IκBβ. G. FL-LRH-1 demonstrates a similar MARCoNI trace to apo LRH-1 LBD when not complexed with agonist. (H) Heatmap comparing coregulator recruitment by 10CA and 6N to FL-LRH-1. "10CA+" and "6N+" refer to the addition of excess ligand immediately before the MARCoNI versus only complexing with ligand before size exclusion chromatography (see Methods).

Table S3.3. Modified colitis disease activity score.

Scores	0	1	2	3	4
Weight loss (%)	0	>0%	≥5%	≥10%	≥15%
Diarrhea	Normal	Mild	Loose	Moderate	Liquid
Hematochezia	None	FOBT±	FOBT+	Blood++	Blood+++
Sick appearance	Normal	Mild uncleanness	Moderate uncleanness	Hunched, Slow moving	Lethargic

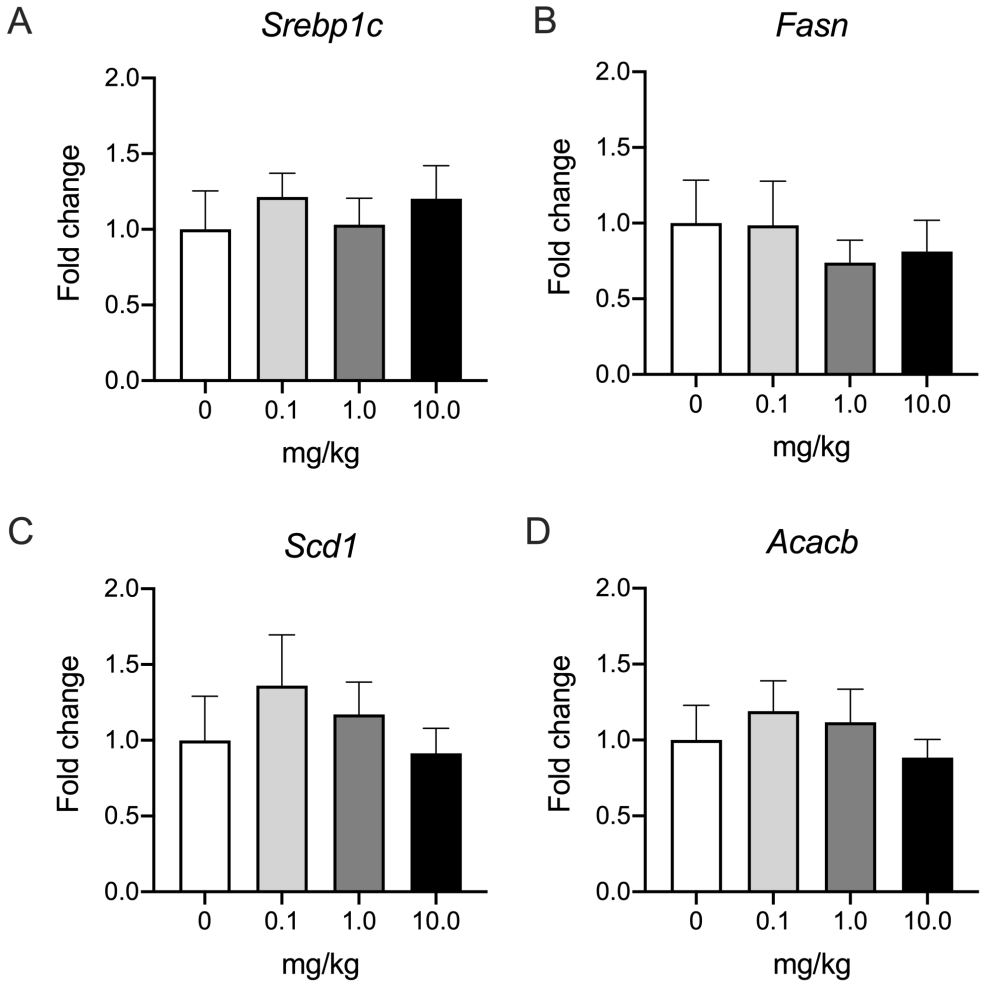


Figure S3.10. 10CA treatment does not activate mouse LRH-1. Nanostring gene expression analysis in wildtype mice reveals that 10CA does not activate mouse LRH-1 and induce expression of LRH-1 target genes. Mice were treated with 5 doses of the indicated dosage of 10CA over 3 days, each 12 hours apart (see Methods for further detail). Shown here are *Srebp1c* (*A*), *Fasn* (*B*), *Scd1* (*C*), and *Acab* (*D*) mRNA expression. Error bars shown are SEM; n = 4-5 mice per group.

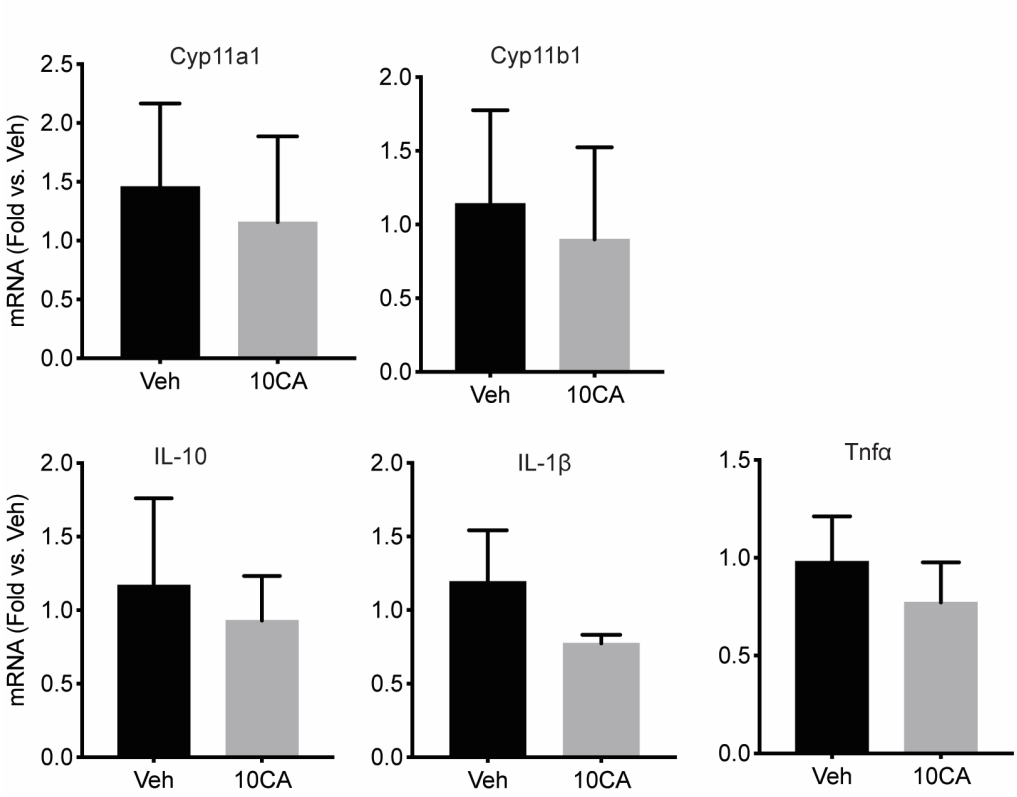


Figure S3.11. 10CA does not alter gene expression in enteroids lacking LRH-1. Enteroids derived from epithelial-specific Lrh-1 knockout mice were treated with DMSO (Veh) or 10CA at 1 uM. Gene expression changes were measured by q-RT-PCR as described in Figure 6.

References

1. van der Veen, J.N. et al. The critical role of phosphatidylcholine and phosphatidylethanolamine metabolism in health and disease. *Biochim Biophys Acta Biomembr* **1859**, 1558-1572 (2017).
2. Chaurio, R.A. et al. Phospholipids: key players in apoptosis and immune regulation. *Molecules* **14**, 4892-914 (2009).
3. Dawaliby, R. et al. Allosteric regulation of G protein-coupled receptor activity by phospholipids. *Nat Chem Biol* **12**, 35-9 (2016).
4. Liu, S. et al. A diurnal serum lipid integrates hepatic lipogenesis and peripheral fatty acid use. *Nature* **502**, 550-4 (2013).
5. O'Donnell, V.B., Rossjohn, J. & Wakelam, M.J. Phospholipid signaling in innate immune cells. *J Clin Invest* **128**, 2670-2679 (2018).
6. Wymann, M.P. & Scheiter, R. Lipid signalling in disease. *Nat Rev Mol Cell Biol* **9**, 162-176 (2008).
7. Li, Z. et al. The ratio of phosphatidylcholine to phosphatidylethanolamine influences membrane integrity and steatohepatitis. *Cell Metab* **3**, 321-31 (2006).
8. Furse, S. & de Kroon, A.I. Phosphatidylcholine's functions beyond that of a membrane brick. *Mol Membr Biol* **32**, 117-9 (2015).
9. Walker, A.K. 1-Carbon Cycle Metabolites Methylate Their Way to Fatty Liver. *Trends Endocrinol Metab* **28**, 63-72 (2017).
10. Chakravarthy, M.V. et al. Identification of a physiologically relevant endogenous ligand for PPARalpha in liver. *Cell* **138**, 476-88 (2009).
11. Kang, H.W., Kanno, K., Scapa, E.F. & Cohen, D.E. Regulatory role for phosphatidylcholine transfer protein/StarD2 in the metabolic response to peroxisome proliferator activated receptor alpha (PPARα). *Biochimica et Biophysica Acta (BBA) - Molecular and Cell Biology of Lipids* **1801**, 496-502 (2010).
12. Musille, P.M. et al. Antidiabetic phospholipid-nuclear receptor complex reveals the mechanism for phospholipid-driven gene regulation. *Nat Struct Mol Biol* **19**, 532-7, S1-2 (2012).
13. Ortlund, E.A. et al. Modulation of human nuclear receptor LRH-1 activity by phospholipids and SHP. *Nat Struct Mol Biol* **12**, 357-63 (2005).
14. Musille, P.M., Kossmann, B.R., Kohn, J.A., Ivanov, I. & Ortlund, E.A. Unexpected Allosteric Network Contributes to LRH-1 Coregulator Selectivity. *J Biol Chem* (2015).
15. Sablin, E.P. et al. Structure of Liver Receptor Homolog-1 (NR5A2) with PIP hormone bound in the ligand binding pocket. *J Struct Biol* (2015).
16. Lee, J.M. et al. A nuclear-receptor-dependent phosphatidylcholine pathway with antidiabetic effects. *Nature* **474**, 506-10 (2011).
17. Mamrosh, J.L. et al. Nuclear receptor LRH-1/NR5A2 is required and targetable for liver endoplasmic reticulum stress resolution. *Elife* **3**, e01694 (2014).
18. 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. *Proc Natl Acad Sci U S A* **104**, 13098-13103.
19. Xu, P., Oosterveer, M.H., Stein, S., Demagny, H., Ryu, D., Moullan, N., Comment, A., Auwerx, J., Schoonjans, K. LRH-1-dependent programming of mitochondrial glutamine processing drives liver cancer. *Genes Dev* **30**, 1255-1260 (2016).

20. Nadolny, C. & Dong, X. Liver receptor homolog-1 (LRH-1): a potential therapeutic target for cancer. *Cancer Biol Ther* **16**, 997-1004 (2015).
21. Cobo-Vuilleumier, N. et al. LRH-1 agonism favours an immune-islet dialogue which protects against diabetes mellitus. *Nat Commun* **9**, 1488 (2018).
22. Bayrer, J.R. et al. LRH-1 Mitigates Intestinal Infammatory Disease by Maintaining Epithelial Homeostasis and Cell Survival. *bioRxiv* (2018).
23. Musille, P.M., Kossmann, B.R., Kohn, J.A., Ivanov, I. & Ortlund, E.A. Unexpected Allosteric Network Contributes to LRH-1 Coregulator Selectivity. *Journal of Biological Chemistry* **291**, 1411-1426 (2015).
24. Wagner, M. et al. Liver receptor homolog-1 is a critical determinant of methyl-pool metabolism. *Hepatology* **63**, 95-106 (2016).
25. Lin, C.J. & Wang, M.C. Microbial metabolites regulate host lipid metabolism through NR5A-Hedgehog signalling. *Nat Cell Biol* **19**, 550-557 (2017).
26. Choi, S. et al. Methyl-sensing nuclear receptor Liver Receptor Homolog-1 regulates mitochondrial function in mouse hepatocytes. *Hepatology* **71**, 1055-1069 (2019).
27. Flynn, A.R., Mays, S.G., Ortlund, E.A. & Jui, N.T. Development of Hybrid Phospholipid Mimics as Effective Agonists for Liver Receptor Homologue-1. *ACS Medicinal Chemistry Letters* **9**, 1051-1056 (2018).
28. Krylova, I.N. et al. Structural analyses reveal phosphatidyl inositols as ligands for the NR5 orphan receptors SF-1 and LRH-1. *Cell* **120**, 343-55 (2005).
29. Mays, S.G. et al. Development of the First Low Nanomolar Liver Receptor Homolog-1 Agonist through Structure-guided Design. *J Med Chem* (2019).
30. Cornelison, J.L. et al. Development of a new class of liver receptor homolog-1 (LRH-1) agonists by photoredox conjugate addition. *Bioorganic & Medicinal Chemistry Letters* **30**(2020).
31. Musille, P.M. et al. Antidiabetic phospholipid-nuclear receptor complex reveals the mechanism for phospholipid-driven gene regulation. *Nature Structural and Molecular Biology* **19**, 532-7, S1-2 (2012).
32. Mays, S.G. et al. Crystal Structures of the Nuclear Receptor, Liver Receptor Homolog 1, Bound to Synthetic Agonists. *Journal of Biological Chemistry* **291**, 25281-25291 (2016).
33. Mays, S.G. et al. Development of the first low nanomolar liver receptor homolog-1 agonist through structure-guided design. *Journal of Medicinal Chemistry* **62**, 11022-11034 (2019).
34. Koppen, A., Houtman, R., Pijenburg, D., Jeninga, E.H., Ruijtenbeek, R., Kalkhoven, E. Nuclear Receptor-Coregulator Interaction Profiling Identifies TRIP3 as a Novel Peroxisome Proliferator-activated Receptor. *Molecular and Cellular Proteomics*, 2211-2226 (2009).
35. Na, S.-Y. et al. IkBb Interacts with the Retinoid X Receptor and Inhibits Retinoid-dependent Transactivation in Lipopolysaccharide-treated Cells. *J Biol Chem* **273**, 3212-3215 (1998).
36. Corneilson, J.L.C., M. L.; Johnson, A. M.; D'Agostino, E. H.; Melchers, D.; Patel, A. B.; Mays, S. G.; Houtman, R.; Ortlund, E. A.; Jui, N. T. Development of a new class of liver receptor homolog-1 (LRH-1) agonists by photoredox conjugate addition. *bioRxiv* (2020).
37. Lu, T. et al. Molecular Basis for Feedback Regulation of Bile Acid Synthesis by Nuclear Receptors. *Molecular Cell* **6**, 507-515 (2000).
38. Geiss, G.K., Bumgarner, R.E., Birditt, B., Dahl, T., Dowidar, N., Dunaway, D.L., Fell, H.P., Ferree, S., George, R.D., Grogan, T., James, J.J., Maysuria, M., Mitton, J.D., Oliveria, P., Osborn, J.L., Peng, T., Ratcliffe, A.L., Webster, P.J.,Davidson, E.H., Hood, L., Dimitrov, K. Direct multiplexed measurement of gene expression with color-coded probe pairs. *Nature Biotechnology* **26**, 317–325 (2008).
39. Ordás, I., Eckmann, L., Talamini, M., Baumgart, D.C. & Sandborn, W.J. Ulcerative colitis. *The Lancet* **380**, 1606-1619 (2012).

40. Coste, A. et al. LRH-1-mediated glucocorticoid synthesis in enterocytes protects against inflammatory bowel disease. *Proc Natl Acad Sci U S A* **104**, 13098-103 (2007).
41. Mueller, M. et al. The nuclear receptor LRH-1 critically regulates extra-adrenal glucocorticoid synthesis in the intestine. *J Exp Med* **203**, 2057-62 (2006).
42. Bouguen, G. et al. Intestinal steroidogenesis controls PPARgamma expression in the colon and is impaired during ulcerative colitis. *Gut* **64**, 901-10 (2015).
43. Bayrer, J.R. et al. LRH-1 mitigates intestinal inflammatory disease by maintaining epithelial homeostasis and cell survival. *Nat Commun* **9**, 4055 (2018).
44. Basak, O. et al. Induced Quiescence of Lgr5+ Stem Cells in Intestinal Organoids Enables Differentiation of Hormone-Producing Enteroendocrine Cells. *Cell Stem Cell* **20**, 177-190 e4 (2017).
45. Chassaing, B., Aitken, J.D., Malleshappa, M. & Vijay-Kumar, M. Dextran sulfate sodium (DSS)-induced colitis in mice. *Curr Protoc Immunol* **104**, 15 25 1-15 25 14 (2014).
46. Chassaing, B., Aitken, J.D., Malleshappa, M. & Vijay-Kumar, M. Dextran sulfate sodium (DSS)-induced colitis in mice. *Curr Protoc Immunol* **104**, 1-14 (2014).
47. Treede, I. et al. Anti-inflammatory effects of phosphatidylcholine. *J Biol Chem* **282**, 27155-64 (2007).
48. Olson, A., Diebel, L.N. & Liberati, D.M. Exogenous phosphatidylcholine supplementation improves intestinal barrier defense against Clostridium difficile toxin. *J Trauma Acute Care Surg* **77**, 570-5; discussion 576 (2014).
49. Chen, M. et al. Phosphatidylcholine regulates NF-κB activation in attenuation of LPS-induced inflammation: evidence from in vitro study. *Animal Cells and Systems* **22**, 7-14 (2017).
50. Schneider, H., Braun, A., Fullekrug, J., Stremmel, W. & Ehehalt, R. Lipid based therapy for ulcerative colitis-modulation of intestinal mucus membrane phospholipids as a tool to influence inflammation. *Int J Mol Sci* **11**, 4149-64 (2010).
51. MacDonald, J.I., Specher, H. Phospholipid Fatty Acid Remodeling in Mammalian Cells. *Biochim Biophys Acta, Lipids Lipid Metab* **1084**, 105-121 (1991).
52. Fernandez-Marcos, P.J., Auwerx, J. & Schoonjans, K. Emerging actions of the nuclear receptor LRH-1 in the gut. *Biochim Biophys Acta* **1812**, 947-55 (2011).
53. Gensler, L.S. Glucocorticoids: complications to anticipate and prevent. *Neurohospitalist* **3**, 92-97 (2013).
54. D'Agostino, E.H. et al. Development of a Versatile and Sensitive Direct Ligand Binding Assay for Human NR5A Nuclear Receptors. *ACS Medicinal Chemistry Letters* (2019).
55. Aarts, J.M. et al. Robust array-based coregulator binding assay predicting ERalpha-agonist potency and generating binding profiles reflecting ligand structure. *Chem Res Toxicol* **26**, 336-46 (2013).
56. Conway, J.R., Lex, A. & Gehlenborg, N. UpSetR: an R package for the visualization of intersecting sets and their properties. *Bioinformatics* **33**, 2938-2940 (2017).
57. Otwinowski, Z., Minor, W. Processing of X-Ray Diffraction Data Collected in Oscillation Mode. *Methods in Enzymology* **276**, 307-326 (1997).
58. Adams, P.D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* **66**, 213-21 (2010).
59. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* **60**, 2126-32 (2004).
60. Schrodinger, LLC. The PyMOL Molecular Graphics System, Version 1.3r1. (2010).
61. Laskowski, R.A. & Swindells, M.B. LigPlot+: multiple ligand-protein interaction diagrams for drug discovery. *J Chem Inf Model* **51**, 2778-86 (2011).

62. Case, D.C., 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, K. M.; Monard, G.; Nguyen, H.; Omelyan, I.; Onufriev, A.; Pan, F.; Qi, R.; Roe, D. R.; Roitberg, A.; Sagui, C.; Simmerling, C.; Botello-Smith, W.; Swails, J.; Walker, R.; Wang, J.; Wolf, R.; Wu, X.; Xiao, L.; York, D.; Kollman, P. AMBER 2017. *University of Californa, San Franscisco* (2017).
63. Maier, J.A.M., C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C. ff14SB: Improving the accuracy of protein side chain and backbone parameters from ff99SB. *Journal of chemical theory and computation* **11**, 3696-3713 (2015).
64. Wang, J.W., W.; Kollman, P. A.; Case, D. A. Antechamber: an accessory software package for molecular mechanical calculations. *J Am Chem Soc* **222**, U403 (2001).
65. Ryckaert, J.-P.C., G.; Berendsen, H. J. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *Journal of Computational Physics* **23**, 327-341 (1977).
66. Roe, D.R.C.I., T. E. PTRAJ and CPPTRAJ: software for processing and analysis of molecular dynamics trajectory data. *Journal of chemical theory and computation* **9**, 3084-3095 (2013).

CHAPTER 4: CRYSTAL STRUCTURE OF STEROIDOGENIC-FACTOR 1 IN COMPLEX WITH HIGH-AFFINITY SYNTHETIC AGONIST

Emma H. D'Agostino[@*], Michael L. Cato[@*], Jeffery L. Cornelison[#], Alyssa M. Johnson[#], Nathan T. Jui[#], Eric A. Ortlund[@]

Departments of [@]Biochemistry and [#]Chemistry, Emory University, Atlanta, Georgia

[*]Denotes equal contribution

Steroidogenic factor-1 is a phospholipid-regulated nuclear receptor, closely related to liver receptor homolog-1, that controls steroidogenesis in the adrenal glands and gonads and metabolism from the hypothalamus. While there is significant interest in targeting this receptor for the treatment of adrenal cancer and obesity, drug development has been hindered by the lack of crystal structures in complex with any synthetic small molecules. Our previous work with LRH-1 offers the exciting possibility to target SF-1. The work represents the first crystal structure of SF-1 in complex with a synthetic agonist, and the synthetic agonist here is the best reported for SF-1. This chapter will be submitted for publication in 2021.

---

[a]E.H.D, M.L.C., and E.A.O. participated in research design. E.H.D., M.L.C., J.L.C., and A.M.J. performed experiments. E.H.D., M.L.C., J.L.C., A.M.J., and E.A.O. analyzed data. E.H.D. wrote the manuscript. E.H.D. conducted all binding and thermal shift assays and solved the crystal structure.

Abstract

Steroidogenic factor-1 (SF-1) is a phospholipid-regulated nuclear receptor expressed in the adrenal glands, gonads, and hypothalamus that controls steroidogenesis and metabolism. There is significant therapeutic interest in SF-1, as it is an oncogene in adrenocortical cancer and may be targetable in obesity. However, the endogenous phospholipid ligand has not been identified for this receptor and drug development has been difficult. While synthetic modulators have been described, no crystal structures have been reported of SF-1 in complex with synthetic compounds, preventing the establishment of structure-activity relationships. Here, we report the first crystal structure of SF-1 in complex with a synthetic agonist. This compound, recently discovered in our labs, is the best agonist reported for SF-1, with low nanomolar affinity and potency. This structure offers the opportunity to link structure and activity, facilitating design of more potent agonists for SF-1 that also discriminate between this receptor and its close homolog, liver receptor homolog-1.

Introduction

Steroidogenic factor-1 (SF-1) is one of 48 human members of the nuclear receptor (NR) family, composed of ligand-regulated transcription factors that control diverse biological processes including development, metabolism, and steroidogenesis. NR activity is controlled by specific, high-affinity lipophilic ligands, making them excellent drug targets, and over one-third of NRs have already been targeted in the clinic.[1] SF-1 belongs to the NR5A subfamily, a class of monomeric, phospholipid-binding NRs.[2] SF-1 is potential therapeutic target for the treatment of adrenocortical cancer and obesity, but its largely hydrophobic binding pocket and the lack of available crystal structures in complex with synthetic small molecules have hindered the development of new modulators.[3]

SF-1 is expressed in the adrenal cortex, gonads, and ventromedial nucleus of the hypothalamus (VMH), and it is necessary for the development of each organ where it is expressed.[3, 4] In the adrenal

glands and gonads, it is a master steroidogenic regulator, controlling: the expression of every adrenal and gonadal steroidogenic gene; the uptake, transport, and *de novo* synthesis of cholesterol to supply steroidogenesis; and the expression of receptors for adrenocorticotropic and follicle-stimulating hormones to stimulate adrenal and gonadal steroidogenesis.[5] SF-1's control of steroidogenesis has implicated it as a likely oncogene in adrenocortical carcinoma, a rare and deadly cancer.[6-8] In the VMH, a region critical for the maintenance of energy homeostasis, SF-1 regulates metabolism. The precise genes controlled by SF-1 in the VMH are unclear, but loss of SF-1 in this region leads to obesity due to increased food intake, decreased energy expenditure, and impaired glucose tolerance and insulin sensitivity.[9] The processes that SF-1 controls therefore make it an attractive therapeutic target in adrenocortical carcinoma and obesity.

The endogenous ligands for the NR5A receptors are believed to be phospholipids (PLs). The first crystal structures of both SF-1 and the only other human NR5A receptor, liver receptor homolog-1 (LRH-1), revealed bacterial phospholipids in the binding pocket.[10, 11] Subsequent studies identified medium-chained, saturated phosphatidylcholine species which could activate the NR5As in cells and activate LRH-1 *in vivo*.[12] The NR5As can also bind phosphatidylinositol species with high affinity.[13, 14] However, there has not yet been a ligand identified which activates SF-1 *in vivo*. Even if a specific PL were identified, PLs do not make ideal therapeutic candidates; they are challenging to work with in the laboratory setting and are rapidly metabolized and remodeled *in vivo*, limiting their specificity. Thus, there is significant interest in the development of synthetic modulators both to inactivate and activate this receptor.

Synthetic modulators have been challenging to develop for SF-1. The ligand binding pocket (LBP) is largely hydrophobic, leaving few regions for scaffolds to specifically target. Furthermore, only recently was a simple, direct binding assay developed that makes it possible to characterize positive results from large-scale screens.[14] There have been a handful of antagonists and one agonist scaffold

reported for SF-1, but no crystal structures have been solved to date of SF-1 in complex with any synthetic modulators, preventing the establishment of a structure-activity relationship to further their development.[12, 15-17]

Recently, our labs synthesized the most effective LRH-1 agonist to date, with 43 nanomolar potency and 180 picomolar affinity (unpublished). This agonist, 6N-10CA, is a hybrid of earlier small molecules which target both polar residues in the deep binding pocket and PL-binding residues at the pocket mouth.[18-20] We show here that 6N-10CA is also the best agonist reported for SF-1, with low nanomolar affinity and potency, and the stability induced by this molecule allowed us to solve the first crystal structure of SF-1 in complex with a synthetic agonist. Surprisingly, the structure shows that the LBP is co-occupied by a bacterial PL, a phosphatidylethanolamine (PE) species, as a result of the soaking process used to introduce 6N-10CA into the pocket, but the electron density for both ligands is unambiguous. As we have shown previously with 6N and 6N-10CA for LRH-1, the 6N-10CA sulfamide makes an extensive hydrogen bond network with the pocket interior (Mays 2019 and unpublished).[19] This structure offers the first opportunity to directly examine the interactions made by an SF-1 agonist and will accelerate small molecule development for this exciting therapeutic target.

Results

Agonist design. We have previously reported extensive modifications to the SF-1/ LRH-1 dual agonist RJW100.[18-21] Substitution of a sulfamide moiety off the hexahydropentalene core improved affinity and potency two orders of magnitude towards LRH-1, resulting in the first low-nanomolar LRH-1 agonist (Figure 4.1a).[14, 19] This dramatic improvement resulted from an expansive hydrogen bond network deep in the LRH-1 binding pocket initiated by the sulfamide group.[19] While the sulfamide compound, termed 6N, was potent and high affinity, it did not activate LRH-1 to a greater extent than the parent RJW100 compound. Thus, we next sought to increase activation by increasing the

alkyl chain length and appending phospholipid-mimicking groups to access phospholipid-coordinating residues at the pocket mouth (Figure 4.1b).[18, 20] This strategy was successful, resulting in improved activation and potency towards LRH-1, and the best compound from this series, termed 10CA, activates LRH-1 *in vivo* in a murine model of ulcerative colitis.[18, 20] Most recently, we combined these two design strategies into a "hybrid" compound termed 6N-10CA to access both the deep pocket and pocket mouth interactions, resulting in the best LRH-1 agonist to date (Figure 4.1c). The modifications have a synergistic effect; 6N-10CA has picomolar affinity and potency (K_i: 180 pM, 95% confidence interval [120, 280]; EC_{50}: 43 nM [27, 69]) and more effectively alters LRH-1 target gene expression in qPCR assays than its single component agonists (data unpublished). While our work has focused on LRH-1, given that the parent molecule, RJW100, is a dual SF-1/ LRH-1 agonist, we reasoned that its derivatives would also activate SF-1. We have previously demonstrated that a subset of our RJW100 derivatives bind SF-1, though with lower affinity than to LRH-1, and shown that 6N activates SF-1 with high potency.[14, 19] A high-resolution crystal structure of agonist-bound SF-1 would be highly valuable, both for SF-1 therapeutic development and for characterization of potential cross-targeting between LRH-1 and SF-1 given the nearly identical ligand-binding pockets. We hypothesized that 6N-10CA would be high affinity and stabilizing enough to yield such a structure a structure.

6N-10CA is a potent, stabilizing SF-1 agonist. We began by comparing the functional effects of 6N-10CA with its constituent compounds, 6N and 10CA (Figure 4.2). We found that the 6N and 10CA modifications had a synergistic effect on SF-1 binding in our fluorescence polarization assay, resulting in a ~60-fold increase in affinity over 6N and a ~4-fold improvement over 6N (Figure 4.1d). The 60.4 nM K_i of 6N-10CA is the highest affinity reported for an SF-1 synthetic agonist, though it does not reach the picomolar affinity of 6N-10CA for LRH-1. We found mixed effects for in-cell activation, measured by luciferase (Figure 4.1e). 6N-10CA was 3-fold less potent than 6N and

19-fold more potent than 10CA, with an EC_{50} of 170 nM. 6N-1CA had a maximum activation in between 6N and 10CA of 1.7, similar to the level achieved in LRH-1 luciferase assays (unpublished). Finally, we determined the effect of these three compounds on the stability of SF-1 using thermal melting (Figure 4.1f). Compared to the phospholipid control, all three compounds selected for a stable conformer of SF-1: 6N and 6N-10CA both resulted in a shift in melting temperature of ~6.5 ºC, whereas 10CA shifted the melting temperature by ~2 ºC. In contrast, 6N-10CA has a synergistic effect in thermal melting for LRH-1, shifting the melting temperature by 11 ºC versus ~9 ºC for 6N and 5.5 ºC for 10CA, indicating that the mechanism of action may differ between the two receptors (Mays 2019 and data unpublished).[19] Taken together, these data demonstrate that the sulfamide and carboxylic acid moieties combine to improve potency and affinity, creating the best SF-1 agonist to date.

Analysis of SF-1-6N-10CA crystal structure. The strong potency, high affinity, and significant stabilization led us to pursue crystallization studies with 6N-10CA to visualize the binding mode and understand the mechanism of action. We solved the structure of SF-1 bound to 6N-10CA and a fragment of the coactivator TIF2 to a resolution of 2.59 Å, using a CysLite version of SF-1 lacking surface cysteines (Table 4.1, Figure 4.2).[13] Surprisingly, 6N-10CA is not the only ligand in the ligand-binding pocket. The pocket is co-occupied by a bacterial phospholipid, mostly likely dipalmitoylphosphoethanolamine (DPPE; 16:0/16:0), which has been co-crystallized with SF-1 several times previously as a bacterial phospholipid from purification.[11, 22, 23] Of the two molecules in the asymmetric unit, 6N-10CA and DPPE are present in chain B, and chain A is fully occupied by DPPE (Figure 4.2a-b). The partial occupancy of 6N-10CA is most likely a result of the soaking method used to generate the structure (see Methods). Nevertheless, there is unambiguous electron density indicating the presence of both 6N-10CA and DPPE in the binding pocket (Figure 4.2 c-f). Ligand omit maps omitting 6N-10CA, DPPE, or all ligands demonstrate that the ligand density is

only satisfied when both 6N-10CA and DPPE are present (Figure 4.2c-e, S4.1). The need for both ligands is particularly clear due to the strong density from the 6N-10CA sulfamide group and the DPPE phosphate. In contrast, an omit map generated for chain A shows that DPPE fully satisfies the ligand density (Figure 4.2f).

Although the co-occupied ligands are overlapping, this structure still yields useful information about the interactions made with SF-1. As expected based on previous structures, the DPPE phosphate is coordinated by hydrogen bonds with Y436, K440, and the α-amino group of G341 in chains A and B (Figure 4.3a). The carboxylic acid moiety of 6N-10CA, however, is distant from these same phosphate-coordinating residues at the pocket mouth, suggesting that the 6N-10CA "tail" may not contact the same residues as a phospholipid (Figure 4.3b). This is surprising given our previous work with similar compounds bound to LRH-1 showing the carboxylic acid moiety engaging in hydrogen bonds with the phosphate-coordinating residues (Figure 4.3c).[20] Superposition of the LRH-1 and SF-1 6N-10CA structures reveals that the alkyl tails take dramatically different paths in the pocket, resulting in a distance of 3.7 Å between the carboxylic acid carbons (Figure 4.3c). However, fitting the SF-1 6N-10CA alkyl tail during structure refinement was challenging. Its overlap with the tail of DPPE prevented manipulation towards the established LRH-1 tail path and may have obscured its true positioning.

The 6N-10CA interactions in the deep pocket are nearly identical to those of 6N and LRH-1 (Figure 4.3d).[19] The sulfamide moiety directly interacts with M268 and engages in a water-mediated hydrogen bond network with R313, H310, and the α-amino group of L324. These are all interactions made by the sulfamide group with LRH-1. Missing is a direct interaction with the α-amino group of V326 (V406 in LRH-1). Additionally, LRH-1 T352 is critical for activation by 6N, but not within hydrogen-bonding distance in the structure; the analogous T272 is a similar distance of 3.4 Å from the sulfamide in the SF-1 structure. In sum, 6N-10CA binds SF-1 very similarly to

LRH-1, particularly in the deep pocket, and the co-occupancy with DPPE complicates interpreting the alkyl chain path.

Discussion

SF-1 is a compelling therapeutic target in the settings of obesity and adrenocortical cancer due to its regulation of metabolism and steroidogenesis. Furthermore, while SF-1 is a critical regulator of metabolism through the VMH, its precise role in this region is unknown, and a strong agonist which could target the brain would help to elucidate its transcriptional program.

Here, we characterize a dual LRH-1/ SF-1 agonist, 6N-10CA recently synthesized in our labs (unpublished). 6N-10CA is the best agonist to date for both receptors, with 60.4 nM affinity, 170 nM EC_{50}, and 6 ºC stabilization in by thermal shift for SF-1 (Figure 4.1). The high affinity and stabilization provided by 6N-10CA enabled the solution of the first SF-1 crystal structure with any non-phospholipid small molecule (Figure 4.2). This 2.59 Å structure showed a surprising co-occupancy with the bacterial phospholipid DPPE; SF-1 co-purifies with *E. coli* PLs, and displacing them for crystallization is difficult. We obtained the structure by soaking 6N-10CA into crystals known to contain lipid, and the soaking procedure only partially replaced DPPE. Despite the incomplete ligand exchange, the 6N-10CA-SF-1 structure still yielded valuable information about the molecular interactions between this agonist class and SF-1. The interactions in the deep pocket were nearly identical to those of 6N and 6N-10CA with LRH-1, with the sulfamide moiety forming an extensive hydrogen bond network with M268, H310, R313, and L324 (Mays 2019 and unpublished).[19] Only the LRH-1 V406 hydrogen bond was absent, as the α-amino group of V326 is too far (3.4 Å) to form a hydrogen bond with the sulfamide in the SF-1 structure. A residue critical for activation of LRH-1, T272 (T352 in LRH-1), was 3.4 Å from the sulfamide (3.9 Å in LRH-1); although this is not within hydrogen-bonding distance, it is within interaction distance and may be similarly important for SF-1 activation. Another residue, M268 (M345 in LRH-1), which directly

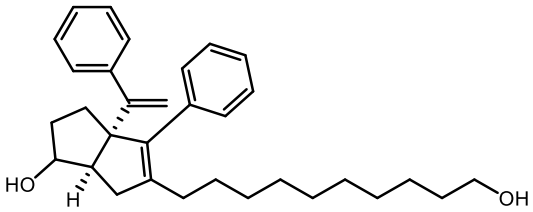
engages the sulfamide group in both the SF-1 and LRH-1 structures but was not found necessary for LRH-1 activity, is another candidate residue for functional investigation. While the deep pocket interactions were nearly identical to those of LRH-1, the electron density of 6N-10CA diverged at the SF-1 pocket mouth. Based on the multiple structures we have previously solved of similar agonists bound to LRH-1 we hypothesized that 6N-10CA would engage the phosphate-coordinating residues Y436, K440, and the α-amino group of G341 (Figure 4.3c).[20] Instead, the 6N-10CA tail diverged from the LRH-1 path and the carboxylic acid moiety is positioned outside of hydrogen-bonding distance with these residues in our structure. However, the definitive path of the 6N-10CA tail was obfuscated by the overlapping DPPE tail, and the positioning was uncertain. Further functional studies with mutagenesis or molecular dynamics simulations will determine whether interactions with the deep pocket and pocket mouth are necessary for the activity of 6N-10CA. Together, these results demonstrate a high affinity, potent, strongly stabilizing SF-1 agonist, likely with a similar mechanism of activation to LRH-1.

With a highly effective agonist and the first synthetic small molecule-bound crystal structure of SF-1 in-hand, new avenues of therapeutic development are now feasible. 6N-10CA would be a useful tool for elucidating SF-1's particular role in altering gene expression in the VMH. Given the therapeutic potential of SF-1 in obesity, altering 6N-10CA to cross the blood-brain barrier and target the VMH *in vivo* is now feasible, with structure-guided design. This structure may also enable improvements on agonist efficacy. 6N-10CA is nearly four times less potent and two orders of magnitude lower affinity than for LRH-1, indicating differences in mechanisms of activation and room for improvement towards SF-1. The ability to directly compare SF-1- and LRH-1-6N-10CA structures also offers the potential to develop more selective compounds. While global activation of LRH-1 is unlikely to be detrimental, unnecessary activation of SF-1 in the adrenal glands would lead to overproduction of adrenal steroids. Thus, using this structure to select agonists towards LRH-1 and

away from SF-1 would further LRH-1 agonist development in the settings of diabetes, obesity, and inflammatory bowel diseases. Finally, SF-1 antagonists are sought for the treatment of adrenocortical cancer, and this structure may yield strategies to modify 6N-10CA for receptor destabilization and antagonism.

Methods

Chemical synthesis. Compounds 6N and 10CA were synthesized as previously described.[18-20] The synthesis and characterization for compound 6N-10CA were as follows:

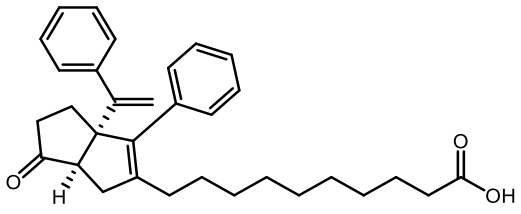


5-(10-hydroxydecyl)-4-phenyl-3a-(1-phenylvinyl)-1,2,3,3a,6,6a-hexahydropentalen-1-ol (3): A slight modification of the procedure of Flynn et al. was used. Prior to use in the reaction, all reagents were dried by azeotropic removal of water using benzene. A dry round bottom flask containing bis(cyclopentadienyl)zirconium(IV) dichloride (1.403 g, 4.8 mmol, 1.2 equiv) under nitrogen, was dissolved in anhydrous, degassed tetrahydrofuran (THF, 5 mL/mmol enyne) and cooled to -78 °C. The resulting solution was treated with *n*-BuLi (3.84 mL, 9.6 mmol, 2.4 equiv.) and the light yellow solution was stirred for 50 minutes. A solution of **tert-butyldimethyl((7-phenylhept-1-en-6-yn-3-yl)oxy)silane** (1.202 g, 4.0 mmol, 1.0 equiv) in anhydrous, degassed THF (5 mL/mmol) was added. The resulting salmon-colored mixture was stirred at -78 °C for 45 minutes, the cooling bath removed, and the reaction mixture was allowed to warm to ambient temperature with stirring (2.5 hours total). The reaction mixture was then cooled to -78 °C for 15 minutes and **tert-butyl((10,10-dibromodecyl)oxy)diphenylsilane** (2.492 g, 4.4 mmol, 1.1 equiv) was added as a solution in anhydrous THF (5 mL/mmol) followed by freshly prepared lithium diisopropylamide (LDA, 4.4

mL, 4.4 mmol, 1.0 M, 1.1 equiv.). After 30 minutes, a freshly prepared solution of lithium phenylacetylide (14.4 mmol, 3.6 equiv.) in anhydrous THF (2 mL/mmol) was added dropwise and the resulting rust-colored solution was stirred at -78 °C for 1 hour. The reaction was quenched with methanol and saturated aqueous sodium bicarbonate and allowed to warm to room temperature, affording a light yellow slurry that stirred overnight. The slurry was then poured onto water and extracted with ethyl acetate four times. The combined organic layers were washed with brine, dried with Na_2SO_4, filtered, and concentrated *in vacuo* to afford a crude mixture. The resulting crude mixture was dissolved in 200 mL of 1:1 DCM:MeOH in a round bottom flask then 0.5 mL of concentrated HCl added. The resulting solution was stirred at room temperature for 2.5 hours before concentrating *in vacuo* and subjecting to silica gel chromatography (5-50% EtOAc/hexanes eluent) to afford the title compound as a yellow oil and 1.7:1 mixture of diastereomers used in the next step without separation. (1.47 g, 80% over 2 steps).

Exo diastereomer: **^{1}H NMR** (600 MHz, $CDCl_3$) δ 7.37 – 7.23 (m, 8H), 7.20 (t, *J* = 8.0 Hz, 2H), 5.07 (d, *J* = 1.4 Hz, 1H), 4.99 (d, *J* = 1.4 Hz, 1H), 3.96 – 3.93 (m, 1H), 3.64 (t, *J* = 6.6 Hz, 2H), 2.36 (dd, *J* = 16.9, 9.3 Hz, 1H), 2.29 (dd, *J* = 9.3, 1.8 Hz, 1H), 2.13 – 1.98 (m, 4H), 1.75 – 1.63 (m, 2H), 1.56 (p, *J* = 6.8 Hz, 3H), 1.43 – 1.17 (m, 14H).

Endo diastereomer: **^{1}H NMR** (600 MHz, $CDCl_3$) δ 7.37 – 7.23 (m, 8H), 7.20 (t, *J* = 7.5 Hz, 2H), 5.07 (d, *J* = 1.4 Hz, 1H), 4.93 (d, *J* = 1.4 Hz, 1H), 4.18 (td, *J* = 8.9, 5.6, 1H), 3.64 (t, *J* = 6.6 Hz, 2H), 2.62 (dd, *J* = 17.5, 2.1 Hz, 1H), 2.48 (td, *J* = 8.7, 2.0 Hz, 1H), 2.13 – 1.98 (m, 4H), 1.84 (dq, *J* = 10.0, 4.9 Hz, 1H), 1.75 – 1.63 (m, 2H), 1.56 (p, *J* = 6.8 Hz, 2H), 1.43 – 1.17 (m, 14H).

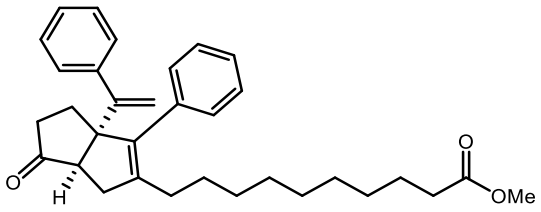


10-(-6-oxo-3-phenyl-3a-(1-phenylvinyl)-1,3a,4,5,6,6a-hexahydropentalen-2-yl)decanoic acid (4): To a solution of **3** in acetonitrile (592 mg, 1.3 mmol, 0.1 M) was added tetrapropylammonium perruthenate (45.3 mg, 0.13 mmol, 0.1 equiv.), *N*-methylmorpholine *N*-oxide (2.29 g, 12.9 mmol, 10 equiv.), and water (0.24 mL, 12.9 mmol, 10 equiv.) and stirred at room temperature overnight. The reaction solution was then filtered through a pad of silica with 99:1 EtOAc:AcOH to collect the title compound as a yellow oil (608 mg, quant.).

^{1}H NMR (500 MHz, $CDCl_3$) δ 7.40 – 7.19 (m, 10H), 5.22 (d, *J* = 1.5 Hz, 1H), 5.11 (d, *J* = 1.4 Hz, 1H), 2.46 (d, *J* = 7.8 Hz, 1H), 2.36 – 2.26 (m, 4H), 2.16 – 1.95 (m, 5H), 1.91 (dd, *J* = 16.5, 7.8 Hz, 1H), 1.61 (p, *J* = 7.5 Hz, 2H), 1.46 – 0.97 (m, 12H). Carboxylic acid proton (-COO**H**) not observed.

^{13}C NMR (126 MHz, $CDCl_3$) δ 223.0, 179.9, 153.3, 145.0, 142.6, 137.5, 136.8, 129.1, 128.4, 128.2, 127.7, 127.2, 127.1, 115.4, 65.6, 55.7, 38.9, 37.6, 34.2, 30.1, 29.8, 29.4, 29.3, 29.1, 28.5, 27.7, 24.8.

HPLC method A, **LRMS** (ESI, APCI) *m/z*: calc'd for $C_{32}H_{39}O_3$ $(M+H)^+$ 471.3, found 470.8.

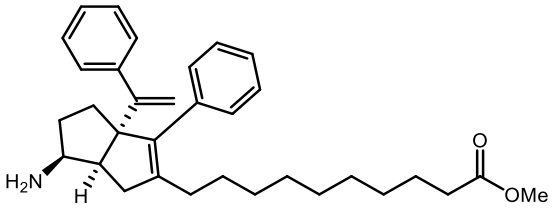


methyl 10-(6-oxo-3-phenyl-3a-(1-phenylvinyl)-1,3a,4,5,6,6a-hexahydropentalen-2-yl)decanoate (5): To a solution of **4** in methanol (945 mg, 2 mmol, 0.1 M) was added 5 drops of concentrated HCl and stirred at room temperature overnight. Reaction solution was then concentrated *in vacuo* and filtered through a pad of silica to collect the title compound as a yellow oil (930 mg, 96%).

^{1}H NMR (500 MHz, $CDCl_3$) δ 7.40 – 7.19 (m, 10H), 5.22 (d, *J* = 1.3 Hz, 1H), 5.11 (d, *J* = 1.3 Hz, 1H), 3.66 (s, 3H), 2.46 (d, *J* = 7.7 Hz, 1H), 2.34 – 2.25 (m, 4H), 2.16 – 1.95 (m, 5H), 1.91 (dd, *J* = 16.5, 7.8 Hz, 1H), 1.60 (p, *J* = 7.5 Hz, 2H), 1.33 – 1.10 (m, 12H).

^{13}C NMR (126 MHz, $CDCl_3$) δ 222.7, 174.4, 153.3, 144.9, 142.6, 137.5, 136.7, 129.0, 128.3, 128.2, 127.7, 127.1, 127.1, 115.3, 65.5, 55.6, 51.5, 38.8, 37.6, 34.2, 30.0, 29.7, 29.4, 29.33, 29.27, 29.2, 28.4, 27.7, 25.0.

HPLC method A, **LRMS** (ESI, APCI) *m/z*: calc'd for $C_{33}H_{41}O_3$ $(M+H)^+$ 485.3, found 484.9.



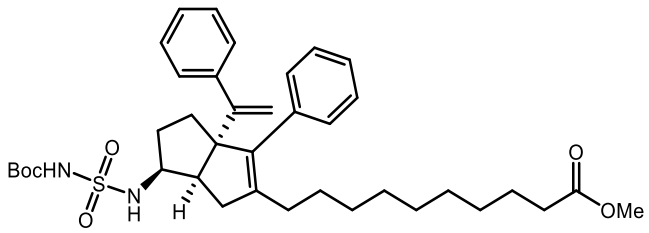
methyl 10-(6-amino-3-phenyl-3a-(1-phenylvinyl)-1,3a,4,5,6,6a-hexahydropentalen-2-yl)decanoate (6): To a screw top test tube charged with a stir bar was added **5** (350 mg, 0.72 mmol, 1.0 equiv.) and ethanol (3 mL) and sealed. Ammonia (7 M in methanol, 0.52 mL, 3.61 mmol, 5.0 equiv.) then titanium(IV) isopropoxide (0.33 mL, 1.08 mmol, 1.5 equiv.) were added via syringe and stirred at room temperature for 6 hours. The test tube cap was then removed and sodium borohydride (82 mg, 2.16 mmol, 3 equiv.) added portion-wise. The resulting solution was stirred at room temperature overnight before being quenched with EtOAc, saturated aqueous potassium sodium tartrate, and 2 M aqueous sodium hydroxide. The resulting slurry was then sonicated in the reaction tube for 10 minutes before adding to a separatory funnel. The aqueous layer was then drained and remaining EtOAc washed with 2 x 20 mL of aqueous potassium sodium tartrate and 2M sodium hydroxide then 20 mL water and 20 mL brine. The remaining organic layer was then

dried over Na_2SO_4, filtered, and concentrated *in vacuo* to give the title compound as a yellow oil (283 mg, 81%).

^{1}H NMR (500 MHz, $CDCl_3$) δ 7.35 – 7.22 (m, 8H), 7.20 (t, *J* = 2.0 Hz, 1H), 7.18 (t, *J* = 1.5 Hz, 1H), 5.07 (d, *J* = 1.4 Hz, 1H), 4.93 (d, *J* = 1.5 Hz, 1H), 3.66 (s, 3H), 3.31 (td, *J* = 8.7, 5.8 Hz, 1H), 2.49 – 2.40 (m, 2H), 2.29 (t, *J* = 7.6 Hz, 2H), 2.12 – 2.00 (m, 4H), 1.85 – 1.78 (m, 1H), 1.72 – 1.67 (m, 2H), 1.60 (p, *J* = 7.5 Hz, 2H), 1.42 – 1.15 (m, 12H). Amine protons (-N**H$_2$**) not observed.

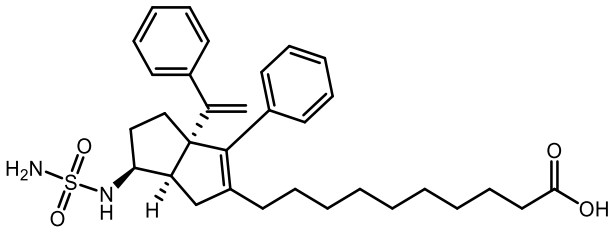
^{13}C NMR (126 MHz, $CDCl_3$) δ 174.5, 155.1, 144.3, 143.0, 139.5, 137.2, 129.9, 127.9, 127.8, 127.7, 126.8, 126.7, 115.3, 69.6, 55.3, 51.6, 34.4, 34.3, 33.3, 30.0, 29.9, 29.52, 29.50, 29.4, 29.3, 28.1, 25.1.

HPLC method A, **LRMS** (ESI, APCI) *m/z*: calc'd for $C_{33}H_{44}NO_2$ $(M+H)^+$ 486.3, found 485.8



methyl 10-(6-((N-(tert-butoxycarbonyl)sulfamoyl)amino)-3-phenyl-3a-(1-phenylvinyl)-1,3a,4,5,6,6a-hexahydropentalen-2-yl)decanoate (7): To a solution of *tert*-butyl alcohol (47 mg, 0.64 mmol, 1.1 equiv.) in anhydrous DCM (0.6 mL) in an oven-dried flask under nitrogen at 0 °C was added neat chlorosulfonylisocyanate (0.050 mL, 0.58 mmol, 1.0 equiv.) and stirred for 45 minutes, warming to room temperature in that time. The resulting solution was then added via syringe to a solution of **6** (283 mg, 0.58 mmol, 1.0 equiv.) and triethylamine (0.12 mL, 0.87 mmol, 1.5 equiv.) in anhydrous DCM (0.6 mL) under nitrogen in an oven-dried flask at 0 °C. The reaction was then stirred and warmed to room temperature over 3 hours before diluting with DCM and

washing with 2 x 10 mL 0.5 M aqueous HCl, 10 mL water and 10 mL brine. The organic layer was then dried over Na_2SO_4, filtered, and concentrated *in vacuo* to give crude material. This material was subjected to silica gel chromatography (10-40% EtOAc/hexanes) to collect material taken crude to the next step.



10-(3-phenyl-3a-(1-phenylvinyl)-6-(sulfamoylamino)-1,3a,4,5,6,6a-hexahydropentalen-2-yl)decanoic acid (1): A 20 mL scintillation vial was charged with a stir bar and **7** (160 mg, 0.24 mmol) and cooled to 0 °C. A 3:1 v/v solution of dioxane and concentrated HCl (2 mL) was then added and allowed to warm to room temperature and stirred for 24 hours before heating to 40 °C for 14 hours. The reaction solution was then diluted with EtOAc and washed with 3 x 5 mL 0.5 M aqueous HCl, 5 ml water, and 5 mL brine. The organic layer was then dried over Na_2SO_4, filtered, and concentrated *in vacuo* to give the title compound as a brown oil (94 mg, 29% over 2 steps).

^{1}H NMR (500 MHz, $CDCl_3$) δ 7.35 – 7.22 (m, 8H), 7.19 (t, *J* = 2.0 Hz, 1H), 7.17 (t, *J* = 1.6 Hz, 1H), 5.10 (d, *J* = 1.3 Hz, 1H), 4.95 (d, *J* = 1.4 Hz, 1H), 4.84 (d, *J* = 7.8 Hz, 1H), 4.72 (s, 2H), 3.78 (dtd, *J* = 11.2, 8.3, 6.0 Hz, 1H), 2.62 (td, *J* = 8.9, 2.1 Hz, 1H), 2.42 (dd, *J* = 17.7, 2.1 Hz, 1H), 2.34 (t, *J* = 7.4 Hz, 2H), 2.16 (dd, *J* = 17.6, 9.0 Hz, 1H), 2.11 – 2.00 (m, 2H), 1.99 – 1.92 (m, 1H), 1.75 – 1.68 (m, 2H), 1.62 (p, *J* = 7.4 Hz, 2H), 1.53 – 1.42 (m, 1H), 1.43 – 1.18 (m, 12H). Carboxylic acid proton (-COO**H**) not observed.

^{13}C NMR (126 MHz, $CDCl_3$) δ 179.3, 154.3, 143.8, 143.01 139.3, 136.8, 129.8, 128.0, 127.8, 127.0, 126.9, 115.7, 68.9, 57.2, 47.5, 35.6, 34.0, 32.6, 31.8, 29.9, 29.7, 29.2, 29.1, 29.0, 28.9, 27.9, 24.6.

HPLC method B, **LRMS** (ESI, APCI) *m/z*: calc'd for $C_{32}H_{43}N_2O_4S$ $(M+H)^+$ 551.3, found 551.8.

Purification – Wildtype SF-1. E. coli strain BL21(DE3)-pLysS was transformed with SF-1 LBD (amino acids 218-461) in the pLIC-His vector and cultured at 37 °C to OD_{600} of 0.6 in Lysogeny Broth medium in the presence of chloramphenicol and ampicillin. Protein expression was induced with 0.5 mM isopropyl-1-thio-D-galactopyranoside (IPTG) for four hours at 32 °C. Cell pellets were lysed in 125 mL NiA (500 mM NaCl, 25 mM imidazole, 5% glycerol, 20 mM Tris HCl pH 7.4, 0.5 mM TCEP) with lysozyme, phenylmethylsulfonyl fluoride (PMSF), and DNase followed by sonication. Lysate was clarified by centrifugation in a Sorvall RC 6+ centrifuge at 16,000 x g for 45 minutes. Supernatant was flowed over a 5 mL HisTrap FF column (GE Healthcare, Little Chalfront, UK) and protein was eluted with NiB (NiA with 500 mM imidazole). To homogenize the lipid population, SF-1 was incubated overnight with dilauroylphosphatidylcholine (DLPC). Size exclusion chromatography (SEC) into assay buffer (150 mM NaCl, 20 mM Tris HCl pH 7.4, 5% glycerol) was used as a final purification step; protein was concentrated to ~3 mg/mL, flash frozen, and stored at -80 °C for use in assays.

Purification – CysLite SF-1. CysLite SF-1 (amino acids 218-461, C30S, C195S) in the pLic-His vector was used for crystallization. This protein was purified as described for wildtype SF-1 through the HisTrap column; after elution from the HisTrap column, the 6X-His tag was cleaved overnight using tobacco etch virus protease. Cleaved protein was flowed over a second HisTrap column and the flowthrough was collected, concentrated to ~3 mg/mL, flash frozen, and stored at -80 °C for use in crystallization.

Crystallization. Cleaved SF-1 CysLite was incubated overnight with 4-fold molar excess compound 6N, and SEC was used to remove bacterial lipids and excess compound and exchange into crystallization buffer (150 mM NaCl, 100 mM ammonium acetate (pH 7.4), 1 mM DTT, and 1 mM

EDTA, 2 mM 3-[(3-cholamidopropyl)dimethylammonio]-1propanesulfonic acid (CHAPS)). An additional 2-fold molar excess of 6N was added to SF-1 collected after SEC. SF-1 was concentrated to~5 mg/mL, 4-fold molar excess of the Tif2 peptide (H_3N-KENALLRYLLDK-CO_2^-) was added, and the complex was incubated at room temperature for two hours. Crystals were seeded with LRH-1-RJW100 crystals, grown as previously described.[21] Crystals were grown using hanging drop vapor diffusion at 4 °C with 0.05 mM Na acetate (pH 4.6), 0-25% glycerol, and 5-11% PEG 4000, with 2-4 uL drops. The crystals contained bacterial phospholipid rather than 6N; thus, soaking was used to exchange the phospholipids. 6N-10CA (100 mM DMSO) was diluted to 2.5 mM in mother liquor, and 0.5 uL volumes were added to drops containing crystals for a 2-day soak. Crystals were flash-frozen in liquid nitrogen using mother liquor with a cryoprotectant of 30% glycerol.

Structure Determination. Data were collected remotely from Argonne National Laboratory (South East Regional Collaborative Access Team, Lemont, IL) using the 22ID beamline. Data were processed and scaled using HKL2000[24] and phased by molecular replacement using Phaser-MR in Phenix[25] with PDB 1ZDT as the search model. Coot[26] and Phenix.refine[25] were used for model building and refinement, respectively. Figures were constructed using Pymol.[27] Ligplot+[28] was used to identify residues interacting with the ligands.

Fluorescence Polarization. All assays were conducted in black, polystyrene, non-binding surface 384-well plates (Corning Inc., Corning, NY) with 30 μL volumes in assay buffer. Binding affinity for 6N-FAM was determined using 10 nM 6N-FAM and protein concentrations ranging from 1^{-10}–5^{-5} M. Plates were incubated overnight at 4 °C and centrifuged at 2,000 x *g* for 2 minutes before polarization measurement. Polarization was monitored on a Neo plate reader (Biotek, Winooski, VT) at an excitation/emission wavelength of 485/528 nm. Nine technical replicates were conducted over three experiments and compiled binding data were baseline-corrected to wells with no protein and fit with a one-site binding curve in GraphPad Prism version 7 (GraphPad, Inc., La Jolla, CA). For competition

assays, 10 nM 6N-FAM (0.8 times the affinity of SF-1 for 6N-FAM) and 25 nM SF-1 (60% of the forward binding B_{max}) were used. Competitor ligand concentration ranged from 2^{-11}-2^{-4} M, and competitor ligand volume was kept constant to maintain constant DMSO in each well (6.7% v/v). Eight technical replicates were performed over two experiments, and GraphPad Prism version 7 was used to analyze compiled data using a one-site, fit K_i curve, with normalization to 6N competition.

Luciferase. Hela cells were purchased from Atlantic Type Culture Collection. Hela cells were cultured in MEMα medium supplemented with 10% charcoal-stripped fetal bovine serum. Cells were maintained under standard culture conditions. Hela cells were seeded into 96-well culture plates at 7500 cells/ well and transfected with reporter plasmids and either pcDNA empty vector or full-length SF-1-pcDNA. Reporter plasmids consisted of (1) the pGL3 basic vector containing a portion of the SHP promoter containing the SF-1 response element cloned upstream of firefly luciferase and (2) a constitutively active vector encoding *Renilla* luciferase used for normalization. Transfections utilized the FugeneHD transfection reagent at a ratio of 3 ul Fugene: 1 ug DNA. Cells were treated with agonists 24 hours after transfection at concentrations indicated in the figure legends. 24 hours after treatment, luminescence was quantified using the DualGlo kit from Promega on a BioTek Neo plate reader. Values for EC_{50} and E_{max} were calculated by fitting data to a three-parameter dose-response curve in GraphPad Prism, v7.

Differential Scanning Fluorimetry. Purified protein, pre-exchanged with DLPC (0.2 mg/mL), was combined with 6N-10CA overnight at 4 °C in assay buffer. SYPRO orange dye was added to the complexes the next day, at a final dilution of 1:1000. Complexes were heated at a rate of 0.5 °C/ minute on a StepOne Plus thermocycler, using the ROX filter for fluorescence detection. The melting temperature (T_m, 50% unfolding) was calculated using the Bolzman equation (GraphPad Prism, V7). Assays were conducted with nine technical replicates over three experiments.

Figures

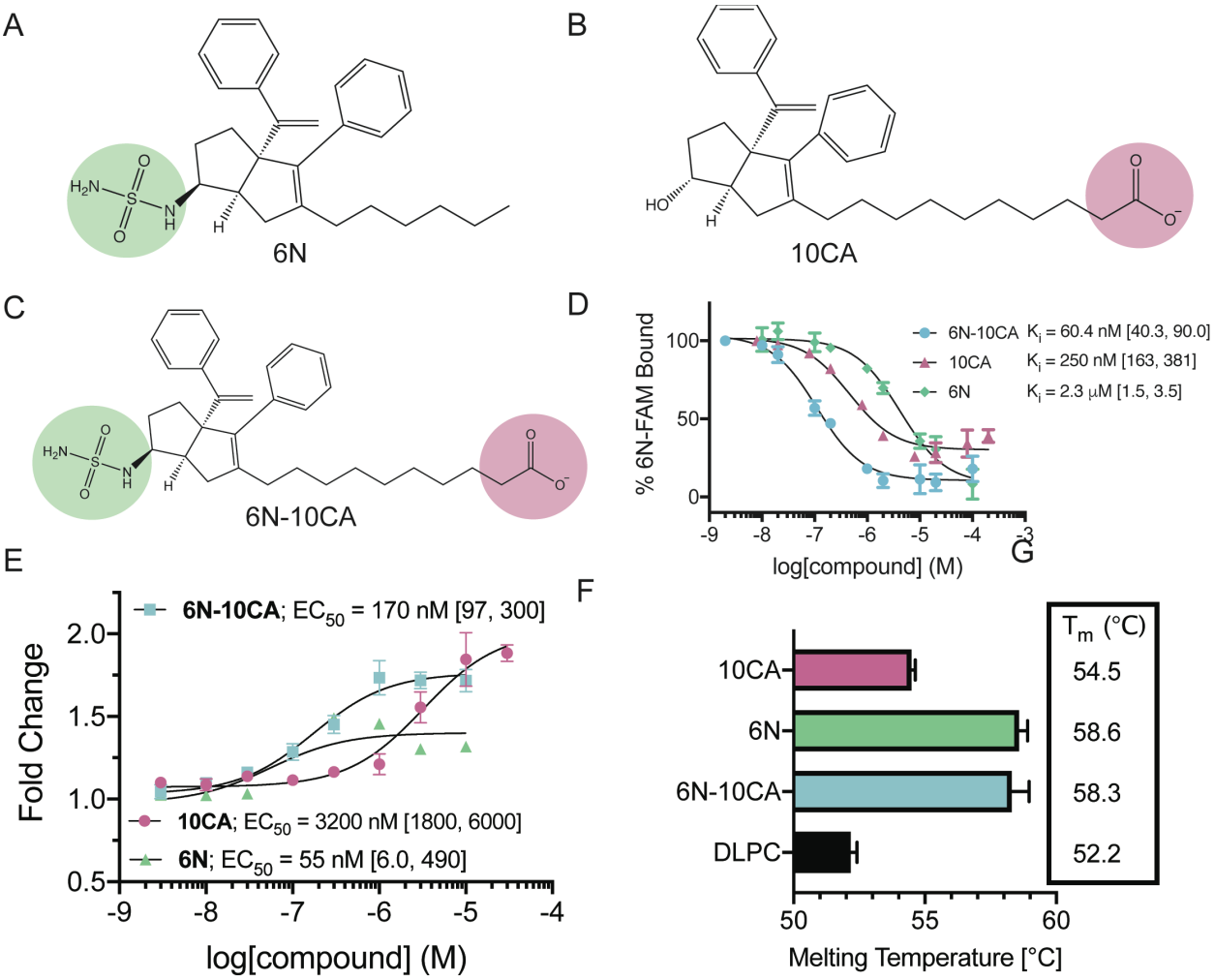


Figure 4.1. 6N-10CA is the highest affinity, most potent SF-1 agonist to date. (A-C) Molecular structures of the previously published 6N[19] (*A*) and 10CA[20] (*B*); the 6N sulfamide (green) and 10CA carboxylic acid (pink) moieties were combined to yield the "hybrid" 6N-10CA (*C*). (D) The sulfamide and carboxylic acid modifications have a synergistic effect on SF-1 binding, yielding the tightest binding agonist to date. Experiments were conducted twice in quadruplicate; error bars are SEM; values in brackets are 95% confidence intervals. (E) 6N-10CA is more potent 10CA, but not 6N, in luciferase reporter assays, with similar an intermediate ~1.7-fold maximum activation. Experiments were conducted thrice in triplicate; error shown is SEM; values in brackets are 95% CIs. (F) Thermal melting reveals that 6N-10CA and 6N stabilize SF-1 to a similar degree compared to DLPC control

but more than 10CA. Experiments were conducted thrice in triplicate; error bars are SEM from the combined replicates.

Table 4.1: X-ray data collection and refinement statistics.

Data collection	SF-1 + 6N-10CA + TIF2
Space group	$P3_1 21$
Cell dimensions	
a, *b*, *c* (Å)	73.4, 73.4, 194.2
α, β, γ (□)	90, 90, 120
Resolution (Å)	38.59 - 2.59 (2.68 - 2.59)
R_{pim}	0.035 (0.453)
$I / \sigma I$	20.3 (1.21)
$CC_{1/2}$	99.0 (67.0)
Completeness (%)	99.8 (99.2)
Redundancy	12.4 (11.2)
Refinement	
Resolution (Å)	2.59
No. reflections	19623
R_{work} / R_{free} (%)	22.9/28.2
No. atoms	
Protein	3934
Water	11
B-factors	
Protein	85.2
Ligand	183.0
Water	68.4
R.m.s. deviations	
Bond lengths (Å)	0.002
Bond angles (°)	0.46
Ramachandran favored (%)	95.6
Ramachandran outliers (%)	0.21
PDB accession code	Not yet deposited

Values in parentheses indicate highest resolution shell.

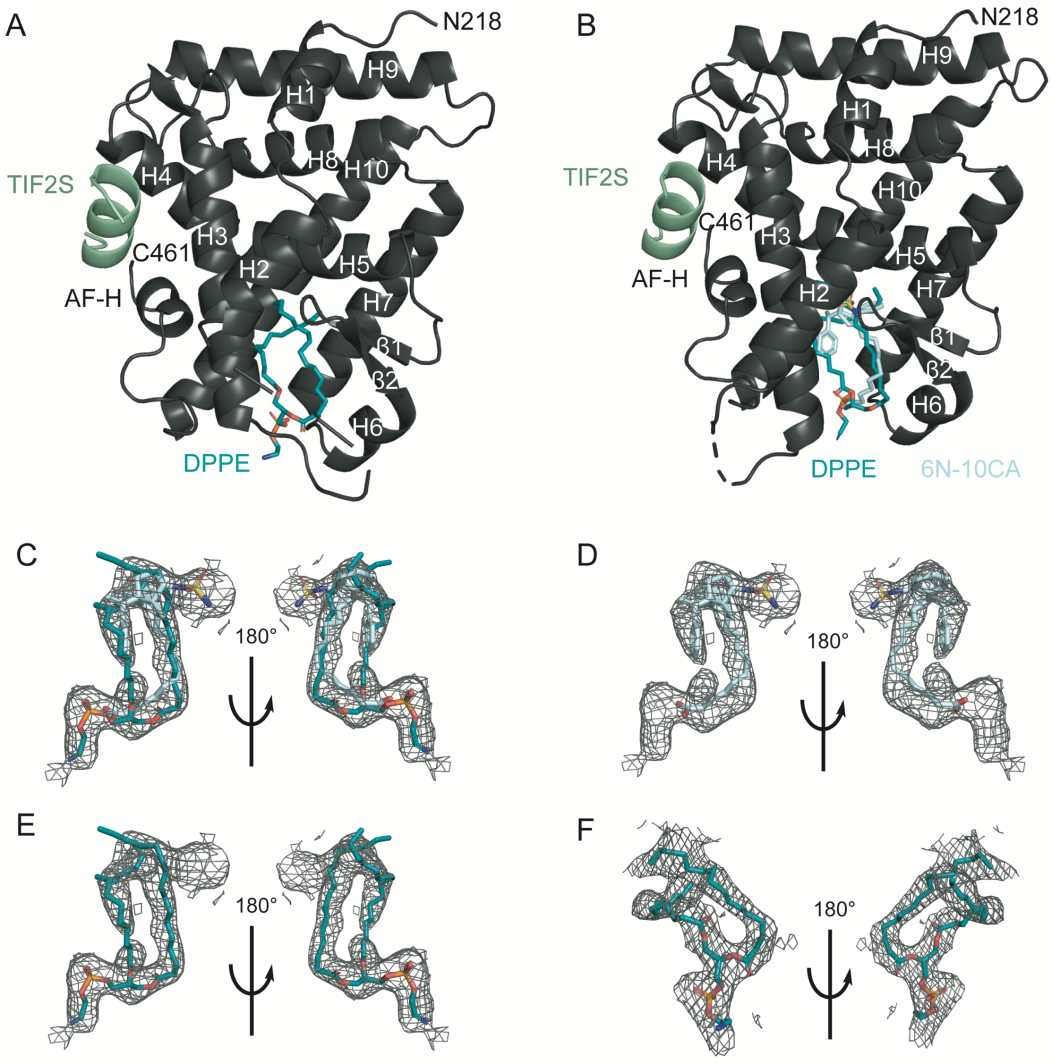


Figure 4.2. 6N-10CA and DPPE co-occupy chain B, but not chain A. (A) In Chain A of the SF-1-6N-10CA structure (not yet deposited in the PDB), the bacterial phospholipid DPPE dominated occupancy of the ligand-binding pocket and 6N-10CA was not detected. (B) In Chain B, 6N-10CA and DPPE co-occupy the ligand-binding pocket, with the tail of DPPE overlapping with the 6N-10CA acyl tail. When all ligands are omitted from the map, 6N-10CA and DPPE together (C) satisfy the ligand density for chain B, but not 6N-10CA (D) or DPPE (E) alone. (E) In contrast, in chain A, DPPE satisfies the ligand density completely. Omit maps are F_o-F_c contoured at $\sigma = 2.0$.

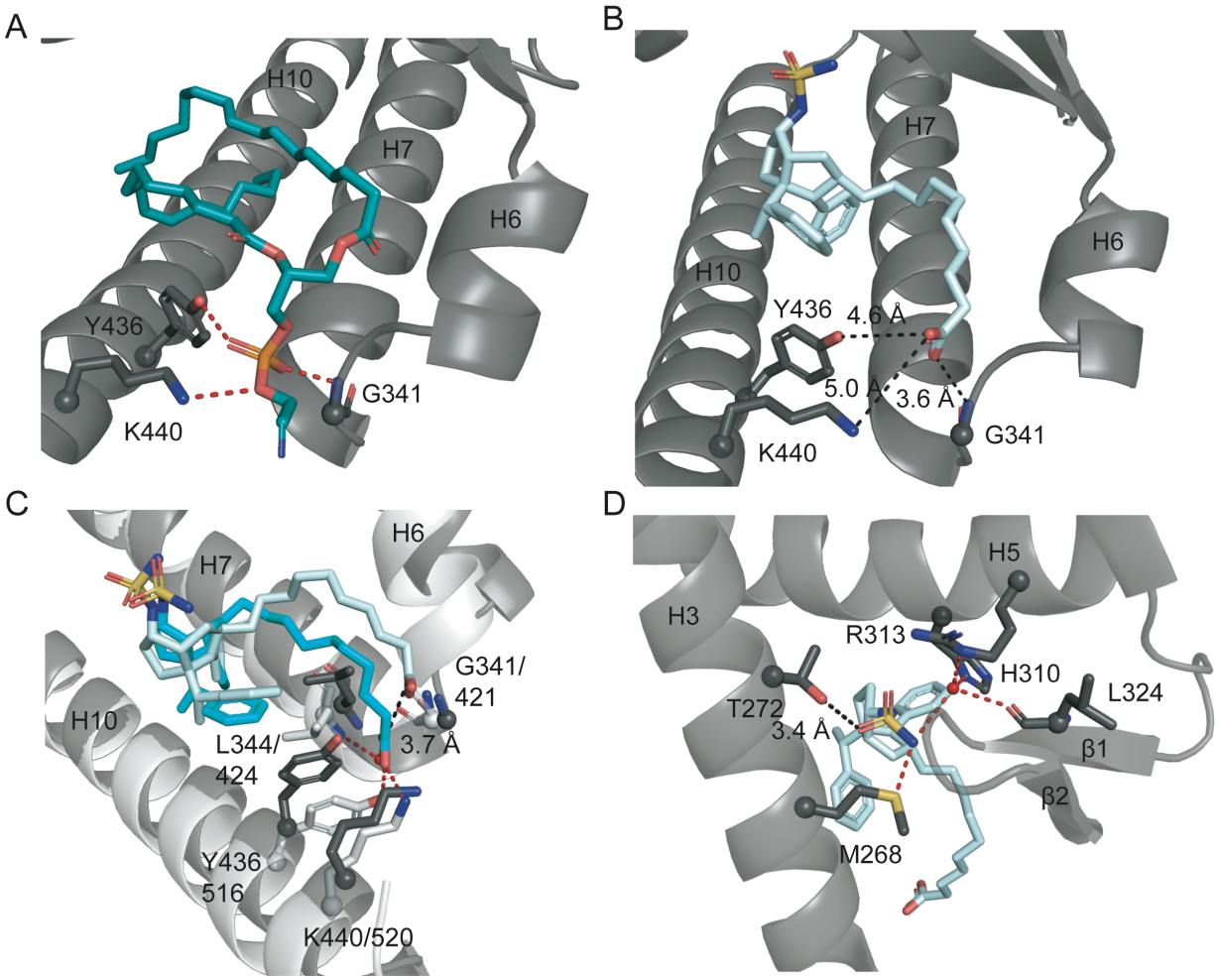
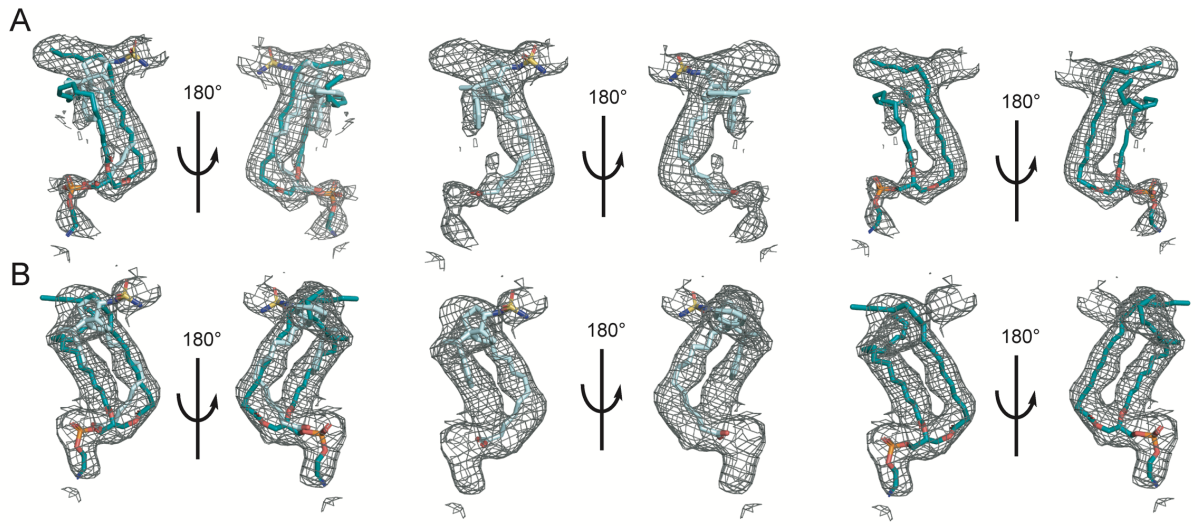


Figure 4.3. DPPE and 6N-10CA ligand binding pocket-interactions. (A) As seen in previous structures, the phosphate group of DPPE is coordinated by a hydrogen bond network with Y436, K440, and the a-amino group of G341.[11, 22, 23] (B) Surprisingly, the 6N-10CA carboxylic acid moiety is not within hydrogen bonding distance of the phosphate-coordinating residues at the pocket mouth in the structure. (C) Comparison of the SF-1 and LRH-1 6N-10CA structures reveals that the SF-1 6N-10CA tail (ligand in ice blue) is shifted considerably from the LRH-1 6N-10CA tail (sky blue; PDB 6OQY). The terminal carbon is shifted away from helices 7 and 10 by ~4A, preventing hydrogen bonding with Y436, K440, and the a-amino group of G341. SF-1 protein backbone is shown in dark grey, LRH-1 in light grey; SF-1 residue numbers are given first. (D) 6N-10CA engages several polar residues deep in the binding pocket, including a direct interaction with M268 and water-mediated

interactions with H310, R313, and L324. There is no hydrogen bond with T272, a residue known to be critical for sulfamide activation of LRH-1.



Supplemental Figure 4.1. Omission of 6N-10CA or DPPE confirms ligand co-occupancy. (A) Omit map with 6N-10CA omitted shows that only 6N-10CA and DPPE together fully satisfy the ligand density, not 6N-10CA (middle) or DPPE (right) alone. (B) DPPE omit map further confirms that only 6N-10CA and DPPE combined satisfy all features of the ligand density. Map is F_o-F_c, contoured at 2σ.

Acknowledgements: The authors are grateful to the beamline staff at Argonne National Laboratory, South East Regional Collaborative Team, for support during remote collection of crystal diffraction data.

Funding and additional information: This work was supported in part by the National Institutes of Health under the following awards: T32GM008367-27 (E.H.D.), 1F31DK122745-01A1 (M.L.C.), R01DK095750 (E.A.O.), R01DK114213 (E.A.O., N.T.J., J.W.C). E.H.D was supported by the National Science Foundation Graduate Research Fellowship.

Conflict of Interest: The authors declare no conflicts of interest.

References

1. Rask-Andersen M, Almen MS, Schioth HB. Trends in the exploitation of novel drug targets. Nat Rev Drug Discov. 2011;10(8):579-90. Epub 2011/08/02. doi: 10.1038/nrd3478. PubMed PMID: 21804595.
2. Weikum ER, Liu X, Ortlund EA. The nuclear receptor superfamily: A structural perspective. Protein Sci. 2018;27(11):1876-92. Epub 2018/08/16. doi: 10.1002/pro.3496. PubMed PMID: 30109749; PMCID: PMC6201731.
3. Meinsohn MC, Smith OE, Bertolin K, Murphy BD. The Orphan Nuclear Receptors Steroidogenic Factor-1 and Liver Receptor Homolog-1: Structure, Regulation, and Essential Roles in Mammalian Reproduction. Physiol Rev. 2019;99(2):1249-79. Epub 2019/02/28. doi: 10.1152/physrev.00019.2018. PubMed PMID: 30810078.
4. Parker KL, Schimmer BP. Steroidogenic factor 1: a key determinant of endocrine development and function. Endocr Rev. 1997;18(3):361-77. Epub 1997/06/01. doi: 10.1210/edrv.18.3.0301. PubMed PMID: 9183568.
5. Val P, Lefrancois-Martinez AM, Veyssiere G, Martinez A. SF-1 a key player in the development and differentiation of steroidogenic tissues. Nucl Recept. 2003;1(1):8. Epub 2003/11/05. doi: 10.1186/1478-1336-1-8. PubMed PMID: 14594453; PMCID: PMC240021.
6. Lalli E. Adrenocortical development and cancer: focus on SF-1. J Mol Endocrinol. 2010;44(6):301-7. Epub 2010/03/05. doi: 10.1677/JME-09-0143. PubMed PMID: 20200142.
7. Duregon E, Volante M, Giorcelli J, Terzolo M, Lalli E, Papotti M. Diagnostic and prognostic role of steroidogenic factor 1 in adrenocortical carcinoma: a validation study focusing on clinical and pathologic correlates. Hum Pathol. 2013;44(5):822-8. Epub 2012/11/20. doi: 10.1016/j.humpath.2012.07.025. PubMed PMID: 23158211.
8. Creemers SG, Hofland LJ, Korpershoek E, Franssen GJ, van Kemenade FJ, de Herder WW, Feelders RA. Future directions in the diagnosis and medical treatment of adrenocortical carcinoma. Endocr Relat Cancer. 2016;23(1):R43-69. Epub 2015/10/18. doi: 10.1530/ERC-15-0452. PubMed PMID: 26475053.
9. Kinyua AW, Yang DJ, Chang I, Kim KW. Steroidogenic Factor 1 in the Ventromedial Nucleus of the Hypothalamus Regulates Age-Dependent Obesity. PLoS One. 2016;11(9):e0162352. Epub 2016/09/07. doi: 10.1371/journal.pone.0162352. PubMed PMID: 27598259; PMCID: PMC5012732.
10. Ortlund EA, Lee Y, Solomon IH, Hager JM, Safi R, Choi Y, Guan Z, Tripathy A, Raetz CR, McDonnell DP, Moore DD, Redinbo MR. Modulation of human nuclear receptor LRH-1 activity by phospholipids and SHP. Nat Struct Mol Biol. 2005;12(4):357-63. Epub 2005/02/22. doi: 10.1038/nsmb910. PubMed PMID: 15723037.
11. Wang W, Zhang C, Marimuthu A, Krupka HI, Tabrizizad M, Shelloe R, Mehra U, Eng K, Nguyen H, Settachatgul C, Powell B, Milburn MV, West BL. The crystal structures of human steroidogenic factor-1 and liver receptor homologue-1. Proc Natl Acad Sci U S A. 2005;102(21):7505-10. Epub 2005/05/18. doi: 10.1073/pnas.0409482102. PubMed PMID: 15897460; PMCID: PMC1140416.
12. Whitby RJ, Stec J, Blind RD, Dixon S, Leesnitzer LM, Orband-Miller LA, Williams SP, Willson TM, Xu R, Zuercher WJ, Cai F, Ingraham HA. Small molecule agonists of the orphan nuclear receptors steroidogenic factor-1 (SF-1, NR5A1) and liver receptor homologue-1 (LRH-1, NR5A2). J Med Chem. 2011;54(7):2266-81. Epub 2011/03/12. doi: 10.1021/jm1014296. PubMed PMID: 21391689; PMCID: PMC4151520.
13. Blind RD, Sablin EP, Kuchenbecker KM, Chiu HJ, Deacon AM, Das D, Fletterick RJ, Ingraham HA. The signaling phospholipid PIP3 creates a new interaction surface on the nuclear receptor SF-1. Proc Natl Acad Sci U S A. 2014;111(42):15054-9. Epub 2014/10/08. doi: 10.1073/pnas.1416740111. PubMed PMID: 25288771; PMCID: PMC4210282.

14. D'Agostino EH, Flynn AR, Cornelison JL, Mays SG, Patel A, Jui NT, Ortlund EA. Development of a Versatile and Sensitive Direct Ligand Binding Assay for Human NR5A Nuclear Receptors. ACS Med Chem Lett. 2020;11(3):365-70. Epub 2020/03/19. doi: 10.1021/acsmedchemlett.9b00442. PubMed PMID: 32184971; PMCID: PMC7074214.
15. Del Tredici AL, Andersen CB, Currier EA, Ohrmund SR, Fairbain LC, Lund BW, Nash N, Olsson R, Piu F. Identification of the first synthetic steroidogenic factor 1 inverse agonists: pharmacological modulation of steroidogenic enzymes. Mol Pharmacol. 2008;73(3):900-8. Epub 2007/12/07. doi: 10.1124/mol.107.040089. PubMed PMID: 18055761.
16. Madoux F, Li X, Chase P, Zastrow G, Cameron MD, Conkright JJ, Griffin PR, Thacher S, Hodder P. Potent, selective and cell penetrant inhibitors of SF-1 by functional ultra-high-throughput screening. Mol Pharmacol. 2008;73(6):1776-84. Epub 2008/03/13. doi: 10.1124/mol.108.045963. PubMed PMID: 18334597; PMCID: PMC3228235.
17. Corzo CA, Mari Y, Chang MR, Khan T, Kuruvilla D, Nuhant P, Kumar N, West GM, Duckett DR, Roush WR, Griffin PR. Antiproliferation activity of a small molecule repressor of liver receptor homolog 1. Mol Pharmacol. 2015;87(2):296-304. Epub 2014/12/05. doi: 10.1124/mol.114.095554. PubMed PMID: 25473120; PMCID: PMC4293447.
18. Flynn AR, Mays SG, Ortlund EA, Jui NT. Development of Hybrid Phospholipid Mimics as Effective Agonists for Liver Receptor Homologue-1. ACS Med Chem Lett. 2018;9(10):1051-6. Epub 2018/10/23. doi: 10.1021/acsmedchemlett.8b00361. PubMed PMID: 30344916; PMCID: PMC6187417.
19. Mays SG, Flynn AR, Cornelison JL, Okafor C. Development of the First Low Nanomolar Liver Receptor Homolog-1 Agonist through Structure-guided Design2019;62(24):11022-34. doi: 10.1021/acs.jmedchem.9b00753. PubMed PMID: 31419141.
20. Mays SG, D'Agostino, E.H., Flynn, A.R., Huang, X., Wang, G., Millings, E.J., Okafor, C.D., Patel, A., Cato, M.L., Cornelison, J.L., Melchers, D., Houtman, R., Moore, D.D., Calvert, J.W., Jui, N.T., Ortlund, E.A. Tapping into a phospholipid-LRH-1 axis yields a powerful anti-inflammatory agent with in vivo activity against colitis. BioRXIV. 2020. Epub Sept 2 2020. doi: https://doi.org/10.1101/2020.09.01.278291.
21. 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):25281-91. Epub 2016/10/04. doi: 10.1074/jbc.M116.753541. PubMed PMID: 27694446; PMCID: PMC5207232.
22. 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, Ingraham HA. Structural analyses reveal phosphatidyl inositols as ligands for the NR5 orphan receptors SF-1 and LRH-1. Cell. 2005;120(3):343-55. Epub 2005/02/15. doi: 10.1016/j.cell.2005.01.024. PubMed PMID: 15707893.
23. Li Y, Choi M, Cavey G, Daugherty J, Suino K, Kovach A, Bingham NC, Kliewer SA, Xu HE. Crystallographic identification and functional characterization of phospholipids as ligands for the orphan nuclear receptor steroidogenic factor-1. Mol Cell. 2005;17(4):491-502. Epub 2005/02/22. doi: 10.1016/j.molcel.2005.02.002. PubMed PMID: 15721253.
24. Otwinowski Z, Minor, W. Processing of X-Ray Diffraction Data Collected in Oscillation Mode. Methods in Enzymology. 1997;276:307-26.
25. 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, Zwart PH. PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta crystallographica Section D, Biological crystallography.

2010;66(Pt 2):213-21. Epub 2010/02/04. doi: 10.1107/S0907444909052925. PubMed PMID: 20124702; PMCID: 2815670.
26. Emsley P, Cowtan K. Coot: model-building tools for molecular graphics. Acta crystallographica Section D, Biological crystallography. 2004;60(Pt 12 Pt 1):2126-32. Epub 2004/12/02. doi: 10.1107/S0907444904019158. PubMed PMID: 15572765.
27. Schrodinger, LLC. The PyMOL Molecular Graphics System, Version 1.3r1. 2010.
28. Laskowski RA, Swindells MB. LigPlot+: multiple ligand-protein interaction diagrams for drug discovery. J Chem Inf Model. 2011;51(10):2778-86. doi: 10.1021/ci200227u. PubMed PMID: 21919503.

CHAPTER 5: DIMERIZATION OF STEROIDOGENIC-FACTOR 1

Emma H. D'Agostino,[*] Anamika Patel,[*] Michael L. Cato,[*] Eric A. Ortlund[*]

[*]Department of Biochemistry, Emory University, Atlanta, Georgia, 30322, United States

While our initial interest in SF-1 stemmed from its similarity to LRH-1 and our ability to modulate its activity with our library of small molecules, we were quickly pulled in a second direction. The NR5A receptors are canonically monomers – they are, in fact, the only nuclear receptors which act exclusively as monomers. However, our work with this receptor reveals that SF-1 may also dimerize. This story still has open questions, but we have now shown through several experiments that SF-1 acts as both a monomer and a dimer.

---

[a]E.H.D, A.P., and E.A.O. participated in research design. E.H.D., A.P, and M.L.C. performed experiments.

Abstract

Steroidogenic factor-1 (SF-1) is a nuclear receptor (NR) in the NR5A subfamily, one of the few subfamilies which binds DNA as monomers and the only which acts exclusively as monomers. The NR5A receptors harbor an extension in the DNA-binding domain (DBD), the Ftz-F1 box, which extends the canonical DNA recognition element by three base pairs and permits monomeric DNA binding. SF-1 is a critical regulator of steroidogenesis in the adrenal glands and gonads and metabolism in the hypothalamus. While this subfamily has not been previously reported to dimerize, we show here that SF-1 can dimerize *in vitro* and in cells, and that this dimerization is sensitive to a recently reported, high-affinity agonist. This work adds SF-1 to the list of NRs that act dually as monomers and dimers and demonstrates a need for further investigation into the functional significance of the dimer, as the oligomeric state of NRs often determines the direction of gene regulation.

Intro

Nuclear receptors (NRs) are a family of transcription factors containing 48 human members. The vast majority of NRs bind to DNA as dimers, with each subunit binding six base pairs of DNA.[1] There are exceptions to the general rule: the NRs in subfamily four were initially discovered to bind DNA as monomers with an eight-base pair response element (RE), and the glucocorticoid receptor was recently reported to function as a monomer when repressing gene transcription.[2, 3] There is also precedence for NRs to function in multiple oligomeric states; the NR4A family members, while initially reported to operate solely as monomers, were shortly thereafter discovered to function as type 2 NRs in concert with the retinoid X receptor (RXR), and then as homodimers.[4, 5] The glucocorticoid receptor, while canonically a dimer and, indeed, a founding member of the NR family and the dimer paradigm, functions as both a dimer and monomer in a context-dependent manner. There may be other NRs that function in multiple oligomeric states.

The NR5A receptors, steroidogenic factor-1 (SF-1, NR5A1) and liver receptor homolog-1 (LRH-1, NR5A2) are unique among NRs as they function only as monomers.[6] SF-1 is expressed in the adrenal glands, gonads, and hypothalamus whereas LRH-1 is expressed in the liver, pancreas, colon, gonads, and breast, and their roles are, broadly, to regulate steroidogenesis, development, and metabolism.[6] The NR5As contain a C-terminal extension of the conserved NR DNA-binding domain (DBD) termed the Ftz-F1 box, giving the ability to bind three additional base pairs of DNA and stabilizing monomeric DNA binding.[7] This monomeric binding was initially described through studies with the fruit fly, silkworm, and mouse NR5A orthologs.[7] The DBDs of the NR5A orthologs were confirmed to bind to the previously-reported NR5A RE, and no dimerization was detected by electrophoretic mobility shift assay for any receptor. Shortly thereafter, murine SF-1 was reported together with nerve growth factor 1B (NGF1-B), a member of the NR4A subfamily, as a monomeric NR.[3] The DBD was also used for these studies, and another gel shift assay showed no dimerization. Numerous crystal and NMR structures of the individual NR5A DNA and ligand-binding domains (LBDs) do not show dimerization, although the SF-1 LBD structures tend to contain two copies in the asymmetric unit.[8-14] Individual NR domains often do not dimerize even for well-characterized dimeric receptors; unsurprisingly, there is no literature describing the behavior of the full-length NR5A receptors *in vitro*, as this work is only in its early stages in the NR field as a whole. Limited in-cell ChIP-seq experiments have also identified the monomeric nonamer consensus response element.[15, 16] However, these experiments were not performed with NR5A ligands, and the SF-1 RE search was structured to search for a 9-base pair RE, and thus would not have found a dimeric RE which would be at least 18 base pairs long.

We conducted a large-scale screen to identify NR-NR and NR-lipid transport protein interactions and, given the widely accepted view of the NR5A receptors as monomers, were surprised to see SF-1-SF-1 as one of the strongest interactions in the entire screen. Here, we show

that SF-1 dimerizes in cells and in several assays *in vitro*, as both the LBD alone and full-length protein, and that this dimerization is modified by synthetic agonists. This discovery introduces a new paradigm to the NR5A family and uncovers new questions about the mechanism of action of NR gene regulation.

Results

Split luciferase assay reveals SF-1 homodimer. To identify NR-NR and NR-lipid transport protein interactions in an unbiased manner, we conducted a large protein-protein interaction screen using a split nanoluciferase system (Figure 5.1a). In this assay, each potential interaction partner is fused to one half of the nanoluciferase protein. If two interaction partners come within 10 nm, the nanoluciferase halves form a functional enzyme, allowing readout of a quantifiable luciferase signal. Unexpectedly, this screen identified the SF-1 homodimer as one of its strongest interactions at two different transfection concentrations (Figure 5.1b). Out of every SF-1 interaction, the SF-1 homodimer was by far the strongest interaction as well (Figure 5.1c). Because this screen only identifies interactions, but not their functional significance, we chose to further interrogate the SF-1-SF-1 homodimer.

The SF-1 LBD dimerizes in a ligand-dependent manner. We first examined the SF-1 LBD for dimerization. In our experience with NRs, individual purified domains do not typically dimerize. SF-1 purified from *E. coli* is bound to a mixture of bacterial phospholipids and appears as a single peak at the expected elution volume for a monomer via size exclusion chromatography (Figure 5.2a). Similarly, with a moderate SF-1 synthetic agonist, 6N, SF-1 elutes as a single peak at the monomer elution volume with a slight shoulder (Figure 5.2b-c).[17] However, with a higher-affinity agonist, 10CA, SF-1 elutes with two peaks, one at the dimer elution volume and one at the monomer elution volume (Figure 5.2b, d) (see Chapter 4). The oligomeric states from the size exclusion chromatography were confirmed with analytical ultracentrifugation (Figure 5.2d). The unusual presence of a dimer with a

purified LBD suggested the possibility of a strong dimer in agreement with the nanoluciferase screen and encouraged us to pursue further studies.

Full-length SF-1 dimerizes in vitro. To more completely understand the behavior of the SF-1-SF-1 homodimer, we sought to purify the full-length receptor (FL-SF-1) for further study (Figure 5.3). Full-length NRs are challenging to purify, but to understand the full mechanism of dimerization, working with individual domains is inadequate. Our lab has recently made great progress in full-length NR receptor expression and purification. Expression of FL-SF-1 in *E. coli* is modest, but adequate for purification (Figure 5.3g). The purification scheme involves a nickel column (Figure 5.3a) followed by cleavage of the His-SUMO tag. To strip DNA content from the receptor, a tandem ion exchange strategy is used, with a Q column to capture DNA/SUMO tag (Figure 5.3b) and an SP column to capture FL-SF-1 (Figure 5.3c). FL-SF-1 can be challenging to purify by size exclusion chromatography, which is an area for future focus (Figure 5.3d-f). However, despite significant degradation, protein with >95% purity can be obtained from the SP column for *in vitro* assays (Figure 5.3e).

To confirm the activity of FL-SF-1, we used fluorescence polarization assays to test ligand and DNA binding (Figure 5.4). FL-SF-1 bound to the 6N-FAM ligand with single-digit nanomolar affinity, a full order of magnitude tighter than the LBD alone (Figure 5.4a).[18] FL-SF-1 also bound to a single copy of the *CYP7A1* response element with single-digit nanomolar affinity (Figure 5.4b). Surprisingly, the DNA-binding assay illustrated a two-site binding curve, indicative of dimerization. The response element in this assay is only 12 base pairs; the SF-1 DNA-binding domain requires 9 base pairs of DNA, so it is unlikely that two copies of SF-1 are bound on the same DNA molecule. However, the FL-SF-1 is likely dimerizing, perhaps induced by the presence of DNA, and further study is warranted.

Ligand disrupts SF-1-SF-1 interaction in cells. Given the ligand-dependence of the LBD dimerization, we returned to the split nanoluciferase assay to interrogate the effect of ligand on the SF-1 homodimer in cells. We used 10CA, which induced the LBD dimer, and the 6N-10CA agonist described in Chapter 4 as it is the best SF-1 agonist to date. Although we hypothesized that these strong agonists would enhance the in-cell interaction, they had the opposite effect, strongly disrupting the SF-SF-1 dimer (Figure 5.5). While this result appears to be in direct opposition to the LBD reliance upon ligand for dimerization, SF-1 likely has access to its endogenous phospholipid ligand in mammalian cells, and this ligand may induce the dimer more strongly than our synthetic agonists. This result emphasizes the importance of FL-SF-1 and in-cell studies to capture the impact of DBD allostery and interactions with endogenous ligands and coregulator proteins.

Discussion

SF-1 was initially reported together with Nurr77, a member of the NR4A family, as a monomeric receptor. The NR4A family was shortly thereafter reported to act as homo- and heterodimers on DNA, and with this work we add SF-1 to the list of NRs which forms both a monomer and a dimer. We first identified the SF-1-SF-1 dimer in a large, unbiased NR-lipid transport protein screen, where SF-1-SF-1 was one of the strongest interactions in the entire screen (Figure 5.1). SF-1 interacted more strongly with itself than with any other protein in the screen. SF-1 also interacted with several other NRs and lipid transport proteins in this screen, and while not all are likely to be functionally relevant (for instance, SF-1 and LRH-1 are not coexpressed *in vivo*) there are many potential avenues for future work. We then sought to verify this surprising interaction *in vitro*. We showed that the LBD purifies as a monomer in the presence of bacterial phospholipids and a weak agonist, but that a strong agonist induces a dimer (Figure 5.2). The LBD also may have some dimer content even without a strong agonist, as the SEC curves show a slight shoulder; further AUC studies will clarify

this possibility. Future studies with the LBD will focus on identifying the dimerization interface and determining whether the interface is consistent across the LBD, FL-SF-1, and in cells.

We successfully purified FL-SF-1 for the first time; this purified full-length NR is a powerful tool for *in vitro* studies, as the individual domains may not fully recapitulate activity (Figure 5.3). We show that FL-SF-1 is pure and active and binds ligand and DNA with single-digit nanomolar affinity (Figure 5.4). We also show that FL-SF-1 does not dimerize in the presence of our 6N-FAM probe (consistent with the LBD), but that it forms a two-site binding curve on DNA, indicative of dimerization (Figure 5.4). Interestingly, the DBD alone does not exhibit a two-site binding curve, highlighting the need to study FL-SF-1 rather than individual domains and the importance of allostery between NR domains (data not shown). There is still much to understand about the behavior of FL-SF-1, including the effects of various ligands, characterization of the dimer interface, and the effects of different sequences and lengths of DNA.

Finally, we returned to the split nanoluciferase assay used in the initial screen which identified the SF-1 homodimer to test the effect of ligand on dimer formation. Contrary to our results with purified LBD in which ligand induced dimer formation, strong agonists disrupted the SF-1 homodimer in cells (Figure 5.5). While this result was unexpected, behavior in cells is often different from the behavior of purified protein. Agonists could induce DNA binding, and the subsequent allosteric effects could disrupt the dimer. Agonists likely lead to the recruitment of coregulator proteins, which could sterically clash with dimer formation. Additional studies will be needed to understand the functional consequences of SF-1 homodimer formation.

There are many potential future directions of this work, which will be further explored in Chapter 6. They include 1) identifying the dimerization interface, 2) determining whether dimerization is occurring on DNA and, if so, on which response elements, 3) determining whether dimerization is necessary for transcriptional activation in cells. The oligomeric state of SF-1 may also imply a novel

function, analogous to the dimer *versus* monomer of GR determining activation or repression of transcription.[2] SF-1 has thus far only been shown to activate transcription, and the dimer may repress transcription or regulate an as-yet-undiscovered set of genes. The experiments here have shown across multiple systems and methods that SF-1 is not exclusively a monomer as previously reported, and that further experimentation is warranted to fully understand its complete mechanism of transcriptional regulation.

Methods

Split luciferase screen. HEK 293T cells were purchased from Atlantic-Type Culture Collection and cultured in minimal essential media without phenol red (Gibco) with the addition of 10 % fetal bovine serum. Cells were maintained under standard culture conditions. Cells were plated in a 1536-well, white, solid-bottom plate and transfected with either 0.5 or 5 ng of each component of the split luciferase system, that is, a nuclear receptor or lipid transport protein fused at the N-terminus to half of the Nanoluciferase enzyme, using linear polyethylenimine (PEI) at a total DNA:PEI ratio of 1:3. After 48 hours, furimazine (NanoGlo; Promega, Madison, WI) was directly added to the cells and luminescence was measure immediately using an Envision Multilabel plate reader (PerkinElmer, Waltham, MA). Protein-protein interaction (PPI) scores were reported as a ratio between PPI luminescence divided by empty vector pair luminescence. Assays were performed in quadruplicate.

Protein Purification: Wildtype LBD. SF-1 LBD (amino acids 218-461) in the pLIC-His vector was used to transform *E. coli* strain BL21(DE3) and cultured in 6 L Lysogeny Broth media in the presence of ampicillin at 37 °C to an OD_{600} of 0.6. Protein expression was induced with 0.5 mM isopropyl-1-thio-D-galactopyranoside (IPTG) for 4 h at 32 °C. Cell pellets were lysed in buffer containing 20 mM Tris HCl pH 7.4, 500 mM NaCl, 0.5 mM TCEP, 5% glycerol, lysosome, DNAse, and phenylmethylsulfonyl fluoride (PMSF) followed by sonication (2 minutes, with 1 second on/ off intervals). Lysate was clarified by centrifugation for 45 min at 16,000 x g in a Sorvall RC 6+

centrifuge. The lysate was flowed over a 5 mL HisTrap FF column (GE Healthcare, Little Chalfront, UK) with Buffer A (20 mM Tris pH 7.4, 500 mM NaCl, 0.5 mM TCEP, 5% glycerol, 25 mM imidazole) and SF-1 was eluted with Buffer B (20 mM Tris pH 7.4, 500 mM NaCl, 0.5 mM TCEP, 5% glycerol, 500 mM imidazole). To homogenize the lipid population, SF-1 was incubated overnight with dilauroylphosphatidylcholine (DLPC). Size exclusion chromatography (SEC) into assay buffer (150 mM NaCl, 20 mM Tris HCl pH 7.4, 5% glycerol) was used as a final purification step; protein was concentrated to ~3 mg/mL, flash frozen, and stored at -80 °C for use in fluorescence polarization assays.

Purification – CysLite SF-1. CysLite SF-1 (amino acids 218-461, C30S, C195S) in the pLic-His vector was used for dimerization assays to eliminate the possibility of covalent bonding. This protein was purified as described for wildtype SF-1 through the HisTrap column; after elution from the HisTrap column, the 6X-His tag was cleaved overnight using tobacco etch virus protease. Cleaved protein was flowed over a second HisTrap column and the flowthrough was collected, concentrated to ~3 mg/mL, flash frozen, and stored at -80 °C for use in assays.

Protein Purification: Full-Length. Full-length SF-1 in the PSMT3 vector was used to transform *E. coli* strain BL21(DE3)-pLysS and cultured Terrific Broth. Protein expression was induced with 1 mM IPTG for 4 h at 32 °C. Cell pellets were lysed in buffer containing 20 mM Tris pH 7.4, 750 mM NaCl, 0.5 mM TCEP, 5% glycerol, 1 uM $ZnCl_2$, 5 mM CHAPS lysozyme, DNAse, PMSF, and a protease inhibitor tablet (Pierce). To further lyse cells, a homogenizer followed by sonication (2 minutes, with 1 second on/ off intervals) was used. The lysate was flowed over a 5 mL HisTrap FF column (GE Healthcare, Little Chalfront, UK) with Buffer A (20 mM Tris pH 7.4, 500 mM NaCl, 0.5 mM TCEP, 5% glycerol, 1 uM $ZnCl_2$) and FL-SF-1 was eluted with Buffer B (20 mM Tris pH 7.4, 500 mM NaCl, 0.5 mM TCEP, 5% glycerol, 500 mM imidazole, 1 uM $ZnCl_2$). The His-SUMO tag was cleaved overnight in dialysis Buffer C (20 mM Tris pH 7.4, 500 mM NaCl, 0.5 mM TCEP,

5% glycerol, 1 uM ZnCl2, PMSF). To strip DNA from FL-SF-1, a tandem ion exchange approach was used: the salt concentration was quickly diluted to 250 mM, and the protein-DNA complex was flowed over connected 5-mL Q (to strip DNA content) and SP (to capture protein) columns. The columns were then eluted separately with Buffer D (20 mM Tris pH 7.4, 1 M NaCl, 0.5 mM TCEP, 5% glycerol, 1 uM $ZnCl_2$) using a linear gradient of 250 – 650 mM NaCl over 30 CVs, with a final step to 1 M NaCl for 5 CVs. Captured protein could then be frozen into aliquots and stored at -80 ºC, or complexed with DNA/ ligand in dialysis Buffer E (20 mM Tris pH 7.4, 250 M NaCl, 0.5 mM TCEP, 5% glycerol, 1 uM $ZnCl_2$) using 3,500 MWCO tubing to ensure binding followed by size exclusion chromatography in Buffer F (20 mM Tris pH 7.4, 250 M NaCl, 0.5 mM TCEP, 5% glycerol, 1 uM $ZnCl_2$) as a final purification step.

Fluorescence Polarization. All assays were conducted in black, polystyrene, non-binding surface 384-well plates (Corning Inc., Corning, NY) with 30 μL volumes in assay buffer. Binding affinity for 6N-FAM was determined using 10 nM 6N-FAM and protein concentrations ranging from 1^{-11}–5^{-5} M. Binding affinity for the 12-bp CYP7A1 response element (5'-GTTCAAGGCCAG-3') was determined using 10 nM CYP7A1-FAM and protein concentrations ranging from 1^{-11}–5^{-5} M. Plates were incubated overnight at 4 °C and centrifuged at 2,000 × *g* for 2 minutes before polarization measurement. Polarization was monitored on a Neo plate reader (Biotek, Winooski, VT) at an excitation/emission wavelength of 485/528 nm. Nine technical replicates were conducted over three experiments and compiled binding data were baseline-corrected to wells with no protein and fit with a one-site or two-site binding curve in GraphPad Prism version 7 (GraphPad, Inc., La Jolla, CA). Competition assays were performed in accordance with development guidelines.[19] For SF-1 LBD, 10 nM 6N-FAM (0.8 times the affinity of SF-1 for 6N-FAM) and 25 nM SF-1 (60% of the forward binding B_{max}) were used. Competitor ligand concentration ranged from 2^{-11}-2^{-4} M, and competitor ligand volume was kept constant to maintain constant DMSO in each well (6.7% v/v). Eight technical

replicates were performed over two experiments, and GraphPad Prism version 7 was used to analyze compiled data using a one-site, fit K_i curve, with normalization to 6N competition.

Analytical Ultracentrifugation. Analytical ultracentrifugation experiments were carried out as previously described using a Beckman Coulter ProteomeLabTM XL-A analytical ultracentrifuge equipped with absorbance optics and an eight-hole An-50 Ti analytical rotor.20 Sedimentation velocity experiments were carried out at 10 °C and 50,000 rpm (200,000 × g) using 3-mm two-sector charcoal-filled Epon centerpieces with quartz windows. Each sample was scanned at 0-min time intervals for 300 scans. Protein samples were run at 0.2 mg/ mL. Sedimentation boundaries were analyzed by the continuous sedimentation coefficient distribution (c(s)) method using the program SEDFIT.21

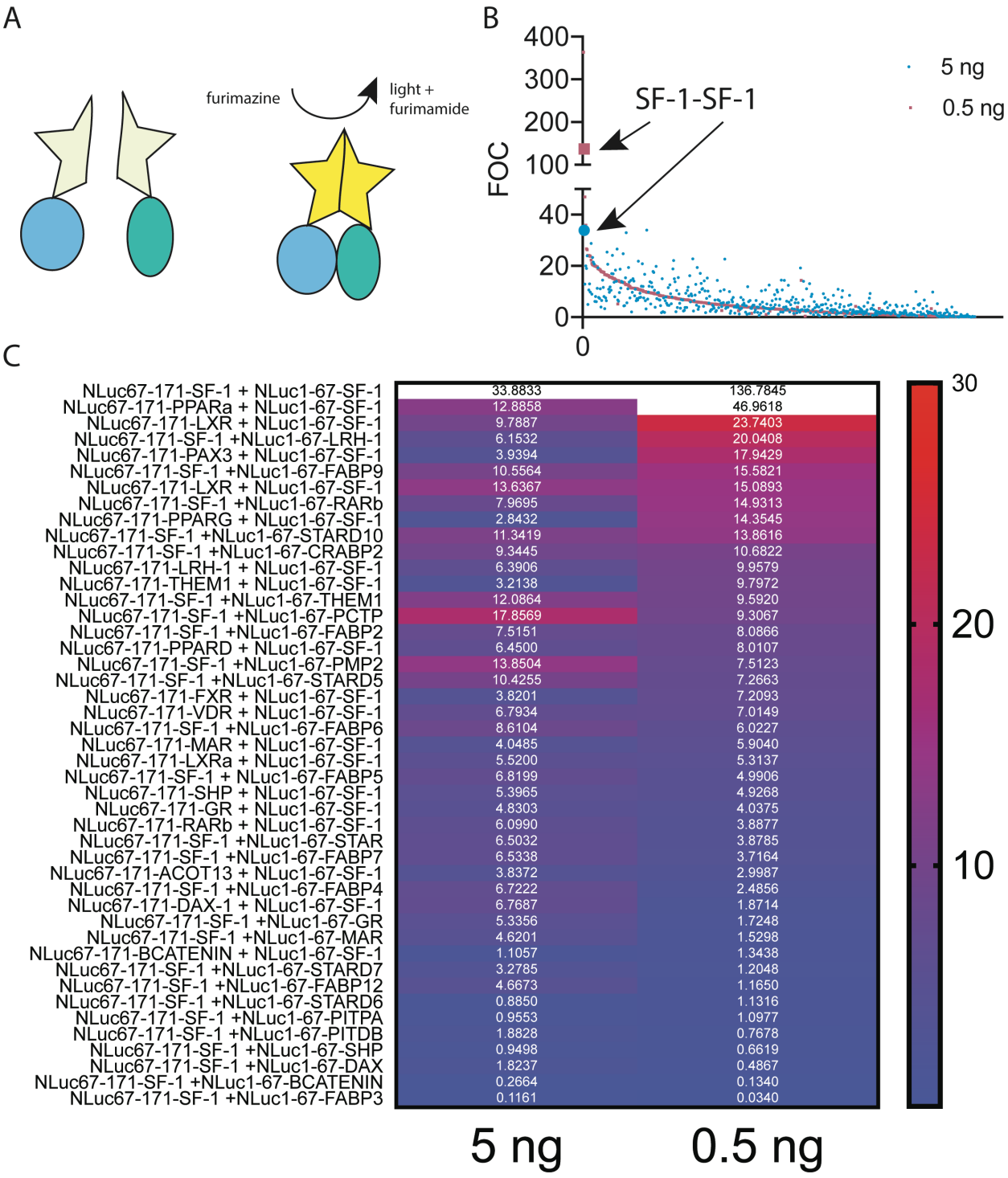


Figure 5.1 Split luciferase screen. (A) In the split luciferase screen, the nanoluciferase enzyme is split in half and each half is attached to the N-terminus of a potential interaction partner. If the partners interact with 10 nm, the enzyme comes together and is able to convert substrate into a fluorescent

signal for quantitative readout. (B) Results of the full NR/ LTP screen; in the entire screen, the SF-1-SF-1 interaction was one of the top hits for both the 5 ng and 0.5 ng transfection amounts. "FOC" is fold over control; each dot represents a NR/ LTP potential interaction partner. (C) Heat map of SF-1-specific screen results. The SF-1-SF-1 interaction was the top result; to better illustrate gradation, values >30 are excluded from heat map coloration. Values are fold over control, and all values > 5 are considered relevant.

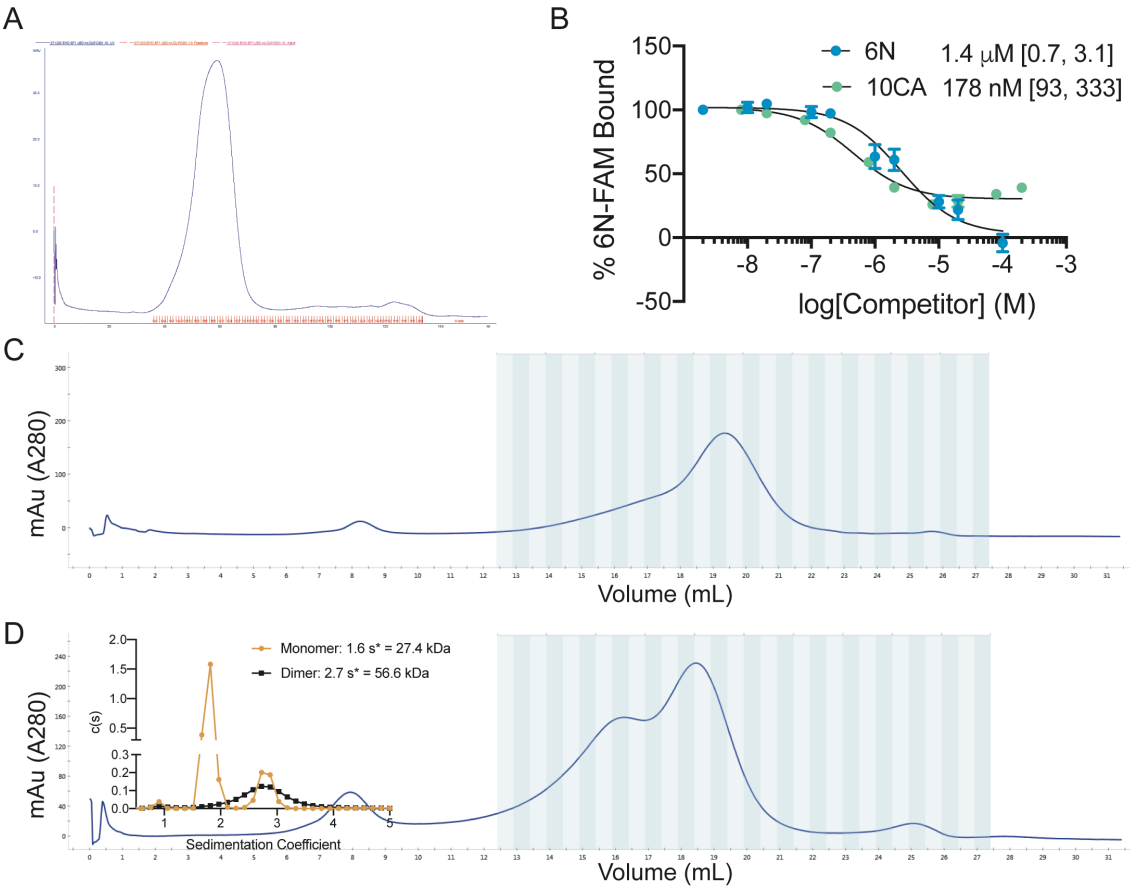


Figure 5.2 SF-1 LBD dimerizes in a ligand-driven manner. (A) SEC chromatogram of SF-1 purified from *E. coli* shows a single peak at the expected elution volume of ~65 mL for a monomer. (B) Fluorescence polarization competition assay reveals that 10CA binds with nearly 100-fold higher affinity than 6N. Values are K_I, with 95% confidence intervals in brackets. (C) SEC chromatogram of SF-1 LBD in complex with 6N agonist shows a single, monomeric peak with expected elution

volume of ~19 mL. (D) SEC chromatogram of SF-1 LBD in complex with 10CA agonist shows two peaks, at ~16 and ~19 mL, indicative of potential dimer induction. *Inset*: Analytical ultracentrifugation result from the putative dimer and monomer peaks in confirms presence of dimer formation driven by 10CA.

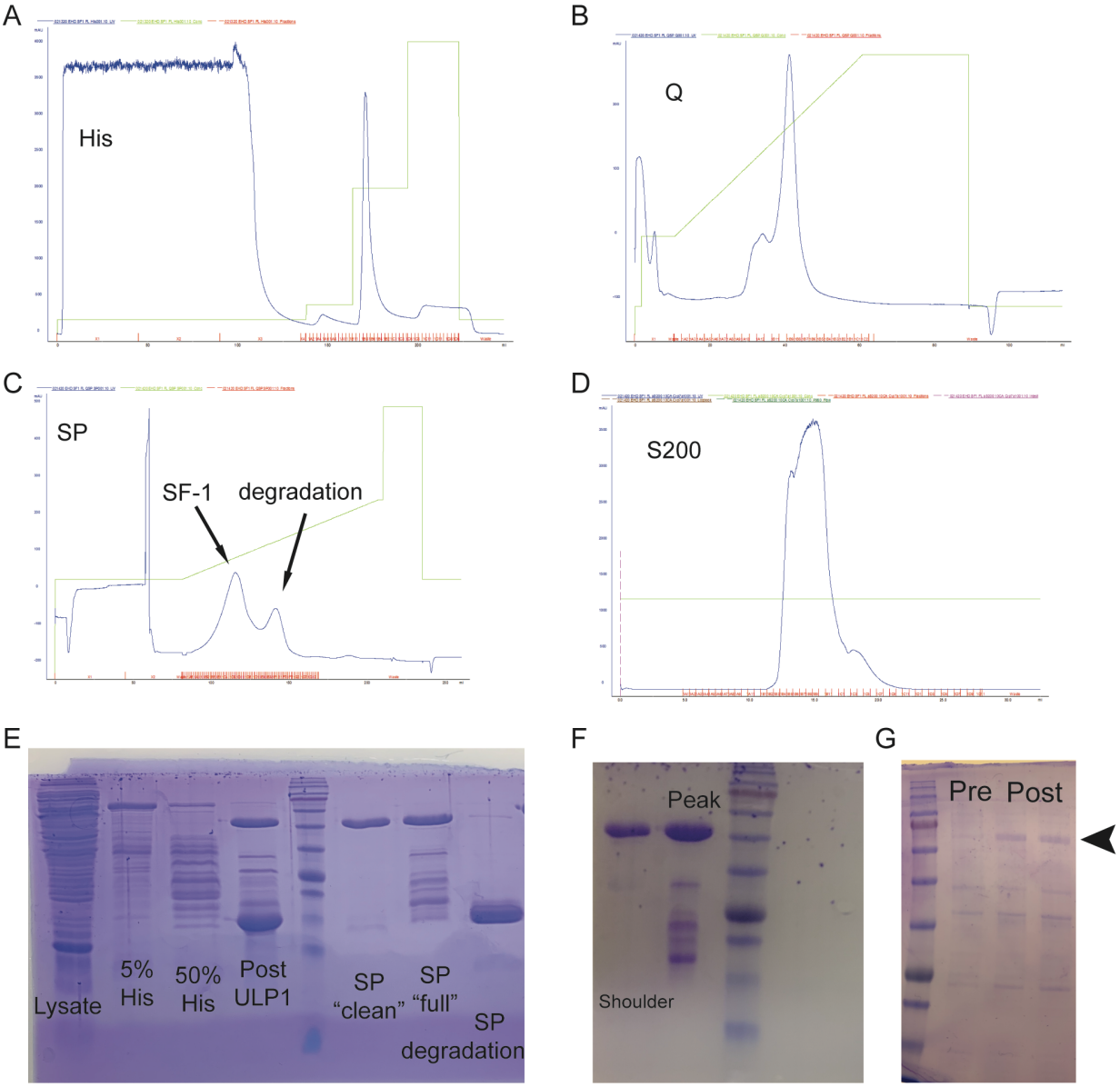


Figure 5.3 SF-1-FL purification. (A) His column, with ~50 mg yield at 50% peak. (B) Q column elution of DNA after tandem ion exchange. (C) SP column elution of pure SF-1-FL (left) and degradation product (right) peaks after tandem ion exchange to strip DNA. (D) Sizing column with SF-1

complexed with DNA; pure SF-1 eluted as a shoulder, and the majority degraded and eluted as a broad peak with high absorbance. (E) First SDS-PAGE gel showing lysate before purification; first peak from the His column; second peak from the His column; results from ULP1 His-SUMO tag cleavage; ladder; SP "clean" meaning a conservative set of fractions taken from the left peak of the SP column to ensure clean protein; SP "full" meaning the rest of the left peak of the SP column; and the second peak from the SP column, which is degraded SF-1. (F) SEC shows that most of the protein degraded, and the only clean protein was a shoulder on the main peak. (G) Results from the expression; ladder, pre-induction, and post-induction from two separate liters of growths. Expression band indicated with arrow.

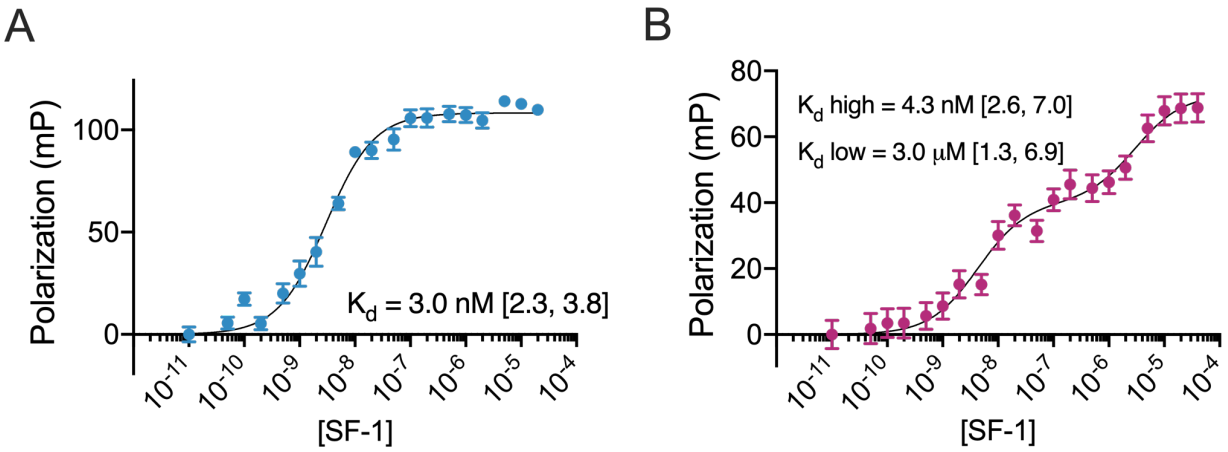


Figure 5.4. FL-SF-1 dimerizes in vitro on DNA. (A) Fluorescence polarization ligand binding assay confirms functional ligand binding. (B) Fluorescence polarization DNA binding to single *CYP7A1* response element confirms functional DNA binding. Two-site curve indicates dimerization. Assays were conducted thrice in triplicate. K_d values are shown with 95% confidence intervals in brackets; error bars are SEM.

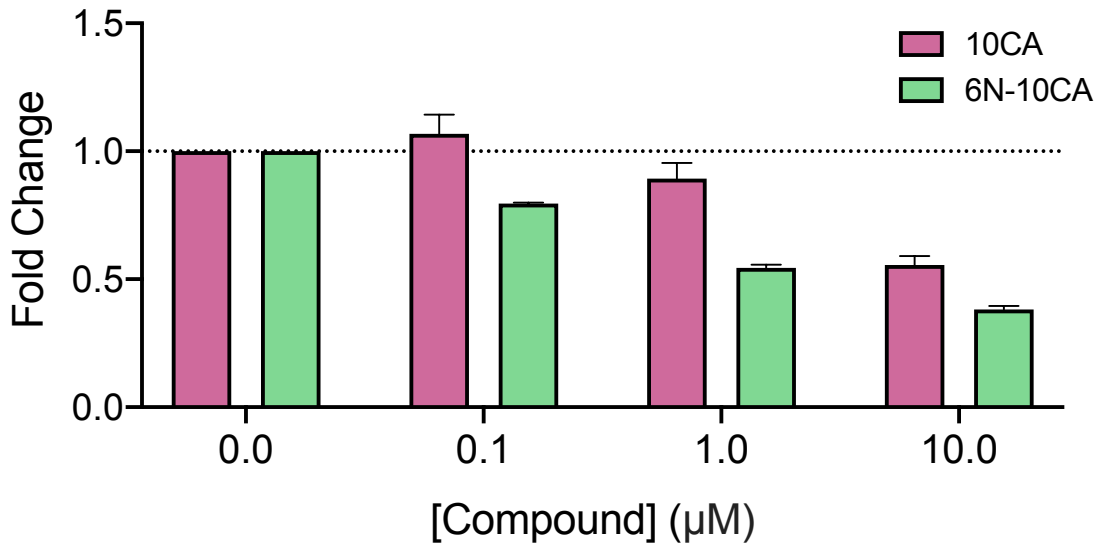


Figure 5.5. Ligand disrupts dimerization in cells. In the split nanoluciferase assay using FL-SF-1 in cells, both 10CA and 6N-10CA disrupt the SF-1-SF-1 interaction. 6N-10CA disrupts the dimer at a lower concentration than 10CA. Assay was conducted twice in triplicate; error shown is SEM.

References

1. Weikum ER, Liu X, Ortlund EA. The nuclear receptor superfamily: A structural perspective. Protein Sci. 2018;27(11):1876-92. Epub 2018/08/16. doi: 10.1002/pro.3496. PubMed PMID: 30109749; PMCID: PMC6201731.
2. Hudson WH, Youn C, Ortlund EA. The structural basis of direct glucocorticoid-mediated transrepression. Nat Struct Mol Biol. 2013;20(1):53-8. Epub 2012/12/12. doi: 10.1038/nsmb.2456. PubMed PMID: 23222642; PMCID: PMC3539207.
3. Wilson TE, Fahrner TJ, Milbrandt J. The orphan receptors NGFI-B and steroidogenic factor 1 establish monomer binding as a third paradigm of nuclear receptor-DNA interaction. Mol Cell Biol. 1993;13(9):5794-804. Epub 1993/09/01. doi: 10.1128/mcb.13.9.5794. PubMed PMID: 8395013; PMCID: PMC360322.
4. Perlmann T, Jansson L. A novel pathway for vitamin A signaling mediated by RXR heterodimerization with NGFI-B and NURR1. Genes Dev. 1995;9(7):769-82. Epub 1995/04/01. doi: 10.1101/gad.9.7.769. PubMed PMID: 7705655.
5. Philips A, Lesage S, Gingras R, Maira MH, Gauthier Y, Hugo P, Drouin J. Novel dimeric Nur77 signaling mechanism in endocrine and lymphoid cells. Mol Cell Biol. 1997;17(10):5946-51. Epub 1997/10/07. doi: 10.1128/mcb.17.10.5946. PubMed PMID: 9315652; PMCID: PMC232442.
6. Meinsohn MC, Smith OE, Bertolin K, Murphy BD. The Orphan Nuclear Receptors Steroidogenic Factor-1 and Liver Receptor Homolog-1: Structure, Regulation, and Essential Roles in Mammalian Reproduction. Physiol Rev. 2019;99(2):1249-79. Epub 2019/02/28. doi: 10.1152/physrev.00019.2018. PubMed PMID: 30810078.
7. Ueda H, Sun GC, Murata T, Hirose S. A novel DNA-binding motif abuts the zinc finger domain of insect nuclear hormone receptor FTZ-F1 and mouse embryonal long terminal repeat-

binding protein. Mol Cell Biol. 1992;12(12):5667-72. Epub 1992/12/01. doi: 10.1128/mcb.12.12.5667. PubMed PMID: 1448096; PMCID: PMC360506.
8. Sablin EP, Blind RD, Krylova IN, Ingraham JG, Cai F, Williams JD, Fletterick RJ, Ingraham HA. Structure of SF-1 bound by different phospholipids: evidence for regulatory ligands. Mol Endocrinol. 2009;23(1):25-34. Epub 2008/11/08. doi: 10.1210/me.2007-0508. PubMed PMID: 18988706; PMCID: PMC2646595.
9. Little TH, Zhang Y, Matulis CK, Weck J, Zhang Z, Ramachandran A, Mayo KE, Radhakrishnan I. Sequence-specific deoxyribonucleic acid (DNA) recognition by steroidogenic factor 1: a helix at the carboxy terminus of the DNA binding domain is necessary for complex stability. Mol Endocrinol. 2006;20(4):831-43. Epub 2005/12/13. doi: 10.1210/me.2005-0384. PubMed PMID: 16339274.
10. Wang W, Zhang C, Marimuthu A, Krupka HI, Tabrizizad M, Shelloe R, Mehra U, Eng K, Nguyen H, Settachatgul C, Powell B, Milburn MV, West BL. The crystal structures of human steroidogenic factor-1 and liver receptor homologue-1. Proc Natl Acad Sci U S A. 2005;102(21):7505-10. Epub 2005/05/18. doi: 10.1073/pnas.0409482102. PubMed PMID: 15897460; PMCID: PMC1140416.
11. Solomon IH, Hager JM, Safi R, McDonnell DP, Redinbo MR, Ortlund EA. Crystal structure of the human LRH-1 DBD-DNA complex reveals Ftz-F1 domain positioning is required for receptor activity. J Mol Biol. 2005;354(5):1091-102. Epub 2005/11/18. doi: 10.1016/j.jmb.2005.10.009. PubMed PMID: 16289203.
12. Li Y, Choi M, Suino K, Kovach A, Daugherty J, Kliewer SA, Xu HE. Structural and biochemical basis for selective repression of the orphan nuclear receptor liver receptor homolog 1 by small heterodimer partner. Proc Natl Acad Sci U S A. 2005;102(27):9505-10. Epub 2005/06/25. doi: 10.1073/pnas.0501204102. PubMed PMID: 15976031; PMCID: PMC1157103.
13. Li Y, Choi M, Cavey G, Daugherty J, Suino K, Kovach A, Bingham NC, Kliewer SA, Xu HE. Crystallographic identification and functional characterization of phospholipids as ligands for the orphan nuclear receptor steroidogenic factor-1. Mol Cell. 2005;17(4):491-502. Epub 2005/02/22. doi: 10.1016/j.molcel.2005.02.002. PubMed PMID: 15721253.
14. 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, Ingraham HA. Structural analyses reveal phosphatidyl inositols as ligands for the NR5 orphan receptors SF-1 and LRH-1. Cell. 2005;120(3):343-55. Epub 2005/02/15. doi: 10.1016/j.cell.2005.01.024. PubMed PMID: 15707893.
15. Doghman M, Figueiredo BC, Volante M, Papotti M, Lalli E. Integrative analysis of SF-1 transcription factor dosage impact on genome-wide binding and gene expression regulation. Nucleic Acids Res. 2013;41(19):8896-907. Epub 2013/08/03. doi: 10.1093/nar/gkt658. PubMed PMID: 23907384; PMCID: PMC3799431.
16. Chong HK, Biesinger J, Seo YK, Xie X, Osborne TF. Genome-wide analysis of hepatic LRH-1 reveals a promoter binding preference and suggests a role in regulating genes of lipid metabolism in concert with FXR. BMC Genomics. 2012;13:51. Epub 2012/02/03. doi: 10.1186/1471-2164-13-51. PubMed PMID: 22296850; PMCID: PMC3295688.
17. Mays SG, Flynn AR, Cornelison JL, Okafor CD, Wang H, Wang G, Huang X, Donaldson HN, Millings EJ, Polavarapu R, Moore DD, Calvert JW, Jui NT, Ortlund EA. Development of the First Low Nanomolar Liver Receptor Homolog-1 Agonist through Structure-guided Design. J Med Chem. 2019;62(24):11022-34. Epub 2019/08/17. doi: 10.1021/acs.jmedchem.9b00753. PubMed PMID: 31419141.
18. D'Agostino EH, Flynn AR, Cornelison JL, Mays SG, Patel A, Jui NT, Ortlund EA. Development of a Versatile and Sensitive Direct Ligand Binding Assay for Human NR5A Nuclear

Receptors. ACS Med Chem Lett. 2020;11(3):365-70. Epub 2020/03/19. doi: 10.1021/acsmedchemlett.9b00442. PubMed PMID: 32184971; PMCID: PMC7074214.
19. Huang X. Fluorescence polarization competition assay: the range of resolvable inhibitor potency is limited by the affinity of the fluorescent ligand. J Biomol Screen. 2003;8(1):34-8. Epub 2003/07/12. doi: 10.1177/1087057102239666. PubMed PMID: 12854996.
20. Patel A, Dharmarajan V, Vought VE, Cosgrove MS. On the mechanism of multiple lysine methylation by the human mixed lineage leukemia protein-1 (MLL1) core complex. J Biol Chem. 2009;284(36):24242-56. Epub 2009/06/27. doi: 10.1074/jbc.M109.014498. PubMed PMID: 19556245; PMCID: PMC2782018.
21. Schuck P. Size-distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and lamm equation modeling. Biophys J. 2000;78(3):1606-19. Epub 2000/02/29. doi: 10.1016/S0006-3495(00)76713-0. PubMed PMID: 10692345; PMCID: PMC1300758.

CHAPTER 6: CONCLUSIONS AND FUTURE DIRECTIONS

Conclusions

LRH-1

Liver receptor homolog-1 has long been recognized as a promising therapeutic target in a number of disease states, including type 1 and 2 diabetes, inflammatory bowel diseases, and obesity. However, targeting the ligand-binding pocket (LBP) of this receptor has been more challenging than expected. Recombinant LRH-1 purified from *E. coli* is contaminated with phospholipids, which have been difficult to displace in binding assays and screens. Removing these phospholipids requires unfolding and refolding the protein, which is difficult and results in significant loss of protein. The LBP is almost entirely hydrophobic, complicating finding a scaffold which can access the few polar residues available for interaction. Because of these difficulties, small molecules screens have only yielded one agonist scaffold which binds directly in the LBP, the RJW100 scaffold which we have further developed in this work. In Chapter 2, we describe our development of one of the only direct binding assays available for LRH-1, enabled by our discovery of a potent synthetic agonist. The development of a direct ligand-binding assay that is simple to perform has greatly accelerated our ability to evaluate novel small molecules. We previously relied on a thermal shift assay which determined the effect of our agonists on the global stability of LRH-1. The degree of stability conferred by an agonist generally correlates with in-cell activity; however, the thermal shift assay does not indicate ligand binding in the ligand-binding pocket or allow the calculation of binding affinity. With the fluorescence polarization competition assay, we can determine whether affinity correlates to in-cell activity and rapidly assess new compounds. We also can now easily compare efficacy across diverse small molecules classes, including phospholipids and agonist and antagonist scaffolds, and confirm binding in the LBP. This represents a significant advancement relative to the mass spectrometry studies which were previously required to detect competitive ligand binding.

The progression of LRH-1 agonists from RJW100 to 10CA, described in Chapter 3, is a significant leap forward. We have advanced from a small molecule with modest activity which did not show *in vivo* activation of LRH-1 to a high-affinity, reasonably potent molecule which activates LRH-1 *in vivo*. The robust structure-activity relationship we have established through the dozens of compounds we have studied in complex with LRH-1, including over half a dozen solved crystal structures and extensive biochemical characterization, has positioned us to further improve our agonists and study them in a variety of *in vivo* contexts as outlined below.

SF-1

The direct ligand binding assay in Chapter 2 was similarly significant for SF-1; few synthetic molecules have been reported for SF-1 and this assay will facilitate compound screening. The assay was immensely helpful for identifying high-affinity candidates for pursuing crystallization studies. One of the challenges in SF-1 small molecule development has been the lack of synthetic molecule-bound crystal structures. In Chapter 4, we report the first crystal structure of SF-1 bound to a synthetic agonist, 6N-10CA, which is also the best agonist reported for SF-1 to date. While there is still more to be learned from this structure (see below), it has already begun to answer long-standing questions in our lab about the comparative binding modes between the human NR5A receptors.

In Chapter 5, we address an established paradigm of the NR5A family. LRH-1, SF-1, and all known NR5A orthologs have been consistently reported in the literature as monomers on DNA, making them unique among NRs. Here, we show that SF-1 dimerizes *in vitro* and in cells and that the dimerization is sensitive to ligand. This dimerization creates a new paradigm in the NR family; there are many unanswered questions about the function of this homodimer which are addressed below.

Future Directions

LRH-1 Small Molecule Development

The discovery of the first agonist to activate LRH-1 *in vivo* in a colitis model represents a significant advancement in our agonist program. The agonist presented in Chapter 4, 10CA, still presents opportunities for improvement in terms of potency and pharmacokinetic properties. We have already developed a significantly more potent compound, 6N-10CA, briefly described in Chapter 5. With this compound in hand, we now turn to improving the solubility and pharmacokinetic properties of 6N-10CA. We are currently in the process of etherizing the alkyl "tail" to improve the partition coefficient (clogP) 3-log fold and eliminating the styrenyl group which is a metabolic liability (Figure 6.1); these modifications are not expected to impact efficacy.

We now have the opportunity to study LRH-1 activation *in vivo*, both in wildtype and disease models. Traditionally, LRH-1 murine studies have required the use of overexpression models, as the mouse ortholog ligand-binding pocket differs significantly from the human, generating a muted ligand-driven response. We have recently developed a humanized mouse using CRISPR editing, and the combination of this mouse model with our new agonists will enable several avenues of study. In wildtype animals, we can characterize our agonists to determine their gene regulation programs; we already have evidence that they drive the recruitment of unique sets of coregulator peptides, which may alter their transcriptional profiles.[1, 2] In models of disease, we are encouraged by the efficacy of 10CA in colitis and will pursue further studies in this area as there are multiple models of colitis, including the T-cell transfer, oxalazone-induced, and Il-10 knockout models. Given the *in vivo* action of 10CA, we would also like to test our small molecules in models of type 1 diabetes and in a high-fat diet induced model of obesity and type 2 diabetes.

SF-1 Small Molecule Studies

Our work with SF-1 has advanced our understanding of how SF-1 interacts with synthetic agonists, particularly in comparison to LRH-1. We have characterized SF-1 with our ~100-compound library

of small molecules and solved the first synthetic agonist-bound crystal structure. We are now well-positioned to leverage this information to expand our understanding of full mechanism of this receptor and develop SF-1-specific modulators.

Further interrogation of the 6N-10CA mechanism of action: we must determine whether 6N-10CA is functionally engaging the pocket mouth because the partial occupancy with DPPE in our structure may have biased this region of the electron density map. To accomplish this, we can introduce substitutions to critical phospholipid-coordinating residues at the pocket mouth, Y436F and K440A, and interrogate the ability of 6N-10CA to activate SF-1 in comparison to the wildtype receptor. A second region that is required for LRH-1 activation with 6N is the sulfamide interaction network. We have established with multiple agonists that T352 is a critical residue for LRH-1 activation, although it is not within hydrogen bonding distance of the sulfamide in the LRH-1 6N or 6N-10CA structures.[3, 4] Similarly, although the sulfamide group was within 3.4 Å of T272 in our SF-1 structure, it was not within hydrogen bonding distance.[3, 4] Thus, to determine whether interaction with SF-1 relies on T272 for activation, we can introduce a T272V substitution. We will also introduce a substitution to the nearby M268; this residue directly interacts with the sulfamide in both the SF-1 and LRH-1 structures, but was not shown to affect LRH-1 activation.[4] We will determine whether SF-1 uses the same residues as LRH-1 or has a different mechanism of activation using luciferase and binding assays. We will also use molecular dynamics simulations to examine the interactions between 6N-10CA and each of these critical residues, as well as the α-amino group of G341. Finally, we wish to further understand the effect of 6N-10CA on the function of SF-1. One measure of NR function is its ability to recruit coregulator proteins, which can be assessed through hydrogen deuterium exchange coupled to mass spectrometry (to determine the stability of the coregulator binding surface in response to ligand binding) and coregulator recruitment assays, including the Microarray Assay for Real-time Coregulator Nuclear receptor Interaction (MARCoNI) previously

employed by our lab. We also would like to pursue qPCR studies in a relevant SF-1 adrenal cell line to better understand the ability of 6N-10CA to affect the SF-1 transcriptional program. Together, these studies will firmly elucidate the mechanism of activation of the receptor.

We can also use the structural and dynamic knowledge gained to dissociate LRH-1 and SF-1 signaling with this agonist class. We now know that our agonist library is not LRH-1-specific and that SF-1 will need to be accounted for in LRH-1 agonist development, as SF-1 adrenal activation is expected to be detrimental. However, while our agonists almost universally activate both receptors, they do not perform equally for these close homologs. For instance, the sulfamide modification drives low nanomolar potency for LRH-1; for SF-1, this modification does not significantly improve affinity compared to RJW100.[5] The best agonist for both receptors, 6N-10CA, has ~300 times the affinity and ~4 times the potency for LRH-1 than SF-1, implying a difference in mechanism that is still unclear. One advantage to the general trend of greater efficacy towards LRH-1 than SF-1 with our agonist library is that with careful dosing, it may be possible to avoid off-target effects towards SF-1. However, this work has shown that we will need to test for SF-1 activation in LRH-1 studies. The experiments outlined above, including hydrogen-deuterium exchange and molecular dynamics simulations, may elucidate mechanistic differences and allow us to develop agonists that select for LRH-1.

The development of a potent SF-1 agonist offers the opportunity to illuminate SF-1's role in specific organ systems. LRH-1 and SF-1 are both expressed in the ovaries and testes, though they are not coexpressed, and their roles in individual cell types are not fully understood.[6] If we can successfully develop selective agonists, we may be able to parse out the different NR5A roles. SF-1 (and LRH-1) is capable of stem cell reprogramming, and a strong agonist may enhance its reprogramming ability.[7] A relatively uncharacterized region of SF-1 activity is the ventromedial nucleus of the hypothalamus (VMH), a critical region for feeding behavior and energy homeostasis.[8] While it is clear that SF-1 is

critical for function of this region, the precise transcriptional program has not been defined, and a potent agonist is an invaluable tool for defining the receptor's gene targets in the VMH.[9] We can use our agonists to study SF-1 in *ex vivo* VMH neurons, and if 6N-10CA can be modified to cross the blood-brain barrier and access the VMH, we can also characterize SF-1's ability to counteract metabolic disease.

Another key area for future SF-1 small molecule studies is in antagonist development. SF-1 is a defined oncogene in adrenocortical cancer, a rare and deadly disease for which novel treatments are critically needed.[10, 11] We have generated two strategies for inactivating LRH-1: traditional antagonists, with the goal of destabilizing helix 12, and PROTAC (proteolysis targeting chimeras) degraders, which recruit ubiquitin ligases and lead to receptor degradation. These strategies have shown early success and we are prepared to pursue full structure-activity relationship studies (unpublished). Given the similarity of LRH-1 and SF-1 and the cross-reactivity of our agonists, we believe that these inactivating compounds will also cross-react with SF-1. We plan to test the binding affinity of the small set of compounds with SF-1 as well as their ability to destabilize SF-1 in a thermal melting assay and decrease the receptor's transcriptional output by qPCR in a relevant adrenal cell line. We may encounter difficulties with antagonist development as our initial set of compounds rely on the sulfamide moiety to maintain affinity, but the ability to crystallize SF-1 will facilitate the progression of this work.

SF-1 Dimer

The discovery that SF-1 has the capacity to dimerize was a surprise that has left many open questions and future directions.

Our first goal will be to identify the dimerization interface. Given that the LBD appears to behave differently than FL-SF-1 in cells, it will be important to identify the interface for both purified LBD and FL-SF-1 and verify that the interface is the same for both constructs. A combination of

hydrogen-deuterium exchange and crosslinking mass spectrometry can be used to identify the interface(s), which may be tested by mutagenesis studies *in vitro* and in cells.

Studies with purified FL-SF-1 are still preliminary. Obtaining protein at >95% for *in vitro* assays is an important achievement, and now further optimization of the purification scheme will allow characterization of FL-SF-1 in complex with different ligands and DNA elements. In cells, FL-SF-1 appears to have an opposite response to ligand than the purified LBD (that is, ligand disrupts the dimer rather than enhancing it); is this also true for purified FL-SF-1? Are there DNA elements which induce the dimer, as seen in fluorescence polarization for the *CYP7A1* element? Are there DNA elements which disrupt the dimer? The oligomeric state of FL-SF-1 can be analyzed similarly to the LBD using size exclusion chromatography and analytical ultracentrifugation.

The next set of questions to address in this project revolves around where the dimer is occurring both in terms of cellular location (*i.e.* nuclear versus cytoplasmic) and whether it exists on versus off of DNA. Previous ChIP-seq studies have only searched for a monomeric, 9-base pair response element for SF-1.[12] It is possible that reanalysis of this study with a search for a larger element of at least 18 base pairs could yield a different result; dimeric elements may be present. Using the protein complementation assay, it may be possible to visualize the interaction in cells, localize it, and determine whether it is nuclear or cytoplasmic. A different protein complementation assay, such as split GFP, may be better suited for such a localization study. The antibody used for ChIP-seq may or may not be suitable for immunofluorescence studies; NR antibodies are notoriously challenging to work with, but this method of in-cell localization is another potential avenue for investigation.

Finally, the functional consequence of the dimer is critical; does SF-1 form this dimer to activate transcription, or perhaps to repress it? Is the dimer formed on DNA, or off of it, perhaps as a sponge to prevent transcriptional activation? The in-cell assay we performed in which strong agonist disrupts the dimer may suggest that the dimer may not be present as part of a transcriptional

complex, perhaps explaining why the dimer has not been detected in previous SF-1 studies. We first must identify the dimerization interface and verify dimer disruption in purified protein and in the split nanoluciferase assay. The next step will be to perform luciferase assays with the dimer mutants to determine whether there is a functional consequence to disallowing dimer formation. If a dimeric response element is identified in the ChIP-seq reanalysis, we will also analyze this response element in luciferase. Together, all of the studies suggested here will introduce a new paradigm of NR function into this diverse family.

Figures

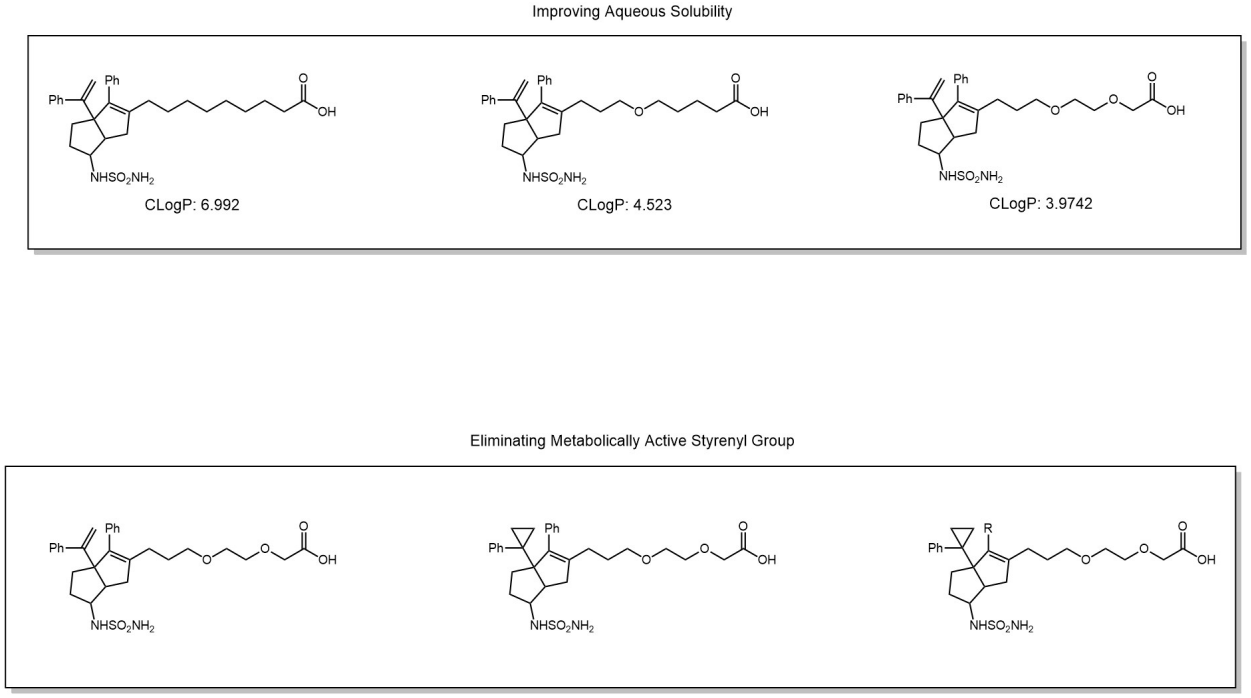


Figure 6.1 LRH-1 agonist design strategy. Top: The current best agonist, 6N-10CA, has a CLogP of 6.992; etherizing the alkyl tail is predicted to improve this parameter 3-log fold to 3.97. Bottom: 6N-10CA has a styrenyl group which is a metabolic liability; shown are strategies to improve

References

1. Cornelison JL, Cato ML, Johnson AM, D'Agostino EH, Melchers D, Patel AB, Mays SG, Houtman R, Ortlund EA, Jui NT. Development of a new class of liver receptor homolog-1 (LRH-1) agonists by photoredox conjugate addition. Bioorg Med Chem Lett. 2020;30(16):127293. Epub 2020/07/08. doi: 10.1016/j.bmcl.2020.127293. PubMed PMID: 32631515.
2. Mays SG, D'Agostino, E.H., Flynn, A.R., Huang, X., Wang, G., Millings, E.J., Okafor, C.D., Patel, A., Cato, M.L., Cornelison, J.L., Melchers, D., Houtman, R., Moore, D.D., Calvert, J.W., Jui, N.T., Ortlund, E.A. Tapping into a phospholipid-LRH-1 axis yields a powerful anti-inflammatory agent with in vivo activity against colitis. BioRXIV. 2020. Epub Sept 2 2020. doi: https://doi.org/10.1101/2020.09.01.278291.
3. 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):25281-91. Epub 2016/10/04. doi: 10.1074/jbc.M116.753541. PubMed PMID: 27694446; PMCID: PMC5207232.
4. Mays SG, Flynn AR, Cornelison JL, Okafor C. Development of the First Low Nanomolar Liver Receptor Homolog-1 Agonist through Structure-guided Design2019;62(24):11022-34. doi: 10.1021/acs.jmedchem.9b00753. PubMed PMID: 31419141.
5. D'Agostino EH, Flynn AR, Cornelison JL, Mays SG, Patel A, Jui NT, Ortlund EA. Development of a Versatile and Sensitive Direct Ligand Binding Assay for Human NR5A Nuclear Receptors. ACS Med Chem Lett. 2020;11(3):365-70. Epub 2020/03/19. doi: 10.1021/acsmedchemlett.9b00442. PubMed PMID: 32184971; PMCID: PMC7074214.
6. Fayard E, Auwerx J, Schoonjans K. LRH-1: an orphan nuclear receptor involved in development, metabolism and steroidogenesis. Trends Cell Biol. 2004;14(5):250-60. Epub 2004/05/08. doi: 10.1016/j.tcb.2004.03.008. PubMed PMID: 15130581.
7. Meinsohn MC, Smith OE, Bertolin K, Murphy BD. The Orphan Nuclear Receptors Steroidogenic Factor-1 and Liver Receptor Homolog-1: Structure, Regulation, and Essential Roles in Mammalian Reproduction. Physiol Rev. 2019;99(2):1249-79. Epub 2019/02/28. doi: 10.1152/physrev.00019.2018. PubMed PMID: 30810078.
8. King BM. The rise, fall, and resurrection of the ventromedial hypothalamus in the regulation of feeding behavior and body weight. Physiol Behav. 2006;87(2):221-44. Epub 2006/01/18. doi: 10.1016/j.physbeh.2005.10.007. PubMed PMID: 16412483.
9. Kinyua AW, Yang DJ, Chang I, Kim KW. Steroidogenic Factor 1 in the Ventromedial Nucleus of the Hypothalamus Regulates Age-Dependent Obesity. PLoS One. 2016;11(9):e0162352. Epub 2016/09/07. doi: 10.1371/journal.pone.0162352. PubMed PMID: 27598259; PMCID: PMC5012732.
10. Lalli E. Adrenocortical development and cancer: focus on SF-1. J Mol Endocrinol. 2010;44(6):301-7. Epub 2010/03/05. doi: 10.1677/JME-09-0143. PubMed PMID: 20200142.
11. Creemers SG, Hofland LJ, Korpershoek E, Franssen GJ, van Kemenade FJ, de Herder WW, Feelders RA. Future directions in the diagnosis and medical treatment of adrenocortical carcinoma. Endocr Relat Cancer. 2016;23(1):R43-69. Epub 2015/10/18. doi: 10.1530/ERC-15-0452. PubMed PMID: 26475053.
12. Doghman M, Figueiredo BC, Volante M, Papotti M, Lalli E. Integrative analysis of SF-1 transcription factor dosage impact on genome-wide binding and gene expression regulation. Nucleic Acids Res. 2013;41(19):8896-907. Epub 2013/08/03. doi: 10.1093/nar/gkt658. PubMed PMID: 23907384; PMCID: PMC3799431.