Summer 8-8-2017

Computational Studies of Liver Receptor Homolog 1 in the Presence of Small Molecule Agonists: Allosteric Communication and Virtual Screening for New Potential Drug Candidates

Bernard Scott Jr
Georgia State University, bscott22@student.gsu.edu

COMPUTATIONAL STUDIES OF LIVER RECEPTOR HOMOLOG 1 IN THE PRESENCE OF SMALL MOLECULE AGONISTS: ALLOSTERIC COMMUNICATION AND VIRTUAL SCREENING FOR NEW POTENTIAL DRUG CANDIDATES

by

BERNARD SCOTT

Under the Direction of Ivaylo Ivanov, PhD

ABSTRACT

Liver Receptor Homolog 1 (LRH-1) is a nuclear receptor whose dysfunction is affiliated with diseases such as diabetes and cancer. Recent investigations demonstrate that higher levels of activation and modulation of its activity can be achieved through its interaction with phospholipids (PLs) and synthetic small molecules. We employed molecular dynamics (MD) simulations to understand more about the structural basis of LRH-1's activity when bound to small molecule agonist RJW100 as well as the RJW100 derivative 65endo. We find that RJW100 and derivative 65endo can trigger allosteric communication in LRH-1 despite the RJW100 scaffold inducing motions that differ from those induced by PLs. We also provide supporting evidence that a key threonine residue and a water network are important for RJW100's ability to activate LRH-1. Finally, in a campaign to identify new LRH-1 lead compounds, virtual screening was performed against RJW100, 65endo, and a second RJW100 derivative, 8AC.

INDEX WORDS: Allosteric Modulation, Virtual Screening, Transcriptional regulation, Dynamical Network.

COMPUTATIONAL STUDIES OF LIVER RECEPTOR HOMOLOG 1 IN THE PRESENCE OF SMALL MOLECULE AGONISTS: ALLOSTERIC COMMUNICATION AND VIRTUAL SCREENING FOR POTENTIAL NEW DRUG CANDIDATES

by

BERNARD SCOTT

A Thesis Submitted in Partial Fulfillment of the Requirements for the Degree of

Master of Science

in the College of Arts and Sciences

Georgia State University

2017

COMPUTATIONAL STUDIES OF LIVER RECEPTOR HOMOLOG 1 IN THE PRESENCE OF SMALL MOLECULE AGONISTS: ALLOSTERIC COMMUNICATION AND VIRTUAL SCREENING FOR POTENTIAL NEW DRUG CANDIDATES

by

BERNARD SCOTT

Committee Chair: Ivaylo Ivanov

Committee: Donald Hamelberg

Alfons Baumstark

Electronic Version Approved:

Office of Graduate Studies

College of Arts and Sciences

Georgia State University

August 2017

ACKNOWLEDGEMENTS

I would really like to thank Dr. Ivanov for his guidance over the 2 years spent in the lab. Considering that I do not come from a physics or computational background, I am glad that he took me in and even gave me some freedom in terms of projects I wanted to work on. In addition to this, he is an awesome teacher that had the patience to break down concepts. I still have a lot of learning and growing to do, but working under his supervision has helped achieve both personal and intellectual growth. I would also like to really thank Dr. Hamelberg and Dr. Baumstark for serving on my committee and encouraging me through this process.

I would really like to thank Brad, Kathleen, Patrick, Tom, and Chunli for advising, helping with technical difficulties, and just being patient with me when possible. I want to especially thank Brad and Kathleen who also made amazing teachers that pushed me to really think and at least attempt to grasp ideas I didn't believe I was capable of. They really set the bar for me. I have had several ups and downs and could still use a lot of improvement, but both represent something I should strive towards in the future studies I pursue. I truly do appreciate all of the help and support from everyone in the Ivanov group,

I would also like to thank the Ortlund group for providing the necessary material and insight to complete my project. I learned a lot from working on this project and am very thankful for being allowed as a collaborator.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

Liver Receptor Homolog-1	LRH-1
Phospholipid	PL
Small Heterodimer Partner	SHP
Transcription Intermediary Factor 2	TIF2
Ligand Binding Pocket	LBP
Beta-H6 Region	β-H6
Activation Function Region	AF2
Activation Function Helix	AF-H
Alternate Function Region	AF
molecular dynamics	MD
Principal Component Analysis	PCA
dilauroylphosphatidylcholine	DPLC
Helix	H
Retinoid X Receptor	RXR

1 INTRODUCTION

Liver Receptor Homolog 1 (LRH-1) belongs to the subfamily V of nuclear receptors (*NR5A*). Like many other nuclear receptors, LRH-1 functions as a transcription factor [1]. LRH-1 has been shown to be important in regulating many metabolic processes. Of special importance is the association of its dysregulation with metabolic disorders and cancers. This primarily occurs through its role in facilitating the transcriptional activity of metabolic genes involved in gluconeogenic pathways such as fatty acid synthase and glucocorticoid kinase [1]. It also plays a pivotal role in the regulation of genes involved in cellular differentiation and is consequently implicated in several cancers. LRH-1's transcriptional activity is modulated through several mechanisms which include ligand binding, post-translational modifications, and binding to co-activating and repressing proteins (Fig. 1.1) [2].

1.1 LRH-1 Dysregulation and Disease

LRH-1 has been linked to liver disease through its role in the regulation of glucose metabolism. It is highly expressed in liver and is known to directly influence the synthesis of bile acids by promoting increased expression of Cytochrome p450s, namely Cyp7a1 and Cyp8b [2] and other targets that control cholesterol uptake. Its regulation of glucose metabolism stems from its role in regulating expression glucokinase enzyme. LRH-1 knockout mice have diminished ability to induce glycolysis as well as glycogen and fatty acid synthesis [3].

Treatment of LRH-1 with phospholipid DLPC has been shown to have anti-diabetic effects in mice [3]. LRH-1 is known for its role in regulating intermediary glucose metabolism through its regulation of genes such as glucokinase [4]. LRH-1 has also been linked to pathways involved in resolving Endoplasmic Reticulum [ER] stress with some recent findings suggesting that diabetes is sometimes associated with unresolved ER stress [5]. LRH-1 is known for

promoting embryonic development [1]. Its involvement in such processes makes it unsurprising that its dysregulation can result in various cancers including pancreatic, lung, intestinal, and breast cancer [1,6,7,8,9,10, 11] where LRH-1 has been found to be abnormally expressed in 45% of breast carcinomas [6].

1.2 Allostery: A Brief Overview

Allostery is defined as the interaction of two topologically distinct sites as triggered by a binding event at one of the sites [12]. Models of allostery that account for the dynamical behavior of macromolecules have been developed [13]. Among them is the ensemble model of allostery [14] which proposes that all proteins exhibit an ensemble of conformations and that introduction of a perturbation (such as ligand binding) triggers a shift in the population distribution of the ensemble [14]. Allosteric phenomena are often associated with large conformational changes, however there are several cases where more subtle motions are responsible for allosteric communication within a macromolecular system [15]. Use of a network theory framework provides us with tools to account for how smaller changes contribute to allosteric phenomenon [16,17,18,19].

Allosteric targeting provides a promising alternative to targeting orthosteric binding sites. This especially applies to cases in which a family of enzymes or receptors have a highly conserved orthosteric site. Allosteric sites within proteins of the same family or subfamily are typically less conserved than the corresponding orthosteric site and present an opportunity to achieve selective targeting [20]. This approach also has the advantage of modulating a target's activity as opposed to completely abolishing it [20].

1.3 Structural Features of LRH-1

Like other nuclear receptors [21], LRH-1 functionality is largely modulated via its interaction with a suite of small co-regulating peptides capable of either repressing or activating LRH-1's regulation of transcriptional activity. Our investigations focus on LRH-1's interactions with the co-activator Transcription Intermediary Factor 2 (TIF2. Aliases include GRIP2 and SRC2). TIF2 is a co-activating peptide which typically recognizes nuclear receptors in a ligand dependent manner [22]. Co-regulating peptides such as TIF2 exploit LXXLL motifs to access the N-terminal binding clefts of LRH-1 which is consistent with recognition motifs of other nuclear receptors such as Retinoid X Receptor (RXR) [23]. Several important conformational changes occur upon ligand binding to nuclear receptors such as the reorientation of Helix 12, the Activation Function Helix (AF-H). This reorientation facilitates the completion of a hydrophobic binding surface for co-regulator proteins [24]. However, a full understanding of how ligand binding facilitates transcription activating states of LRH-1 has yet to be achieved. The ligand binding pocket (LBP) includes helices 2 and 3 along with the β-sheet-H6 Region (β-H6, yellow) which encompasses β-sheets 1 and 2 as well as helix 6. These two regions combined in addition to helices 5 and 10 comprise the Alternate Function (AF) domain whereas regions of LRH-1 engaging the co-regulator (purple) are designated as the Activation Function Domain (AF2) (Fig. 1.2) [23].

1.3.1 Features that Allow for Ligand Independent Activation of LRH-1

LRH-1 can upregulate transcription in absence of a ligand and studies with murine LRH-1 revealed an empty ligand binding pocket (LBP) [24]. These studies highlighted key structural differences from ligand activated nuclear receptors. These differences include features that allow for post-translational modifications such as phosphorylation and sumoylation [24,25,26].

They also include an extended and rigid Helix 2 that packs against Helix 3 in LRH-1. Retinoid X Receptor and many other nuclear receptors feature Helices 1 and 2 that are separated by a highly disordered flexible loop [24]. Helix 1 in LRH-1 is positioned closer to Helix H9, strategically placing an N-terminal proline residue on a different face of H1 than homologous receptors. These differences are proposed to result in more extensive contacts between Helix 1 and other portions of the protein, which is thought to be crucial to LRH-1's constitutive activation and recruitment of co-activating peptides in the absence of a ligand [24].

However, the study also foreshadowed future findings that LRH-1 is responsive to ligand binding. An overlay of LRH-1 on top of RXR (Fig. 1.3) demonstrates that mLRH-1's activated conformation was similar to ligand activated RXR. This suggested that LRH-1 was at least capable of accommodating sizeable ligands such as retinoids and cholesterol derivatives [24]. However, the ligand independent nature of activation was reaffirmed through evidence derived from mutagenesis studies. Methionine and Tryptophan mutants were made of LBP inner facing residues Ala368 from H1 and Ala532 from H11. These mutations were selected to impair binding of ligands to the LBP. The results showed enhanced LRH-1 facilitated transcription of SHP in both Ala532 mutants where LRH-1 retained the ability to recruit co-repressing peptides in the absence of a ligand [24].

Further mutagenesis experiments provided evidence that Helix 2 may play a significant role in co-regulator recruitment. Mutation of three glutamine residues led to diminished activity of LRH-1 as well as diminished responsiveness to co-regulators TIF2 and SHP [24]. It is speculated that H2's importance rests in its ability to assist in stabilization of the AF-H [24].

LRH-1's co-regulator binding cleft may not be optimized for recruitment of co-activators such as TIF2. Notably, LRH-1 has many features that differentiate its co-activator binding cleft

from others including an AF-H with bulkier residues and variants of residues that contribute to a charged clamped interaction between the AF-H and co-regulator proteins in other nuclear receptors. One such example is Gln398 of LRH-1 which aligns with lysine or arginine residues found in other receptors. This difference may be associated with the reduced ability of LRH-1 to discriminate between co-activators and repressors *versus* other nuclear receptors. Mutagenesis studies replacing bulkier residues of LRH-1's AF-H with the smaller residues found in RXR resulted in increased transcriptional activity in response to over-expression of common co-activator proteins [24].

1.3.2 Discovery of Ligand Activation

Despite originally having been thought to be an orphan nuclear receptor, evidence showed that it is not when phospholipids (PLs) were identified in LRH-1's LBP indicating that LRH-1 may be a lipid metabolism sensing receptor [27]. Treatment of mice with dilauroylphosphatidylcholine (DLPC) resulted in lower serum glucose levels [3]. DLPC binds LRH-1's LBP with a unique binding mode (Fig. 1.4) which features the insertion of the fatty acyl tail into the interior of the protein exploiting several hydrophobic contacts as well as the placement of polar head group near the "mouth" of the protein which is solvent exposed. The polar head group participates in key polar contacts with Tyr516 and Lys520 [28].

LRH-1 discriminates by tail-length more so than headgroup composition [29]. As shown in Figure 1.5, LRH-1 can bind many types of PL headgroups but displays a clear preference for medium fatty acyl tail lengths, lengths between 8 and 16 carbons [29].

1.3.3 Allosteric Communication and PL Induced Dynamics in LRH-1

LRH-1 activates transcription upon binding medium-tailed PLs such as DLPC and represses transcription upon binding of select long-tailed PLs. Computational studies suggest

that LRH-1 ligand binding increases co-activator access by triggering an allosteric pathway that allows for communication between the LBP and the Activation Function region primarily through Helix 5 which bridges the two remote regions of LRH-1. This allosteric communication is characterized by the highly correlated motions that take place between the two regions upon ligand binding. In the presence of phospholipids, notable structural changes occur at both the ligand binding pocket and the co-regulator binding cleft. The statuses of the LBP and the co-regulator binding cleft can either be in agreement or disagreement [29]. An agonist bound to LBP and co-activator bound in co-regulator binding cleft would characterize statuses in agreement whereas an agonist bound LBP with a co-repressor bound would characterize disagreeing statuses.

LRH-1's ability to selectively bind co-regulators is in part dictated by whether the statuses of the two regions agree or disagree with one another [29]. Crystal structures from various LRH-1 complexes demonstrate a possible association. One indication of this link is the fidelity of the charge clamp interaction in transcription repressing structures. De-activated LRH-1 bound to Small Heterodimer Partner (SHP) a co-repressing protein and *Escherichia coli (E. coli)* PLs at the LBP as well as apo LBP features an intact charge clamp interaction between SHP and the AF-H [28,29,30]. LRH-1·TIF2 when complexed with DLPC (agreement status) features a diminished distance between LRH-1's Glu534 and TIF2 versus activated structures with disagreeing statuses as well as the more optimal positioning of Glu534 that allows for the contact with TIF2 (Figure 1.6).

Principle Component Analysis (PCA) demonstrated that DLPC induced a notable "breathing" motion at the mouth of LRH-1 (PC2) and between H10 and H9 (PC1). This is consistent with Hydrogen Deuterium Exchange mass spectrometry studies that show that β-H6 is

more dynamical than when LRH-1 is bound to longer tailed PLs [29]. This finding also extends to potentially antagonistic PLs (*E. coli* PLs) that induce structural changes more associated with the inactivated conformation [29]. A crystal structure was identified with LRH-1 in complex with *E. coli* PLs and a fragment of SHP. This complex appears to place the opening of the LBP, including H2, H3 and β-H6 region in a more open conformation than the TIF2 bound complex (Fig. 1.6A). In addition, H9 and H10 assume a more open conformation in the deactivated structure. When the Principal Components are plotted against each other, the DLPC-bound structure exhibits a single density-populated cluster of conformational subspace along PC2 whereas activated LRH-1 with disagreeing statuses (Fig. 1.7, *C versus D*) seems to feature two clusters along PC2 [29]. The results suggest that agreement status of the two remote domains influence the topology of the co-regulator binding cleft, which could in part explain variances in co-regulator selectivity.

Dynamical Network Analysis results (via suboptimal path analysis) showed that, when bound to PLs, agreement statuses did correlate well with the amount of communication (in this case the number of suboptimal paths) between the LBP and the co-regulator binding cleft. Figure 1.8A demonstrates how DLPC binding to LRH-1·TIF2 induces much stronger communication in comparison to the *E. coli* PL binding where one sees a diminished number of suboptimal paths travelling between the two remote regions [29].

1.3.4 Insight Gained in Early Drug Discovery Efforts

GSK8470, a small molecule agonist of LRH-1, provided interesting structural information and served as a template for future candidates [31]. An X-ray crystal structure with LRH-1·TIF2 and GSK8470 in complex reveals a mode of binding that exploits hydrophobic contacts. Among these hydrophobic contacts is a face to face pi-pi interaction between the

aniline moiety of GSK8470 and His390 (Fig. 1.9) [31]. Attempts to derivatize this lead compound led to discovery of another agonist, RJW100 which features a hydroxyl functionality on the bicyclic core of the molecule [30].

RJW100 was crystallized with LRH-1·TIF2 (Fig. 1.10) as well and its binding mode contains several features that differentiate it from GSK8470 [31]. The proposed pi-pi interaction is preserved, but instead has an edge-face configuration. Of note is a potential water-mediated interaction between the hydroxyl of RJW100 and Thr352. This interaction suggests that addition of a polar moiety to the original scaffold induces productive allosteric communications between the LBP and the co-regulator binding cleft.

1.3.5 Current Experimental Insight into Activation of LRH-1 by Synthetic Agonists.

Studies have attempted to elucidate the structural basis of small molecule activation of LRH-1 [32]. RJW100 destabilized helices 10 and 3, and stabilizes AF-H. Despite the stability of the AF-H, Glu534 becomes disordered in comparison to DLPC bound LRH-1 (Fig. 1.11) [32]. This implies that RJW100 is unable to achieve selective co-activator recruitment as well as DLPC. Key placement of structural waters likely plays a role in the binding of LRH-1 by small molecules. Mutagenesis studies reveal a diminished ability of RJW100 to activate LRH-1 upon T352V mutation (Fig. 1.12) to disrupt the water network. Simulations additionally demonstrate the stability of crystallographic waters within the LBP of wildtype (wt) LRH-1 as well as destabilization in the mutant [32].

Two RJW100 derivatives, 65endo and 8AC have been synthesized. 65endo replaces the hydroxyl functional group with a sulfonamide whereas 8AC retains the hydroxyl group while extending and capping the aliphatic tail with a polar ester group (Fig. 1.13). Both compounds were designed strategically to enhance bio-availability and to exploit key structural features of

LRH-1 to trigger increased transcriptional activity relative to RJW100. RJW100's hydroxyl group was replaced with a sulfonamide yielding 65endo. This modification was intended to enhance binding to the polar patch within LRH-1's LBP. The incorporation of the ester functional group to yield 8AC was intended to allow for contacts at the mouth of the binding pocket (mimicking PL agonists such as DLPC) while maintaining contacts within the interior via the hydroxyl group.

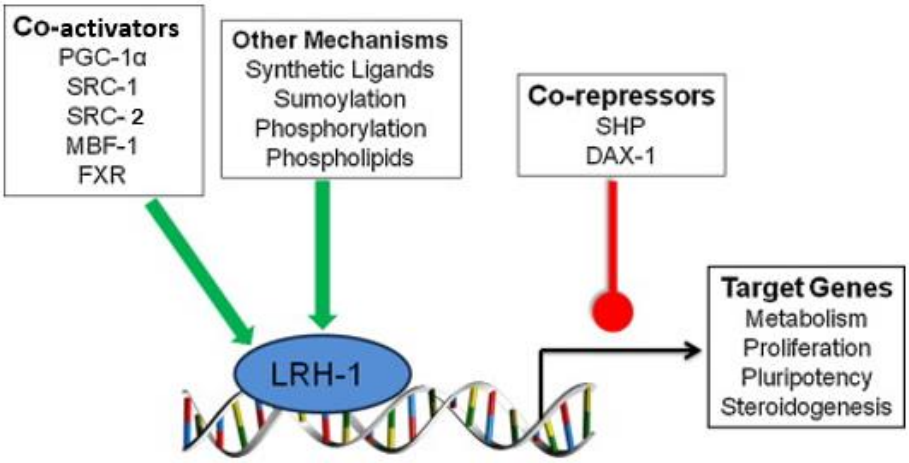

Figure 1.1. Mechanisms for regulating LRH-1 activity.
LRH-1-mediated transcriptional activity can be modulated through several mechanisms including ligand binding, post-translational modifications, co-activator binding, and co-repressor binding. LRH-1 activation facilitates transcription of genes involved in several processes including metabolism, cell growth, and cell differentiation. Figure has been adapted with permission from Nadolny, C. and X. Dong, *Liver receptor homolog-1 (LRH-1): a potential therapeutic target for cancer*. Cancer Biol Ther, 2015. 16(7): p. 997-1004.

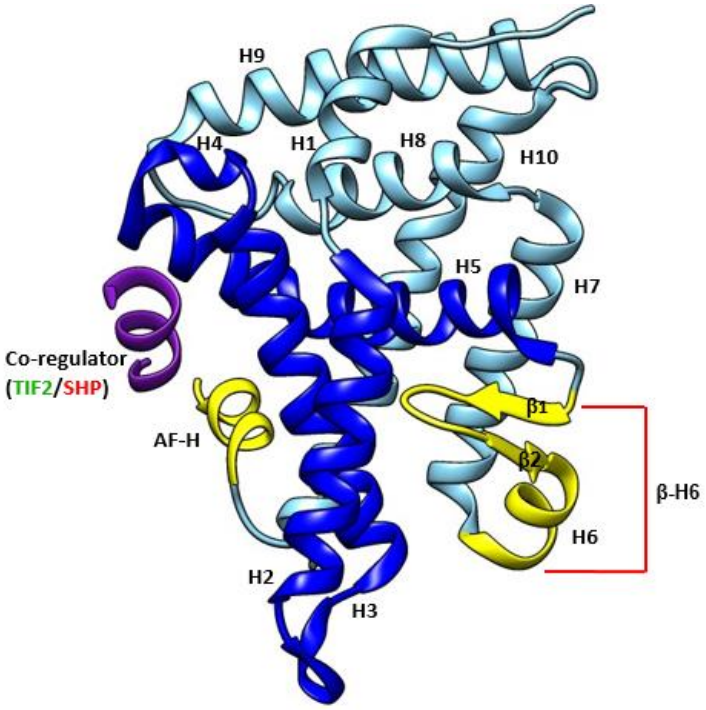

Figure 1.2. Ribbon diagram highlighting key structural features of LRH-1.
Yellow, β-H6 region and Activation Function Helix (AF-H). Blue, H2 and H3. Purple, Co-regulator either co-activator TIF2 or co-repressor SHP. (pdb 4DOS) [32].

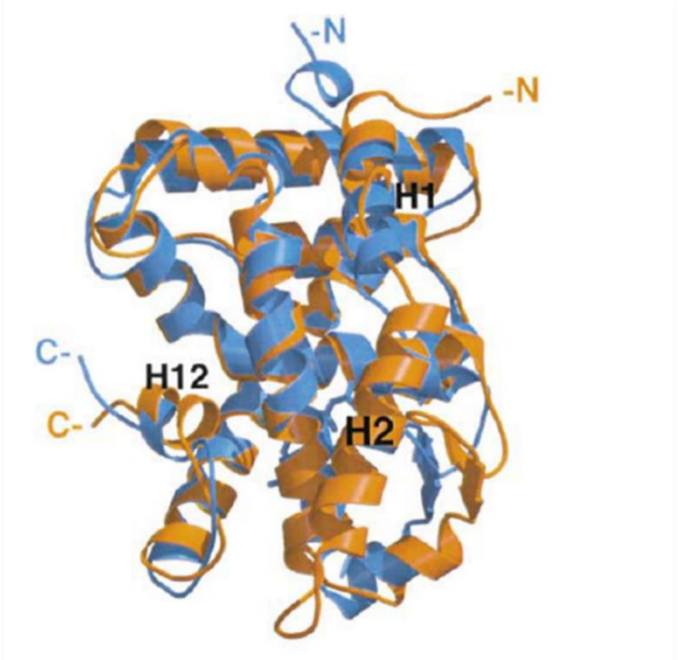

Figure 1.3. Overlay of LRH-1 onto Retinoid X Receptor (RXR).
LRH-1 (orange) shares many topological features with RXR (blue), but notable differences include an elongated Helix 2. This figure has been adapted with permission from Sablin, E.P., et al., *Structural Basis for Ligand-Independent Activation of the Orphan Nuclear Receptor LRH-1*. Molecular Cell, 2003. **11**(6): p. 1575-1585. [28]

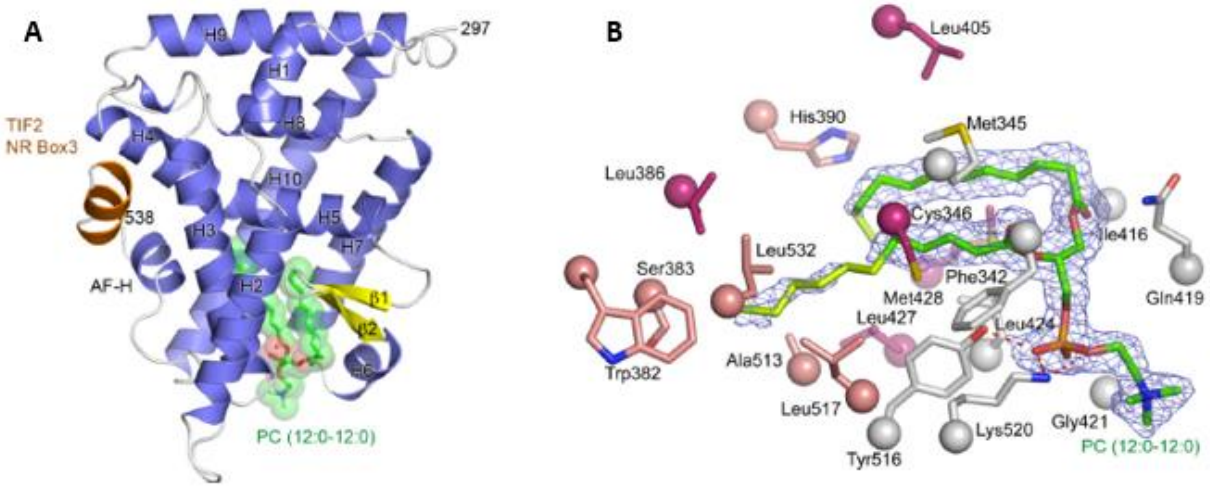

Figure 1.4. Structures of LRH-1·TIF2 in complex with agonist DLPC.
A) Ribbon Diagram of LRH-1·TIF2 complexed with DLPC (within green electron density) (pdb 4dos). B) DLPC and sidechains lining LRH-1 LBP. Figure has been reproduced with permission from Musille, P.M., et al., *Antidiabetic phospholipid-nuclear receptor complex reveals the mechanism for phospholipid-driven gene regulation.* Nat Struct Mol Biol, 2012. **19**(5): p. 532-7, S1-2. [22]

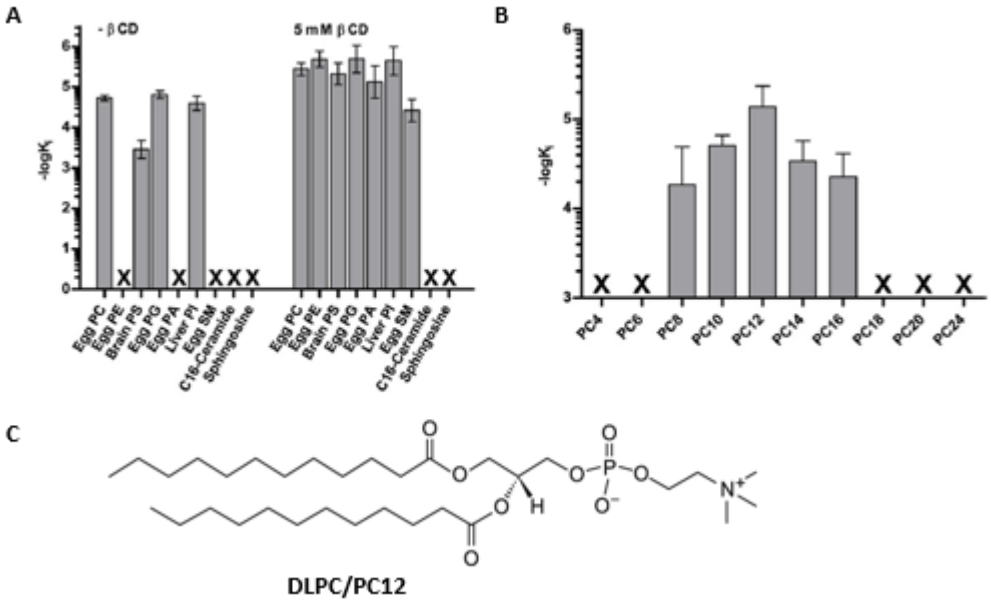

Figure 1.5. Measurements of LHR-1 Phospholipid binding preferences.
A) Binding data from investigations of PL headgroup preferences of LRH-1. B) Binding data investigating fatty acyl tail length preferences of LRH-1. C) Structure of DLPC. Binding assays results are derived from competition assays described in [29]. This figure has been adapted from Musille, P.M., et al *Unexpected Allosteric Network Contributes to LRH-1 Co-regulator Selectivity*. J Biol Chem, 2016. **291**(3): p. 1411-26. [29]

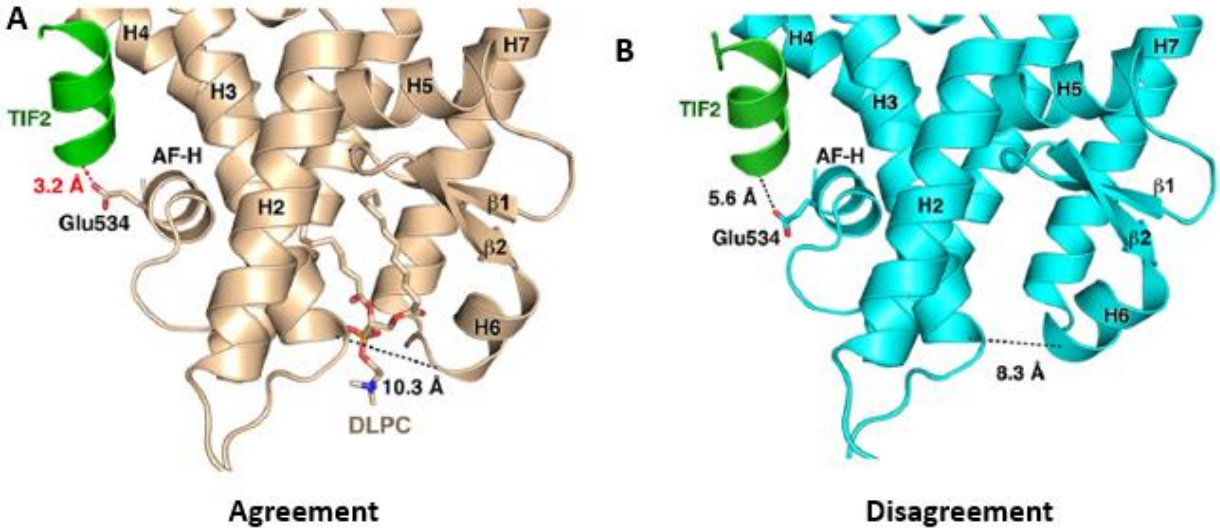

Figure 1.6. Ribbon diagrams of LRH-1·TIF2 systems with varied LBP status. A) LRH-1 bound to TIF2 co-activator (green) and DLPC, an agreement state. Red dashed lines indicate distance between E534 in the Activation Function Helix and Tif-2. B) apo LRH-1 bound to TIF2 co-activator (green) where black dashed lines indicate distance between E534 in the Activation Function Helix and Tif-2. DLPC binding increases the width of at the mouth of the LBD. Figure has been adapted from Musille, P.M., et al., *Unexpected Allosteric Network Contributes to LRH-1 Co-regulator Selectivity*. J Biol Chem, 2016. **291**(3): p. 1411-26. [29]

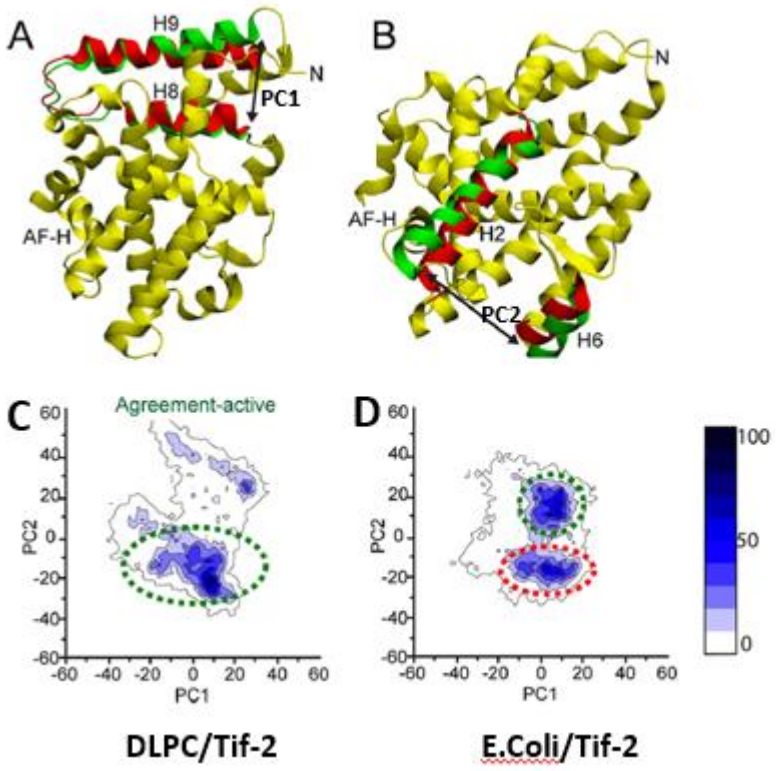

Figure 1.7. Correlated motions in LRH-1 when bound to Phospholipids. A) Principal Component Analysis of LRH-1 systems Principal Component 1 (PC1) features an opening motion of the activated state at Helix 8 and 9 versus de-activated state. B) Principal Component 2 (PC2) features an opening motion at the opening of LRH-1's Ligand Binding Pocket of the activated form of LRH-1. In both PCs, transcription activating structures (green) have larger interhelical distances (between Helices 8 and 9 in PC1, and helices 2 and 6 in PC2) than repressed (red) structures. C) Principal Component 1 and 2 plotted against each other for LRH-1·Tif-2 complexed to DLPC. D) LRH-1·Tif-2 complexed to an *E.coli* Phospholipid [11]. The higher the density, the more frames (out of 10000) characterized as occupying the indicated region of conformational subspace. This figure has been adapted from Musille, P.M., et al., *Unexpected Allosteric Network Contributes to LRH-1 Co-regulator Selectivity*. J Biol Chem, 2016. **291**(3): p. 1411-26 [29].

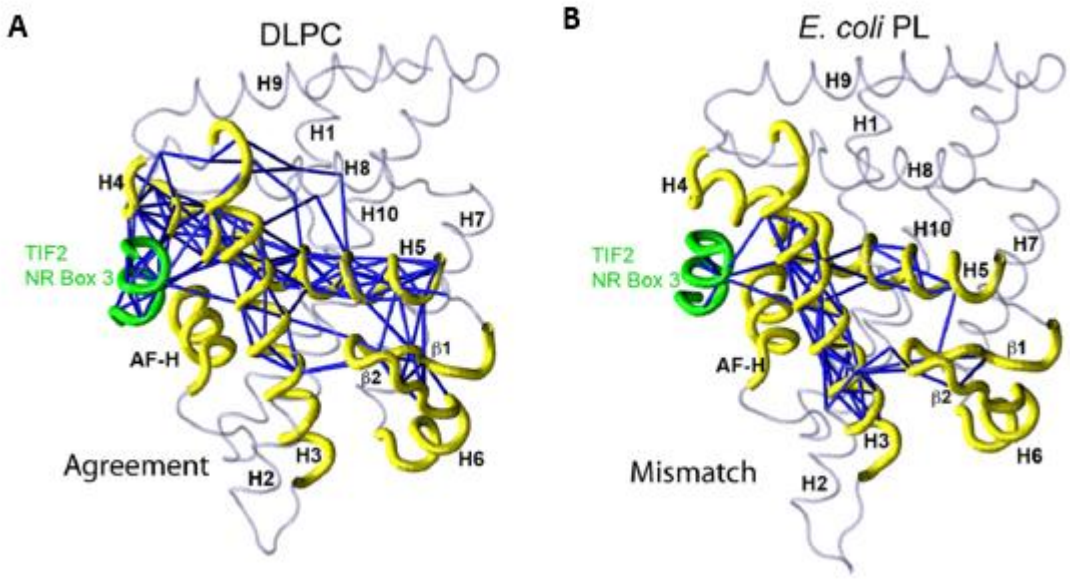

Figure 1.8. Allosteric communication between β-H6 and co-activator TIF2. A) Suboptimal Paths (blue) that occur upon binding of DLPC to LRH-1·TIF2 complex (pdb 4DOS). B) Suboptimal Paths (blue) that occurs upon the binding of *E. coli phospholipids* to LRH-1·TIF2 complex. Figure has been adapted from Musille, P.M., et al., *Unexpected Allosteric Network Contributes to LRH-1 Co-regulator Selectivity.* J Biol Chem, 2016. **291**(3): p. 1411-26. [29]

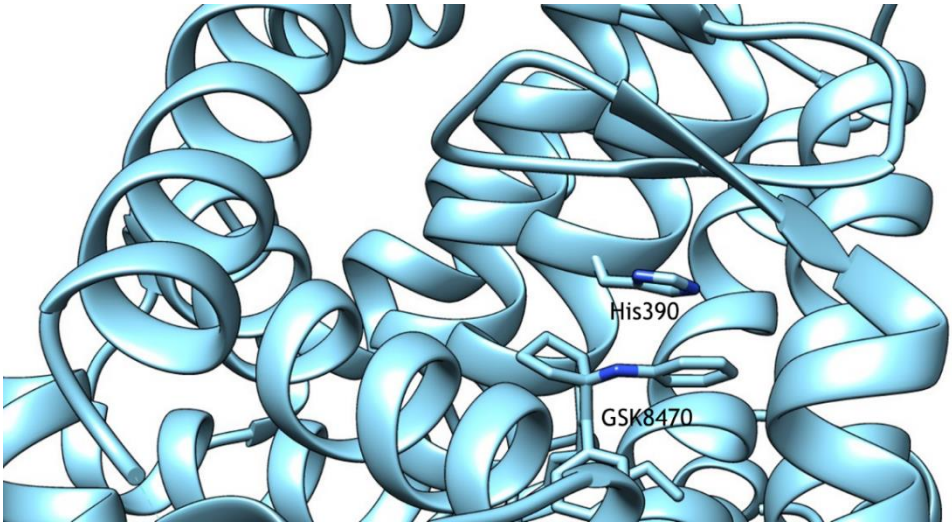

Figure 1.9. Ribbon diagram featuring structure of LRH-1·TIF2 complexed to GSK8470.
Crystal Structure of GSK8470 bound to LRH-1 (pdb 3PLZ) His390 is labelled to emphasize a potential face to face interaction between the aniline moiety of GSK8470 and His390 [31].

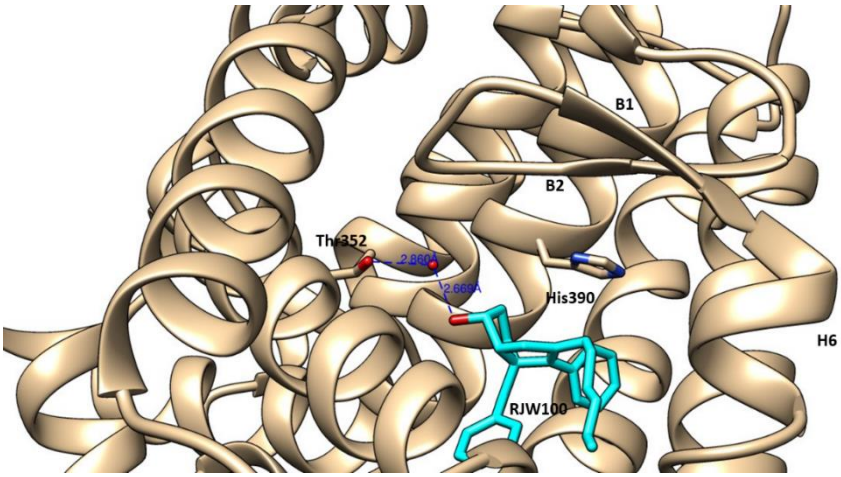

Figure 1.10. Ribbon diagram featuring tructure of LRH-1·TIF2 in complex with RJW100.

Crystal structure featuring RJW100 bound to LRH-1(pdb 5L11). Distances are displayed in blue to highlight a potential water-mediated contact between the hydroxyl function of RJW100 and Thr352 of LRH-1. His390 is labelled to emphasize a potential contact with the arene group of RJW100 [32].

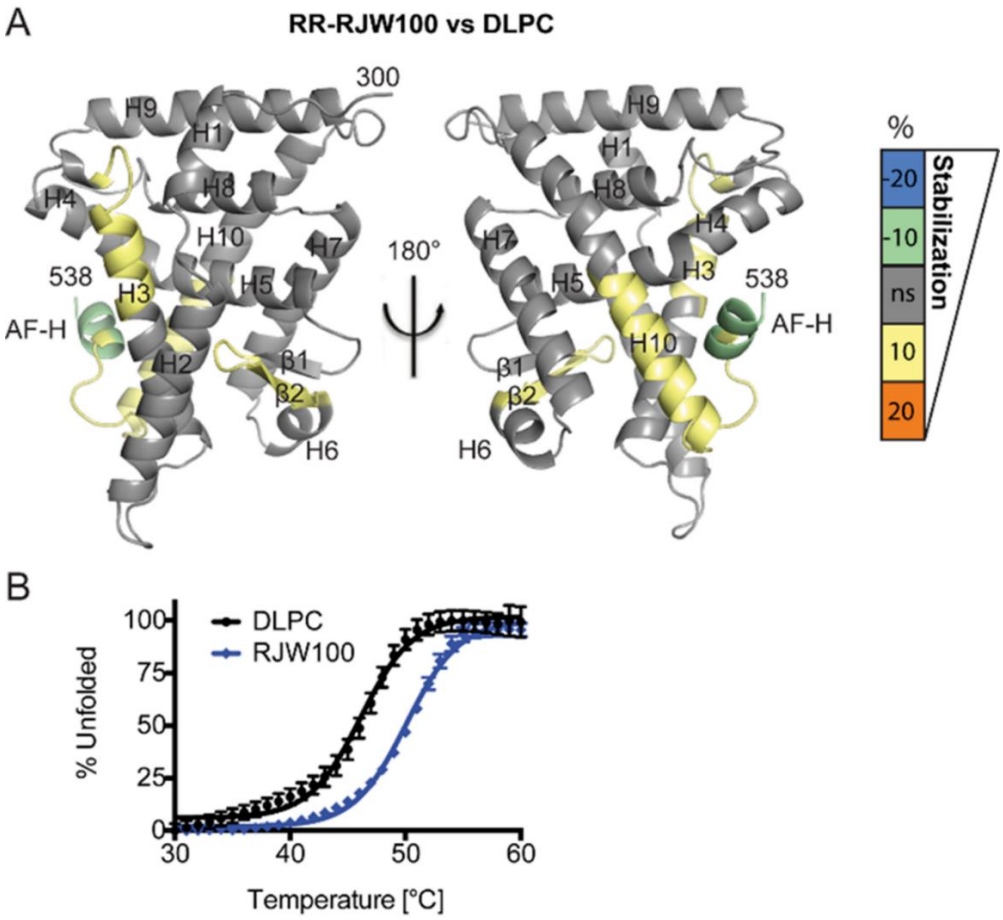

Figure 1.11. Experimentally measured dynamical and thermodynamic profile of LRH-1·TIF2 when complexed to DLPC and RJW100.
A) Hydrogen deuterium exchange mass spectrometry data. The scale is colored by % exchange with deuterium. Stability decreases as % deuterium exchange increases. B) Melting temperature as a function of % protein unfolded. Figure has been adapted from Mays, S.G., et al., *Crystal Structures of the Nuclear Receptor, Liver Receptor Homolog 1, Bound to Synthetic Agonists.* J Biol Chem, 2016. 291(49) p.25281-25291 [32]

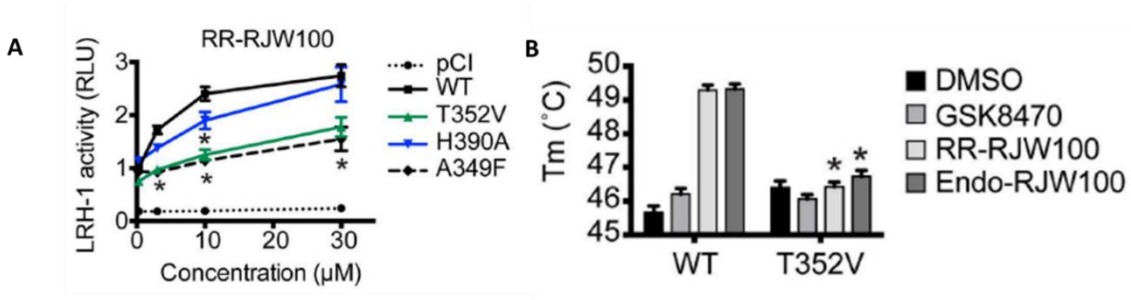

Figure 1.12. Measurement of activity levels and thermostability of LRH-1·TIF2 systems complexed with small molecule agonists.
A) LRH-1·TIF2 activity levels measured in Relative Light Units (RLU). Solid black (wildtype), green (T352V mutant), blue (H390A mutant). Dashed black (A349F mutant). B) Comparing thermostability of wildtype LRH-1·TIF2 to T352V mutant when bound to agonists GSK8470, RJW100, and endo-RJW100. Figure has been adapted from Mays, S.G., et al., *Crystal Structures of the Nuclear Receptor, Liver Receptor Homolog 1, Bound to Synthetic Agonists.* J Biol Chem, 2016. **291**(49) p. 25281-25291 [32].

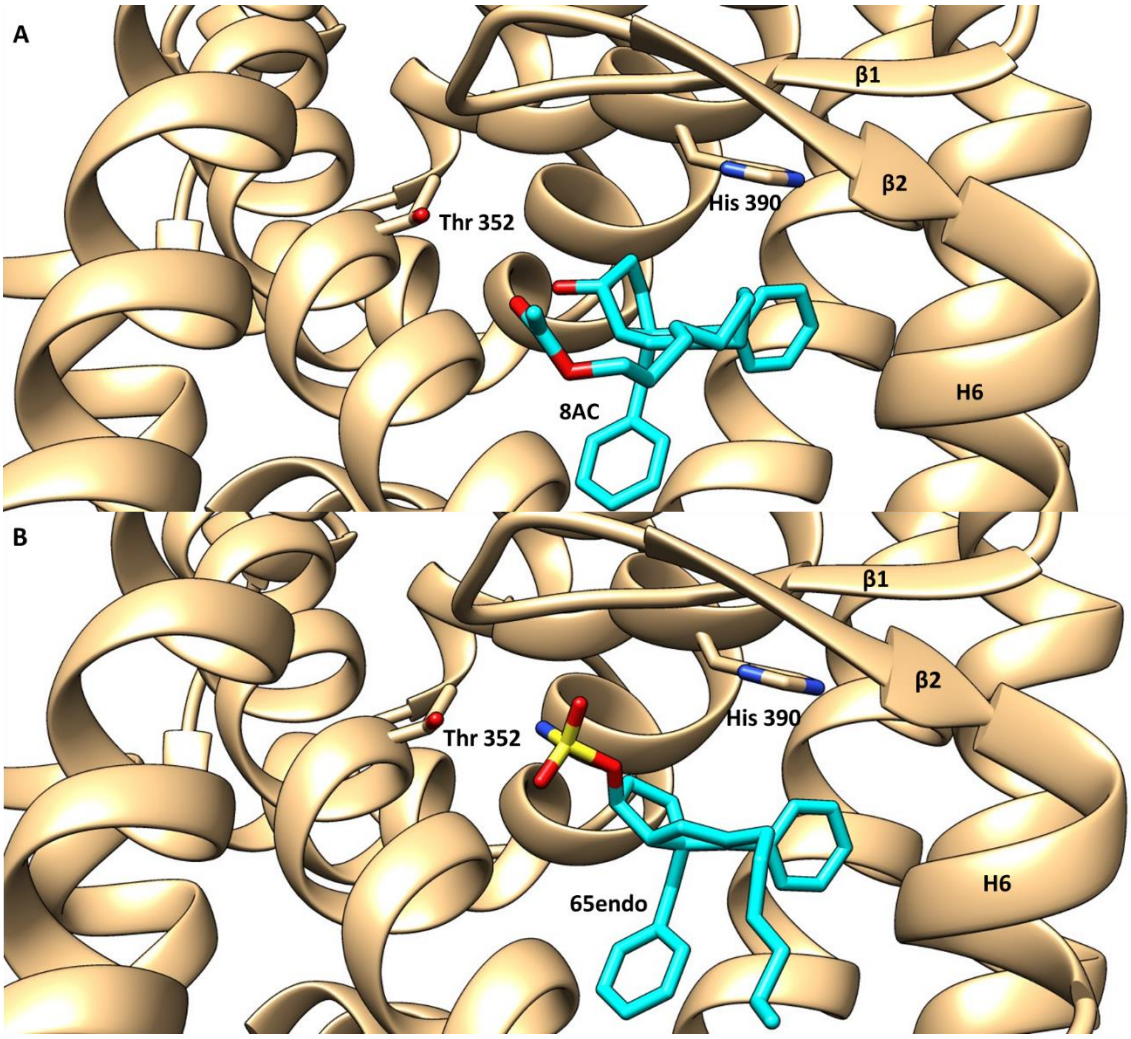

Figure 1.13. Ribbon diagram featuring the structure of LRH-1·Tif-2 in complex with RJW100 derivatives.
A) LRH-1 bound to 8AC. B) LRH-1 bound to 65endo.

1.4 References

1. Fayard, E., J. Auwerx, and K. Schoonjans, *LRH-1: an orphan nuclear receptor involved in development, metabolism and steroidogenesis.* Trends in Cell Biology. **14**(5): p. 250-260
2. Nadolny, C. and X. Dong, Liver receptor homolog-1 (LRH-1): a potential therapeutic target for cancer. Cancer Biol Ther, 2015. 16(7): p. 997-1004.
3. Musille, P.M., et al., *Antidiabetic phospholipid-nuclear receptor complex reveals the mechanism for phospholipid-driven gene regulation.* Nat Struct Mol Biol, 2012. **19**(5): p. 532-7, S1-2
4. Oosterveer, M.H., et al., *LRH-1-dependent glucose sensing determines intermediary metabolism in liver.* J Clin Invest, 2012. **122**(8): p. 2817-26.
5. Mamrosh, J.L., et al., *Nuclear receptor LRH-1/NR5A2 is required and targetable for liver endoplasmic reticulum stress resolution.* Elife, 2014. **3**: p. e01694.
6. Thomas, S.E., et al., *Unravelling the story of protein misfolding in diabetes mellitus.* World J Diabetes, 2011. **2**(7): p. 114-8.
7. Jiang, W., et al., *MicroRNA-376c suppresses non-small-cell lung cancer cell growth and invasion by targeting LRH-1-mediated Wnt signaling pathway.* Biochem Biophys Res Commun, 2016. **473**(4): p. 980-6.
8. Bianco, S., et al., *LRH-1 governs vital transcriptional programs in endocrine-sensitive and -resistant breast cancer cells.* Cancer Res, 2014. **74**(7): p. 2015-25.
9. Benod, C., et al., *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): p. 16927-31.
10. Chand, A.L., et al., *The orphan nuclear receptor LRH-1 promotes breast cancer motility and invasion.* Endocr Relat Cancer, 2010. **17**(4): p. 965-75.

11. Clyne, C.D., et al., *Liver receptor homologue-1 (LRH-1) regulates expression of aromatase in preadipocytes.* J Biol Chem, 2002. **277**(23): p. 20591-7.

12. Kramer, H.B., et al., *LRH-1 drives colon cancer cell growth by repressing the expression of the CDKN1A gene in a p53-dependent manner.* Nucleic Acids Res, 2016. **44**(2): p. 582-94.

13. Changeux, J.-P., *50 years of allosteric interactions: the twists and turns of the models.* Nat Rev Mol Cell Biol, 2013. **14**(12): p. 819-829.

14. Kern, D. and E.R.P. Zuiderweg, *The role of dynamics in allosteric regulation.* Current Opinion in Structural Biology, 2003. **13**(6): p. 748-757.

15. Motlagh, H.N., et al., *The ensemble nature of allostery.* Nature, 2014. **508**(7496): p. 331-339.

16. Tsai, C.-J., A.d. Sol, and R. Nussinov, *Allostery: Absence of a change in shape does not imply that allostery is not at play.* Journal of molecular biology, 2008. **378**(1): p. 1-11.

17. Dokholyan, N.V., *Controlling Allosteric Networks in Proteins.* Chemical Reviews, 2016. **116**(11): p. 6463-6487.

18. Allosteric Communication Occurs via Networks of Tertiary and Quaternary Motions in Proteins.

19. O'Rourke, K.F., S.D. Gorman, and D.D. Boehr, *Biophysical and computational methods to analyze amino acid interaction networks in proteins.* Computational and Structural Biotechnology Journal, 2016. **14**: p. 245-251.

20. Nussinov, R. and C.-J. Tsai, *Allostery in Disease and in Drug Discovery.* Cell. **153**(2): p. 293-305.

21. Millard, C.J., et al., *An evolving understanding of nuclear receptor coregulator proteins.* J Mol Endocrinol, 2013. **51**(3): p. T23-36.

22. Voegel, J.J., et al., *The coactivator TIF2 contains three nuclear receptor-binding motifs*

and mediates transactivation through CBP binding-dependent and -independent pathways. The EMBO Journal, 1998. **17**(2): p. 507-519.

23. Savkur, R.S. and T.P. Burris, *The coactivator LXXLL nuclear receptor recognition motif.* The Journal of Peptide Research, 2004. 63(3): p. 207-212.
24. Sablin, E.P., et al., *Structural Basis for Ligand-Independent Activation of the Orphan Nuclear Receptor LRH-1*. Molecular Cell. **11**(6): p. 1575-1585.
25. Lee, Y.K., et al., *Phosphorylation of the hinge domain of the nuclear hormone receptor LRH-1 stimulates transactivation.* J Biol Chem, 2006. **281**(12): p. 7850-5.
26. Priest, C. and P. Tontonoz, *SUMOylation Places LRH-1 in PROXimity to Lipid Metabolism*. Cell Metabolism. 20(4): p. 558-559.
27. Krylova, I.N., et al., *Structural analyses reveal phosphatidyl inositols as ligands for the NR5 orphan receptors SF-1 and LRH-1*. Cell, 2005. **120**(3): p. 343-55.
28. Ortlund, E.A., et al., *Modulation of human nuclear receptor LRH-1 activity by phospholipids and SHP*. Nat Struct Mol Biol, 2005. **12**(4): p. 357-63.
29. Musille, P.M., et al., *Unexpected Allosteric Network Contributes to LRH-1 Co-regulator Selectivity.* J Biol Chem, 2016. **291**(3): p. 1411-26.
30. Zhang, T., et al., *Molecular dynamics simulation study for LRH-1: interaction with fragments of SHP and function of phospholipid ligand.* Proteins, 2008. **70**(4): p. 1527-39.
31. Whitby, R.J., et al., *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): p. 2266-81.
32. Mays, S.G., et al., *Crystal Structures of the Nuclear Receptor, Liver Receptor Homolog 1, Bound to Synthetic Agonists.* J Biol Chem, 2016. **291**(49): p. 25281-25291.

2 METHODS AND MATERIALS

2.1 Virtual Screening

2.1.1 Tanimoto Scoring with ROCS and EON

A portion of our drug discovery efforts employed ligand based virtual screening techniques to identify new scaffolds that can bind to the same pocket with similar overall shape and charge distribution. An ensemble of ligand conformations was generated for all compounds in NCI database through the program OMEGA [1], and the query molecules (binding poses of each agonist as found in LRH-1·TIF2 X-ray crystal structures) and generated conformers are aligned and scored with ROCS [2]. After structural alignment, EON [3] can be used to compare electrostatic potential maps of the same molecules using Tanimoto measurements to aid in discovery of potential agonists.

$$Shape\ Tanimoto = \frac{Shape\ Overlap(A,B)}{Shape(A)+Shape(B)+Overlap\ Shape(A,B)} \quad \text{Eq. 2.1}$$

$$Electrostatic\ Tanimoto = \frac{EField\ Overlap(A,B)}{Efield(A)+Efield(B)+Efield(A,B)} \quad \text{Eq. 2.2}$$

Tanimoto measurement is used in the comparison of sets of 3D parameters, and it involves comparing the overlap of those properties. Shape functions (volumes) for A and B are compared with the overlap between the two shape functions after alignment to give Shape Tanimoto values which approach 0.0 for dissimilar and 1.0 for identical (eq. 2.1) [2]. Eon Tanimoto (eq. 2.2) uses Poisson-Boltzmann electrostatic field functions to compare A and B assigning a penalty for compounds with fields with opposite signs, such that the minimum score is -1/3 in the case of a field with perfectly opposite electrostatic fields [3].

2.1.2 Molecular Docking

Docking considers both the shape, electrostatics, bond length, bond angles, and Van der Waals forces of a small molecule and attempts to find the most energetically favorable binding pose within a user-specified search region. One drawback of standard small molecule docking is that it cannot account for how motions within the region of interest can affect binding of a molecule as well as an intrinsic difficulty in accounting for solvation patterns and the role of solvent in binding [4]. Autodock vina [5] was employed to score top ranking compounds resulting from our ligand based virtual screening efforts. Given the high volume of compounds (10,000 per lead), this study was limited to rigid docking to the ligand binding domain of LRH-1.

2.2 Molecular Dynamics

Molecular Dynamics (MD) simulation is a very powerful tool often applied for investigation of biological and biochemical systems. These simulations reveal conformational and dynamic characteristics at the atomic level, assisting in the interpretation of data obtained from experiments.

2.2.1 Amber Force Field

The potential energy (V) in MD simulations is a function of the molecular parameters to describe intra- and inter- molecular forces between atoms in the system. To make computational calculations tractable, approximations are made to construct a force field, which ultimately consists of several terms to describe the contributions of bonding and non-bonding interactions to the potential energy as demonstrate below [6]:

$$V(R)_{total} = V(R)_{bonded} + V(R)_{non\text{-}bonded} \qquad \text{Eq. 2.3}$$

The bonded interactions correspond to bond length, angle, and torsion. The non-bonded terms include both Columbic potential to account for electrostatic interactions and the Lennard-Jones

(LJ) potential to describe van der Waals interactions. Forcefield parameters are experimentally-derived or calculated using *ab initio* methods [7]. For example, The Amber Forcefield which is described as [6]:

$$U(r_1, \cdots, r_N) = \sum_{bonds} \frac{k_r}{2}\left(r - r_{eq}\right)^2 + \sum_{angles} \frac{k_\theta}{2}(\theta_i - \theta_{i0})^2 + \sum_{torsions} \frac{V_n}{2}[1 + \cos(n_\theta - \gamma_i] + \sum_{i<j}\left[\left(\frac{A_{ij}}{R_{ij}}\right)^{12} - \left(\frac{B_{ij}}{R_{ij}}\right)^{6}\right] + \sum_{atom\ pairs} k\frac{q_i q_j}{\varepsilon R_{ij}} \quad \text{Eq. 2.4}$$

The above potential energy function includes the first four terms for intra-molecular interactions and the last two terms for inter-molecular interactions. Parameters K_r, K_θ, and V_n are the associated force constants for bond, angle, and torsion terms. $r - r_{\mathbf{i}}$ represents the change in bond length from its equilibrium length, $\theta - \theta_{\mathbf{i}}$ the angle from equilibrium between three consecutive atoms, the third term (n) the multiplicity, ϕ the dihedral angle, δ the phase shift, and $\varphi - \varphi_{\mathbf{i}}$ describes the deviations of the out-of-plane angle from an equilibrium angle [7].

The very last two terms describe intermolecular non-bonded interactions. The first term of intermolecular non-bonded interactions employs the Lennard-Jones (LJ) to account for the attractive and repulsive forces between two particles [7], where r_{ij} is the distance between these two particles, A is the attractive force, B the repulsive force Lennard-Jones well depth, and R_{ij} is the distance between two particles at the minimum potential which has been fitted from experiment data or quantum calculations. The second term employs a Coulombic potential to describe electrostatic interaction, where r_{ij} is the distance between atom i and j, q_{i} and q_{j} are the point charges of them respectively, and ε_{i} is the effective dialectic constant (permittivity of free space) [7].

2.2.2 Newton's Equations

After the determination of initial coordinates and velocities for a system of all atoms, a set of classical Newton's equations of motion (eq. 2.5) are integrated using numerical methods to obtain a trajectory charting the systems' motions over a user specified time period.

$$F = ma = m\frac{d^2R}{dt^2} \qquad \text{Eq. 2.5}$$

In the eq. 2.5, F is the force acting on an atom, m is the mass of the atom, a is the acceleration, and R is the position of the atom. The acceleration may also be expressed as the second derivative of the position of the atom with respect to time, t. Force can be rewritten as the gradient of potential energy:

$$F = -\Delta U(R) = -\frac{dU}{dR} \qquad \text{Eq. 2.6}$$

Combination of eq. 2.5 and eq. 2.6 leads to equation 2.7 which describes of the relationship of the derivative of the potential energy to changes in position with respect to time:

$$-\frac{dU}{dR} = m\frac{d^2R}{dt^2} \qquad \text{Eq. 2.7}$$

Once forces acting on atoms are calculated, the accelerations of all atoms can be obtained along with new positions and velocities derived within a short timestep. The potential energy function is used to calculate the forces on the atom in accordance to new positions. All data is saved over a specified time interval allowing for creation of a simulation trajectory. The fastest motions of macromolecular systems correspond to about 10 fs, so a timestep on the order of the fs is usually chosen [8].

To reduce computational expenses the SHAKE [9] algorithm constrains bonds associated with hydrogen in the system. This method uses the Verlet integration algorithm [10]

which employs the Taylor Expansions shown in eq. 2.10 and 2.11:

$$r_{n+1} = r_n + v_n\Delta t + \frac{1}{2}\left(\frac{F_n}{m}\right)\Delta t^2 + O(\Delta t^3) \quad \text{Eq. 2.8}$$

$$r_{n-1} = r_n - v_n\Delta t + \frac{1}{2}\left(\frac{F_n}{m}\right)\Delta t^2 - O(\Delta t^3) \quad \text{Eq. 2.9}$$

In equations 2.10 and 2.11, r is the position at the nth timestep, r_n is the position at its previous timestep, and r_{n+1} is the position at its next timestep. $O(\Delta t^3)$ is the function of Δt with n order or higher order. When 2.10 and 2.11 are summed one obtains equation 2.12:

$$r_{n+1} = 2r_n + \left(\frac{F_n}{m}\right)\Delta t^2 + O(\Delta t^4) \quad \text{Eq. 2.10}$$

The current position and force exerted on an atom at nth timestep as well as its previous position at nth -1 timestep is then used to update to a new atomic position at the nth +1 timestep [10].

2.2.3 Long Range Interactions

Dealing with long range interactions is difficult, especially with respect to Coulombic interactions where applying hard cut-offs introduces discontinuities into the forcefield calculations that can produce artifacts in simulations [11].

$$E_{dir} = -\frac{1}{2}\sum_n \sum_{i,j=1}^{N} \frac{q_i q_j \mathrm{erf}(B|r_j - r_i + n|)}{|r_j - r_i + n|} \quad \text{Eq. 2.11}$$

$$E_{rec} = \frac{1}{2\pi v}\sum_{m\neq 0} \frac{e^{\left(\frac{\pi^2 m^2}{\beta^2}\right)}}{m^2} S(\boldsymbol{m})S(-\boldsymbol{m}) \quad \text{Eq. 2.12}$$

$$E_{dir} = -\frac{1}{2}\sum_n \sum_{(i,j)=\epsilon M}^{N} \frac{q_i q_j \,\mathrm{erf}(B|r_j - r_i|)}{|r_j - r_i|} - \frac{\beta}{\sqrt{\pi}}\sum_{i=1}^{N} q_i^2 \quad \text{Eq. 2.13}$$

β represents the Ewald parameters, V is the volume of the unit cell, m is a lattice vector (m1, m2, m3) in reciprocal space, and $S(\mathrm{m})$ is the lattice structure factor.

$$S(\boldsymbol{m}) = \sum_{i=1}^{N} q_j e^{(2\pi i \boldsymbol{m}\cdot r_j)} \quad \text{Eq. 2.14}$$

The scale of Ewald summation is $O(N^2)$, where N is the number of atoms in the system. The Particle Mesh Ewald (PME) method reduces computational expenses by decreasing the scale of Ewald summation to $O(Nlog(N))$ [11]. In PME, a three-dimensional grid is placed over the system, atomic charges are mapped to the grid, and then fast Fourier transforms are used to sum these grid points [11].

2.2.4 Controlling Temperature using Langevin dynamics

In MD simulations, the total energy of the system is conserved. When the total number of atoms (N) and the volume of the unit cell (V) are fixed, the MD simulation is considered being conducted in what is called a microcanonical ensemble, a collection of different microstates belonging to the same macroscopic system. MD simulations can also be conducted in NVT characterized by fixed atom number, volume, and temperature or NTP ensemble which is characterized by fixed atom number, temperature, and pressure. In each case, temperature control is required and these investigations use Langevin dynamics to keep all systems at a constant temperature [11].

This method assigns a random force and a friction force for each step which is then related to fluctuation-dissipation theorem [12] to ensure sampling of the NVT data. The dynamics of atoms are described by Langevin's equation [11] which is represented by the following equation:

$$m\ddot{r}_i = -\nabla U - m\Gamma\dot{r}_i + W_i(t) \qquad \text{Eq. 2.15}$$

where m is the mass of the particle, U represents the particle interaction potential, Γ is determined from a Gaussian distribution, and $W_i(t)$ changes with a given temperature and times steps.

2.2.5 Pressure control

The NPT ensemble portion of the simulation allows the volume to fluctuate but fixes the pressure.

$$P = \frac{2}{3V}(E_{kinetic} + \Xi) \qquad \text{Eq. 2.16}$$

The virial theorem is employed to calculate pressure [13]. V is the volume of the unit box, E_{kinetic} is the kinetic energy, and Ξ is the inner virial tensor for pair-additive potentials defined by equation 2.19:

$$\Xi = -\frac{1}{2}\sum_{i<j} r_{ij} . f(r_{ij}) \qquad \text{Eq. 2.17}$$

Here, $f(r_{ij})$ is a force on particle *i* induced by particle *j*, and r_{ij} is the distance between the two particles. When the system is treated in an isotropic manner, the pressure becomes a scalar and can be expressed as [13]:

$$P = \frac{Tr(P)}{3} \qquad \text{Eq. 2.18}$$

Adjusting the inner virial Ξ by scaling distances of the inter particle allows for the correction of the Pressure during the simulation. A piston (σ) introduced into the equation of motion for each atom facilitates variation of the volume of the unit cell when employing a Nosé-Hoover Langevin piston barostat [13].

2.3 MD Analysis Techniques

2.3.1 Principal Component Analysis(PCA)

Principal Component Analysis is a statistical technique that serves to reduce the dimensionality of data needed to describe motions in proteins. Principal Component Analysis involves construction of a covariance matrix from atomic coordinates with covariance being defined using the following equation:

$$C_{ij} = (\vec{r_i}(t) - \langle \Delta\vec{r_i}(t) \rangle)(\vec{r_j}(t) - \langle \Delta\vec{r_j}(t) \rangle) \quad \text{Eq. 2.19}$$

Where C_{ij} is the covariance between atoms i and j, $\vec{r}$ is the displacement of atoms i and j, $\langle\ \rangle$ is represents a time averaged value. This value is positive when the atoms (or degrees of freedom) move in the same direction, negative when they move in opposite directions and is zero when these motions are independent from each other. The covariances are placed into a matrix which is then diagonalized. The eigenvectors (modes) derived from this process are then ranked by their associated eigenvalues. The eigenvectors describe the motions (directionality) of degrees of freedom within the system whereas the eigenvalues describe the magnitude of the covariances [14]. With this method, the first few modes typically describe a substantial portion of the protein's longest timescale motions which, for many systems, are those of the highest biological relevance [14].

2.3.2 Dynamical Network Analysis

To investigate allosteric communication between the Ligand Binding Domain and the co-activator Tif-2 in the presence of RJW100 and its derivatives, we employed Dynamical Network Analysis. Dynamical network analysis employs graph and network theory to describe coordinated motions within each system. Nodes selected for these studies were non-adjacent alpha carbons, Cα. Edges are drawn between each of the nodes based upon a distance matrix constructed per frame of the simulation. Within the distance matrix, edges are drawn between atoms that are within 4.5A of each other for 75% of the simulation's duration. The edge distance, d_{ij}, is determined by the log of pairwise correlations, -log[C_{ij}], where C_{ij} is described by the following equation:

$$C_{ij} = \frac{\langle \Delta\vec{r_i}(t) \cdot \Delta\vec{r_j}(t) \rangle}{\sqrt{\langle \Delta\vec{r_i}(t)^2 \rangle \langle \Delta\vec{r_j}(t)^2 \rangle}} \quad \text{Eq. 2.20}$$

where:

$$\Delta\vec{r}_i(t) = \vec{r}_i(t) - \langle\Delta\vec{r}_i(t)^2\rangle \qquad \text{Eq. 2.21}$$

With $\Delta\vec{r}_i(t)$ being the position of the atom that corresponds to the ith node [15].

2.3.2.1 Suboptimal Paths

These are paths most frequently used to transmit a signal. However, certain perturbations to the structure (changes to the environment) can result in the protein re-organizing itself in a way that creates another set of slightly longer paths (suboptimal paths) to transmit a signal. This is primarily why we employ suboptimal paths in these investigations. These pathways are likely to be very responsive to the binding of different effectors (small molecule agonists) which function as the source of perturbation to the system that induces rewiring [16].

2.3.3 Grid Inhomogeneous Solvation Theory

To further explore the potential role of crystallographic water molecules located in LRH-1's ligand binding pocket, we employed a method called Grid Inhomogeneous Solvation Theory (GIST) to probe the energetics of the water [17]. GIST allows a user to explore a specific area of interest within the protein whose solvent dynamics one is interested in. Changes in solvation patterns are often a key component of ligand recognition motifs. In this investigation, we sought to determine whether key water molecules located in the binding pocket along with RJW100 were not only dynamically stable as demonstrated by our collaborators, but energetically stable. Such stability would hint at a potential contribution in LRH-1's ability to recognize ligands.

This method discretizes equations of Inhomogeneous Solvation theory onto a grid in a volume of interest. Localized entropies, enthalpies, and free are calculated within each voxel.

Inhomogeneous Solvation Theory transforms integrals over molecular coordinates to integrals over distribution functions yielding expressions for various thermodynamic quantities expressed in terms of correlation functions [17].

These quantities ultimately contribute to the following expression:

$$\Delta G = \Delta E_{sw} + \Delta E_{ww} + \Delta S_{Trans} + \Delta S_{orient} \quad \text{Eq. 2.22}$$

This theory calculates solvation entropies in the following way:

$$\Delta S_{solv} = \Delta S_{ww} + \Delta S_{sw} \quad \text{Eq. 2.23}$$

ΔS_{sw} describes solute-water (sw) correlations and ΔS_{ww} water-water (ww) correlations. The solute-water term is defined as:

$$S^{\omega}(\boldsymbol{r}) \equiv \frac{-k_B}{8\pi^2} \int g_{sw}(\boldsymbol{r}, \omega) \ln g_{sw}(\boldsymbol{r}, \omega) \, \boldsymbol{dr} \boldsymbol{d}\omega \quad \text{Eq. 2.24}$$

This truncation only accounts for solute water correlations. K_B is Boltzmann's constant, $g_{sw}(r,\omega)$ describes the solute water pair correlation with the solute as the frame of reference. $1/8\pi$ is the normalization factor of the orientation integrals [17].

One may approximate ΔS_{solv} as ΔS_{sw} because the correlation integral g approaches unity for bulk solvent, and the solute-solvent correlation function approaches unity with increasing distance from the solute as well. Given this, the integrand decays to zero as distance increases between the solvent and solute and thus reasonable to treat S_{solv} as a local integral around the solute [17].

This allows the solute-water term to be broken down into orientational and translational entropies:

$$\Delta S_{sw} = \Delta S_{Trans} + \Delta S_{orient} \quad \text{Eq. 2.25}$$

Gist calculates these values by analyzing a volume that includes solute and a solvent region of interest. It discretizes these values into grid boxes k of volume V_k centered on a location r_k. Translational entropies can be calculated as follows:

$$\Delta S_{sw}^{trans}(\boldsymbol{r}_k) \equiv -k_B \rho^0 \int g(\boldsymbol{r_k}) \ln g(\boldsymbol{r}) d\boldsymbol{r} \qquad \text{Eq. 2.26}$$

Where this integral is over gridbox k and g(r) is treated as uniform over each grid box, and ρ^0 is the density of oxygens per voxel, where $g(r_k)$ is estimated based upon the amount of simulation frames through the following equation:

$$g(\boldsymbol{r_k}) = \frac{1}{\rho^0 N_f} \sum_{i=1}^{N_f} \frac{n_{i,k}}{V_k} \qquad \text{Eq. 2.27}$$

Here N_f is the frame indexing and n_{ik} is the number of waters per within box k per frame.

The total translational entropy is calculated via summation of $\Delta S_{sw}^{trans}(\boldsymbol{r}_k)$:

$$\Delta S_{sw}^{trans} = \sum \Delta S_{sw}^{trans}(\boldsymbol{r}_k) \qquad \text{Eq. 2.28}$$

Orientational entropies are calculated via the following equation:

$$\Delta S_{sw}^{orient}(\boldsymbol{r}_k) \equiv \rho^0 \int_k g(\boldsymbol{r})\, S^{\omega}(\boldsymbol{r}_k) d\boldsymbol{r} \qquad \text{Eq. 2.29}$$

With S^{ω}, the localized orientational entropies described as the following:

$$S^{\omega}(\boldsymbol{r^k}) \equiv \frac{-k_B}{V_k 8\pi^2} \int_k \int_r g_{sw}(\omega|\boldsymbol{r}) \ln[g_{sw}(\omega|\boldsymbol{r})] \boldsymbol{d\omega} d\boldsymbol{r} \qquad \text{Eq. 2.30}$$

and total orientational entropy described like so:

$$\Delta S_{sw}^{orient} = \sum_k n_{k,avg} S_k^{\omega`} \qquad \text{Eq. 2.31}$$

where $n_{k,avg}$ is the mean number of water molecules found in box k per frame of the simulation:

$$n_{k,avg} = \sum_{\boldsymbol{i=1}}^{\boldsymbol{N_f}} \frac{n_{i,k}}{N_f} \qquad \text{Eq. 2.32}$$

These values in combination with calculated energies summed over grid box k as described by the equation below:

$$\Delta E_{sw} = \sum E_{sw}(\boldsymbol{r}_k) \quad \text{Eq. 2.33}$$

Where:

$$E_{sw}(\boldsymbol{r}_k) = \int_{\boldsymbol{k}} \Delta E_{sw}^{loc}(\boldsymbol{r}) d\boldsymbol{r} \quad \text{Eq. 2.34}$$

can be used to calculate free energy values as described in [17].

2.4 Experimental Procedure

Model Construction

LRH-1·TIF2 complexes bound to agonists RJW100 (pdb 5L11) and 65endo were constructed. The crystal structures for these complexes were provided by the Ortlund group at Emory University. Chimera modeler was employed to model in the density between residues 527 and 530 using pdb 4DOS as the template structure.

MD Simulations and Selected Analysis Techniques

Wildtype LRH-1·TIF2 systems complexed with RJW100 and 65endo were simulated. In addition, we simulated a T352V mutant of LRH-1·TIF2 complexed with RJW100. All systems were solvated in a 0.15M NaCl, TIP3P waterbox with 10Å distances between the edges of the solute and the box. Simulations were run using pmemd in Amber14 [6] with the amber ff14SB and the general amber force field (GAFF) which is designed to parameterize organic molecules (drug-like compounds which contain O, S, N, P, H, and C) [18]. Key torsions in each ligand were optimized in Guassian 09 [19]. Point charges were assigned to each agonist using Antechamber in Ambertools and the AM1-BCC charge model. Bonded and short-range interactions were evaluated every 2fs, with long-range electrostatics evaluated with Particle Mesh Ewald method [11]. Short-range nonbonded interactions were evaluated using a 10Å cutoff. SHAKE [9] was employed to fix bonds between hydrogens and heavy atoms. Each system was subjected to 20,000 steps of minimization with the first 200 cycles employing

steepest decent and the remaining switching to conjugate gradient minimization. This process was evenly split (10,000 steps each). Solvent molecules were first minimized followed by the whole system. This was then followed by 500ps of dynamics in the NVT ensemble, during which the system was heated from 0 to 300 K with positional restraints applied to the solute (protein and small molecules). Finally, 8 NPT ensemble simulations were performed at 1atm for 2ns each, with each stage releasing the positional restraints, first on the protein, followed by each agonist. Sidechains were released before backbone atoms. 200ns of unrestrained NPT was performed. PCA and Dynamical Network analysis was performed on 10000 evenly spaced frames from the simulations and GIST was performed on 1000 frames of the same simulation.

Principal Component Analysis was performed on all backbone heavy atoms of LRH-1 bound to RJW100 and Tif-2 (pdb 5L11). Dynamical Network Suboptimal path analysis was limited to non-adjacent alpha Carbons (Cα) for the complexes bound to RJW100 and 65endo. The offset employed for suboptimal path generation was set to 50.

GIST Analysis involved placement of a Grid over LRH-1·Tif-2 and RJW100 with dimensions of 10^6 $\dot{A}^3$, with 0.5$\dot{A}$ spacing per voxel. The grid was centered on the geometric center of RJW100. One analysis of energetics restricted the area of interest to those within a 4.5$\dot{A}$ volume of agonist RJW100 using the procedure described in [17]. Another analysis was performed by grouping solvent properties in the region of interest as described in [17]. In this case, analysis was performed by imposing restrictions only including solvent with a water density(gO) above 1.0 and total (summed water-water and water-solute) enthalpy value less than -0.5 kcal/mol/$\dot{A}^3$ and integrating (summation of grid voxels) to find thermodynamic values as described in [17]. From here, thermodynamic values were derived for regions corresponding to water molecules of interest.

Virtual Screening Protocol

Our virtual screening protocol is outlined in scheme 1. Crystal structure binding poses of RJW100, 65endo, and 8AC served as templates. Openeye OMEGA was employed to generate conformers for each template [1]. All templates were screened with ROCS based upon shape overlap with NCI database compounds. The top 10000 were scored using Tanimoto Combo score and then docked to LRH-1 with Autodock vina [5]. Derivatives 8AC and 65endo are more polar and thus were also screened for electrostatic similarity after screening for shape similarity. To optimize overlap based on both electrostatics and structure, the top 10000 compounds in said cases were ranked through equal weighting and summation of associated Eon Tanimoto Combo and Tanimoto Combo scores [2,3]. The resulting set of compounds was also docked to corresponding LRH-1 structures. Structural waters of each structure were retained to more accurately represent specific water contacts within the binding pocket and to reproduce the binding poses found in crystal structures.

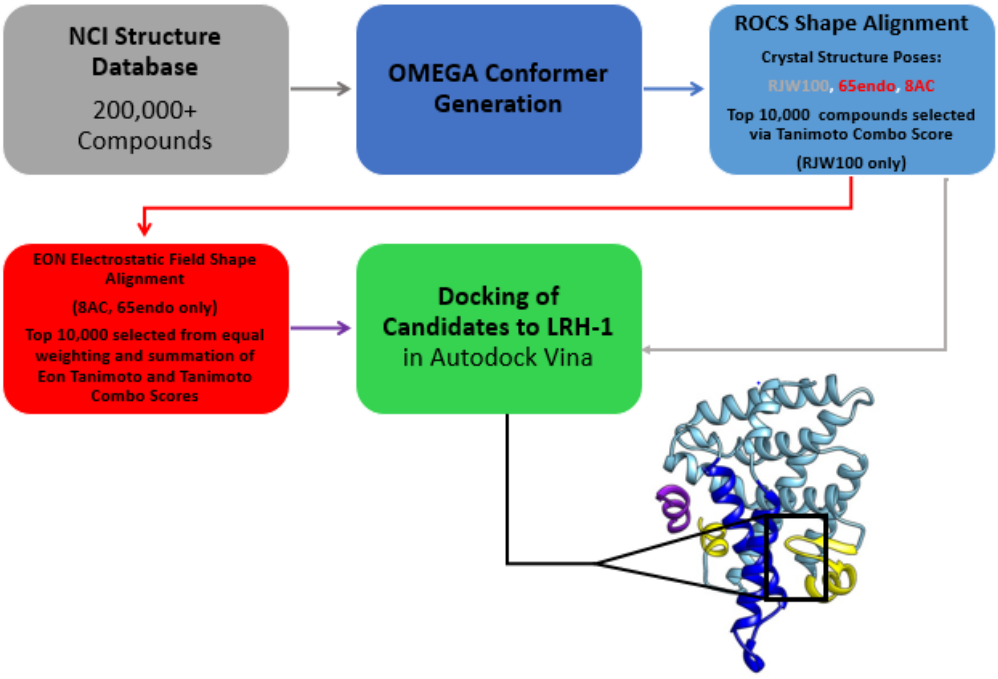

Figure 2.1. Protocol for Virtual Screening of LRH-1.
OMEGA is employed to generate several conformations of compounds from the NCI structure database which are then screened against crystal structure binding poses of each agonist. Compounds were screened against RJW100 only using ROCS shape alignment. Screening against 65endo and 8AC included both ROCS shape alignment and EON electrostatic field alignment. Candidates were docked to LRH-1·TIF2 LBP using Autodock vina

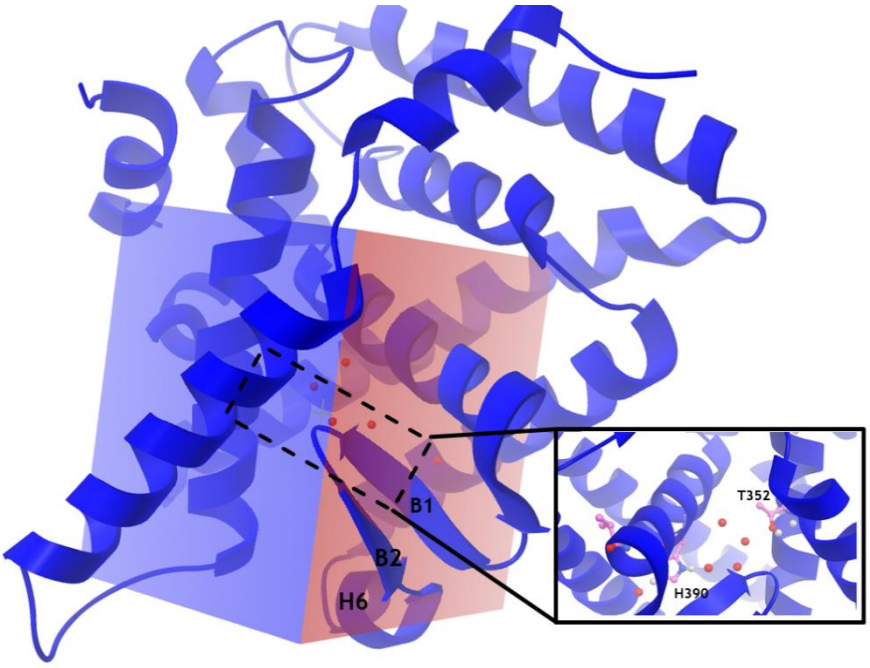

Figure 2.2. Search spaced employed for docking of candidates identified in the ligand based component of virtual screening.
The search space encompasses the entirety of the LBP. Water molecules were retained in the LBP.

2.5 References

1. Hawkins, P.C.D., et al., *Conformer Generation with OMEGA: Algorithm and Validation Using High Quality Structures from the Protein Databank and Cambridge Structural Database.* Journal of Chemical Information and Modeling, 2010. **50**(4): p. 572-584.
2. Hawkins, P.C.D., A.G. Skillman, and A. Nicholls, *Comparison of Shape-Matching and Docking as Virtual Screening Tools.* Journal of Medicinal Chemistry, 2007. **50**(1): p. 74-82.
3. EON 2.2.0.5: OpenEye Scientific Software, Santa Fe, NM. http://www.eyesopen.com.
4. Mohan, V., et al., *Docking: successes and challenges.* Current pharmaceutical design, 2005. **11**(3): p. 323-333.
5. Trott, O. and A.J. Olson, *AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading.* J Comput Chem, 2010. **31**(2): p. 455-61.
6. Case, D.A., et al., *The Amber biomolecular simulation programs.* Journal of Computational Chemistry, 2005. **26**(16): p. 1668-1688.
7. Chen, J. and K.N. Houk, Molecular Modeling: Principles and Applications By Andrew R. Leach. Addison Wesley Longman Limited: Essex, England, 1996. 595 pp. ISBN 0-582-23933-8.
8. Jo, J.C. and B.C. Kim, *Determination of proper time step for molecular dynamics simulation.* Bulletin of the Korean Chemical Society, 2000. **21**(4): p. 419-424.

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

10. Verlet, L., *Computer "Experiments" on Classical Fluids. I. Thermodynamical Properties of Lennard-Jones Molecules.* Physical Review, 1967. **159**(1): p. 98-103.

11. Darden, T., D. York, and L. Pedersen, *Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems.* The Journal of Chemical Physics, 1993. **98**(12): p. 10089-10092.

12. Hünenberger, P.H., *Thermostat algorithms for molecular dynamics simulations.* Advanced computer simulation, 2005: p. 130-130.

13. Kubo, R., *The fluctuation-dissipation theorem.* Reports on Progress in Physics, 1966. **29**(1): p. 255.

13. Feller, S.E., et al., *Constant pressure molecular dynamics simulation: The Langevin piston method.* The Journal of Chemical Physics, 1995. **103**(11): p. 4613-4621.

14. David, C.C. and D.J. Jacobs, *Principal Component Analysis: A Method for Determining the Essential Dynamics of Proteins.* Methods in molecular biology (Clifton, N.J.), 2014. **1084**: p.193-226.

15. Stone, J., et al., *Dynamical Network Analysis.* 2012.

16. Bhattacharyya, M., S. Ghosh, and S. Vishveshwara, Protein structure and function: looking through the network of side-chain interactions. Current Protein and Peptide Science, 2016. 17(1): p. 4-25.

17. Nguyen, C.N., T.K. Young, and M.K. Gilson, *Grid inhomogeneous solvation theory: Hydration structure and thermodynamics of the miniature receptor cucurbit[7]uril.* The Journal of Chemical Physics, 2012. **137**(4): p. 044101.

18. Wang, J., et al., *Development and testing of a general amber force field.* Journal of Computational Chemistry, 2004. **25**(9): p. 1157-1174.

19. Frisch, M., et al., *G09 Gaussian Inc.* Wallingford, CT, 2009.

3 RESULTS AND DISCUSSION

3.1 PCA and Cross-correlation Analysis

We conducted PCA studies on LRH-1·TIF2 in presence of RJW100. PCA allows us to specifically focus on major motions that occur upon binding of RJW100. We analyzed the first four modes, identifying two principle components of interest. A breathing motion at the mouth (characterized by H2, H3, H7, and β-H6) of the LBP, between H2 and H6 designated as PC2, and a breathing motion between H8 and H9 as PC1 as implied by the porcupine plots (Fig. 3.1C, D). Of note in the porcupine plots is how PC1 (Fig. 3.1C) features H2, H3, and β-H6 converging inward towards one another other as well as an outward swing of H9 relative to H8 which mimics the profile of DLPC complexed to LRH-1·TIF2 (Fig. 1.7A) [1]. Figure 3.2D shows β-H6 and H2 and H3 swinging outward with respect to each other. Plotting these PCs *versus* one another allows for another way to visualize the result (Fig. 3.1A). PC1 is clustered in one region of the PC1-PC2 subspace around a centroid of 10 whereas PC2 is divided into two distinct clusters clustered around centroids of 10 and -10. These patterns are not consistent with LRH-1·TIF2 systems possessing an agreement status as identified in [1]. However, it is possible that the difference in the patterns observed reflect the differing binding modes of the compounds. The contacts that DLPC makes with the mouth of LRH-1's LBD may explain the single cluster of PC1-PC2 conformational space (Fig. 1.7C) [1]. RJW100 does not engage in these same contacts and instead binds deep inside of the LBP. This suggests that highly concentrated subspace at the mouth of the LBP is not a requirement in transcription-activated conformations when LRH-1 is bound to non-PL agonists.

Despite the dissimilarities with PL studies, these results are consistent with hydrogen deuterium exchange (HDX) mass spectrometry studies comparing the dynamics of the DLPC

complex to RJW100 (Fig. 1.11). In comparison to DLPC, RJW100 binding results in rapid exchange with deuterium in β-H6 and H10, which all contribute to the mouth of LRH-1's binding pocket [2]. This rapid exchange suggests that the mouth of the LBP is more dynamical *versus* DLPC and our PCA results indicate that two distinct conformations exist about the mouth of the LBP.

High levels of LRH-1-mediated transcriptional activity require selective co-activator recruitment and HDX studies of DLPC binding *versus* longer tailed PLs imply that some destabilization of the LBP (specifically β-H6) is associated with enhanced selectivity [3]. Computational studies reveal that DLPC binding results in highly correlated motions between H4-7 and anti-correlated motions between these helices and the AF-H (Fig. 3.2A) and it is suggested that this profile of correlated motions predicts high levels of activation [4]. Figure 3.1B, featuring the RJW100 complex shows a different pattern where there are no particularly highly correlated or anti-correlated motions between the same regions. Instead, the cross-correlation patterns that characterize RJW100 binding resemble systems with disagreement statuses (Fig. 3.2, *B versus C*) [1] except that motions in H3-5 and H7-9 as well as those in H1-H3 and the AF-H are more highly correlated than those in both the Apo LRH-1·TIF2 and the DLPC complexed systems. These results reinforce the idea that the profile of correlated motions in DLPC complexed LRH-1 are not a requirement to enhance levels of transcriptional activity significantly above basal levels, but instead may be an essential characteristic of activated structures that trigger very high levels of transcriptional activity [2].

3.2 Investigation of Solvent Energetics within LRH-1's Ligand Binding Pocket

RJW100 binding affinity is largely attributed to the ability of Thr352 to anchor the ligand via a water-mediated contact between itself and the hydroxyl moiety of RJW100 [2]. The crystal

structure features four water molecules within the binding pocket, with one of the four contacting the ligand as described. Our computational studies provide some preliminary evidence that the four proximal waters identified in the crystal structure are energetically stable in those positions which bolsters other studies that document their stability [2]. Grid Inhomogeneous Solvation Theory (GIST) [3] was employed to elucidate whether the waters molecules occupy thermodynamically favored positions within the LBP. We evaluated thermodynamic parameters of key water molecules and Figure 3.3 demonstrates that energetically favorable positions for water occupancy exist in regions corresponding to the structural water molecules. The red wireframes indicate regions (voxels within the grid) within LRH-1 that have ΔH_{sw} of -1.5 kcal/mol/$\dot{A}^3$. When GIST analysis was performed on a T352V mutant, there was complete absence of voxels corresponding to energetically favorable waters within the LBP. The four waters labelled alphabetically are proximal to RJW100 and their thermodynamic properties are of high interest. The water molecule labelled A participates in the bridging interaction between RJW100's hydroxyl moiety and Thr352 so is of special interest.

To analyze the energetics of a substantial portion of LRH-1's LBP, solvation thermodynamic parameters in voxels within a distance of 4.5$\dot{A}$ of RJW100 heavy atoms were integrated and Table 3-1 presents this data and reveals the dominance of enthalpic contributions to a favorable free energy (-61 kcal/mol).

Table 3-1 Integrated thermodynamic data of solvent within 4.5Å of RJW100.

All units are in kcal/mol.

ΔH	**TΔS**	**ΔG**
-71	-10	-61

The data in Table 3-2 shows a favorable Gibbs free energy for the four regions proximal to RJW100. This data was obtained from a group of waters that correspond to 1.0 oxygen density (every voxel with density equal to or greater than bulk solvent), but had enthalpies less than or equal to -0.5 kcal/mol.

Table 3-2 Thermodynamic data of solvent groups in closest proximity to RJW100.
All units are in kcal/mol**.**

Group	**ΔH**	**TΔS**	**ΔG**
A	-7	-1	-6
B	-13	-2	-11
C	-3	-1	-4
D	-6	-1	-5

The data indicates that these regions account for a substantial portion of the energetically favorable positions for water occupation within the LBP. The favorable free energies are due to favorable enthalpic contributions. Group A and D's relative ranking can be rationalized by inspecting potential contacts within the binding pocket. The structure shows that the water

corresponding to group D may engage in a contact with His390 and that the group corresponding to A engages with Thr352 as well as RJW100's hydroxyl moiety during the simulation. Given the interactions of A, it is unsurprising that its position is at least slightly more favored than D's. However, it is surprising that group A does not correspond to the region with the most thermodynamically favored properties. Instead, group B does by a substantial margin having a free energy nearly two times more favorable than group A (-10 kcal/mol *versus* -5 kcal/mol). This is interesting considering that the water-mediated contact with Thr352 is speculated to be one of the most important components contributing to RJW100's ability to bind and activate LRH-1. This suggests that any water associated with the region of the LBP corresponding to B is better anchored than the other three sites and is also consistent with the observation that the water-mediated contact with Thr352 is somewhat labile. Water molecules facilitating this contact are exchanged during the simulation (water leaves the position and is then replaced by another).

Inspection of the X-ray crystal structure of TIF2·LRH-1 in complex with 65endo reveals the displacement of two of the crystallographic waters including the one that facilitates contact with Thr352 in the RJW100 complexed structure. The water molecule corresponding to group B remains. This observation supports the relative rankings reported in Table 3-2. Furthermore, these results predict that 65endo more effectively binds to LRH-1. Its ability to displace two of the water molecules implies that there are a set of compensatory interactions between 65endo and LRH-1's LBP that allow for the two waters to be displaced. The sulfonamide functional group replacing the hydroxyl of RJW100 extends into the space occupied by group A and C (with C corresponding to the least favored region among the four). In the case that a new

structural scaffold is discovered to activate LRH-1, potential modifications can be explored more strategically based upon solvent properties as suggested in [3].

3.3 Investigation of RJW100 Induced Allosteric Communication

We investigated the suboptimal paths generated between the LBP and TIF2 as the number of suboptimal paths allows us to gauge the strength of communication as well as key structural features contributing to communication between them. In each figure, blue lines represent edges drawn between nodes (alpha carbons) in suboptimal paths.

Our studies with LRH-1 in complex with RJW100 reveal a similar network to that triggered upon DLPC binding in terms of structural components traversed by the paths. Specifically, suboptimal paths heavily traverse helix 5. RJW100 induces the generation of 73 suboptimal paths (Table 3-3). By visual inspection of Figure 3.4A, the amount of communication triggered is lower than DLPC complexed LRH-1·TIF2 (Fig. 3.4B), and features diminished contribution of helix 3 to the network. Despite the differing structures and binding modes, these results suggest that non-PL activators can induce a similar communication network to that induced upon DLPC binding. However, the strength of communication triggered by RJW100 appears to be intermediate between DLPC complexed LRH-1·TIF2 (agreement status) and systems with disagreement statuses such LRH-1·TIF2 complexed with *E. coli* PLs (Fig. 1.8B).

In addition to this study, we performed a suboptimal path analysis on the T352V mutant. Figure 3.5 features both systems and shows that the T352V mutant (Fig. 3.5B) features somewhat diminished communication. While the two appear similar, data reveals that RJW100 induces only 54 suboptimal paths in the mutant structure.

Table 3-3 Number of Suboptimal Paths generated upon binding of agonists.
Both wildtype and T352 mutant for RJW100 complexed LRH-1·TIF2 are presented.

Complex	RJW100 wt	RJW100 T352V	65endo wt
Number of SOPs	73	54	2375

This suggests that the threonine does contribute to the ability of RJW100 to not only bind LRH-1, but to trigger communication. In the context of LRH-1-mediated transcriptional activity, this is also consistent with studies that show diminished, but not abolished transcriptional activation by the T352V mutant complexed to RJW100 in comparison to wildtype LRH-1 (Fig. 1.12). RJW100 triggers a 2.5-fold activation in wildtype protein but this diminishes to about a 1.7-fold activation when bound to the T352V mutant [2].

This result is interesting in conjunction with the analysis of thermodynamic data for key water molecules participating in the proposed water network. Thr352 for example belongs to H3 and His390 to H5, which function as the primary structural feature that propagates the signal. Motions at the mouth of the protein that include helices 2 and 3 may partially explain how DLPC is able to induce communication. Perhaps RJW100, through its exploitation of the water network is also able to do this, albeit in a manner that does not result in full activation. The water-mediated contact between RJW100's hydroxyl function and Thr352 may serve to link the local motions at helix 3 to those at helix 5 through via the identified water network. Given the indirect nature of these contacts, it is no surprise that RJW100 is unable to trigger activation as well as an agonistic PL such as DLPC, but these contacts may provide an alternative to DLPC's binding mode that explains the relative success of the bicyclo-octene scaffolds.

3.3.1 RJW100 Derivative 65endo Enhances Allosteric Communication

We also probed the nature of the allosteric network in the presence of RJW100 derivative 65endo which features a sulfonamide moiety in lieu of the hydroxyl function. Differences

between the network triggered by 65endo and RJW100 can provide rationale for future modifications to RJW100, producing additional compounds with increased efficacy and binding affinity.

Indeed, the 65endo agonist induces a very high level of communication between the LBP and co-regulator binding cleft upon binding. The communication is much stronger than that which occurs upon RJW100 binding as 65endo triggers 2375 suboptimal paths (Table 3-3). This is corroborated by experimental evidence that 65endo increases activity of LRH-1 relative to RJW100. The more polar sulfonamide group may be triggering activation by making more direct contacts with the polar patch within the interior of the LBP. As shown in Figure 3.6, the set of suboptimal paths generated upon 65endo binding share similarities with RJW100 and DLPC bound systems with most of the communication being propagated through helix 5. However, like RJW100, the network differs from DLPC's in that the paths generated upon 65endo binding do not traverse helix 3.

Analysis of frequently traversed nodes reveal many similarities but some subtle differences in how the signal is propagated in each small molecule bound complex as indicated by the histograms in Figure 3.7. The data confirms that the signal is propagated through helix 5 with Leu378 and Trp382 being heavily traversed in all three networks. Both mutant and wildtype systems with RJW100 bound feature Arg393 as a heavily traversed node. Arg393 is located closest to the LBP (Fig. 3.7D) and in both RJW100 bound systems appears to be the primary, if not only receptor of paths originating from β-H6. When bound to 65endo, however, the data corroborates what is shown in Figure 3.6B where divergent patterns exist in paths near the LBP. Data for 65endo in Figure 3.6C shows two nodes of H5 instead of a single node (as is the case with the RJW100 complexes) being heavily traversed, those associated with Leu391 and

Gln394. However, patterns observed in paths in closer proximity to TIF2 are very similar in RJW100 and 65endo complexes. Both employ Met375 and Leu378, but the 65endo complex features more suboptimal paths traversing Leu378 than Met375, the converse of the pattern observed for the RJW100 complex. 65endo binding also results in suboptimal paths that heavily traverse Leu391 and Gln394, a pattern not seen in either of the RJW100 complexes. This results in the 65endo complex featuring an expanded profile of heavily traversed nodes. The increased number of suboptimal paths generated in addition to the slightly wider distribution of heavily frequented nodes implies that 65endo induces a more robust profile of correlated motions and that there is less discrimination between nodes employed to propagate the signal.

The crystal structure of the 65endo complex indicates a diminished distance between Glu534 of the AF-H and TIF2 co-activator versus RJW100 complex as shown in figure 3.8. This is similar to a phenomenon observed in the DLPC complexed with LRH-1·TIF2 (Fig. 1.6) [1]. Figure 3.8A shows that Glu534 is positioned 3.53Å away from the amide bond of Leu744 in TIF2 and an orientation more similar to the DLPC complex (which allows for contact with TIF2) whereas Figure 3.8B shows the RJW100 complex with a distance of 4.36Å as well as an orientation not conducive to making a contact with TIF2 much like systems with disagreement statuses. This comparison also presents another example of RJW100 binding inducing changes in LRH-1 that are intermediate between systems in agreement and disagreement.

The recurrence of certain patterns in transcription activating structures implies that the positioning of Glu534 in LRH-1·TIF2 crystal structures provides some predictive power in determining how effectively an agonist can trigger allosteric communication. However, experimental data shows that the AF-H is highly dynamical in the presence of DLPC. Because LRH-1 does not form a true charge clamp with co-activating peptides, it may also be instructive

to gain insight from elsewhere. One may expect, for example, a profile of correlated motions that more closely resembles that of transcription activated systems with agreement statuses identified in [1]. Another component generally considered when investigating allostery in proteins is larger conformational changes whether they be large shifts in position of secondary structural features or reorientation of sidechains [4-6].

In totality, this result suggests that modifications to the RJW100 bicyclo-octene scaffold at the hydroxyl function could be an optimal method to design agonists that further enhance transcriptional activation (though modifications of this moiety require caution because it could destabilize the compound, which is a major drawback of GSK8470) by exploiting contacts in the interior of the LBP in a manner that creates more direct contacts between the agonist and LRH-1. Further studies could potentially explore the fidelity and organization of the water network when 65endo is bound as the increased number of donor and acceptor atoms resulting from the modification not only presents the possibility of direct contacts, but water-mediated contacts as well. These investigations in addition to any further studies conducted to explore coordinated motions such as PCA or cross-correlation analysis could serve to partially explain the enhanced communication observed.

3.4 Virtual Screening

3.4.1 RJW100 as Lead Compound Yields Hydrophobic Pharmaceutical Targets of LRH-1

Through use of the X-ray structure provided by collaborators, we used RJW100 complexed to LRH-1·TIF2 as a template for a shape-based virtual screening effort employing Openeye software. Representative structures (top 12 highest performers as based upon Autodock vina docking score) are presented in Figure 3.9 below along with the docking score. As expected, these compounds mostly possess a polycyclic, hydrophobic core. Many of these

structures have very similar scaffolds. Specifically, 26684, 30879, and 102943 have scaffolds that only feature tricyclic systems linked to one another. In addition, most of these very top scoring compounds in terms of donor and acceptor capabilities lack diversity. For example, 9 compounds only feature groups capable of accepting a hydrogen bond (109840, 119443, 114934, 125037, 55220, 26684,1 02943, 128597, 400595). This is unideal when seeking scaffolds with features highly complementary to those of the polar patch within LRH-1's LBP, however molecules with any hydrogen bond acceptor or donor capabilities are more likely to bind the target selectively, so this is a promising result overall. We have hope that some of the candidates will be revealed as bio-active in screening which would allow for scaffold hopping away from the RJW100 architecture.

3.4.2 Virtual Screening of Polar Derivatives

The structural component of our virtual screening effort for RJW100 derivatives yielded interesting results especially when the two derivatives are compared *versus* one another. 8AC screening resulted in subtle changes to the profile of the top performing compounds (in context of the docking score) shown in Figure 3.10 with two of these compounds, 26684 and 30789 (which is nearly identical to 30798 from the RJW100 set being identical to top performers derived from RJW100 as lead. The overlap in these results can at least be partially attributed the fact that 8AC overall, has a very similar binding mode and shape to that of RJW100. However, one would expect the electrostatic profile of 8AC to be different overall, and this does seem to give rise to features seen in compounds such as 726771 and 719649. These two compounds appear to reflect both the shape and electrostatic profile of 8AC in the sense that 8AC features two polar regions that are distal from one another (the hydroxyl group located within the bicyclic core and the ester that terminates the lipophilic tail) and 726771 and 719649 separate their most

polar regions by a linker. A promising feature of this set is that 7 compounds including 368290, 319072, 726771, 116643, 30789, 658172, and 154683 contain a mixture of hydrogen bond donor and acceptor atoms.

Top performers yielded from the screening of 65endo (Fig. 3.11) share no overlap with those of RJW100 or 8AC (though 116653 and 116643 are very similar), but outside of this, results are somewhat disappointing in terms of structural diversity. The profile of these candidates is similar to the RJW100 set except that most compounds are not completely lipophilic (completely lacking hydrogen bond acceptors and donors). However, five of these compounds are only capable of accepting hydrogen bonds: 255296, 106668, 723898, 368003, and 230368. The most promising feature of this set of compounds is the inclusion of three very polar compounds: 375090, 291853, and 723898. Compounds with this feature align with the goal to find scaffolds with high solubility. Another key feature is the notable shift in the docking scores in this set of compounds to a slightly lower range indicating more favorable binding of these compounds relative to those in other sets. However, great care must be taken when interpreting these docking scores which only serve to rank compounds relative to each other in a specific context.

The protocol employed in this investigation adds complexities to interpretation of the results and the ability to compare between sets. A key reason for this is the retention of the water molecules in the ligand binding pocket of each crystal structure (Fig. 2.1). Typical protocols for rigid docking include the stripping of all water molecules, ligands, and salt molecules. The arrangement of the crystallographic waters differs at least slightly in each complex (with the biggest difference seen in the crystal structure complexed with 65endo where two waters found in RJW100 complexed LRH-1 have been displaced). We chose to retain the waters as their

retention resulted in reproduction of crystal structure binding poses for each ligand (with a somewhat disordered aliphatic tail in each case). Given this, the scores and potential binding poses of compounds in each set only reflect the binding pose that can be achieved with the specific water arrangement associated with the LBP of the structure from which it was derived.

It may be possible that a new protocol could be employed where each complex is completely stripped before docking compounds yielded from the ligand based component of the screening protocol. However, mutagenesis studies show the importance of Thr352 in activation of LRH-1 by GSK8470 whose binding mode features no water-mediated contacts. This implies that the existence of waters in the LRH-1 LBP are important for the activation by non-PLs in general [3].

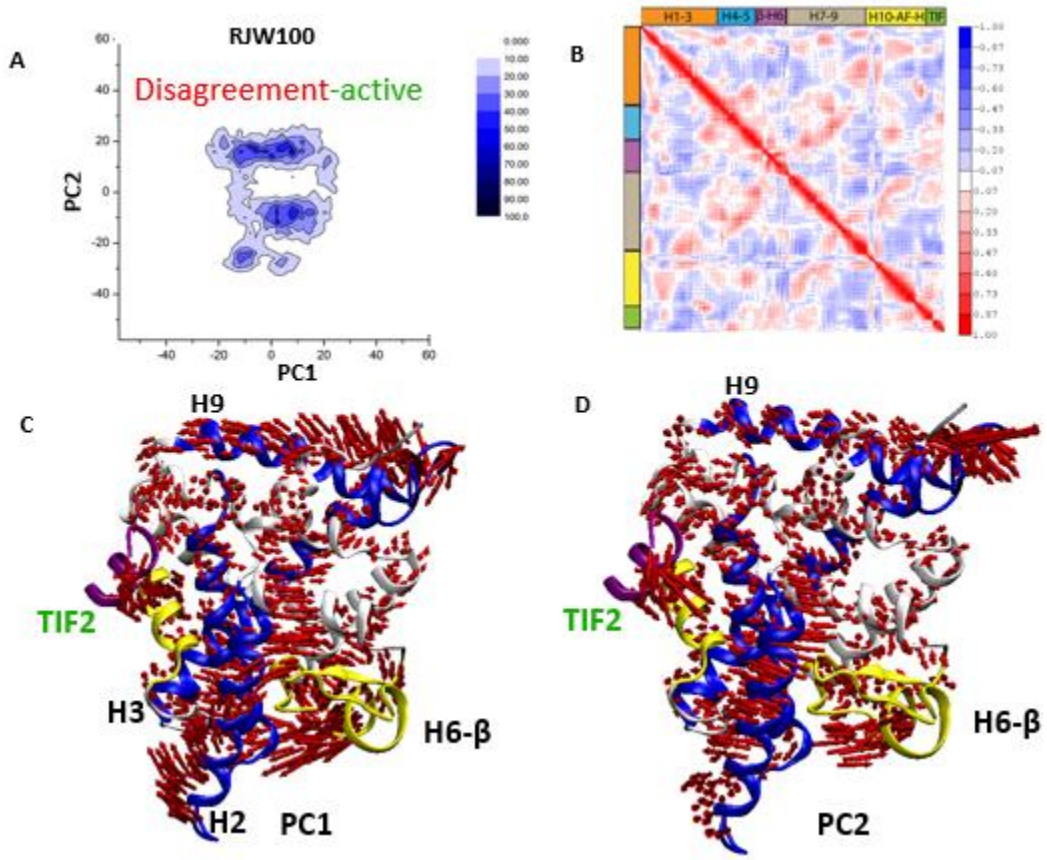

Figure 3.1. Correlated Motions in LRH-1·TIF2 when complexed to RJW100.
A) Two chosen principal components plotted versus one another. The higher the density, the more snapshots (frames out of 10000) occupying that portion of the subspace. B) Cross correlation plot of LRH-1·TIF2 complexed with RJW100. As indicated by the scale, blue areas indicate anti-correlated motions and red indicates correlated motions. C) Porcupine representation of PC1 modes projected onto LRH-1·TIF2 structure complexed with RJW100. The arrowheads indicate direction of motion and length of arrows indicate magnitude of motion. D) Porcupine representation of PC2 modes projected onto LRH-1·TIF2 structure complexed with RJW100.

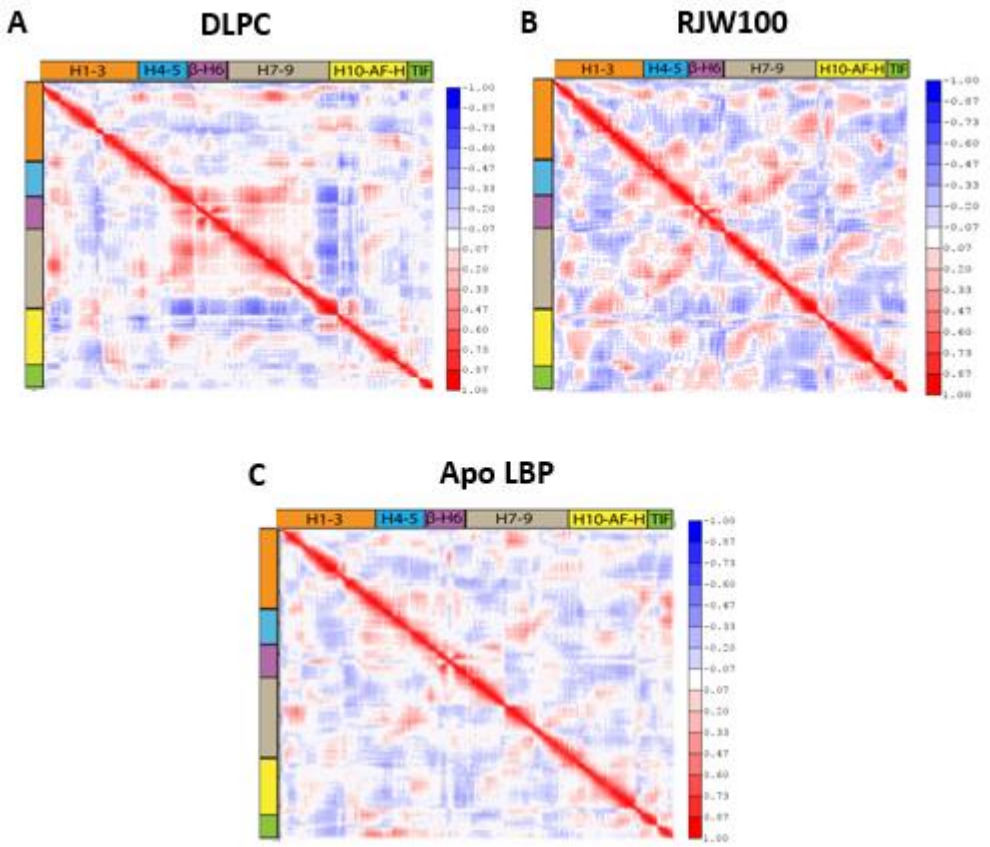

Figure 3.2. Cross Correlation plots of LRH-1·Tif-2 with varied LBP statuses.
A) DLPC. B) RJW100. C) Apo LBP. Figures A) and C) have been adapted from *Musille, P.M., et al., Unexpected Allosteric Network Contributes to LRH-1 Co-regulator Selectivity.* J Biol Chem, 2016. **291**(3): p. 1411-26 [1].

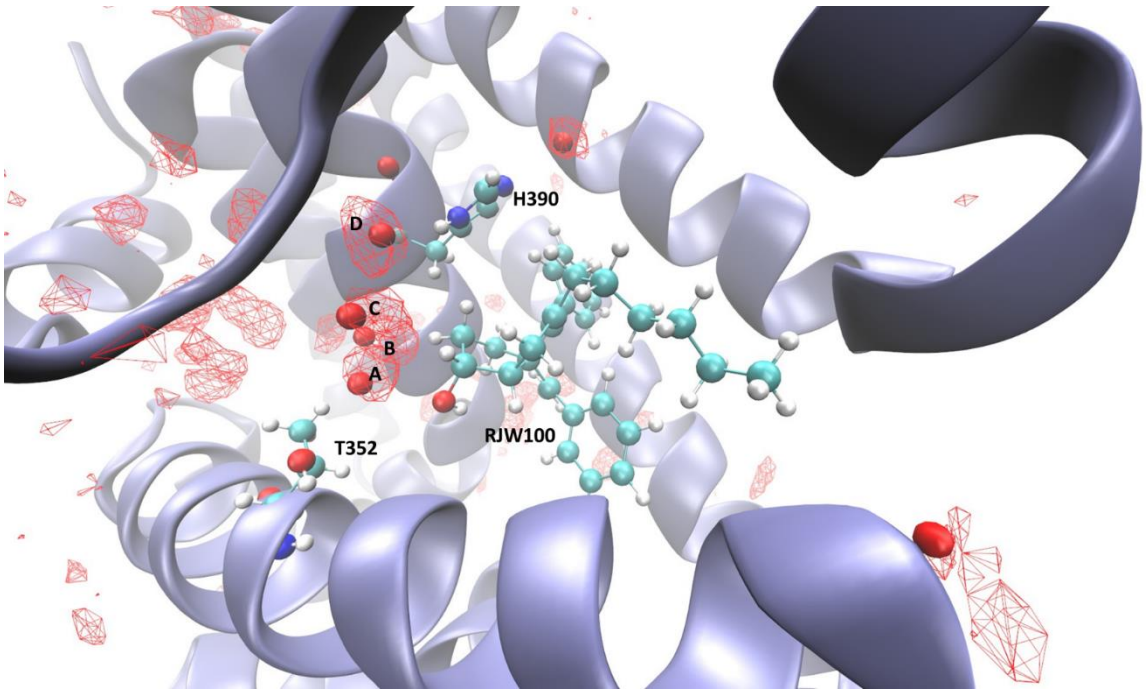

Figure 3.3. Areas of energetically favored water-solute interactions within LRH-1·TIF2 LBP in complex with RJW100.
A, B, C, and D (wireframes) correspond to the regions of water density that were integrated to derive thermodynamic parameters in table 3.1. The original crystal structure (pdb 5L11) of RJW100 bound LRH-1·TIF2 was overlaid on top of GIST [7] water density data to show the crystallographic waters of interest and how they correspond to these densities.

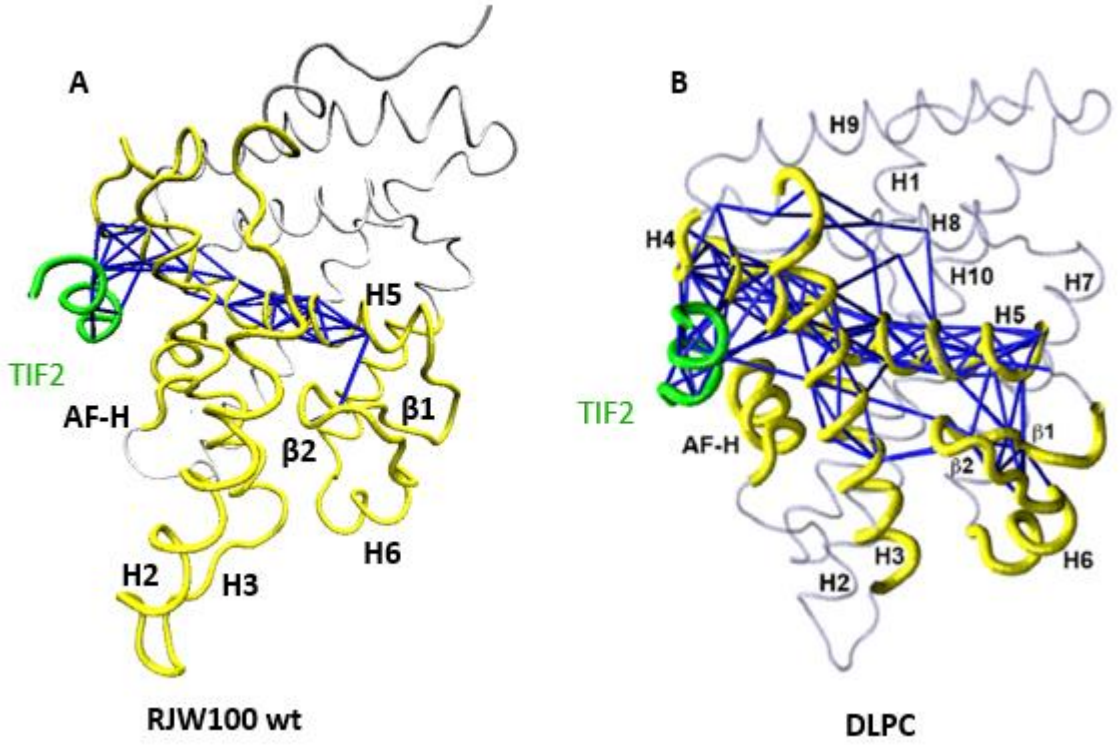

Figure 3.4. Comparison of Suboptimal Paths between LBD and TIF2 co-activator.
A) LRH-1·TIF2 complexed to RJW100. B) LRH-1·TIF2 complexed to DLPC as adapted from Musille, P.M., et al., *Unexpected Allosteric Network Contributes to LRH-1 Co-regulator Selectivity.* J Biol Chem, 2016. **291**(3): p. 1411-26 [1].

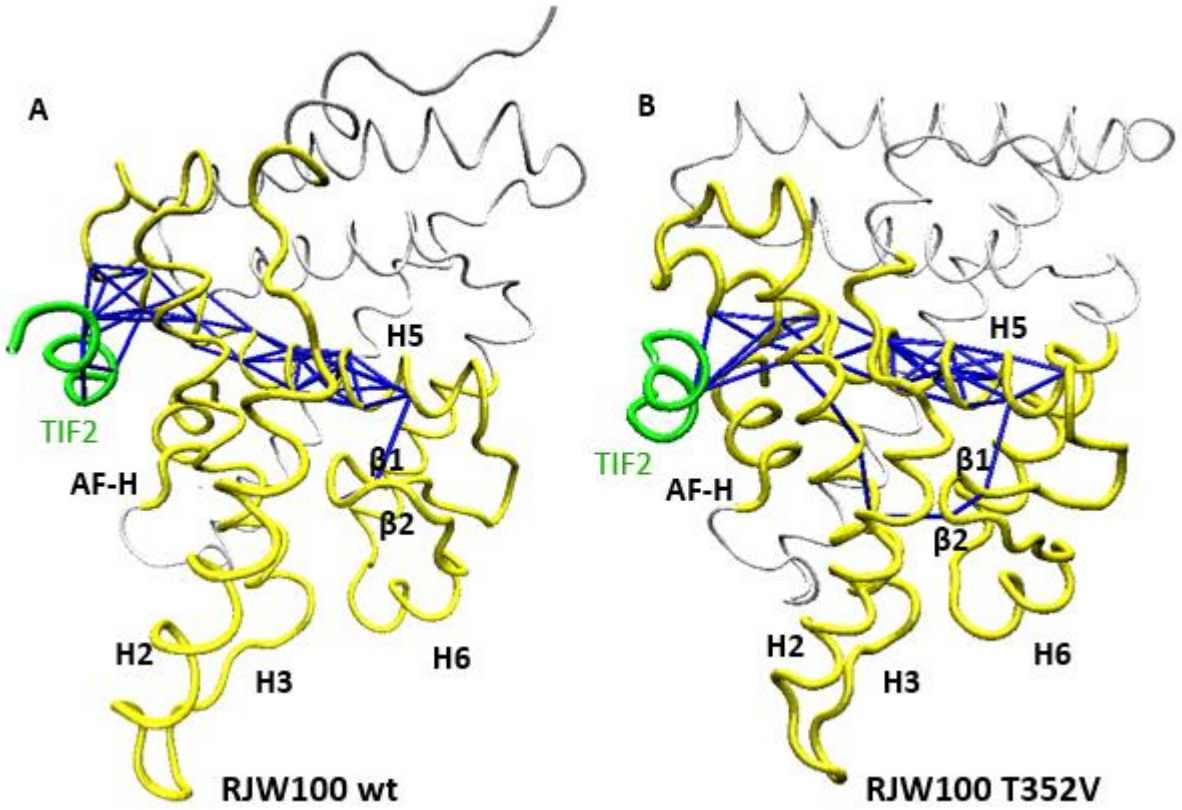

Figure 3.5. Comparison of Suboptimal Paths between LBD and TIF2 co-activator.
A) RJW100 bound LRH-1·TIF2. B) RJW100 bound to LRH-1·TIF2 T352V mutant.

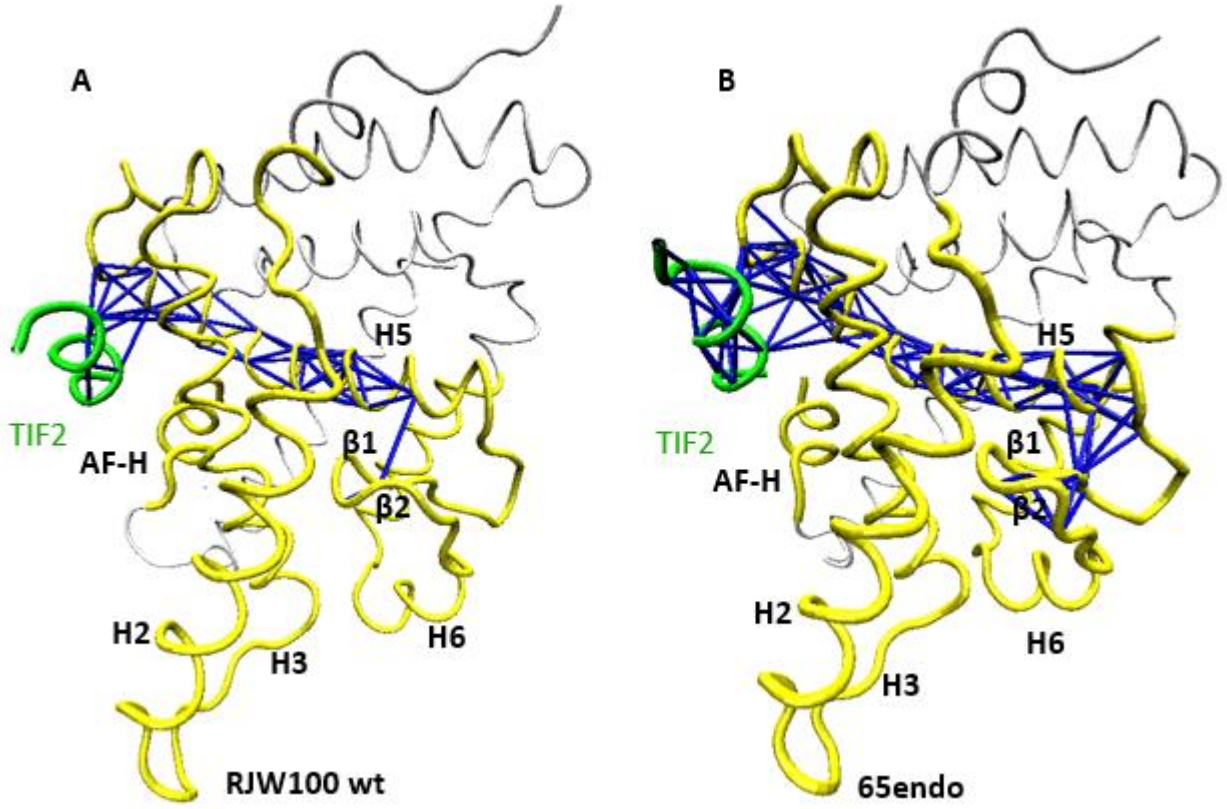

Figure 3.6 Comparison of Suboptimal Paths (blue) between LBD and co-activator TIF2(green).

A) RJW100 complexed to LRH-1·TIF2. B) 65endo complexed to LRH-1·TIF2. There is a noticeable increase in the amount of suboptimal paths available upon binding of 65endo.

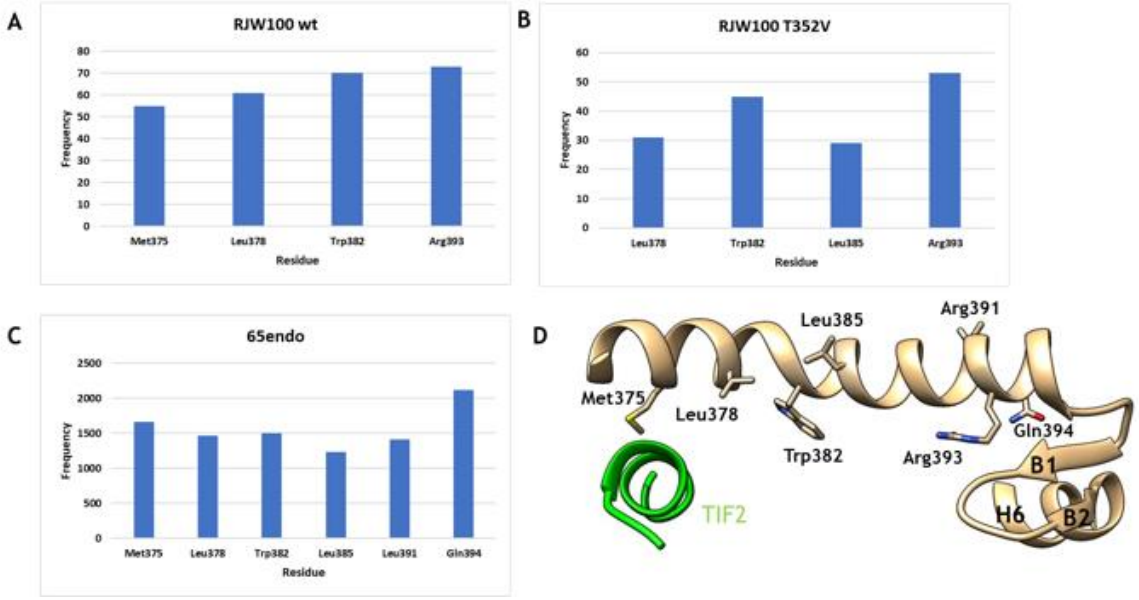

Figure 3.7. Heavily traversed nodes in suboptimal paths.
A) RJW100 complexed to wildtype LRH-1·TIF2. B) RJW100 complexed to T352V mutant. C) 65endo to wildtype LRH-1·TIF2. D) A representation of LRH-1 featuring helix 5, β-H6 (source), and TIF2 co-activator (sink) and sidechains of corresponding residues in highly traversed nodes indicated in histograms A-C.

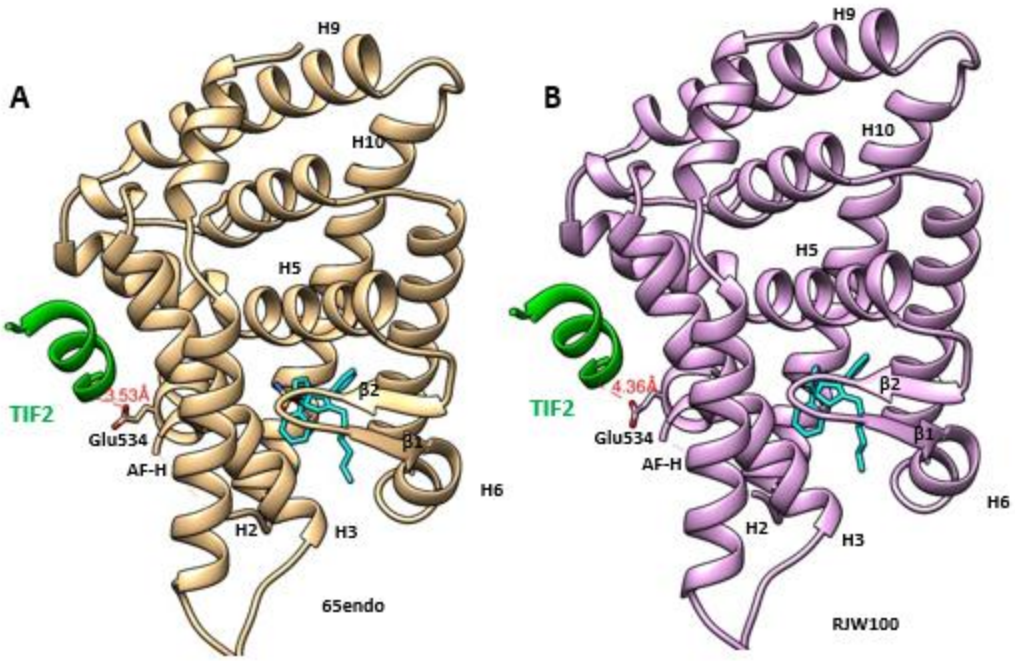

Figure 3.8. Comparison of 65endo and RJW100 complexed crystal structures. A) 65endo complex with distance between Glu534 and TIF2 (green) co-activator peptide displayed in red. B) RJW100 (cyan) complex with the distance between Glu534 and TIF2 co-activator peptide displayed in red.

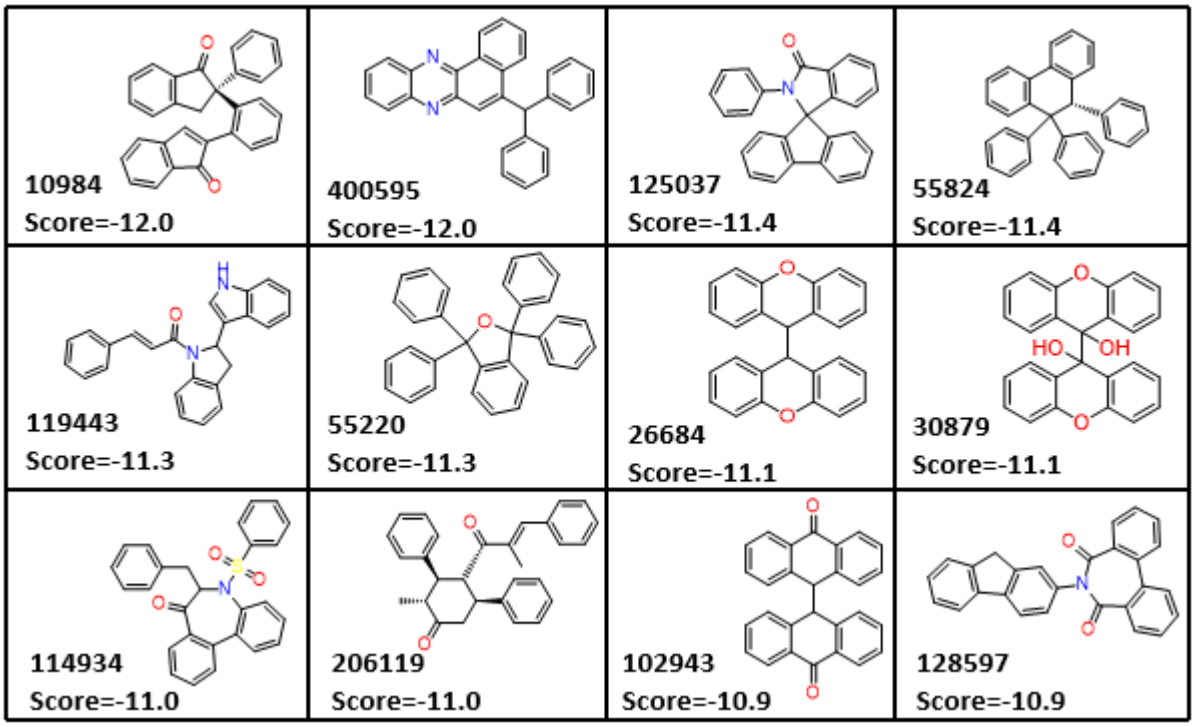

Figure 3.9. Top 12 scoring candidates from the virtual screening of RJW100. All scores are measured in kcal/mol.

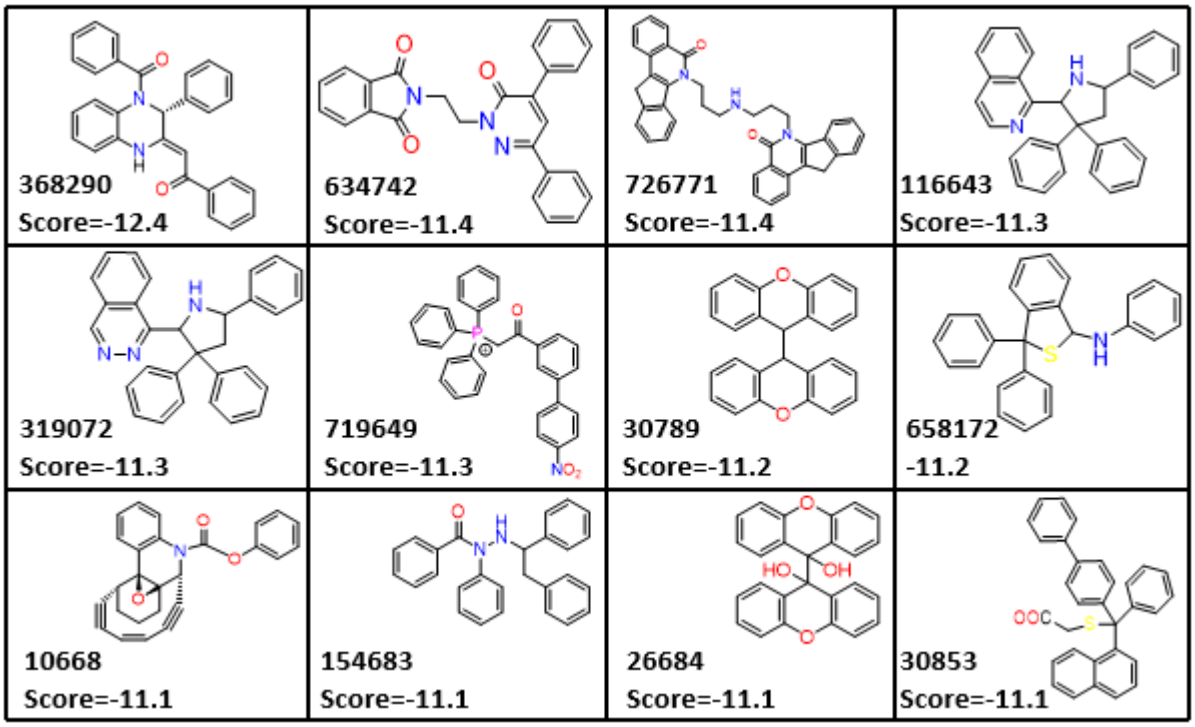

Figure 3.10. Top 12 scoring candidates from the virtual screening of 8AC. All scores are measured in kcal/mol.

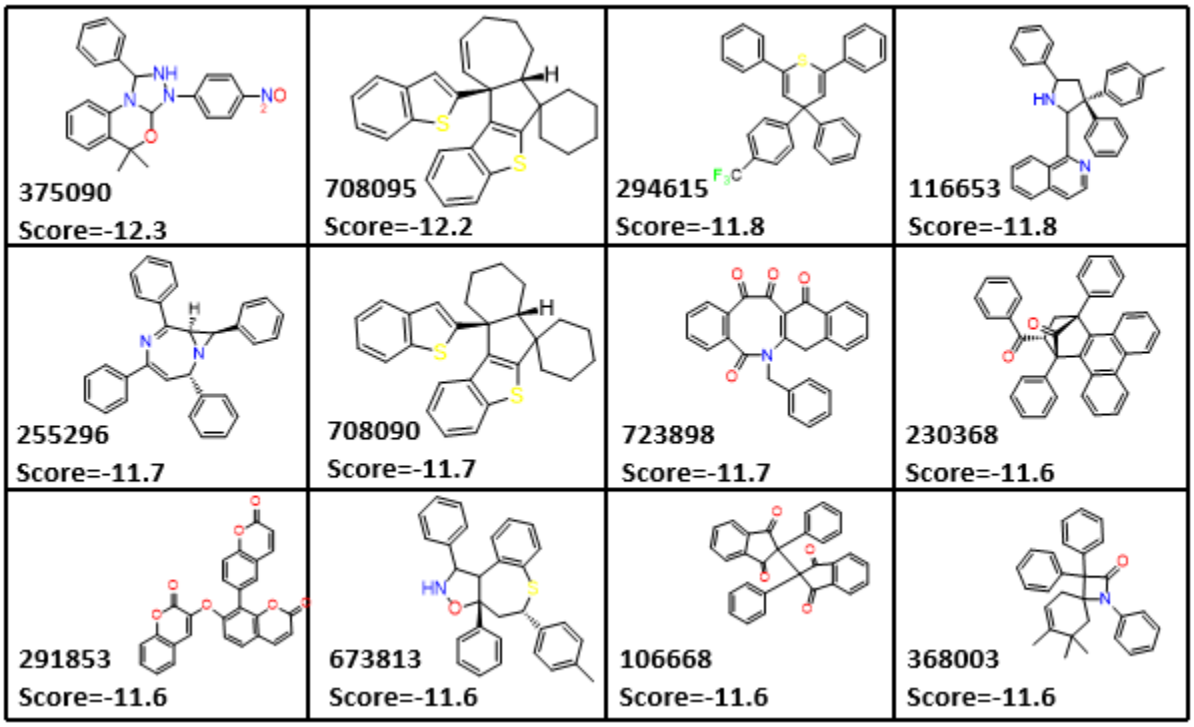

Figure 3.11. Top 12 scoring candidates from the virtual screening of 65endo. All scores are measured in kcal/mol

3.5 References

1. Musille, P.M., et al., *Unexpected Allosteric Network Contributes to LRH-1 Co-regulator Selectivity.* J Biol Chem, 2016. **291**(3): p. 1411-26.
2. Mays, S.G., et al., *Crystal Structures of the Nuclear Receptor, Liver Receptor Homolog 1, Bound to Synthetic Agonists.* J Biol Chem, 2016. **291**(49): p. 25281-25291.
3. Musille, P.M., et al., *Antidiabetic phospholipid-nuclear receptor complex reveals the mechanism for phospholipid-driven gene regulation.* Nat Struct Mol Biol, 2012. **19**(5): p. 532-7, S1-2.
4. Dokholyan, N.V., *Controlling Allosteric Networks in Proteins.* Chemical Reviews, 2016. **116**(11): p. 6463-6487.
5. Daily, M.D. and J.J. Gray, *Allosteric Communication Occurs via Networks of Tertiary and Quaternary Motions in Proteins.* PLoS Computational Biology, 2009. **5**(2): p. e1000293.
6. Kern, D. and E.R.P. Zuiderweg, *The role of dynamics in allosteric regulation.* Current Opinion in Structural Biology, 2003. **13**(6): p. 748-757.
7. Nguyen, C.N., T.K. Young, and M.K. Gilson, *Grid inhomogeneous solvation theory: Hydration structure and thermodynamics of the miniature receptor cucurbit[7]uril.* The Journal of Chemical Physics, 2012. **137**(4): p. 044101.

4 GENERAL CONCLUSIONS

Our computational studies synthetic agonists bound to LRH-1 agree with the observations of collaborators in that the structural underpinnings of their abilities to activate LRH-1 are much more complex than PL agonists such as DLPC. Our initial analysis of correlated motions of RJW100 complexed to LRH-1·TIF2 and essential dynamics demonstrate that RJW100 does not induce the characteristic agreement state of the LBD and AF2 region that occurs upon DLPC binding. However, despite this inability to trigger the same pattern of long time scale motions, it remains capable of triggering productive motions leading to a mild level of communication between the two remote domains in comparison to DLPC. We provide additional evidence that a key residue on the interior of the LBD, Thr352 plays a role in facilitating an alternative mode through which an agonist such as RJW100 can activate LRH-1. This threonine residue seems to stabilize key crystallographic waters which are speculated to create the alternative mode of activation that occurs upon binding of RJW100 and similar scaffolds. Our studies show that four of the waters proximal to RJW100 are thermodynamically stable within the binding pocket. The energetics provide some insight into how maintenance of the water network is achieved when in the presence of RJW100. Specifically, the water adjacent to the water that mediates contact between RJW100's hydroxyl moiety and Thr352 may play a central role in maintaining this network along with Thr352. Our studies also demonstrate that the T352V mutant features diminished allosteric communication upon RJW100 binding, consistent with diminished but not eliminated activation demonstrated in experimental studies.

We also provide evidence that the increased activity of polar derivative 65endo versus RJW100 could be partially attributed to its ability to enhance allosteric communication relative to RJW100. The modification of the hydroxyl function to the sulfonamide may allow for 65endo

to make direct contact with residues in the LBD. The nature of these contacts needs further study, but given some structural evidence contained within the crystal structure, it is possible that 65endo induces correlated motions that are more similar to those induced by DLPC binding than RJW100. This result also provides some evidence that strategic design of RJW100 derivatives could very well benefit from focusing upon modifications that aim to exploit features at the interior of the pocket. Furthermore, the information we gained from probing energetics of solvent within the LBP suggests that 65endo may bind more favorably to LRH-1.

Lastly, we developed a protocol combining ligand based and structure based methods for virtual screening of RJW100, derivative 65endo, and another agonist, 8AC. The inclusion of an electrostatic component to the screening of the latter two appears to have aided in generating a somewhat different set of structures than those from RJW100, but the structural profile of the proposed top binding candidates could likely be improved. Very top scoring compounds from docking seemed to have more architectural uniformity with RJW100 employed as lead whereas the 8AC seemed to be contain more structures with features that may more effectively exploit the polar patch in the interior of LRH-1. Overall, 65endo surprisingly yielded a set of top performing compounds that did not directly overlap with the other two sets but was overall more similar in profile to RJW100 in some ways, but also features some compounds that are much more polar than those in other sets. We await results of experimental screening of these compounds.