



The Development and Characterization of Bile Acid-Derived Agonists of FXR Towards NASH Treatment

Thesis submitted for the degree of Doctor of Philosophy

School of Biological Sciences

Dannielle Kydd-Sinclair

September 2019

Abstract

Over the last few decades, bile acids (BAs) have emerged as important hormone-like signalling molecules. Bile acids have been shown to exert their effects through the binding and activation of the Farnesoid X Receptor (FXR), which regulates various pleiotropic target genes underlying bile acid homeostasis, inflammation and lipid, glucose and cholesterol metabolism. FXR plays a pivotal role in the aetiology of various liver pathologies and metabolic syndromes and, as such, is an attractive therapeutic target. Use of the semi-synthetic BA analogue drug, obeticholic acid (OCA), has been impeded due to its activation of other BA receptors, and due to the promiscuity of systemic FXR activation, which leads to unfavourable and counterintuitive effects. Work described here exploited structural data about the ligand binding domain (LBD) of FXR to design novel, BA-derived agonists. Computational molecular docking approaches were used to determine the putative binding mechanisms of novel compounds, and with supplementary biological activity data, it was postulated that lead candidates implement a unique binding mode. Efforts were devoted to obtaining a 3D structure of the LBD in complex with lead novel compounds and associated cofactors in an effort to validate ligand binding interactions, and to gain further insight into how ligand-mediated structural conformations determine functional activation of FXR. Two lead novel compounds were confirmed as bona fide agonists of FXR, with improved potency compared to OCA, and these were able to recruit coactivators essential for transcriptional activity. Furthermore, these novel agonists were shown to regulate the expression of genes targeted by FXR, both *in vitro* and *in vivo*. Taken collectively, the results presented herein suggested that one of the lead compounds may achieve some FXR-target selectivity by its ability to induce conformations of FXR that preferentially recruit specific coactivators. Despite their common BA scaffold, the two lead compounds displayed differential regulation of target genes *in vivo* compared to OCA and, importantly, this work supports scope for these compounds to be further developed as pharmacological agents for certain diseases.

Declaration

I confirm that this is my own work and the use of all material from other sources has been properly and fully acknowledged. This thesis has not been presented previously for any other degree.

Dannielle Kydd-Sinclair

September 2019

'The important thing in science is not so much to obtain new facts as to discover new ways of thinking about them'

– William Lawrence Bragg

Acknowledgements

First and foremost, I would like to thank the God for the countless blessings that I have received since starting my PhD, in the form of challenges, lessons and opportunities. I am eternally grateful for the people that have been placed in my life to inspire and teach me new things. I would like to express my heart-felt gratitude to Prof. Kimberly Ann Watson, for all the advice, encouragement and useful discussions over the past 3 years. I'd like to thank her for her confidence in me, even when I had little in myself and for allowing me to be independent, to make mistakes and to figure things out for myself. Thank you for encouraging me to develop every aspect of my career. I'd like to also extend my thanks to Dr Alex Weymouth-Wilson for his unwavering support and for believing in the project. A very special thanks goes to Dr Gemma Packer for all of her invaluable help throughout the project, in designing and synthesizing the compounds, in ensuring that I had everything I needed to carry out the experiments without any problems, in supplementing my very limited Chemistry knowledge, and of course, in helping to (re)draw the compounds presented in this thesis. I would further like to thank my PhD committee members, Prof. Jon Gibbins and Prof. David Leake for their interest, inputs and fresh perspectives.

I wish to express my gratitude to Prof. Markus Rudolph and Prof. Steve Soisson for sharing their greatly appreciated advice for working with this protein and for disclosing some of their unpublished methods. I would like to thank everybody at the Oxford Protein Production Facility and Diamond Light Source at the Research Complex in Harwell for all their assistance and direction when setting up the high throughput expression and crystallisation screens. I wish to gratefully acknowledge the assistance given to me by Irene Boz and Charlie Collingham in the tissue culture experiments. A very special thanks goes to Mrs Lee Young, who worked tirelessly to ensure that I always had enough conical flasks whenever they were needed. And most importantly, I'd like to thank her for her friendship and the many gifts she has given to me over the last 3 years. I'd like to thank the rest of the research group, Dr Ged Baltulionis, Dr Amy Danson, Irene Feliciotti and again, Charlie, for all their support, friendship, and help in getting me through the good times and bad. I'd also like to thank the rest of my colleagues and friends in the office, lab and university for making this journey a thousand times more enjoyable.

A special mention goes to my grandparents for all the sacrifices they made in order to provide us with better opportunities, and also my parents, Noreen and Eddie, for teaching me that with hard work and faith, I can achieve anything. I'd like to thank my brother and 'sisters' for their endless encouragement and for always keeping me grounded. I'd like to thank my fiancé, José, for always believing in me, for encouraging me to pursue all of my dreams, and for being there at every step, on every high, and every low of this journey.

Finally, I'd like to thank the University of Reading and New Zealand Pharmaceuticals (NZP-UK) for funding this research.

Oral and Poster Presentations at Conferences

Kydd-Sinclair, D., Watson, K.A., 2017. A structural approach to the development of novel Bile Acid-derived FXR agonists. Presented at University of Reading, School of Biological Sciences PhD symposium, Reading, June 2017 (poster presentation)

Kydd-Sinclair, D., Watson, K.A., 2017. A structural approach to the development of novel Bile Acid-derived FXR agonists. Presented at South West Structural Biology Consortium, Cardiff, July 2017 (poster presentation)

Kydd-Sinclair, D., Watson, K.A., 2017. A structural approach to the design of FXR ligands. Presented at the 24th Congress and General Assembly of IUCr (International Union of Crystallography), Hyderabad, India, August 2017 (poster presentation)

Kydd-Sinclair, D., Watson, K.A., 2018. The structure-based design of novel Bile Acid analogue agonists of FXR. Presented at British Crystallography Association Spring Meeting, Warwick, April 2018 (poster presentation)

Kydd-Sinclair, D., Watson, K.A., 2018. The structure-based design of novel bile acid- derived agonists of FXR. Presented at University of Reading, School of Biological Sciences PhD symposium, Reading, June 2018 (oral presentation) – **Awarded prize for best 2nd year biomedical sciences oral presentation**

Kydd-Sinclair, D., Watson, K.A., 2018. The structure-based design of novel Bile Acid analogue agonists of FXR. Presented at South West Structural Biology Consortium, Exeter, July 2018 (poster presentation)

Kydd-Sinclair, D., Watson, K.A., 2019. The structure-based development of novel Bile Acid- derived agonists of FXR. Presented at INSTRUCT Biennial Structural Biology Conference, Madrid, May 2019 (poster presentation)

Kydd-Sinclair, D., Watson, K.A., 2019. From computers to cures: The structure-based design of novel FXR agonists for NASH treatment. Presented at University of Reading, School of Biological Sciences PhD symposium, Reading, June 2019 (oral presentation) – **Awarded prize for best 3rd year biomedical sciences oral presentation**

Kydd-Sinclair, D., Watson, K.A., 2019. The structure-based development of novel Bile Acid- derived agonists of FXR. Presented at South West Structural Biology Consortium, Reading, July 2019 (oral presentation) – **Awarded prize for 2nd best oral presentation**

Table of Contents

Abstract.....	i
Declaration.....	ii
Acknowledgements.....	iv
Oral and Poster Presentations at Conferences.....	v
List of Abbreviations.....	xiii
Amino Acid Abbreviations	xvi
List of Figures.....	xvii
List of Tables.....	xx

1. Introduction

1.1 Bile Acids.....	2
1.1.1 Physico-chemical Properties of Bile Acids	2
1.1.2 Synthesis and Enterohepatic Circulation of Bile Acids	5
1.2 The Role of Bile acids	9
1.2.1 The Classic Role of Bile Acids	9
1.2.2 The Role of Bile Acids as Signalling Molecules.....	9
1.2.2.1 Nuclear Receptors as Targets of Bile Acids	9
1.2.2.2 G-protein Coupled Bile Acid Receptor – TGR5	11
1.3 The Farnesoid X Receptor	12
1.3.1 Characterization of FXR.....	12
1.3.2 FXR Gene Structure and Tissue Expression	13
1.3.3 FXR Structure and Activation.....	14
1.3.4 Biological Functions of FXR	15
1.3.4.1 BA Metabolism.....	16
1.3.4.2 Lipid, Cholesterol and Glucose Metabolism.....	17
1.3.4.3 Inflammation.....	18
1.3.5 FXR and Disease	18
1.3.5.1 Intestinal Diseases.....	19
1.3.5.2 Cholestasis.....	19
1.3.5.3 Atherosclerosis.....	20
1.3.5.4 Carcinogenesis.....	20
1.3.5.5 Liver Fibrosis	21
1.3.5.6 Non-alcoholic Fatty Liver Diseases	21
1.4 Non-alcoholic Steatohepatitis	22

1.4.1 NASH Pathogenesis.....	22
1.4.2 NASH Treatment	25
1.5 Current FXR agonists.....	25
1.6 Research Aims and Objectives.....	30
2. Materials and Methods	
2.1 Materials	33
2.1.1 Equipment and Software.....	33
2.1.2 Molecular Biology Materials	34
2.2 Methods	39
2.2.1 The Farnesoid X Receptor Gene	39
2.2.2 Structure Prediction and Construct Design.....	39
2.2.3 Analysis of FXR Crystal Structures.....	39
2.3. Molecular Biology Methods	39
2.3.1 Primer Design.....	39
2.3.2 Polymerase Chain Reaction (PCR)	40
2.3.3 Agarose Gel Electrophoresis.....	40
2.3.4 Purification of PCR Products.....	40
2.3.5 Estimation of DNA Concentration.....	41
2.3.6 IN-Fusion Cloning	41
2.3.7. Heat Shock Transformation of <i>E. coli</i> Cells	41
2.3.8. Purification of Plasmid DNA	42
2.3.9 Preparation of Glycerol Stocks	42
2.3.10 DNA Sequencing.....	42
2.4. Production of FXR LBD	42
2.4.1 Expression of FXR LBD	42
2.4.2 Protein Purification by Immobilized Metal Affinity Chromatography.....	43
2.4.3 Additional Purification by Size Exclusion Chromatography	43
2.5 Protein Analysis and Quantification.....	43
2.5.1 Analysis of Proteins by Sodium Dodecylsulfate – Polyacrylamide Gel Electrophoresis (SDS-PAGE)	44
2.5.2 Analysis of Proteins by Western Blot	44
2.5.3 Estimation of Protein Concentration	44
3. Use of Computational Approaches for the Structural Analysis of FXR and Molecular docking of Novel Ligands	

3.1 Introduction.....	46
3.1.1 Structure- Based Drug Design.....	46
3.1.1.1 Molecular Docking	47
3.1.2 Ligand-Based Drug Design.....	48
3.2 Aims and Objectives.....	50
3.3 Materials and Methods.....	52
3.3.1 Homology Modelling.....	52
3.3.2 Initial Analysis of FXR LBD – Ligand Bound Conformations.....	52
3.3.3 Analysis of the FXR Ligand Binding Pocket.....	52
3.3.4 Molecular Modelling and Docking.....	53
3.3.4.1 Compound Library.....	53
3.3.4.2 Receptor Structure Optimization	64
3.3.4.3 Docking Procedure	64
3.3.5 Three-Dimensional Quantitative Structure Activity Relationship (3D-QSAR)	65
3.3.5.1 3D-QSAR Using FLAP	65
3.3.4.2 3D-QSAR Using Volsurf+	66
3.4 Results	67
3.4.1 The Structure of FXR and Analysis of the LBD.....	67
3.4.2 <i>In Silico</i> Docking of Compounds into FXR LBD.....	76
3.4.2.1 Validation of the Docking Method	76
3.4.2.2 Docking Novel BA Derivatives into FXR LBD	77
3.4.2.3 Docking Compounds into Other Structures of FXR LBD.....	94
3.4.3 3D-QSAR of Novel FXR agonists.....	96
3.4.4 <i>In Silico</i> Docking of Compounds into TGR5	99
3.4.4.1 Homology Modelling of TGR5.....	99
3.4.4.2 Docking Novel Compounds into TGR5.....	100
3.5 Discussion	102
3.5.1 The Structural Plasticity of FXR.....	102
3.5.2 Ligand Binding Mechanisms of FXR	103
3.5.3 Docking Novel Compounds into the FXR LBD	105
3.5.4 Proposed Binding Mechanism of Lead Compounds	106
3.5.5 Limitations of a Static Docking Procedure	108
3.5.6 Ligand-based Approaches	109
3.5.7 Conclusion	110

4. Structure Determination of FXR	
4.1 Introduction	113
4.1.1 Methods of Structure Determination	113
4.1.1.1 Macromolecular X-ray Crystallography (MX)	114
4.1.1.2 Nuclear Magnetic Resonance (NMR) Spectroscopy	116
4.1.1.3 Cryo-Electron Microscopy (cryo-EM)	117
4.1.2 Structural Determination of Nuclear Receptors	119
4.2 Aims and Objectives	121
4.3 Methods	122
4.3.1 Sequence of FXR	122
4.3.2 High Throughput Cloning and Expression Screens	122
4.3.3 Optimization and Scale up of His₆-FXR LBD Expression	123
4.3.4 Additional Purification by Anion Exchange Chromatography (AEC)	124
4.3.5 Degradation Tests	124
4.3.6 Thrombin Cleavage	124
4.3.7 Concentration	125
4.3.8 Crystallization Screening	125
4.3.9 Densitometry	126
4.4 Results	127
4.4.1 High Throughput Cloning and Small-Scale Expression Screens	127
4.4.1.1 Rational Construct Design	127
4.4.1.2 Cloning	128
4.4.1.3 Protein Expression and Purification	129
4.4.2 Scale-up and Yield Optimization	132
4.4.3 Removal of Contaminants	134
4.4.4 Challenges with Protein Precipitation	138
4.4.5 Design of a New Construct	139
4.4.6 Protein Expression and Purification of the New His₆-FXR LBD Mutant	139
4.4.7 Concentration and Crystallization of FXR LBDmut1	147
4.5 Discussion	150
4.5.1 High Throughput Expression Screening of FXR Domains	150
4.5.2 Yield Optimization	152
4.5.3 Purity Optimization	154
4.5.4 Preventing Protein Precipitation	157
4.5.5 Crystallization	159

4.5.6 Conclusion	161
5. Activity and Specificity of Lead Compounds	
5.1 Introduction.....	165
5.1.1 Nuclear Receptor Interactions with Coregulators	165
5.1.2 Epigenetic Actions of Coregulators.....	167
5.1.3 Specificity of Nuclear Receptor-Coregulator Interactions	169
5.1.4 Selective Modulation of Nuclear Receptors.....	170
5.2 Aims and Objectives.....	171
5.3 Materials and Methods.....	172
5.3.1 LanthaScreen™ Assay (principle).....	172
5.3.2 Compound Affinity Assays	172
5.3.3 Coactivator Affinity Assays	173
5.3.4 TR-FRET Data Analysis.....	175
5.4 Results	176
5.4.1 The Affinity of Lead Compounds for FXR	176
5.4.2 Jed561-Induced Interactions Between FXR LBD and Coregulators	182
5.4.3 The Affinity of Different Coactivators for FXR LBD	185
5.5 Discussion	187
5.5.1 Lead Compounds are Bona Fide Agonists for FXR	187
5.5.2 Effect of Fluorination on Activity of the Novel Compounds.....	188
5.5.3 Different Compounds Display Varying Degrees of Efficacy	189
5.5.4 Coactivator Affinity for FXR may be Dependent on the Ligand Occupying the Binding Pocket	190
5.5.5 Conclusion	192
6. Function of Lead Compounds	
6.1 Introduction.....	196
6.1.1 FXR Signalling in Health and Disease.....	196
6.1.1.1 FXR and Bile Acid Homeostasis.....	196
6.1.1.2 FXR and Lipid Metabolism	197
6.1.1.3 FXR and Glucose Metabolism	198
6.1.1.4 FXR and NASH.....	199
6.1.2 Experimental Models of NASH	204
6.2 Aims and Objectives.....	205

6.3 Methods.....	207
6.3.1 Cell Culture and Treatment with Compounds.....	207
6.3.2 Maintenance and Treatment of Mice	208
6.3.2.1 Single Dose Administration of Compounds in Wild Type Mice	209
6.3.2.2 Daily Administration of Compounds in Wild Type Mice	209
6.3.2.3 Production of NASH Mouse Model and Treatment with Compounds	209
6.3.3 RNA Extraction from Adherent Cells.....	209
6.3.4 RNA Extraction from Mouse Livers	210
6.3.5 Analysis of RNA Quantity, Purity and Integrity.....	210
6.3.6 DNase Treatment of RNA.....	211
6.3.7 Complementary DNA (cDNA) synthesis	211
6.3.8 Quantitative Real-time Polymerase Chain Reaction (qPCR) (Principle).....	212
6.3.9 Primer Selection.....	212
6.3.9.1 Selection of Reference Genes.....	212
6.3.9.2 Optimization of Primers for Reference and Target Genes	213
6.3.10 Quantitative PCR.....	215
6.3.10.1 Relative Quantification	215
6.3.11 Statistical Analysis	216
6.4 Results	217
6.4.1 Effect of Lead Compounds <i>in vitro</i>	217
6.4.1.1 Selection of Appropriate Reference and Target Genes	217
6.4.1.2 Effect of Jed441 in Huh7 cells.....	221
6.4.1.3 Effect of Jed441 in HepG2 cells.....	226
6.4.1.4 Effect of Jed561 in Huh7 cells.....	231
6.4.1.5 Effect of Jed561 in HepG2 cells.....	235
6.4.1.6 Evaluation of Ligand Specificity in HepG2 Cells	238
6.4.1.7 Effect of Fluorination of Steroid Backbone on FXR-mediated Activity.....	243
6.4.2 Effect of Lead Compounds <i>in vivo</i>	247
6.4.2.1 Selection of Appropriate Reference and Target Genes for Mouse Liver RNA ...	247
6.4.2.2 Short Term Effect of Jed Compounds on FXR Target Genes in Wild Type Mice	251
6.4.2.3 Effect of Continual Treatment with Jed561 on FXR Target Genes in Wild Type Mice	257
6.4.2.4 Effect of Jed561 in a mouse model of NASH	261
6.5 Discussion	267
6.5.1 Effect of Jed441 <i>in vitro</i>	267
6.5.2 Effect of Jed561 <i>in vitro</i>	268

6.5.3 Short Term Effects of Compounds <i>in vivo</i>	269
6.5.4 Long Term Effects of Compounds <i>in vivo</i>	271
6.5.5 Effects of Jed561 in a NASH Mouse Model	273
6.5.6 Conclusions.....	275
7. General Discussion and Recommendations for Future Work.....	277
8. References	292
Appendix 1. Primers used for In Fusion cloning FXR domain constructs	331
Appendix 2. Amino Acid Sequence of FXR Isoform 1(+)	332
Appendix 3. Vector Maps of Constructs used in this Thesis	333
Appendix 4. RNAseq data of coactivator mRNA expression from Expression Atlas data base, as provided by Genotype Tissue Expression (Gtex) Project	339
Appendix 5. RNAseq data from HepG2 and Huh7 cells under basal conditions	340
Appendix 6. RNAseq data from C57BLJ adult Mice Livers under normal physiological conditions	341
Appendix 7. Validation of primers for use on cDNA from Huh7 and HepG2 cells	342
Appendix 8. Validation of Primers for Use on Mouse Liver cDNA	344
Appendix 9. Analysis of 'NASH' mice: body weights and histology scores – data provided by Saretius Ltd.....	346

List of Abbreviations

A	Å	Angström
	aa	Amino acid
	ABST	Apical sodium dependent bile acid transporter
	ACTA2	Smooth muscle actin alpha 2
	ADME	Absorption, Distribution, Metabolism and Excretion
	AEC	Anion exchange chromatography
	AF	Activation function
	AKR1D1	Aldo-keto reductase family 1, member D1
	Amp	Ampicillin
	APOA1	Apolipoprotein A 1
	APOC (1/2/3)	Apolipoprotein C 1, 2 or 3
	AR	Androgen receptor
	ARF1	ADP-ribosylation factor 1
	ATP	Adenosine triphosphate
	AU	Absorbance units
B	BA	Bile acid
	bp	Base pairs
	BSEP	Bile salt export pump
C	C	Carbon
	CA	Cholic acid
	CADD	Computer-aided drug design
	cAMP	Cyclic Adenosine monophosphate
	CAR	Constitutive Androstane receptor
	CBP	CREB binding protein
	CCL2	C-C motif chemokine ligand 2
	CCL4	Carbon tetrachloride
	CCR2	CCL2 receptor
	CDCA	Chenodeoxycholic acid
	cDNA	Complementary DNA
	CMC	Critical micellar concentration
	COL(1/3)A1	Collagen type 1 or 3 alpha 1 chain
	CoMFA	Comparative molecular field analysis
	CoMSIA	Comparative molecular similarity indices analysis
	COX2	Cyclooxygenase 2
	CREB	cAMP response element binding protein
	CRP	cAMP regulatory protein
	CV	Column volume
	CYP27A1	Sterol 27-hydroxylase
	CYP7A1	Cholesterol 7-alpha hydroxylase
	CYP7B1	Oxysterol 7-alpha hydroxylase
	CYP8B1	Sterol 12-alpha hydroxylase
D	DBD	DNA-binding domain
	DCA	Deoxycholic acid
	DMEM	Dulbecco's Modified Eagle Medium
	DMSO	Dimethyl sulfoxide
	DNA	Deoxyribonucleic acid
	dNTPs	Deoxynucleotide triphosphates
	DRIP	Vitamin D receptor interacting proteins
	DTT	Dithiothreitol
E	ECL	Enhanced chemiluminescence
	ECM	Extra cellular matrix
	EM	Electron microscopy
	EMEM	Eagle's Minimum Essential Medium

	ER	Oestrogen receptor
F	F	Fluorine atom
	FBP-1	Fructose-1,6-bisphosphatase 1
	FBS	Foetal bovine serum
	FDA	U.S Food and Drug Administration
	FFA	Free fatty acid
	FGF	Fibroblast growth factor
	FPLC	Fast protein liquid chromatography
	FRET	Fluorescence resonance energy transfer
	FXR	Farnesoid X Receptor
G	G6P-ase	Glucose 6-phosphatase
	GLP-1	Glucagon-like peptide 1
	GLUT-4	Glucose transporter type 4
	GOI	Gene of interest
	GSK3	Glycogen synthase kinase-3
	GST	Glutathione s-transferase
H	HCC	Hepatocellular Carcinoma
	HDL	High density lipoprotein
	HDX-MS	Hydrogen/deuterium exchange-mass spectrometry
	HFD	High fat diet
	HNF4A	Hepatocyte nuclear factor 4 alpha
	HSC	Hepatic stellate cell
	HSD3B7	Hydroxy- δ -5-steroid dehydrogenase, 3 β - and steroid δ -isomerase 7
	HTP	High throughput
I	IBABP	Ileal bile acid binding protein
	ICAM	Intercellular adhesion molecule 1
	IL1 ($\alpha/6$)	Interleukin 1 alpha or 6
	IMAC	Immobilised metal affinity chromatography
	iNOS	Inducible nitric oxide synthase
	IPTG	Isopropyl- β -D-thiogalactoside
K	kDa	Kilo Daltons
L	LB	Luria-Bertani broth
	LBD	Ligand Binding Domain
	LCA	Lithocholic acid
	LDL	Low density lipoprotein
	LDLR	LDL receptor
	LFD	Low fat diet
	LPS	Lipopolysaccharide
	LRH-1	Liver receptor homolog-1
	LXR	Liver X Receptor
M	MIFs	Molecular interaction fields
	MMFF94	Merck molecular force field
	MOPs	3-(N-morpholino)propanesulphonic acid
	mRNA	Messenger RNA
	MRP4	Multi-drug resistance associated protein 4
	MTP	Microsomal triglyceride transfer protein
	MW	Molecular weight
	MWCO	Molecular weight cut off
	MX	Macromolecular crystallography
N	NAFLD	Non-alcoholic Fatty Liver Disease
	NASH	Non-alcoholic steatohepatitis
	NCOR	Nuclear co repressor
	NDB	N-Benzyl-N-(3-(tert-butyl)-4-hydroxyphenyl)-2,6-dichloro-4-(dimethylamino) Benzamide
	NFκB	Nuclear factor kappa-light chain enhancer of activated B cells

	Ni-NTA	Nickel- Nitrilotriacetic acid
	NMR	Nuclear magnetic resonance
	NTCP	Sodium taurocholate Co-transporting protein
	NR	Nuclear receptor
O	OATP	Organic anion transporting polypeptide
	OCA	Obeticholic acid
	OD	Optical density
	OH	Hydroxyl group
	OSTA/B	Organic solute transporter alpha/beta
P	PBC	Primary biliary cholangitis
	PCR	Polymerase chain reaction
	PDB	Protein Data Bank
	PDK4	Pyruvate dehydrogenase kinase 4
	PEPCK	Phosphoenolpyruvate carboxykinase
	pg	Preparation grade
	PGC1α	PPAR coactivator 1 alpha
	pI	Isoelectric point
	PLS	Partial least squares
	PLTP	Phospholipid transfer protein
	PMSF	Phenylmethanesulfonyl fluoride
	POI	Protein of interest
	PPAR (α/ γ)	Peroxisome proliferator activated receptor (alpha or gamma)
	PSC	Primary sclerosis cholangitis
	PXR	Pregnane X Receptor
Q	(3D)-QSAR	Three-dimensional Quantitative structure activity relationship
	qPCR	Quantitative Polymerase chain reaction
R	RIP	Nuclear receptor interacting protein
	RIP14	RXR-interacting protein 14
	RMSD	Root mean square deviation
	RNA	Ribonucleic Acid
	ROS	Reactive oxygen species
	RPM	Rotations per minute
	RXR	Retinoid X Receptor
S	SAXS	Small angle X-ray scattering
	SDS	Sodium dodecyl sulphate
	SDS-PAGE	Sodium dodecyl sulphate – Polyacrylamide gel electrophoresis
	SEC	Size exclusion chromatography
	SEM	Standard error of the mean
	SHP	Small heterodimer partner
	SlyD	FKBP-type peptidyl-prolyl cis-trans isomerase
	SOC	Super optimal broth with catabolite repression
	SMRT	Silencing mediator of retinoic acid and thyroid hormone receptor
	SRC	Steroid Receptor coactivator
	SUMO	Small ubiquitin-like modifier
T	T2DM	Type 2 Diabetes Mellitus
	TBST	Tween 20- Tris buffered saline
	TCEP	Tris(2-carboxyethyl)phosphine
	TGFβ1	Transforming growth factor beta 1
	TGR5	G-Protein coupled bile acid receptor
	TEM	Transmission electron microscopy
	TIMP1	Tissue inhibitor of metalloproteinase 1
	Tm	Melting temperature
	TNFα	Tumour necrosis factor alpha

	TR-FRET	Time resolved FRET
U	TRAP	Thyroid hormone receptor- associated protein
	UDCA	Ursodeoxycholic acid
	UK	United Kingdom
	U.S	United States
	UV	Ultraviolet
V	VDR	Vitamin D Receptor
	VdW	Van der Waals
	VLDL	Very low density lipoprotein
	VLDLR	VLDL Receptor
	v/v	Volume per volume
W	WT	Wild type
	w/v	Weight per volume
X	X-gal	5-bromo-4-chloro-3-indolyl- β -D-galactopyranoside
Y	YWHAZ	Tyrosine 3-monooxygenase/Tryptophan 5-monooxygenase activation protein zeta

Amino Acid Abbreviations

A	Alanine	M	Methionine
C	Cysteine	N	Asparagine
D	Aspartate	P	Proline
E	Glutamate	Q	Glutamine
F	Phenylalanine	R	Arginine
G	Glycine	S	Serine
H	Histidine	T	Threonine
I	Isoleucine	V	Valine
K	Lysine	W	Tryptophan
L	Leucine	Y	Tyrosine

List of Figures

1. Introduction

<i>Figure 1. 1 Structures of the most abundant bile acids in humans, including their conjugate groups.....</i>	4
<i>Figure 1. 2 Bile acid synthesis pathways.</i>	6
<i>Figure 1. 3 Overview of enterohepatic circulation of bile acids after biosynthesis and conjugation.</i>	8
<i>Figure 1. 4 Genomic and tertiary protein structure of FXR. (A)</i>	15
<i>Figure 1. 5 FXR signalling actions on hepatocytes in the liver and enterocytes in the intestines.</i>	16
<i>Figure 1. 6 The pathogenesis of NASH..</i>	24

3. Use of Computational Approaches for the Structural Analysis of FXR & Molecular Docking of Novel Ligands

<i>Figure 3. 1 General formula of novel compounds based on bile acid backbone</i>	54
<i>Figure 3. 2 Full length and individual domain FXR structures predicted by the integrated modelling server IntFOLD.</i>	68
<i>Figure 3. 3 Observations of FXR LBD structural flexibility.</i>	73
<i>Figure 3. 4 Observations of ligand binding modes of compounds with different degrees of agonist behaviour</i>	75
<i>Figure 3. 5 Comparisons of GRID molecular interaction fields for the receptor (A) and for the CDCA ligand (B)</i>	76
<i>Figure 3. 6 Validation of the docking procedure.....</i>	77
<i>Figure 3. 7 Docking results of CDCA-derived compounds</i>	79
<i>Figure 3. 8 Docking results of OCA-derived compounds with extensions of their carboxyl 'tails'</i>	83
<i>Figure 3. 9 Comparisons of Jed575 (blue) and its fluorinated derivative, Jed584 (yellow)</i>	83
<i>Figure 3. 10 Addition of cyclobutane to BA side chain</i>	84
<i>Figure 3. 11 Comparisons of nitrile containing side chains.</i>	84
<i>Figure 3. 12 The effect of extensions or substitutions at the C3 position</i>	86
<i>Figure 3. 13 Comparisons of fluorinated OCA-derivatives.....</i>	88
<i>Figure 3. 14 Addition of multiple fluorine atoms to the A ring and their effect on ligand binding position</i>	89
<i>Figure 3. 15 Unique binding mode of Jed441.....</i>	89
<i>Figure 3. 16 Compounds which make polar contacts with helix 3 and the effect of different side chain functional groups on binding modes</i>	90
<i>Figure 3. 17 Effect of fluorination at the C4 position</i>	90
<i>Figure 3. 18 The effect of open A rings and extended side chains on binding position.....</i>	93
<i>Figure 3. 19 Binding poses of compounds when docked into Apo-FXR LBD (from PDB ID: 5Q0K).</i>	95
<i>Figure 3. 20 Comparisons of binding poses of compounds when docked into antagonist bound-FXR LBD (from PDB ID: 4OIV)</i>	96
<i>Figure 3. 21 Correlating molecular descriptors with biological activity of compounds</i>	98
<i>Figure 3. 22 Proposed binding mechanism of BAs in TGR5 and key binding residues.....</i>	100

<i>Figure 3. 23 Compounds docked into TGR5 homology model using (A) automated protomol generation and (B) residue directed protomol generation.</i>	101
---	-----

4. Structure Determination of FXR

<i>Figure 4. 1 Method of protein crystallization by vapour diffusion.</i>	116
<i>Figure 4. 2 High throughput PCR verification of cloned constructs.....</i>	129
<i>Figure 4. 3 SDS-PAGE Analysis of Ni²⁺-NTA purified proteins from constructs.....</i>	131
<i>Figure 4. 4 Scale up and purification of FXR-LBD.....</i>	132
<i>Figure 4. 5 Further scale up and purification of FXR-LBD.....</i>	133
<i>Figure 4. 6 Size Exclusion Chromatography of His₆-FXR LBD.</i>	136
<i>Figure 4. 7 Anion exchange chromatography of His₆-FXR LBD</i>	137
<i>Figure 4. 8 SDS PAGE analysis of protein degradation.</i>	138
<i>Figure 4. 9 Comparisons between expression and purification of His -FXR LBDmut1 in Rosetta (DE3) pLysS and BL21 (DE3) pLysS</i>	141
<i>Figure 4. 10 Size Exclusion Chromatography of His6 -FXR LBDmut1.....</i>	142
<i>Figure 4. 11 Optimisation of conditions for thrombin cleavage reactions..</i>	145
<i>Figure 4. 12 Thrombin cleavage and secondary purification of FXR LBDmut1</i>	146
<i>Figure 4. 13 Final SEC purification of cleaved FXR LBDmut1.</i>	147
<i>Figure 4. 14 Objects produced in crystallization screens of FXR LBD complexed with Jed561 and SRC2-2.</i>	149
<i>Figure 4. 15 Example of diffraction image observed from in situ plate screening on VMXi beamline</i>	150

5. Activity & Specificity of Lead Compounds

<i>Figure 5. 1 Coactivator binding site on FXR LBD (PDB:4QE6).</i>	166
<i>Figure 5. 2 General model of ligand activated NR regulation of target gene transcription.</i> 169	
<i>Figure 5. 3 Principle of the FXR agonist coactivator peptide recruitment assay.</i>	172
<i>Figure 5. 4 TR-FRET FXR Coactivator recruitment assay.</i>	177
<i>Figure 5. 5 Max response values determined by ligand-activated coactivator recruitment.</i> 178	
<i>Figure 5. 6 Ligand-induced recruitment of different coregulators to FXR LBD.</i>	181
<i>Figure 5. 7 The effect of different NR coregulator peptides on Jed561 induced TR-FRET.....</i>	183
<i>Figure 5. 8 Maximal fluorescence emission ratios in assays with different NR coregulator peptides in the presence or absence of Jed561</i>	184
<i>Figure 5. 9 The effect of ligands on the affinity of FXR LBD for different LXLL-containing coactivator peptides.</i>	186

6. Function of Lead Compounds

<i>Figure 6. 1 Limit of detection of qPCR assays using primers for FXR.....</i>	220
<i>Figure 6. 2 Specificity of qPCR assays using reference gene primers</i>	221
<i>Figure 6. 3 FXR and FXR-target gene expression in Huh7 cells following 6 hours incubation with OCA or Jed441 at their respective EC₅₀ and EC₉₀ concentrations.</i>	222
<i>Figure 6. 4 FXR and FXR-target gene expression in Huh7 cells following 24 hours incubation with OCA or Jed441 at their respective EC₅₀ and EC₉₀ concentrations.</i>	224

<i>Figure 6. 5 Expression of genes involved in FXR-mediated lipid metabolism pathways in Huh7 cells following 24 hours incubation with OCA or Jed441 at their respective EC₅₀ and EC₉₀ concentrations.....</i>	225
<i>Figure 6. 6 TGFβ1 mRNA expression in Huh7 cells following (A) 6, and (B) 24 hours incubation with OCA or Jed441 at their respective EC₅₀ and EC₉₀ concentrations</i>	226
<i>Figure 6. 7 FXR and FXR-target gene expression in HepG2 cells following 6 hours incubation with OCA or Jed441 at their respective EC₅₀ and EC₉₀ concentrations</i>	228
<i>Figure 6. 8 FXR and FXR-target gene expression in HepG2 cells following 24 hours incubation with OCA or Jed441 at their respective EC₅₀ and EC₉₀ concentrations</i>	230
<i>Figure 6. 9 TGFβ1 mRNA expression in HepG2 cells following (A) 6, and (B) 24 hours incubation with OCA or Jed441 at their respective EC₅₀ and EC₉₀ concentrations</i>	231
<i>Figure 6. 10 FXR and FXR-target gene expression in Huh7 cells following 6 hours incubation with OCA or Jed561 at their respective EC₅₀ and EC₉₀ concentrations</i>	232
<i>Figure 6. 11 FXR and FXR-target gene expression in Huh7 cells following 24 hours incubation with OCA or Jed561 at their respective EC₅₀ and EC₉₀ concentrations.</i>	233
<i>Figure 6. 12 Expression of genes involved in FXR-mediated lipid metabolism pathways in Huh7 cells following 24 hours incubation with OCA or Jed561 at their respective EC₅₀ and EC₉₀ concentrations.....</i>	234
<i>Figure 6. 13 TGFβ1 mRNA expression in Huh7 cells following (A) 6, and (B) 24 hours incubation with OCA or Jed561 at their respective EC₅₀ and EC₉₀ concentrations.</i>	235
<i>Figure 6. 14 FXR and FXR-target gene expression in HepG2 cells following 6 hours incubation with OCA or Jed561 at their respective EC₅₀ and EC₉₀ concentrations.</i>	236
<i>Figure 6. 15 FXR and FXR-target gene expression in HepG2 cells following 24 hours incubation with OCA or Jed561 at their respective EC₅₀ and EC₉₀ concentrations.....</i>	237
<i>Figure 6. 16 TGFβ1 mRNA expression in HepG2 cells following (A) 6, and (B) 24 hours incubation with OCA or Jed561 at their respective EC₅₀ and EC₉₀ concentrations.</i>	238
<i>Figure 6. 17 In vitro effect of Z-Guggulsterone and Jed441 on FXR target genes</i>	241
<i>Figure 6. 18 In vitro effect of Z-Guggulsterone and Jed561 on FXR target genes</i>	242
<i>Figure 6. 19 Comparison of In vitro effect of Jed561 and Jed692 on FXR target genes.....</i>	245
<i>Figure 6. 20 Comparison of In vitro effect of Jed561 and Jed692 on FXR target genes.....</i>	246
<i>Figure 6. 21 Comparisons of primer sensitivity and efficiencies.....</i>	248
<i>Figure 6. 22 Stability of candidate reference genes across differentially treated mice</i>	250
<i>Figure 6. 23 Expression of FXR target genes following single oral dose of Jed441 in C57BL/6J mice.</i>	253
<i>Figure 6. 24 Expression of FXR target genes following single oral dose of Jed561 in C57BL/6J mice.</i>	254
<i>Figure 6. 25 Expression of FXR target genes following single oral dose of Jed and competitor compounds in C57BL/6J mice.</i>	256
<i>Figure 6. 26 Expression of FXR target genes following continual treatment with Jed compounds in C57BL/6J mice.</i>	259
<i>Figure 6. 27 Expression of genes involved in FXR-mediated lipid metabolism following continual treatment with Jed compounds in C57BL/6J mice.....</i>	260
<i>Figure 6. 28 Effect of Jed561 on fibrosis genes in 'NASH' mice</i>	263
<i>Figure 6. 29 Effect of Jed561 on inflammation genes in 'NASH' mice.....</i>	264
<i>Figure 6. 30 Effect of Jed561 on lipid homeostasis genes in 'NASH' mice.....</i>	265
<i>Figure 6. 31 Effect of Jed561 on CASP8 apoptosis gene in 'NASH' mice..</i>	266

List of Tables

1. Introduction

<i>Table 1. 1 Summary of main biological functions of FXR</i>	18
<i>Table 1. 2 Agonists of FXR and their structures</i>	29

2. Materials & Methods

<i>Table 2. 1 List of equipment</i>	33
<i>Table 2. 2 List of bioinformatics tools and computer software</i>	33
<i>Table 2. 3 List of consumables and reagents</i>	35
<i>Table 2. 4 List of solutions, buffers and media</i>	36
<i>Table 2. 5 List of antibiotics</i>	37
<i>Table 2. 6 List of antibodies</i>	37
<i>Table 2. 7 List of vectors</i>	37
<i>Table 2. 8 List of E. coli strains</i>	37
<i>Table 2. 9 Table of Compounds used for in vitro and in vivo work</i>	38
<i>Table 2. 10 PCR cycling parameters for use with Phusion Flash Master Mix</i>	40

3. Use of Computational Approaches for the Structural Analysis of FXR & Molecular Docking of Novel Ligands

<i>Table 3. 1 CDCA-derived compounds with alterations to A/B ring or side chain</i>	55
<i>Table 3. 2 OCA-derived compounds with substitutions at the tail end (R¹⁷) only</i>	56
<i>Table 3. 3 OCA-derived compounds with extensions at 3C' position and altered tail ends</i>	59
<i>Table 3. 4 OCA-derived compounds with fluorine substitutions on A ring and additions to tail end</i>	61
<i>Table 3. 5 Compounds with open A ring structure</i>	63
<i>Table 3.6 Comparison and characterization of helix movement and ligand binding sites of FXR structures to classic agonist structure 4QE6</i>	70
<i>Table 3. 7 Table summarising polar contacts and VdW interactions between CDCA-derived compounds and FXR</i>	78
<i>Table 3. 8 Table summarising polar contacts and VdW interactions between OCA-derived compounds with extensions of the carboxyl tail only and FXR</i>	80
<i>Table 3. 9 Table summarising polar contacts and VdW interactions between OCA-derived compounds with extensions at their C3 position and FXR</i>	85
<i>Table 3. 10 Table summarising polar contacts and VdW interactions between OCA-derived compounds with substitutions on their A rings and FXR</i>	91
<i>Table 3. 11 Table summarising polar contacts and VdW interactions between BA-derived compounds with an open A ring structure and FXR</i>	93

4. Structure Determination of FXR

<i>Table 4. 1 List of constructs designed for expression of FXR domains with different fusion tags.</i>	127
---	-----

5. Activity & Specificity of Lead Compounds

<i>Table 5. 1 Sequences of fluorescein-labelled coregulator peptides</i>	174
<i>Table 5. 2 EC₅₀ values of compounds tested in FXR coactivator recruitment assays</i>	177
<i>Table 5. 3 EC₅₀ values of compounds tested in FXR coactivator recruitment assays with different coactivator peptides</i>	180
<i>Table 5. 4 EC₅₀ values for Jed561 in FXR coactivator recruitment assays with different LXLL-containing peptides</i>	184

6. Function of Lead Compounds

<i>Table 6. 1 Summary of FXR regulatory effects either by direct transcriptional regulation or downstream effects</i>	202
<i>Table 6. 2 Treatment of Huh7 and HepG2 cells with different compounds for gene expression analysis</i>	208
<i>Table 6. 3 KicQStart® SYBR® Green Primers purchased from Sigma Aldrich tested for use in qPCR reactions</i>	213
<i>Table 6. 4 Comparison of GAPDH and β-actin C_t values across different human hepatic cell line samples</i>	220
<i>Table 6. 5 Comparison of efficiency, sensitivity and reproducibility of primers for reference genes in C57BL/6J mice samples</i>	249

Chapter 1.

Introduction

1.1 Bile Acids

Since circa 1500 B.C., the therapeutic importance of bile has been acknowledged, and bile extracted from black bears has been used in traditional Chinese medicine for thousands of years (Beuers and Boyer, 1994). However, it wasn't until 1848 that the first bile acid, cholic acid (CA), was isolated from ox-gall (Strecker, 1848), and later work into the extraction, identification and characterization of different bile acids, earned Heinrich Weiland the Nobel prize for chemistry in 1927. During this time, bile acids (BAs) were being used in liver tonics and laxatives, but despite this, they were deemed by worldwide medical professionals to have no therapeutic use (Hofmann and Hagey, 2014). Nonetheless, over the following decades, research was carried out into the elucidation of BA metabolism, biosynthesis, bioavailability and circulation, and in 1972 it was first observed that chenodeoxycholic acid (CDCA) could induce gallstone dissolution, sparking a resurgence of interest in the BA field (Danzinger *et al.*, 1972). It was realised that the therapeutic potential of BAs far exceeded their traditional expectations when, at the end of the 20th century, their versatile role as signalling molecules was discovered (Wang *et al.*, 1999; Maruyama *et al.*, 2002). Accordingly, given their underlying role in several different physiological processes, BA biology has become an ever increasingly popular field for academics, pharmaceutical companies and biotechnology industries alike.

To date, various BAs have been approved by the United States Food and Drug Administration (FDA) for the treatment of BA synthesis disorders, peroxisomal disorders, primary biliary cholangitis and for cosmetic treatment of submental fat (Goulis *et al.*, 1999; Lazarević *et al.*, 2019). The use of BAs and BA-analogues is also currently being investigated for the treatment of metabolic disorders, including type 2 diabetes mellitus (T2DM) and fatty liver diseases, inflammatory bowel disease, osteoporosis, and, more recently, neuroinflammation and Parkinson's disease (Thomas *et al.*, 2008; Porez *et al.*, 2012; Schaap *et al.*, 2013; Abdelkader *et al.*, 2016). Furthermore, due to their unique physico-chemical properties, BAs are being viewed by the pharmaceutical industry, as attractive compounds for novel drug delivery systems (Stojančević *et al.*, 2013). As such, the therapeutic potential of these molecules is continually being uncovered.

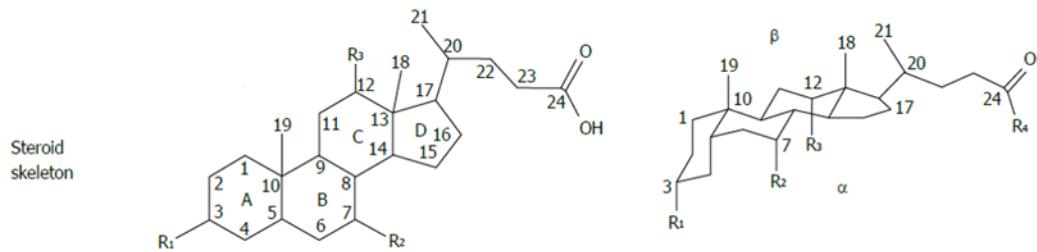
1.1.1 Physico-chemical Properties of Bile Acids

BAs are the major organic solutes that, in addition to phospholipids, cholesterol and bilirubin, constitute the hepatic secretion: bile (Reshetnyak, 2013). They are a group of structurally similar, amphipathic, steroid nucleus-containing molecules, which are produced in the liver as

a result of cholesterol catabolism. It is now known that BAs are responsible for the solubilisation and digestion of dietary lipids, absorption of fat-soluble vitamins, elimination of xenobiotics and maintenance of cholesterol homeostasis, and several of these canonical functions are facilitated by the distinctive structural properties and stereochemistry of these molecules.

In nature, there exists a great variety of chemical structures of BAs, arising from the numerous evolutionary biochemical pathways involved in the breakdown and solubilisation of cholesterol (Hagey *et al.*, 2010). The fact that many BAs were discovered and named prior to the 'classical biochemistry' era, before their structures were determined, creates further complexity to BA chemistry and nomenclature; and to date, there has been no suitable designation which includes every family member (Hagey *et al.*, 2010). In biomedical literature, the term 'bile acids' is generally used to denote 'modern' BAs, or rather Cholanoids; a term comprising acids with 24 carbon atoms, which are predominant in 'modern' mammals (Monte *et al.*, 2009).

All classes of BAs have a four-ring cyclopentanophenanthrene (steroid) nucleus; three rings of which (A,B,C), are six-membered cyclohexanes, and the fourth (D), being a cyclopentane (Figure 1. 1). In most BAs, the steroid nucleus is fully saturated, whereby the double bond at the 5/6th carbon position (C5/6) in cholesterol is isomerized and then reduced during biosynthesis. This produces an A/B ring juncture which can either be curved or flat in relation to the other rings, as a result of the *cis*- or *trans*- steric conformation, but in mammals the nucleus orientation is most commonly *cis*, denoted by a 5 β hydrogen atom (Hofmann and Hagey, 2014). All BAs contain a side chain that supports a carboxyl group (Figure 1. 1). Likewise, all BAs retain the C3 hydroxyl group from cholesterol, which is epimerised to the α configuration, and the addition of hydroxyl groups to the C7 can distinguish primary and secondary BAs. CDCA, which comprises all these features (5 β configuration, C24 carboxyl group and 3 α ,7 α hydroxyl groups), is considered the root, or building block BA and can undergo further modifications to produce all other C24 BAs (Hofmann and Hagey, 2008).



	Name	R₁	R₂	R₃	R₄
Primary	Cholic Acid (CA)	OH	OH	OH	OH
Primary	Chenodeoxycholic Acid (CDCA)	OH (α)	OH (α)	H	OH
Secondary	Deoxycholic Acid (DCA)	OH	H	OH	OH
Secondary	Lithocholic Acid (LCA)	OH	H	H	OH
Secondary	Ursodeoxycholic Acid (UDCA)	OH (α)	OH (β)	H	OH
Primary	Glycocholic Acid	OH	OH	OH	NHCH ₂ COOH
Primary	Taurocholic Acid	OH	OH	OH	NHCH ₂ SO ₃ H

Figure 1. 1 Structures of the most abundant bile acids in humans, including their conjugate groups. BAs retain their 4-ring steroid nucleus. Hydroxyls in the α orientation are situated below and axial to the steroid nucleus, whereas in the β orientation, they are above, and equatorial in relation to the steroid nucleus. Carbon atoms numbered 1-24. Adapted from (Monte *et al.*, 2009).

In CDCA, as with most other BAs, the hydroxyl groups, as well as the carboxyl side chain, are orientated towards the α side of the molecule, creating a concave, hydrophilic, lower face, as seen in Figure 1. 1. Conversely, the β side of the molecule, which contains methyl groups but no substituents, affords hydrophobicity to this convex, upper face. The number of hydroxyl groups at positions along the steroid backbone, and their orientation towards the α - or β - face, in addition to the length of the side chain, can have significant effects on the hydrophobicity of the molecule, and the rank order of hydrophobicity decreases from lithocholic acid (LCA)> deoxycholic acid (DCA)> CDCA> CA (De Aguiar Vallim *et al.*, 2013).

Due to their unique, amphipathic characteristics, in which one side of the molecule is hydrophobic, whilst the other is hydrophilic, when in an aqueous solution, BA anions will spontaneously associate forming micelles. In order to achieve this, the BA concentration needs to exceed a critical value, known as the critical micellar concentration (CMC). However, during digestion, the cooperative association between BA micelles and the lipid membrane, means that mixed micelles, containing up to 50 BA molecules with either biliary phosphatidylcholine or partially ionised fatty acid monoglycerides, can form at concentrations lower than the CMC. In turn, these mixed micelles, can further solubilise dietary cholesterol, fat and other lipophilic

compounds, including liposoluble vitamins, enabling their absorption (Hofmann and Hagey, 2008). There is an inverse correlation between the CMC and the hydrophobic surface area of a BA, and so secondary BAs such as LCA and DCA, have an increased capacity for solubilisation (Hofmann and Hagey, 2008).

1.1.2 Synthesis and Enterohepatic Circulation of Bile Acids

The formation of BAs represents the main route for the elimination of excess cholesterol from the body, and approximately 500mg of cholesterol is catabolized in the liver every day (Thomas *et al.*, 2008). In humans, there are two main multienzyme pathways where insoluble cholesterol is broken down. Both involve processes that lead to the addition of hydroxyl groups to the steroid backbone and the oxidative cleavage of the sterol side chain, ultimately producing a soluble primary BA such as CDCA or CA (seen in Figure 1. 1).

The classical pathway, forming neutral intermediate metabolites, is a cascade of fourteen different enzymes across several intracellular compartments and is restricted to the liver. It results in the production of almost equal ratios of CDCA and CA. The first, and rate-limiting step of this biotransformation, is the conversion of cholesterol to 7α -hydroxycholesterol by the microsomal, cytochrome P450 enzyme, cholesterol 7α -hydroxylase (CYP7A1) (Russell and Setchell, 1992). CYP7A1 is highly regulated on a transcriptional level, and its mRNA has a short half-life, allowing the BA pool to be tightly controlled (Hylemon *et al.*, 2009). The next steps in the biotransformation process involve further modification of the steroid backbone. Isomerisation and reduction of cholesterol's C5/6 double bond creates a *cis*-angled A/B juncture. Furthermore, the epimerization of the hydroxyl group at the C3 position of the A ring, and the addition of a hydroxyl group at the C12 position of the C ring, produce the precursors of either CDCA or CA, 5 β -cholest-3 α ,7 α -diol or 5 β -cholest-3 α ,7 α ,12 α -triol respectively. The side chain of these precursors then undergoes a series of oxidation reactions via mitochondrial enzymes, after which they can become activated to their coenzyme-A esters in the endoplasmic reticulum. This is followed by additional side chain shortening by the β -oxidation actions of 4 different peroxisomal enzymes, producing either CA or CDCA. Finally, prior to secretion into the canalicular lumen, the terminal side chain carboxylic acid of the BA is conjugated as an *N*-acyl amide bond, with either glycine or taurine. This amidation modification has important effects on the ionization and solubility of BAs at physiological pH, and conjugation attenuates their hydrophobic toxicity, minimises passive diffusion, and prevents calcium precipitation and cleavage by pancreatic enzymes (De Aguiar Vallim *et al.*, 2013; Li and Chiang, 2014).

The alternative biosynthetic pathway of BAs, also known as the 'acidic' pathway, due to the formation of acidic intermediate metabolites, accounts for approximately 10% of de novo BA synthesis in normal conditions and predominantly produces CDCA. In this pathway, contrary to the classical pathway, side chain shortening occurs before the modification of the steroid backbone. In addition, the first two enzymes, sterol 27-hydroxylase (CYP27A1) and oxysterol 7 α -hydroxylase (CYP7B1), which oxidise and convert cholesterol into 7 α ,27-dihydroxycholesterol, are widely expressed in various different tissues, and CYP27A1 has the capacity to catalyse multiple oxidation reactions forming the carboxylic acid derivative of the sterol (Monte *et al.*, 2009). Consequently, the oxysterol intermediate, often derived from macrophages, or the brain and lung, must be transported to the liver, for the completion of BA synthesis.

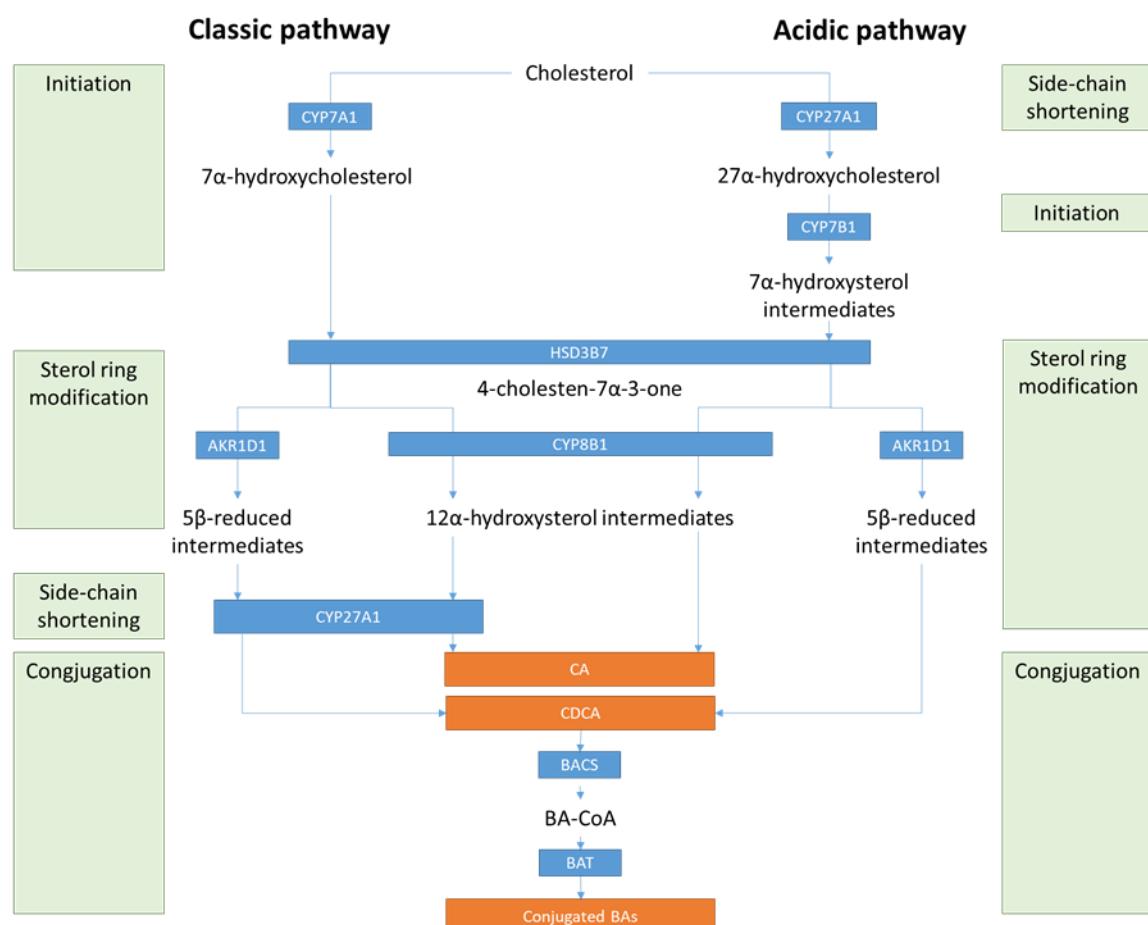


Figure 1. 2 Bile acid synthesis pathways. Cholesterol conversion into primary BAs, cholic acid (CA) and chenodeoxycholic acid (CDCA) occurs via the classic pathway (90%) and the acidic pathway (10%). Steps leading to the synthesis of BAs include hydroxylation at the C7 (initiation by CYP7A1 or CYP7B1), addition of hydroxyl groups at the C12 by CYP8B1, epimerisation of the hydroxyl at the C3 by HSD3B7 and reduction of the double bond by AKR1D1 (sterol ring modification), Side chain shortening by CYP27A1, and conjugation by BACS and BAT. CYP7A1, Cholesterol 7 α -hydroxylase; CYP27A1, sterol 27-hydroxylase; CYP7B1, Oxysterol 7 α -hydroxylase; HSD3B7, Hydroxy- δ -5-steroid dehydrogenase, 3 β - and steroid δ -isomerase 7; CYP8B1, Sterol 12 α -hydroxylase; AKR1D1, Aldo-keto reductase family 1, member D1.

Newly synthesized, conjugated BAs are secreted through transporters on the canalicular membrane, such as bile salt export protein (BSEP), and along with phospholipids and cholesterol, are stored in the gall bladder where they are concentrated during the interdigestive period (Merritt and Donaldson, 2009). Here, they can solubilize cholesterol, preventing its crystallisation and formation of gallstones. After a postprandial stimulus, cholecystokinin release from the duodenum stimulates gall bladder contraction, activating enterohepatic circulation. Subsequently, bile is released into the small intestine where it can execute its digestive role in activating pancreatic lipases, promoting solubilisation and facilitating absorption into the enterocytes (Zhou and Hylemon, 2014). Once its emulsification role has been achieved, BAs are then reabsorbed into the distal ileum by the actions of the apical sodium-dependent bile transporter (ASBT) and the ileal bile acid-binding protein (IBABP), which actively transport BAs across the enterocyte membranes. Finally BAs can be transported by organic solute transporters (OST α /OST β) into the portal vein where they can return to the liver (Monte *et al.*, 2009).

A small percentage of primary BAs are not reabsorbed and are instead transformed further by anaerobic bacteria in the large intestine which act to deconjugate, epimerise and dehydroxylate them, producing secondary BAs. Deconjugation is carried out by a hydrolase enzyme that is widely distributed in several species of intestinal bacteria, and which hydrolyses the *N*-acyl amide bond. Epimerisation and oxidation of 3-, 7- and 12- hydroxyl groups are catalysed by hydroxysteroid dehydrogenase enzymes, which are responsible for the reversible change in stereochemistry from α to β configurations and *vice versa*; for example, converting CDCA to ursodeoxycholic acid (UDCA). However, in terms of secondary BA production, C7 dehydroxylation can be considered the most important bacterial biotransformation, as it is responsible for the production of secondary BAs, LCA and DCA from their primary precursors, CDCA and CA respectively. The 7 α dehydroxylation reaction appears to be restricted to free, deconjugated BAs and despite the high turnover rate, the enzymes are found in a very small percentage of intestinal flora (Ridlon *et al.*, 2006). Nonetheless, the de-conjugation and dehydroxylation of primary BAs decreases their solubility and increases their hydrophobicity. Their increased hydrophobicity is linked to an increased toxicity and secondary BA, LCA, has been shown to contribute to the pathogenesis of cholestatic liver injury and even colon cancer (McGarr *et al.*, 2005). Whilst the majority of LCA is excreted, up to 4% of LCA and most of DCA can be recovered across the colonic epithelium by passive diffusion and returned to the liver where they can be taken up by sodium (Na $^+$)-taurocholate co-transporting polypeptide (NTCP)

and organic anion transporters (OATP), and detoxified by sulfation and amidation (Hofmann, 2004). In the liver, re-conjugated BAs can be subsequently re-secreted into the gall bladder with newly synthesised BAs, where they can re-join enterohepatic circulation.

This cycle (seen in Figure 1.3) occurs between 4 and 12 times a day and is vital for the transport of nutrients from the intestines to the liver, elimination of xenobiotics and excretion of endogenous compounds. Although extremely efficient, the enterohepatic circulation loses approximately 5% of BAs every day by excretion in the faeces, and so *de novo* synthesis from cholesterol is imperative for the replenishment of the BA pool.

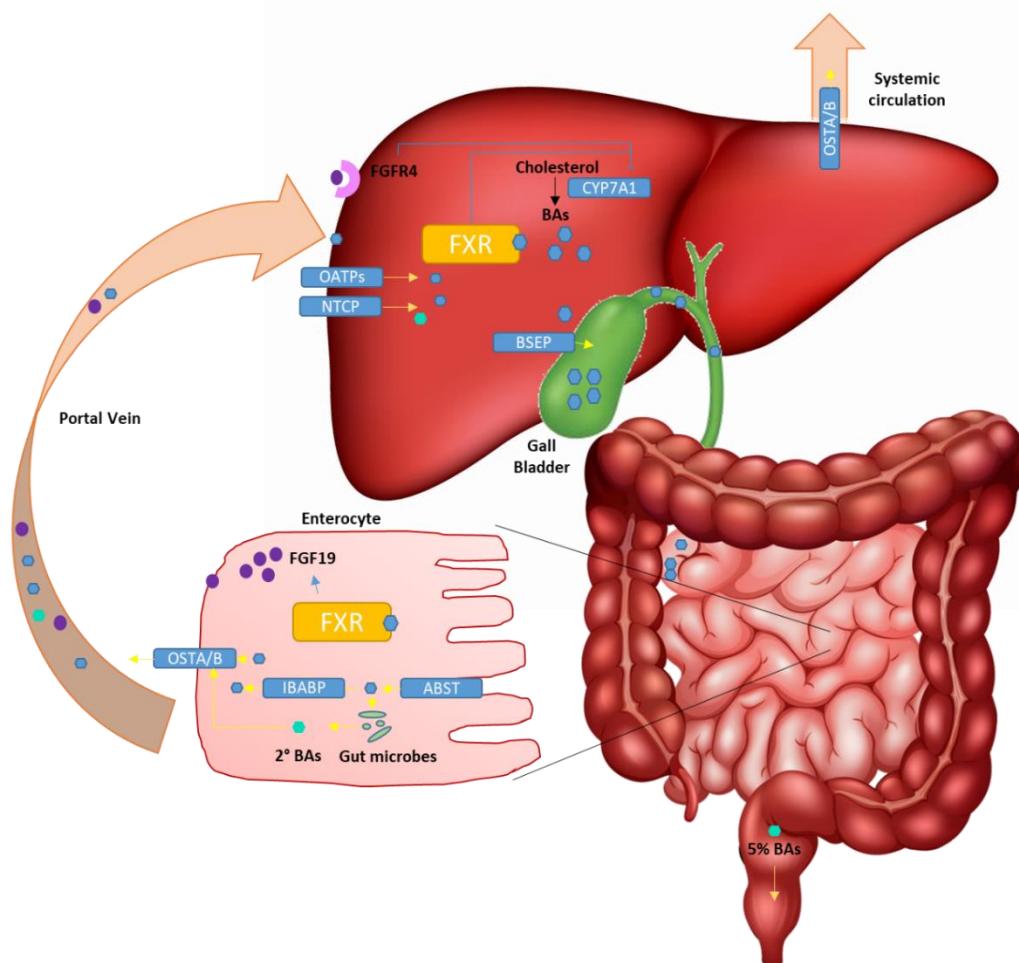


Figure 1.3 Overview of enterohepatic circulation of bile acids after biosynthesis and conjugation. BAs are transported out of hepatocytes and into the gall bladder through transporters such as BSEP, where with cholesterol, phospholipids, bilirubin and other proteins, it forms bile. Bile is released into the small intestines after feeding. After facilitating digestion and absorption, BAs are reclaimed from the intestinal lumen by ABST. Intracellular IBABP creates a BA flux and protects the enterocytes from BA toxicity and OSTa/B transfers BA to the portal vein where it can be recycled to the liver. Similar reuptake mechanisms are present in colonocytes and proximal renal tubules (not shown), to prevent BA loss in the urine and faeces. BAs from the portal circulation can be taken up by NTCP and to a lesser extent, OATPs. FXR, Farnesoid X Receptor; FGF19, Fibroblast Growth Factor 19.

1.2 The Role of Bile acids

1.2.1 The Classic Role of Bile Acids

As mentioned, the canonical role of BAs is in aiding several processes throughout digestion and absorption in the intestines. In summary, BAs can emulsify the insoluble end-products of triglyceride digestion, fatty acids and 2-monoglycerides (Hofmann, 1999). By forming a spherical or cylindrical lattice, whereby the polar lipids are arranged radially with their hydrophilic heads facing the aqueous phase, and the BA molecules lying perpendicular, with their hydrophobic face creating a wedge between the lipid alkyl chains, a mixed micelle can be formed. This formation of mixed micelles greatly increases the aqueous concentration of the fatty acids and monoglycerides, which in turn accelerates diffusion by a factor of 100 (Hofmann, 1999). In addition to the formation of mixed micelles that aid absorption, BAs form mixed micelles with phosphatidylcholine and cholesterol, where they promote the flow of bile from the hepatocytes, to the bile canaliculi and to the gall bladder, solubilizing cholesterol, preventing the formation of gall stones. Finally, another main role of BAs, is that they represent one of the main forms of cholesterol elimination from the body.

1.2.2 The Role of Bile Acids as Signalling Molecules

While their role as digestive surfactants have long been established, over the last two decades, BAs have been recognised as signalling molecules with transduction pathways central to several complex processes. To modulate these processes, which include bile production, insulin sensitivity, energy expenditure, gut motility, immune cell response and BA, glucose and lipid metabolism, BAs activate both intracellular nuclear hormone receptors and cell surface G protein-coupled receptors (Copple and Li, 2016). The functional importance of BA interactions with some of these receptors is yet to be elucidated, but due to the wide distribution of BA receptors throughout the body, it is very likely that new functional roles of BAs will continue to emerge.

1.2.2.1 Nuclear Receptors as Targets of Bile Acids

Whilst their solubilisation properties are central to their role as biological detergents, hydrophobic BAs can be toxic and sometimes lead to the disruption of membranes in cellular compartments, ultimately leading to apoptosis and necrosis. As such, the cellular availability and concentration of BAs needs to be tightly regulated. In order to do this, BAs regulate their own synthesis, metabolism and transport, by the coordination of a network of nuclear receptors (Chiang, 2013).

Nuclear receptors are ligand activated transcription factors that mediate many pleiotropic physiological pathways involved in development, metabolism and reproduction. The receptor contains a DNA binding domain (DBD), which is responsible for recognising and binding to a consensus hormone response element found on its target gene. The C-terminus of the receptor, connected to the *N*-terminal domain by a variable hinge region, contains a ligand binding domain (LBD). The LBD is highly conserved between receptors of the same subfamily and contains a hydrophobic pocket, which is useful in the identification and lodging of small molecule ligands. Binding to an agonist, causes a conformational change in the LBD, and ultimately recruits coactivator proteins to the promoter of the target gene. Once recruited, through a series of chromatin remodelling and post-translational modifications, target gene transcription can be initiated, and the nuclear receptor can modulate downstream effects in direct response to signalling molecules.

BAs have been shown to directly interact with three different nuclear hormone receptors; Farnesoid X receptor (FXR), pregnane X receptor (PXR) and vitamin D receptor (VDR) (Juřica *et al.*, 2016). These receptors are all highly expressed in the liver and intestines, and act as sensors for BAs in enterohepatic circulation, modulating genes involved in BA, nutrient and drug metabolism (Copple and Li, 2016).

1.2.2.1.1 The Farnesoid X Receptor

The identification of FXR as a key BA receptor was crucial in the elucidation of BA metabolism regulation, hence why FXR is one of the most studied targets of BAs (De Aguiar Vallim *et al.*, 2013). FXR is highly expressed in the liver, kidneys, intestine and adrenal glands, while lower levels of expression exist in the heart, adipose and hormone-responsive tissues. The pleiotropic effects arising from FXR activation include regulation of BA metabolism and transport, lipoprotein and glucose metabolism. Endogenous BA can bind to FXR in their free or conjugated forms and several synthetic agonists, both structurally similar and dissimilar to BAs, have been developed, helping to uncover the potential of targeting FXR for the treatment of several human diseases.

1.2.2.1.2 The Pregnan X Receptor

PXR is highly expressed in the intestine and liver, but low levels are expressed in other tissues. Considered to be a xenobiotics sensor, it is a promiscuous receptor that is activated by glucocorticoids, steroids, antifungals, macrolide antibiotics and some herbal extracts, as well as specific BAs. PXR, in coordination with another nuclear receptor, the constitutive androstane receptor (CAR), acts to regulate the expression of phase I detoxifying cytochrome P450 enzymes, phase II conjugation enzymes and phase III uptake and efflux transporters in a process

that promotes the detoxification, transport and elimination of xenobiotics from the body (Lehmann *et al.*, 1998; Zhang *et al.*, 2004). PXR has also been shown to suppress CYP7A1 expression, and thus BA synthesis (Kandel *et al.*, 2016). Similarly, PXR has been shown to regulate the expression of BA conjugation enzymes and transporters. PXR is only activated by LCA and its derivative 3-keto-lithocholic acid. However, LCA, in most circumstances, is present at concentrations too low to elicit a response in PXR. The exception to this being in cases of cholestasis, where the intrahepatic bile ducts rupture, causing reduced bile flow from the liver to intestines. It is thought, therefore, that by increasing the expression of hydroxylating and conjugating enzymes, PXR activation can lead to the conversion of cytotoxic LCA to more hydrophilic, less toxic BAs; serving to limit liver injury during severe cholestasis (Bachs *et al.*, 1992; Marschall *et al.*, 2005).

1.2.2.1.3 The Vitamin D Receptor

VDRs are expressed in multiple tissues throughout the body, including the kidneys, intestines, osteoblasts and macrophages. However, they are not expressed in hepatocytes, but rather in non-parenchymal cells such as Kupffer cells and sinusoidal endothelial cells (Gascon-Barré *et al.*, 2003). The classic endogenous ligand for VDR is vitamin D, which has a role in the regulation of bone and calcium metabolism, cellular growth and differentiation and immunity. As with PXR, LCA and its metabolites can activate VDR, whilst CA, CDCA and DCA cannot (Makishima *et al.*, 2002). Likewise, VDR activation can also promote the expression of detoxification enzymes in the intestines, in addition to mediating BA inhibition of the CYP7A1 enzyme; actions which also provide protection in the gut from BA toxicity (Copple and Li, 2016). Furthermore, with its role in immunity and inflammation, VDR may provide added protective benefits in liver cholestasis.

1.2.2.2 G-protein Coupled Bile Acid Receptor – TGR5

Since 2002, it has been established that BAs are also ligands for a plasma membrane bound G-protein coupled receptor that could activate signal transduction pathways independently of gene expression regulation (Maruyama *et al.*, 2002; Kawamata *et al.*, 2003). This receptor, referred to as membrane-type bile acid receptor (M-BAR), G protein-coupled bile acid receptor (GBAR1) and most commonly, Takeda G-protein receptor 5 (TGR5), is expressed ubiquitously throughout the body, suggesting that BA activity reaches beyond the digestive tract. TGR5 expression is found, but not limited to the heart, lungs, kidneys, stomach, gallbladder, various endocrine glands and non-parenchymal cells, and is involved in a broad range of cell-specific processes (Kawamata *et al.*, 2003). TGR5 is activated by free and conjugated BAs, with the most potent activators being taurine conjugated LCA; and stimulation leads to an increase in cyclic adenosine monophosphate (cAMP), which in turn activates protein kinase A (PKA) leading to

the phosphorylation of target proteins such as cAMP response element binding protein (CREB). TGR5 is thought to be the main component responsible for the immunomodulatory effects of BAs, and it suppresses macrophage activation by inhibiting nuclear factor kappa B (NF κ B)-mediated inflammatory cytokine production, protecting against inflammatory diseases such as atherosclerosis, inflammatory bowel syndrome and fatty liver diseases (Pols *et al.*, 2011; Wang *et al.*, 2011). Furthermore, TGR5 expressed in the sinusoidal endothelial cells in the liver, is thought to regulate cAMP-dependent endothelial nitric oxide synthase, in a mechanism that can scavenge BA-induced reactive oxygen species (ROS), protecting the liver against BA-induced injury (Keitel *et al.*, 2007). TGR5 receptors have a role in promoting smooth muscle relaxation and gall bladder refilling, helping protect against cholesterol gallstone development. In brown adipose tissue, BA-activated TGR5/cAMP signalling regulates mitochondrial oxygen consumption and energy expenditure, whereas cAMP-induced glucagon-like peptide 1 (GLP-1) secretion was shown to play a critical role in regulating glucose homeostasis, insulin secretion, insulin sensitivity and appetite (Thomas *et al.*, 2009).

1.3 The Farnesoid X Receptor

Despite the numerous receptors responsible for the hormone-like signalling properties of BAs, significant interest has been directed at FXR, the master regulator of BA homeostasis and numerous other clinically relevant pathways.

1.3.1 Characterization of FXR

In the early 1990's, two independent studies lead to the discovery of FXR; screening for nuclear receptor DBDs in a rat liver cDNA library by Formann and colleagues, and the isolation of two murine FXR homologs from a yeast two-hybrid screen using human retinoid X receptor (RXR) LBD as a bait protein (Forman *et al.*, 1995; Seol *et al.*, 1995). Two different isoforms of FXR, initially named RXR-Interacting Protein 14 (RIP14), were discovered in mice; expression was detected in the liver and kidney (Seol *et al.*, 1995). It was also originally identified that farnesol, an intermediate of the mevalonate pathway, could activate RIP14, albeit at supraphysiological concentrations, and as such, this orphan receptor adopted the name Farnesoid X Receptor (Forman *et al.*, 1995). Later studies, however, suggested that farnesol could not bind to FXR directly and several groups trying to uncover the mechanisms by which BA homeostasis is controlled, acknowledged BAs as the endogenous ligands of FXR (Makishima, 1999; Parks *et al.*, 1999; Wang *et al.*, 1999). BA binding was confirmed by both ELISA (Makishima, 1999) and mobility shift assays (Wang *et al.*, 1999), and fluorescence resonance energy transfer (FRET)

demonstrated that, upon activation, FXR undergoes a conformational change promoting coactivator recruitment (Parks *et al.*, 1999). Furthermore, cell based assays demonstrated that transcription of BA metabolism genes could be induced by BA-bound FXR (Makishima, 1999; Parks *et al.*, 1999; Wang *et al.*, 1999). Chemical fractionation of the biliary extract revealed that CDCA was the component responsible for the “dramatic” activation of FXR, in a highly specific response; CDCA could not activate other nuclear hormone receptors (Wang *et al.*, 1999). Cell-free FRET studies indicated that FXR-mediated coactivator recruitment could be activated by other BAs such as CA, DCA and LCA, although with a lower efficacy than CDCA (Makishima, 1999; Parks *et al.*, 1999). Conversely, FXR was completely unresponsive to some BAs, such as UDCA, suggesting structure dependant activation (Parks *et al.*, 1999).

1.3.2 FXR Gene Structure and Tissue Expression

There are two genes which encode FXR; *fxrα* (NR1H4), which is highly conserved in many vertebrate species, and *fxrβ* (NR1H5), which exists as a pseudogene in humans. Throughout this thesis, the term ‘FXR’ will be used to refer to FXR α . The *fxrα* gene is composed of 10 introns and 11 exons and encodes four different isoforms, FXR α 1(+), FXR α 1(-), FXR α 2(+) and FXR α 2(-). The different isoforms arise due to a combination of different promoter usage and alternative mRNA splicing. The starting point of mRNA transcription for FXR α 1 isoforms lies upstream of exon 1, whereas FXR α 2 isoforms are driven by a promoter in exon 3, resulting in different starting sequences, and although FXR α 2 variants are 187bp shorter than FXR α 1 transcripts, they encode an additional 37 amino acids (Zhang *et al.*, 2003). Meanwhile, the (+) and (-) isoforms differ in their transcripts with either the random addition or deletion of a four amino acid (MYTG) insertion within exon 5, in the hinge region adjacent to the DNA binding region (Zhang *et al.*, 2003). Exons 8-11, containing the LBD are completely conserved between all four isoforms, suggesting that FXR ligands will bind to any of the isoforms in a nonselective manner (Zhang *et al.*, 2003). Nonetheless, there appears to be differential expression of the FXR isoforms; although all four isoforms may be expressed in a particular cell type, one variant may be significantly dominant. The liver and hepatocyte cells, those with an active steroid metabolism, predominantly express FXR α 1, whereas the colon and intestines and other cells involved in enterohepatic circulation, display specificity towards FXR α 2 isoforms (Huber *et al.*, 2002; Vaquero *et al.*, 2013).

In addition to the variable tissue distribution, the FXR α 1/2 (+) isoforms differ from the (-) isoforms in their ability to modulate certain FXR functions. The addition of the MYTG amino acid insertion in the hinge region, potentially alters the tertiary structure of the protein’s DNA

binding region, limiting the ability of FXR to bind to certain response elements (Zhang *et al.*, 2003).

Whilst studies have shown that FXR α 1(-) appears to be the most potent variant in promoting the upregulation of target genes in both the liver and intestines, it is thought that the overall pattern of isoform expression can have profound effects on the sensitivity and transcriptional response of specific tissues to FXR ligands (Vaquero *et al.*, 2013).

1.3.3 FXR Structure and Activation

As with other nuclear receptors (NRs), the full length FXR protein is organised in modular domains. The *N*-terminal modulatory region contains an activation function 1 (AF1), which is often responsible for the constitutive transcriptional activity of the receptor. Although the AF1 is ligand independent, its transcriptional ability is very weak, and often requires summation in the form of ligand-dependent activation of an alternative activation function at the *C*-terminal. Succeeding the *N*-terminal domain lies the centrally located DBD. The DBD contains two zinc (Zn^{2+}) finger motifs, which recognise and bind to specific nucleotide sequence response elements. NRs are usually found as dimers. They can either exist in the cytosol as a homodimer, which, once a ligand binds, translocates to the nucleus, where it can bind its respective response elements, or the NR can sit in the nucleus, pre-bound to its target DNA, as a heterodimer with its partner receptor RXR. In general, FXR exists as a heterodimer, however homodimers have been observed, particularly on target genes that are negatively regulated by the receptor (Fiorucci *et al.*, 2007).

The LBD is located at the *C*-terminus and acts as an important molecular switch. Its architecture is vital for the creation of a 723\AA^3 hydrophobic cavity that aids ligand binding and for the recognition of BAs. This ligand binding pocket must accommodate the unique structure of BAs; the amphipathic properties of BAs are used for structural recognition and discrimination from other steroid hormones (Mi *et al.*, 2003). FXR crystal structures have shown that similarly to other NRs, the LBD exists as a 12 α -helix bundle that forms a three-layered sandwich, and the amino acids involved in ligand binding, located on helices 3,5,6,7 and 10/11, are conserved between human and rat homologues, (Downes *et al.*, 2003; Mi *et al.*, 2003; Liping Yang *et al.*, 2014). However, in contrast to classic steroid receptors which have the D ring of their respective hormones facing helix 12, BAs bind with their backbone reversed so that the A ring faces the corner and the C24 carboxylic group is situated near the pocket entry point (Mi *et al.*, 2003). When this carboxyl group is substituted or conjugated with glycine or taurine, as commonly seen in naturally occurring BAs, the derivative group lies completely out of the ligand binding pocket, explaining why it does not hinder FXR activation (Mi *et al.*, 2003).

Explained in more detail in Chapter 5, agonistic activation of FXR occurs by a key conformational change that is induced upon ligand binding. The ligand dependent AF2 is located within helix 12 of the LBD, and its conformation determines whether transcriptional coactivators can be recruited to the protein, in a complex that can bind RNA polymerase to initiate the transcription of the genes that lie downstream of the response element. Inactivation by antagonists, however, occur when the active conformation of FXR with coactivator complexes is destabilized, or when the receptor's affinity for a corepressor complex is increased. Using this mechanism, NRS, such as FXR, can act as sensors for various biological molecules, activating numerous downstream pathways in response to increased concentrations of such molecules.

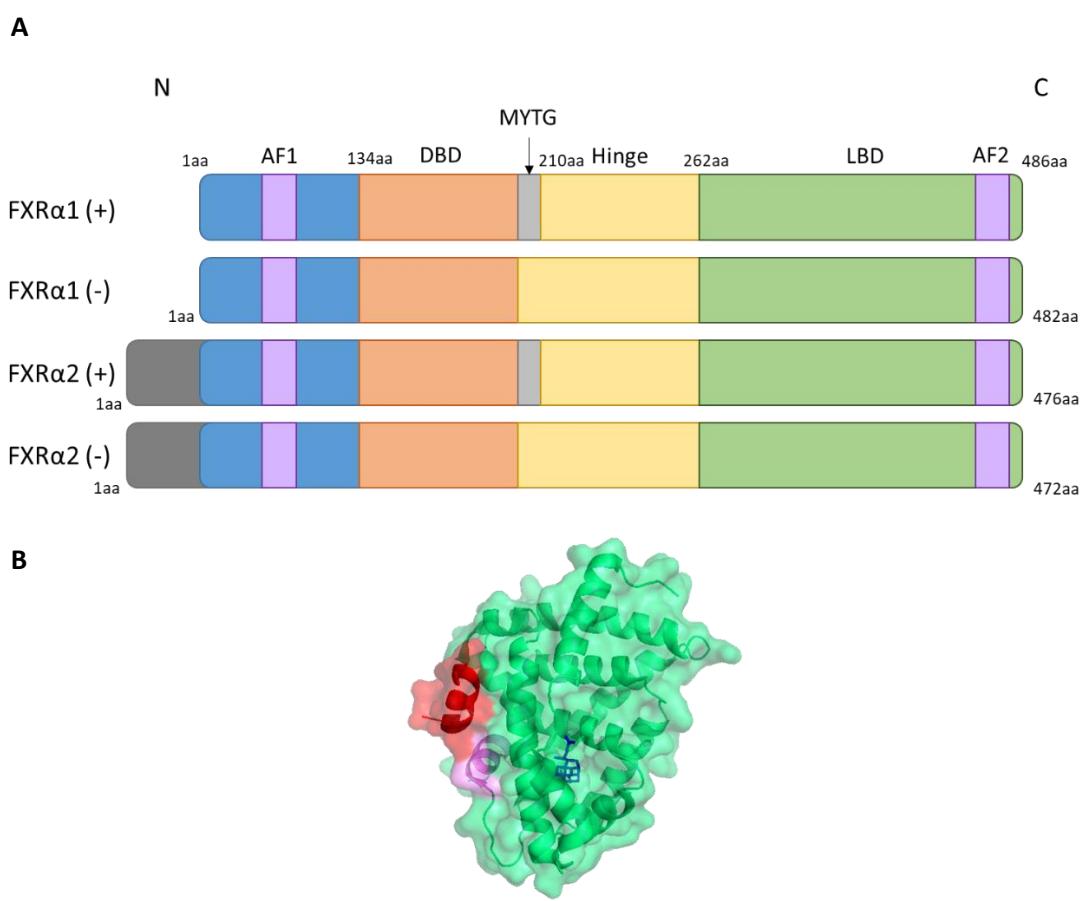


Figure 1. 4 Genomic and tertiary protein structure of FXR. (A) The four protein isoforms of FXR and their modular domains. N-terminal region (blue), DBD (orange), hinge region (yellow), LBD (green), activation functions (purple), splice variations (grey). (B) The tertiary alpha bundle structure of the ligand binding domain, depicting cognate ligand CDCA (blue), AF2 helix 12 (purple) and coactivator peptide helix (red) (PDB ID: 4QE6).

1.3.4 Biological Functions of FXR

FXR is responsible for the regulation of numerous pleiotropic pathways. The detailed signalling cascades contributing to BA, lipid, cholesterol and glucose metabolism are discussed in more depth in Chapter 6 but are briefly outlined here (summarized in Table 1. 1).

1.3.4.1 BA Metabolism

The generation of FXR knock-out mice was essential for its identification as the master regulator of BA homeostasis in hepatic tissues (Kok *et al.*, 2003). BA-dependent inhibition of CYP7A1 and BA synthesis was not observed when mice lacked FXR, and FXR null mice demonstrated more susceptibility to BA toxicity and liver injury in experimentally-induced cholestasis (Wagner *et al.*, 2003). These results confirmed FXR's role in protecting the liver against the detergent effects of BAs. In order to maintain appropriate concentrations of BAs, FXR regulates multiple pathways in both the intestines and liver, whereby gut-liver crosstalk ultimately leads to the inhibition of many of the enzymes required for CDCA or CA synthesis (Figure 1. 5). FXR can also limit hepatic intracellular BA concentration, by upregulating genes for the transporters responsible for biliary secretion, and by downregulating genes responsible for uptake transporters of BAs into the cell. Similarly, intestinal FXR can promote efflux out of enterocytes, by also inhibiting their uptake transporters, and by activating their transport across the basolateral membrane (Moschetta *et al.*, 2004).

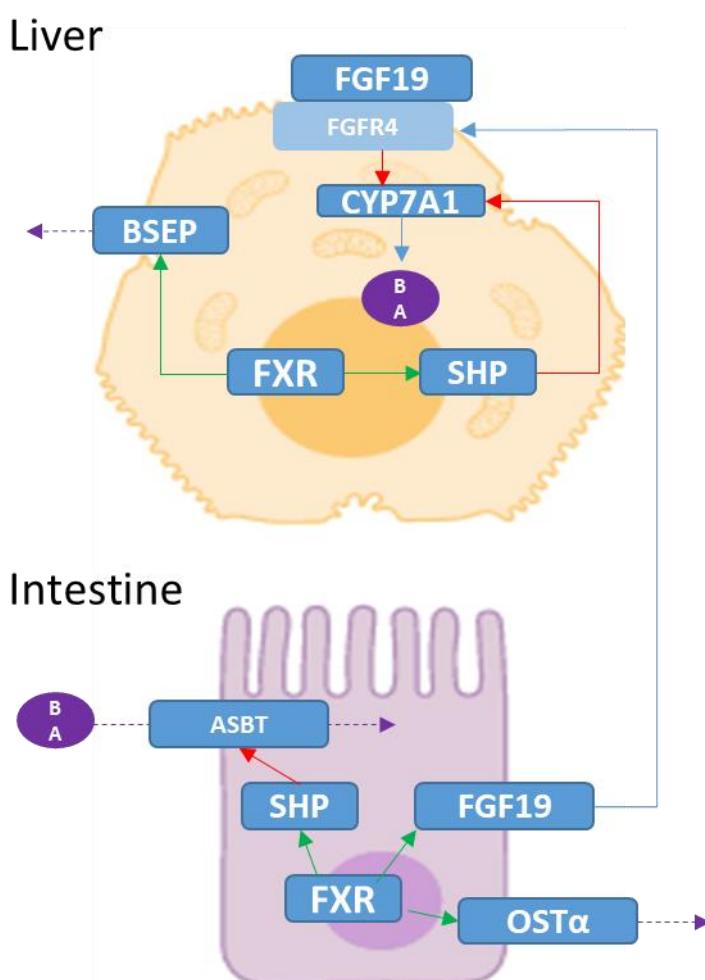


Figure 1. 5 FXR signalling actions on hepatocytes in the liver and enterocytes in the intestines. Activation of FXR leads to the positive regulation (green arrows) of several genes whose functions are to maintain intracellular BA levels by inhibiting (red arrows) their synthesis by CYP7A1, by inhibiting their uptake into the cell, e.g. ASBT, and by promoting their efflux out of the cell, BSEP and OST α . BA transport denoted by purple dotted arrows. ASBT, apical bile salt transporter; BA, bile acid; BSEP, bile salt export pump; CYP7A1, cholesterol 7-alpha hydroxylase; FGF19, fibroblast growth factor 19; FGFR4, fibroblast growth factor receptor 4; FXR, farnesoid x receptor; OST α , organic solute transporter alpha; SHP, small heterodimer partner.

1.3.4.2 Lipid, Cholesterol and Glucose Metabolism

Clinical studies into patients with dyslipidaemia, treated with BA sequestrants, and patients with a CYP7A1 deficiency, highlighted an apparent inverse relationship between the BA pool and triglyceride concentration (Crouse, 1987; Pullinger *et al.*, 2002). A BA receptor was initially implicated in lipid regulation when it was realised that CDCA administration could lower plasma triglyceride concentration in patients being treated for gallstones and for monogenic familial hypertriglyceridemia, whereas UDCA did not have this effect (Leiss and Von Bermann, 1982). The breakthrough came when FXR knockout mice displayed a pro-atherogenic phenotype, with increased plasma triglycerides and cholesterol, as well as the accumulation of free fatty acids in the liver, confirming the direct and pivotal role of FXR in lipid metabolism (Sinal *et al.*, 2000). FXR null mice also identified a role of FXR in reverse cholesterol transport, where excess cholesterol in the periphery is transported back to the liver where it can be catabolised to BAs or secreted into bile for faecal elimination. In addition to cholesterol transport and lipoprotein formation, FXR has been implicated in fatty acid oxidation, and several studies in rodents have uncovered the complex transcriptional modulation of FXR on genes involved in hepatic lipogenesis and triglyceride clearance.

More recent studies have given rise to another unanticipated function of FXR in regulating hepatic glucose homeostasis, whereby the role of FXR is largely determined by nutritional status and interplay between FXR and other receptors. Glucose homeostasis is dependent on the balance between glucose production from non-carbohydrate substrates (gluconeogenesis), glucose consumption in the peripheral tissues where it is catabolized to produce adenosine triphosphate (ATP) (glycolysis), and glucose storage by conversion to glycogen (glycogenesis). These processes are under tight control by insulin activity, which if disrupted, can lead to altered glucose levels, metabolic complications and diseases such as T2DM.

FXR knock out mice displayed mild glucose intolerance and dampened insulin signalling, thus leading to insulin resistance, in the muscle and liver (Ma *et al.*, 2006; Zhang *et al.*, 2006). It is thought that by lowering levels of triglycerides and free fatty acids, FXR activation can increase insulin sensitivity in both the liver and peripheral tissues, such as adipose tissue and skeletal muscle (Zhang *et al.*, 2006). Although other mechanisms are still being elucidated, evidence suggests that FXR can initiate the signalling cascades that are responsible for the reduction of gluconeogenesis and the increase in glycogen storage, as seen in diabetic rats (Ma *et al.*, 2006; Zhang *et al.*, 2006).

1.3.4.3 Inflammation

Growing evidence suggests a role of FXR in regulating cellular inflammatory responses. *In vitro* studies in HepG2 cells demonstrated that FXR could inhibit the major inflammatory hallmark NF κ B at a transcriptional level regardless of whether an endogenous ligand was present. Furthermore, the synthetic FXR agonist, GW4064, was able to prevent downstream NF- κ B signalling both in HepG2 cells and in mice primary hepatocytes (Wang *et al.*, 2008). NF- κ B activation and concomitant inflammation is pivotal to the progression of several disorders, emphasizing the importance of targeting FXR for therapeutic purposes.

Table 1. 1 Summary of main biological functions of FXR

Role	Increased	Decreased
BA Homeostasis	BA secretion	BA synthesis
		BA uptake into hepatocytes
		BA absorption into enterocytes
Lipid Metabolism	Triglyceride clearance	Triglyceride synthesis
	Fatty acid oxidation	VLDL formation
Glucose Metabolism	Insulin signalling	Hepatic gluconeogenesis
	Insulin sensitivity	
	Insulin production	
	Glycogen storage	
Inflammation		NF κ B pathways

1.3.5 FXR and Disease

The physiological importance of its gene targets and its critical role in several metabolic pathways has meant that FXR inevitably plays a pivotal role in the pathogenesis of a wide range of diseases. The continual emergence of evidence suggests that FXR is involved in insulin resistance and many of its manifestations such as diabetes and non-alcoholic steatohepatitis, as well as pathologies concerning inflammation. Furthermore, the elevated serum levels of BAs in patients with hepatic encephalopathy and the recent discovery of FXR mRNA and protein in neurons suggests a potential role of FXR in the development of neurological complications in disorders associated with BA over accumulation (Huang *et al.*, 2016)

1.3.5.1 Intestinal Diseases

Several groups have established antimicrobial properties of BAs (Begley *et al.*, 2005). These properties are due, to some extent, to their activation of FXR. Mice lacking FXR displayed symptoms of a compromised epithelial barrier, including the overgrowth of aerobic bacteria in the ileum. Increased bacterial proliferation and mucosal injury can arise from an obstruction in the bile duct, and precedes the bacterial translocation, epithelial breakdown and systemic infection often seen in patients with impaired bile flow. Epithelial barrier deterioration can lead to inflammatory bowel disease where the intestinal mucosa becomes susceptible to inflammation and, in an attempt to prevent foreign antigens and bacterial toxins, further recruits inflammatory mediators (Shaik *et al.*, 2014). Administration of BAs reverses bacterial overgrowth and FXR activation induces a sophisticated pathway involving the upregulation of enteroprotective genes and the down regulation of NF κ B target genes and proinflammatory cytokines (Inagaki *et al.*, 2006; Vavassori *et al.*, 2009).

1.3.5.2 Cholestasis

Cholestasis arises due to impairments in bile formation, whether due to pregnancy, drug administration or hereditary disorders, or alternatively, because of tumours or gallstones, ultimately resulting in the reduction or obstruction of bile flow. This blockage leads to the accumulation of biliary constituents including cytotoxic BAs, which as previously mentioned, can lead to hepatic injury, fibrosis, inflammation and bile duct hyperproliferation, as seen in the cholestatic diseases primary biliary cholangitis (PBC) and primary sclerosing cholangitis (PSC). To date, endoscopy and surgery are the most popular and effective treatment for obstructive cholestasis. The most common treatment for non-obstructive cholestasis is the administration of hydrophilic BA, UDCA, which reduces the hydrophobicity and thus toxicity of the BA pool, but this is largely considered to be ineffective (Paumgartner and Beuers, 2002). As the master regulator of BA synthesis, detoxification and transport, FXR activation can lead to a decreased and more hydrophilic BA pool, protecting the liver from damage. Evidence for the role of FXR in the development of cholestasis was seen in FXR knockout mice which displayed similar symptoms to a hereditary form of the disease. Furthermore, in mice and rat models of cholestasis, treatment with synthetic FXR agonists, reduced bile duct inflammation and proliferation, suggesting a protective role of FXR activation (Liu *et al.*, 2003; S. Fiorucci *et al.*, 2005). In contrast, FXR knockout mice were also shown to have increased protection against liver injury with an associated increase in the expression of multi-drug resistance associated protein 4 (MRP4), a basolateral transporter responsible for renal BA elimination (Stedman *et al.*, 2006). MRP4 expression is thought to be an adaptive response protecting the liver from BA

accumulation by redirecting their transport to the kidneys for excretion, highlighting the complex network of mechanisms underlying BA homeostasis.

1.3.5.3 Atherosclerosis

FXR suppression of NF κ B signalling is thought to ameliorate vascular inflammation and inhibit the development of atherosclerotic lesions central to many vascular dysfunctions including atherosclerosis (Shaik *et al.*, 2014). FXR is shown to be expressed in vascular smooth muscle cells, atherosclerotic lesions and endothelial cells, albeit at very low levels (Bishop-Bailey *et al.*, 2004). Vascular smooth muscle cells are critical for the maintenance of blood vessels, however inflammation in these cells is thought to contribute to the pathogenesis of atherosclerosis. In addition to smooth muscle cells, macrophages release inflammatory mediators in response to increased uptake of lipoproteins further adding to the development of atherosclerotic plaques. The proinflammatory enzymes responsible for these features, inducible nitric oxide synthase and cyclo-oxygenase 2, lie down stream of NF- κ B. In addition to FXR's negative regulation of cholesterol uptake and vasoconstrictive protein, endothelin-1 expression, FXR activation could be beneficial against the accumulation of lipids, fibrous elements and inflammatory cells that lead to vascular dysfunction (Li *et al.*, 2007; Mencarelli *et al.*, 2009). Furthermore, in addition to FXR's genomic effects on inflammatory cells and lipid altering mechanisms, FXR was shown to be expressed in anuclear platelet cells, with the ability to inhibit their thrombus-forming function by modulating cyclic nucleotide signalling (Moraes *et al.*, 2017). Accordingly, FXR may provide a link between the metabolic and vascular pathways in the development of atherosclerosis. Nonetheless, atherosclerosis is a complex pathology encompassing many systemic and vascular parameters of which FXR's exact effects still need to be fully clarified. FXR activation also results in the unsolicited decrease in High Density Lipoprotein (HDL) levels (Gutierrez *et al.*, 2006), emphasizing the need for selective FXR modulators that can affect systemic inflammation and lipid metabolism without altering HDL levels.

1.3.5.4 Carcinogenesis

Its high levels of expression in enterohepatic tissues, and underlying involvement in several metabolic and inflammatory pathways, undoubtedly means that FXR plays a role in the development of particular cancers. One such cancer, and the fourth leading cause of death in adult cancer patients, is colon cancer (Karsa *et al.*, 2010). The pathogenesis of colorectal cancer occurs due to the combinatorial effects of several genetic mutations, a sedentary lifestyle and the increased fat and carbohydrate consumption, and altered hepatic BA secretion associated with a western diet (Slattery, 2000). Although secondary BAs, such as LCA and DCA, are thought to directly promote tumour formation, particularly in patients with chronic inflammatory bowel

disease (Peterlik, 2008), decreased FXR expression was shown to correlate with increased carcinoma prevalence (Maran *et al.*, 2009). Furthermore FXR null mice displayed increased epithelial cell proliferation and tumour formation thought to be exacerbated by increased Wnt signalling as a result of FXR loss (Modica *et al.*, 2008; Maran *et al.*, 2009; Lax *et al.*, 2012).

Another area of great importance is the role of FXR in hepatocellular carcinoma (HCC). HCC is the fifth most common cancer in the world and stems from liver complications such as hepatitis B and C, and both alcoholic and non-alcoholic fatty liver diseases (Modica *et al.*, 2010). Altered BA metabolism precedes HCC due to the deleterious effects of chronic elevated BA levels (Yang *et al.*, 2007). These effects include inflammation, inflammatory cell production of ROS, subsequent oxidative damage to DNA, apoptotic resistance which leads to chronic liver injury and the ensuing hyperproliferation of hepatocytes potentially as a counteractive mechanism (Perez and Britz, 2009). The role of FXR in preventing tumorigenesis appears to be two fold; on the one hand, tightly regulating BA levels to suppress BA-associated toxicity, and on the other hand, inhibiting NF κ B and other signalling molecules involved in inflammation and the downregulation of tumour suppressor proteins (Jiang *et al.*, 2013). Collectively, the evidence strongly suggests a role of FXR in the protection against hepatic tumorigenesis.

1.3.5.5 Liver Fibrosis

Wound contraction, the increased and altered deposition of extracellular matrix (ECM) and the reduced breakdown of ECM components aids the liver scarring process; hepatic fibrosis. Hepatic stellate cells (HSCs) are the major source of ECM in the liver and undergo a trans-differentiation process during chronic liver disease, changing from a fat-storing, quiescent phenotype to a myofibroblast-like phenotype. FXR, along with several other nuclear receptors, is expressed in HSCs, where it can modulate their activity (Fiorucci *et al.*, 2005). FXR signalling promotes the apoptosis of HSCs and the development of their dormant phenotype, and FXR is considered a potential therapeutic target for treating liver fibrosis.

1.3.5.6 Non-alcoholic Fatty Liver Diseases

Non-alcoholic fatty liver disease (NAFLD) is defined by the accumulation of fat in the liver of patients who do not drink excessive alcohol, and as such, is considered a hepatic manifestation of metabolic syndrome. Metabolic imbalances and dysfunctions in lipid and glucose metabolism, along with augmented inflammatory processes, contribute to the aetiology of the disease. Characterization of FXR knock out mice identified a plethora of pathologies, which span the entire spectrum of NAFLD, highlighting the key role of FXR in the disease (Yang *et al.*, 2007).

1.4 Non-alcoholic Steatohepatitis

NAFLD is one of the most prominent causes of chronic liver disease worldwide, and is fast becoming the primary indication for liver transplantation (Bellentani, 2017). NAFLD describes a spectrum of physiological conditions ranging from simple, lipid accumulation in the liver (steatosis), to non-alcoholic steatohepatitis (NASH), which is characterized by histopathological features such as lobular inflammation, hepatocellular ballooning and injury, and a perisinusoidal fibrosis (Haas *et al.*, 2016). It is mostly asymptomatic and closely associated with obesity and features of metabolic syndrome, including hypertension, dyslipidaemia and central adiposity (Buzzetti *et al.*, 2016). In the UK, NAFLD affects up to 30% of the population (Dyson *et al.*, 2014), and it is considered to be a significant health burden, costing an estimated £35 billion in four European countries, including the UK (Younossi *et al.*, 2018). The increasing prevalence of NAFLD mirrors the increasing prevalence of obesity and T2DM, and it's thought that by 2027, 18 million people will have diagnosed NASH in Japan, the US and the EU5 (UK, France, Germany, Italy, Spain)(Cave *et al.*, 2016; Sumida and Yoneda, 2018).

Of those with NAFLD, at least 10-20% go on to develop NASH, which is accompanied by the predisposition of patients to both hepatic and extrahepatic complications such as fibrosis, cirrhosis, HCC and cardiovascular disease, which can significantly impair life expectancy.

1.4.1 NASH Pathogenesis

Although poorly understood, the pathological progression of steatosis to NASH, is considered to consist of multiple parallel 'hits' acting synergistically to progress the disease. Lipotoxicity, oxidative stress and endoplasmic reticulum stress, are thought to sensitize the liver to additional insults mediated by the innate immune defence systems, leading to cytokine-induced cellular damage (Pacana and Sanyal, 2015). One 'hit' involves hepatic *de novo* lipogenesis, which is activated by insulin resistance, arising from hyperinsulinaemia and carbohydrate rich diets. In addition, as a result of insulin resistance, lipolysis in dysfunctional adipocytes is not deactivated, resulting in the leakage of free fatty acids (FFAs), adipokines, and inflammatory cytokines into circulation. These FFAs accumulate in ectopic tissues, such as the liver, where they are stored as triglycerides. However excessive accumulation of FFAs, beyond the threshold of triglyceride storage, ultimately leads to lipotoxicity due to triglyceride-derived toxic metabolites. Moreover, decreased triglyceride clearance and reduced very low density lipoprotein (VLDL) export contributes to the accumulation of fat in the liver.

These events are counteracted by the mitochondria and peroxisomes which attempt to oxidise the fatty acids. However, these organelles eventually become ‘overwhelmed’ and dysfunctional, resulting in the overproduction of ROS and damage associated molecular pattern molecules. In a subsequent ‘hit’, lipid peroxidation and activation of hepatic macrophages, Kupffer cells, by FFAs and ROS, trigger inflammation and apoptosis, and can activate natural immune defence systems via Toll-like receptors. In addition, ROS, together with oxidised low density lipoprotein (LDL) particles can activate HSCs which initiate fibrogenesis. Similarly, insulin resistance-induced endoplasmic reticulum stress, results in adaptive unfolded protein responses and signalling cascades that perpetuate NASH progression by exacerbating insulin resistance and initiating apoptosis (Cusi, 2012).

Furthermore, recent evidence has implicated the gut-liver axis in the progression of the disease. An altered composition of the gut microbiota plus an increased gut permeability has been seen in NAFLD patients, and inflammasome-mediated dysbiosis is also thought to drive NASH progression (Henao-mejia *et al.*, 2012; Mouzaki *et al.*, 2013). Overgrowth of gut microflora is thought to lead to further production of fatty acids, and bacterial-derived toxin, lipopolysaccharide (LPS), which when paired with the increased permeability in the small intestines, results in increased fatty acid absorption and circulation, driving the production of inflammatory cytokines. Furthermore, NAFLD is associated with changes to the population of bacteria that deconjugate BAs and ultimately affect the species in the BA pool, which can concomitantly affect downstream FXR activation and signalling (Tremaroli and Bäckhed, 2012).

Collectively, shown in Figure 1. 6, the lipotoxic hepatic events, in addition to the indirect effects of inflammatory mediators from adipose tissues, the intestines and immune system, highlight the complex, multifactorial, cross-system nature of this metabolic disorder and importantly, reflect multiple potential therapeutic targets of NASH (Haas *et al.*, 2016).

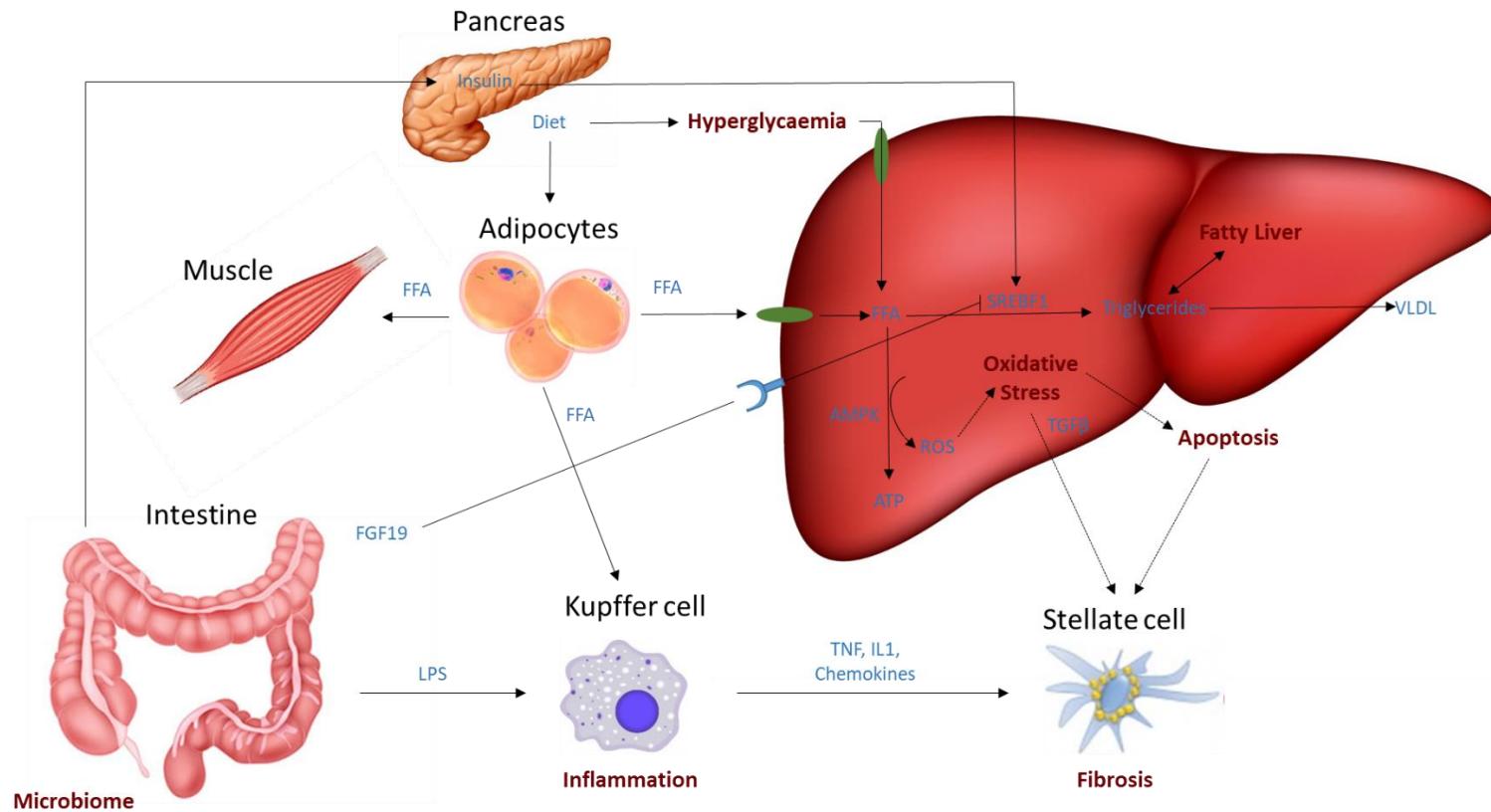


Figure 1. 6 The pathogenesis of NASH. NASH is characterized by fat accumulation, inflammation and fibrosis in the liver. Ongoing lipogenesis results in steatosis, which is exacerbated by the accumulation of free fatty acids (FFAs) and impaired very low density lipoprotein (VLDL) clearance. FFAs from dysfunctional adipocytes can cause lipotoxicity and the activation of Kupffer cell-mediated inflammation. Inflammatory factors can activate stellate cells which in turn initiate fibrogenesis. Oxidation of FFAs in the liver can result in the overproduction of Reactive Oxygen Species (ROS) as a side product, this leads to oxidative stress which further aggravates fibrosis and apoptosis and contributes to NASH progression. AMPK, 5'-adenosine monophosphate-activated protein kinase; ATP, adenosine triphosphate; IL1, interleukin 1; LPS, lipopolysaccharide; SREBF1, sterol regulatory element binding transcription factor 1; TGF β , transforming growth factor Beta; TNF, tumour necrosis factor.

1.4.2 NASH Treatment

Currently, there are no effective medical therapies for NASH treatment or prevention. Most often, weight loss and improved insulin sensitivity through dietary and lifestyle changes are recommended, but as many patients are unable to initiate or maintain these changes, a pharmacological long term solution is needed (Neuschwander-Tetri *et al.*, 2015). Liver transplantation is the only curative treatment for NASH cirrhosis in practice, and even after transplantation, NASH patients are still at a high risk of suffering from cardiovascular complications and mortality, likely associated with metabolic disorders (Vanwagner *et al.*, 2012).

Most treatment strategies target a single feature involved in the pathogenesis of NASH. Vitamin E and antidiabetic thiazolidinediones are the most studied drugs for the treatment of NASH. Although these have a strong antisteatotic response and reduce inflammation, they do not reduce fibrosis (Sanyal *et al.*, 2010). Moreover, concerns about adverse side effects, long term safety and efficacy are yet to be addressed. Whilst improving insulin sensitivity is necessary, it is not completely effective in treating NASH in most patients, hence a drug that can collectively correct insulin resistance, as well as providing anti-inflammatory and anti-fibrotic properties, would be the ideal candidate for the therapeutic treatment of NASH.

Given its functional role in inflammation, lipid and glucose metabolism, FXR ligands are potentially ideal candidates for the treatment of NASH. In addition, FXR's ability to influence hepatic stellate cell-induced fibrosis, as well as activation of cholesterol scavenger receptors (explained in more detail in Chapter 6), means that numerous features of the disease can be targeted in a single receptor. In appreciation of this, significant efforts have been directed at designing or identifying FXR ligands.

1.5 Current FXR agonists

At present, several endogenous BAs, synthetic compounds and natural extracts have been identified as ligands for FXR, and others are continually being discovered (Table 1. 2). Although relatively weak agonists for FXR, knowledge of BAs binding mode to the FXR LBD is crucial for the development of synthetic and semi-synthetic ligands. Furthermore, natural plant extracts such as guggulsterone, cafestol and oleanic acid, also ligands for FXR, provide an insight into gene-specific FXR modulation (Modica *et al.*, 2010).

GW4064 (GlaxoSmithKline, NC, USA) was the first fully synthetic, non-steroidal, high affinity agonist of FXR to be developed, and it was identified by screening a vast library of isoxazole

analogue compounds (Maloney *et al.*, 2000). Since then, it has been used in several animal studies where it has been shown to reduce hepatic inflammation and improve hyperglycaemia, hyperlipidaemia and insulin sensitivity in diabetic, and obese mice (Zhang *et al.*, 2006; Li *et al.*, 2008; Wang *et al.*, 2008). Regardless of being several times more potent than CDCA, with an EC₅₀ of 90nM, GW4064 never advanced to clinical stages after displaying undesirable *in vitro* and *in vivo* properties. The compound was associated with poor rat pharmacokinetics, with a half-life of 3.5 hours and a 10% bioavailability, and its stilbene moiety was unstable and potentially toxic (Maloney *et al.*, 2000). Nonetheless, it is still used as an archetypal standard for FXR agonism in experimental work.

The structure of GW4064 was used as a basis for extensive developments of other nonsteroidal agonists with the aim of improving the physicochemical properties, whilst maintaining the binding affinity. This led to the synthesis of several compounds, many of which, despite being very potent, had limited clinical applicability due to their insolubility and poor absorption, distribution, metabolism and excretion (ADME) properties (Bass *et al.*, 2011; Kinzel *et al.*, 2016). Removal of the stilbene moiety and the addition of different chemical linker groups led to the creation LJN452 (Tropifexor, Novartis), which displayed a favourable safety profile, and to date has been identified as one of the most potent agonists of FXR *in vitro*, with a cell-based EC₅₀ in the range of 1nM (Tully *et al.*, 2017). Tropifexor displayed modestly improved ADME properties, and due to promising results in preclinical animal models, where it significantly regulated FXR target genes, it is currently being tested in phase 2 clinical trials for the treatment of NASH and PBC (Tully *et al.*, 2017).

Throughout the course of drug discovery for FXR ligands, some researchers identified compounds, such as Fexaramine, which, unlike CDCA and GW4064, could induce distinct target gene profiles (Downes *et al.*, 2003). Fexaramine, which is chemically distinct from both BAs and GW4064, was shown to have a more robust effect on specific genes both *in vitro* and *in vivo* compared to CDCA or GW4064. This gene selectivity was thought to arise from its activation of homodimeric FXR and clustering with different coactivator complexes (Downes *et al.*, 2003). Despite showing EC₅₀ values comparable to GW4064 in cell-free assays, Fexaramine was not evaluated for human trials due to its poor absorption when delivered orally, which resulted in preferential activation of FXR in the intestines, rather than the liver and kidney. Nonetheless, activation of intestinal FXR by the gut-restricted Fexaramine, was shown to regulate glucose production, reduce diet-induced weight gain, and reduce global inflammation, suggesting that activation of specific subsets of FXR receptors and target genes, may also prove to be therapeutically beneficial, even without systemic FXR activation (Fang *et al.*, 2015). Whilst

Exenatide received significant interest from media outlets as a promising new 'diet pill' (Macrae, 2015), studies have continued into the use of the compound and its effects on metabolic disorders (Wang *et al.*, 2017).

Despite the numerous non-steroidal FXR agonists identified through the screening of compound libraries and subsequent combinatorial chemistry, a number of semi-synthetic BA-derived agonists, which retain the steroid nucleus, have been developed. Retaining the BA skeleton is advantageous, due to the intrinsic amphipathic properties of BAs, and due to the ability of analogous compounds to enter enterohepatic circulation, without significantly influencing BA population and homeostasis (Lazarević *et al.*, 2019). Furthermore, the fine-tuning of the BA backbone, may help to modulate FXR's activity, and it has been long established that chemical manipulations to both the steroid nucleus and carboxyl side chain, could increase the potency and selectivity of FXR ligands (Modica *et al.*, 2010). Several groups have exploited the BA backbone to develop agonists for FXR, and dual activation of FXR/TGR5 (Lazarević *et al.*, 2019).

The most clinically advanced FXR agonist to date is the semi-synthetic, steroid obeticholic acid (OCA, Intercept Pharmaceuticals, New York, USA), also known as INT-747 and 6 α -ethyl-CDCA, which, as the name suggests, was made by adding an ethyl group to the C6 on the steroid backbone of CDCA. This modification resulted in a 80-fold increased potency of OCA compared with CDCA (Pellicciari *et al.*, 2002). OCA quickly emerged as a promising therapeutic for liver diseases, displaying antifibrotic and anticholestatic effects in rodent models of liver disease (Fiorucci *et al.*, 2005). OCA was the first FXR agonist to be used in human studies, and under the name OCALIVA® (Intercept Pharmaceuticals Inc.), it was recently approved for use in PBC in combination with UDCA, or in monotherapy when UDCA is not tolerated (Mudaliar *et al.*, 2013). A short-term proof of concept study, showed that daily administration of OCA could ameliorate insulin sensitivity in patients with T2DM and NASH, and also was associated with decreased fibrosis markers (Mudaliar *et al.*, 2013). OCA progressed to a randomised clinical trial in NASH patients, where it was shown to ameliorate the NAFLD activity score and decrease the severity of the disease, including hepatocyte ballooning, modular inflammation and again, fibrosis markers. However, it did not significantly reverse NASH; the number of patients displaying NASH resolution was not increased compared to placebo groups. Furthermore, histological improvements seen with OCA, were not significantly better than those seen with other, non FXR-mediated treatments (Neuschwander-Tetri *et al.*, 2015). The trial sparked debate over whether serum levels of hepatic enzymes, traditionally used as surrogate biomarkers indicative of NAFLD, appropriately corresponds with the severity of NASH, which is usually identified by metabolic features, and the relevance of clinical endpoints were questioned. The study also

showed that OCA treatment lead to dyslipidaemia; the observance of increased total serum cholesterol, increased LDL and decreased HDL cholesterol levels, gave rise to safety concerns with long-term OCA treatment (Hegade *et al.*, 2016). Furthermore, OCA was associated with increased fasting insulin concentrations and decreased insulin sensitivity (Neuschwander-Tetri *et al.*, 2015). Paradoxically, one of the adverse effects associated with OCA treatment was the observation of pruritus, which was severe enough to halt the treatment in some patients (Neuschwander-Tetri *et al.*, 2015). The incidence of pruritus was thought to stem from OCA-activation of TGR5 (Alemi *et al.*, 2013), and the promiscuity of this BA analogue highlighted the need for structural modification of the BA backbone to tailor it towards FXR, rather than other BA receptors.

More recently, FXR selectivity has been achieved by the addition of novel hydroxyl groups to the OCA structure (Pellicciari *et al.*, 2016), however, the unfavourable lipid serum profiles associated with OCA remain to be addressed. It is thought that FXR-induced expression of the scavenger receptor B1, causes the increase in hepatic uptake of HDL-cholesterol (Zhang *et al.*, 2010). It has recently been shown that FXR can activate the cholesterol ester transfer protein in humans, but not mice, (Gautier *et al.*, 2013), and this is also thought to account for unfavourable lipoprotein profiles that were not previously observed in mice. As such, the use of global FXR agonists to target certain diseases, such as atherosclerosis, is limited by the unfavourable effects on plasma HDL levels. Likewise, systemic activation of FXR results in increased FGF19, which although is beneficial in its negative feedback inhibition of BA synthesis, can also have deleterious proliferative effects in high levels; overexpression of FGF19 and its receptor FGFR4 in ectopic skeletal muscle has been associated with the formation of tumours (Nicholes *et al.*, 2002). Whether OCA-induced FGF19 elevations can enhance tumour formation remains unknown. However, recently, after several drug-associated deaths, the FDA added a 'black box' label warning to OCA being prescribed by doctors, and it is thought that dose-adaptation and treatment optimization is still required for its use. Research efforts are increasingly focussed on the development of selective FXR modulators that can regulate the functions of individual or small subsets of FXR target genes discretely, in order to target the relevant disease, whilst minimising any side effects associated with chronic treatment (Massafra *et al.*, 2018).

Table 1. 2 Agonists of FXR and their structures

Compound	Class	FXR EC ₅₀ (μ M)	Structure
CDCA	Natural ligand	10 ^a	
GW4064 (GSK)	Non-steroidal	0.07 ^a	
Tropifexor (Novartis)	Non-steroidal	<0.001 ^b	
Fexaramine (Howard Hughes)	Non-steroidal	0.02 ^a	
OCA (Intercept)	Semi-synthetic steroidal	0.1 ^a	

^afrom (Adorini, 2013), ^b from (Tully *et al.*, 2017)

1.6 Research Aims and Objectives

Mounting evidence suggests that FXR has varying, pleiotropic roles, from regulating BA metabolism to mediating inflammation. Accordingly, FXR is a compelling therapeutic drug target for the treatment of diseases associated with the enterohepatic system or altered lipid metabolism and, in particular, those with no current effective treatment, such as NASH. Although FXR agonists are currently being marketed, there is still an ongoing need for the development of highly selective agonists with enhanced activity and reduced adverse side effects.

The hydrophobic nature of FXR's ligand binding pocket requires a ligand that can make key polar contacts, whilst upholding characteristic hydrophobic interactions that can sustain an active receptor conformation. Whilst lipophilic compounds may be potent agonists for FXR, their reduced solubility can be the root of challenging pharmacokinetic properties. BAs, as the cognate ligands for FXR, and due to their congruous amphiphilic structure, provide an appropriate starting scaffold for the design of novel ligands. With the addition of different functional groups to the steroid moiety, research here aims to identify a highly potent agonist of FXR. By doing so, it may be possible to mitigate the adverse effects by only requiring very low doses to be administered to achieve the desired therapeutic effect and subsequently limiting systemic exposure to the compound. Furthermore, if these added extensions to the BA periphery can introduce novel binding mechanisms in the FXR LBD, novel ligands may be able to promote previously undescribed receptor conformations, which may, by virtue of alternative coactivator recruitment, be able to differentially regulate different FXR target genes, steering the compounds towards specific diseases.

Computational methods are frequently used in drug discovery pipelines, and structure-based analysis and molecular docking have proven invaluable in the development of early generations of FXR agonists. In addition to experimental approaches, computer-aided methods may help to streamline this process and may even help to improve our understanding of FXR structure and activation.

The overall aim of this project is to use both computer-aided and experimental wet-lab approaches, encompassing structural, biochemical and molecular techniques to further our understanding of FXR structural dynamics and ligand-mediated activation for the development of a novel, BA-derived agonist with therapeutic potential.

The first objective is to use *in silico* analysis and docking experiments to gain an insight into structural requirements of FXR activation and to discern how prospective ligands bind to the FXR LBD. Using this knowledge, rationale can be applied in the design of appropriate modifications to the ligand scaffold, which could further enhance interactions of the agonist with the receptor. With this input, novel ligands are to be designed by colleagues at NZP UK, who will also be responsible for the synthesis of said compounds and will provide them for proposed work described below.

A second objective of this project is to resolve the crystal structure of FXR to address the fundamental question of how novel ligands can affect structural changes in the FXR LBD, and to postulate how these can subsequently affect function. Expression and isolation, particularly of the FXR LBD, and crystallisation with candidate ligand(s) and cofactors, will help to elucidate the crucial molecular details of the binding modes and interactions, which, in addition to existing structural data for FXR, can be used to inform future development of new compounds.

A third objective is to use cell-free assays and gene expression analysis to observe functional activation of FXR by candidate ligands, at a cellular level. These data will help to evaluate the robustness of computational methods in predicting lead compounds. Furthermore, it will help to determine legitimate activity of the compounds and may allow for the elucidation of gene-specific modulators of FXR. Finally, the testing and observations of lead compound administration in *in vitro* and *in vivo* disease models, may help to confirm the potential of these compounds for use as therapeutics for metabolic liver disease.

Chapter 2. Materials and Methods

2.1 Materials

2.1.1 Equipment and Software

All equipment used in this study is listed in Table 2. 1. The bioinformatics tools and computer software used throughout this work are provided in Table 2. 2.

Table 2. 1 List of equipment

Name	Source
Bio-Robot 8000	Qiagen
Microplate orbital shaker HT9100	Big Bear Automation
96-well magnet type A	Qiagen
Innova 4430 incubator shaker	New Brunswick Scientific
Spectrophotometer CE1020 C1000 series	Cecil Instruments
Avanti J-25 Centrifuge	Beckman Coulter
Allegra 6-K Centrifuge	Beckman Coulter
Spectrafuge 24D Microcentrifuge	Jencons PLS
Digital Sonifier	Branson
FPG12800 pressure cell homogenizer	Stansted Fluid Power
Novex Mini cell gel tank	Invitrogen
Image Scanner III Gel Scanner	GE Healthcare
Transblot SD semi dry transfer cell	Bio rad
Image Quant LAS4000	GE Healthcare
T:Genius Gel imaging system	Syngene
GeneAmp PCR system 2700	Applied Biosystems
StepOne Plus qPCR Machine	Applied Biosystems
Nanodrop 1000 spectrophotometer	Thermofisher
Microplate shaker	Fisherbrand
Digital block heater QBT2	Grant Instruments
Labport vacuum pump	Knf Lab
Autoflow IR direct heat incubator	Nuaire
OMNI Tissue Homogenizer Motor and Soft tissue tips	OMNI International
Nikon TMS inverted microscope	Nikon
AKTA FPLC system	GE Healthcare

Table 2. 2 List of bioinformatics tools and computer software

Tool	Purpose	Sited Accessed	Reference
Protein BLAST	Sequence analysis	blast.ncbi.nlm.nih.gov	(Altschul <i>et al.</i> , 1990)
ClustalW2		ebi.ac.uk/Tools/msa/clustalw2	(Larkin <i>et al.</i> , 2007)
Jalview		-	(Waterhouse <i>et al.</i> , 2009)
IntFold	Tertiary structure prediction	reading.ac.uk/bioinf/IntFOLD	(McGuffin <i>et al.</i> , 2015)
Phyre2		sbg.bio.ic.ac.uk/phyre2	(Kelly <i>et al.</i> , 2015)

ExPASy ProtParam	Protein parameter prediction	web.expasy.org/protparam	(Gasteiger <i>et al.</i> , 2005)
SECRET	Sequence based crystallisability evaluator	mbiljj45.bio.med.uni-muenchen.de:8888/secret/	(Smialowski <i>et al.</i> , 2006)
PyMOL Molecular Graphics system, v2.0		-	Schrödinger, LLC
BioVia Discovery Studio	Visualisation of protein and docked ligands	-	Dassault Systèmes BIOVIA, San Diego: Dassault Systèmes, 2019
LIGPLOT		-	(Wallace <i>et al.</i> , 1995)
SYBYL v8.0	Molecular modelling and docking	-	Tripos Associates Inc., 2007
FLAP 2.2		-	(Baroni <i>et al.</i> , 2007)
ApE v2.0	Plasmid Editor and Viewer	-	M. Wayne Davis, Utah, USA
BLASTp	Primer search and design	ncbi.nlm.nih.gov/tools/primer-blast/	(Ye <i>et al.</i> , 2012)
EMBL-EBI Expression Atlas	Gene and protein expression database	https://www.ebi.ac.uk/gxa/home	(Papatheodorou <i>et al.</i> , 2018)
GraphPad Prism v.7	Analyses and graphical presentation	-	GraphPad Software, CA, USA
IBM SPSS v24	Statistical analyses	-	IBM SPSS Statistics, Armonk, NY: IBM Corp
Image J	Image processing and analysis	https://imagej.nih.gov/ij/index.html	(Abràmoff <i>et al.</i> , 2007)

2.1.2 Molecular Biology Materials

All chemicals and reagents used in this study were supplied by Sigma-Aldrich/MERCK unless otherwise stated and were of the highest purity. Specialty consumables and reagents used in this work are provided in Tables 2.3-2.9.

Table 2. 3 List of consumables and reagents

Name	Source
HisTrap FF column	GE Healthcare
HiTrap TALON column	GE Healthcare
HiTrapQ HP column	GE Healthcare
HiLOAD 26/600 superdex 200pg	GE Healthcare
Ni-NTA magnetic agarose beads	Qiagen
GeBA-Flex Dialysis tubes (3mL, 20mL)	Generon
Amicon Ultra regenerated cellulose centrifugal units 10kDa	Merck Millipore
Pierce Protein Concentrator polyethersulfone 10kDa	Thermofisher
MRC polystyrene 96 well crystallisation plates	Molecular Dimensions
CrystalQuick X2 sitting drop crystallisation plates	MiTegen
MicroAmp fast optical 96 well plates	Thermofisher
96 deep well microplates	BD Falcon
Nunc Microwell polystyrene 96 well plates	Sigma
Corning round bottom, black 384 well plates	Sigma
6x Gel loading dye, purple	New England Biolabs
SYBRsafe DNA gel stain	Invitrogen
Hyperladder 1kb DNA ladder	Bioline
Riboruler high range RNA ladder	Thermofisher
Phusion Flash Master Mix	Thermofisher
AMPure XP PCR Purification system	Agencourt/ Beckman Coulter
Ni-NTA magnetic agarose beads	Qiagen
Isopropyl β-D-1-thiogalactopyranoside (IPTG)	Generon
NuPAGE Bis-Tris 4-12% precast gel	Invitrogen
Low MW range SigmaMarker	Sigma
Wide MW range SigmaMarker	Sigma
Novex sharp prestained protein ladder	Invitrogen
Quick Coomassie	Generon
InstantBlue™ Coomassie protein stain	Expedeon
cComplete™ EDTA-free protease inhibitor cocktail tablets	Roche
Criterion blotting sandwiches	Biorad
Clarity western ECL substrate	Biorad
GeneJet Plasmid mini prep kit	Invitrogen
RNAqueous Total RNA isolation kit	Ambion
iScript CDNA synthesis kit	Biorad
iTaq Universal SYBR green supermix	Biorad
Human reference total RNA	Agilent
Gel filtration marker kit for proteins 12-200kDa	Sigma

Table 2. 4 List of solutions, buffers and media

Solution	Composition
4x Laemmli sample buffer	200mM Tris pH 6.8, 4% (w/v) SDS, 0.4% (w/v) bromophenol blue, 40% (v/v) glycerol, 20% (w/v) tris(2-carboxyethyl)phosphine (TCEP)
1x NuPAGE MOPs running buffer	50 mM MOPS, 50 mM Tris Base, 0.1%(w/v) SDS, 1 mM EDTA, pH 7.7
Semi dry transfer buffer	60mM Tris, 40mM CAPS, pH 9.6 (with 15% methanol for anode), (with 0.1% SDS for cathode)
Tween-20 Tris buffered saline (TBST)	20mM Tris pH 7.5, 150mM NaCl, 0.1% (v/v) Tween-20
Blocking buffer	1x TBST, 5% (w/v) skimmed milk powder
1x TAE agarose gel buffer	40mM Tris, 20mM glacial acetic acid, 1mM EDTA
1x MOPS RNA agarose gel running buffer	40mM MOPS pH7.0, 10mM sodium acetate, 1mM EDTA
Lysis buffer	50mM Tris pH 7.8, 500mM NaCl, 30mM imidazole, 0.2% (v/v) Tween20, (10% (v/v) glycerol)
Ni-NTA wash buffer	50mM Tris pH 7.8, 500mM NaCl, 30mM imidazole
Ni-NTA elution buffer	50mM Tris pH 7.8, 500mM NaCl, 500mM imidazole, 10% (v/v) glycerol
Gel filtration buffer	20mM Tris pH 7.8, 200mM NaCl, 1mM Dithiothreitol (DTT), 10% (v/v) glycerol
Miller's Luria-Bertani broth (LB)	Per litre deionised water: 10g Tryptone, 10g NaCl, 5g Yeast extract
LB agar	Per litre deionised water: 10g Tryptone, 10g NaCl, 5g Yeast extract, 15g agar
SOC medium	2% Tryptone, 0.5% Yeast Extract, 10 mM NaCl, 2.5 mM KCl, 10 mM MgCl ₂ , 10 mM MgSO ₄ , 20 mM glucose
Power broth	Per litre deionised water: 52g Power Broth powder (Molecular Dimensions, UK), 4mL glycerol
Overnight Express™ Instant TB Medium (TB ONEX)	Per litre deionised water: 60g TB ONEX, 10mL glycerol
Eagle's Minimum Essential Medium (EMEM)	EMEM with Earle's salts, non-essential amino acids and sodium bicarbonate, supplemented with 10% and 2mM L-glutamine
Dulbecco's Modified Eagle's Medium (DMEM)	DMEM with low glucose (1000g/L) and sodium bicarbonate, supplemented with 10% FBS and 2mM L -glutamine

Table 2. 5 List of antibiotics

Antibiotic	Stock concentration	Solvent	Working concentration
Carbenicillin	50 mg/mL	Water	50 µg/mL
Cholramphenicol	35 mg/mL	Ethanol	35 µg/mL

Table 2. 6 List of antibodies

Antibody	Type	Dilution	Source
Anti- 6x His tag antibody (mouse monoclonal)	Conjugated (Horse radish peroxidase)	1:1000	Abcam

Table 2. 7 List of vectors

Vector	Antibiotic resistance	Tag	Source
pOPINE	Amp ^R	His ₆ (C-terminal)	Oxford Protein Production Facility
pOPINF	Amp ^R	His ₆ (N-terminal)	
pOPINS3C	Amp ^R	His ₆ , SUMO (both N-terminal)	
pOPINHALO7	Amp ^R	His ₆ , HALO7 (both N-terminal)	
pOPINJ	Amp ^R	His ₆ , GST (both N-terminal)	
pET15b	Amp ^R	His ₆ (N-terminal)	Genscript

Table 2. 8 List of *E. coli* strains

Strain	Genotype	Purpose	Source
One Shot [®] OmniMAX [™] 2 T1	<i>F'</i> { <i>proAB lacIq lacZΔM15 Tn10(TetR) Δ(ccdAB)</i> } <i>mcrA Δ(mrr hsdRMS-mcrBC) Δ 80(lacZ)ΔM15 Δ(lacZYA-argF)U169 endA1 recA1 supE44 thi-1 gyrA96 relA1 tonA panD</i>	Cloning	Invitrogen
NEB5α	<i>fhuA2 (argF-lacZ)U169 phoA glnV44 80 (lacZ)M15 gyrA96 recA1 relA1 endA1 thi-1 hsdR17</i>	Cloning	New England Biolabs
Rosetta (DE3) pLysS	<i>F- ompT hsdSB(rB- mB-) gal dcm λ(DE3) [lacI lacUV5-T7 gene 1 ind1 sam7 nin5] pLysSRARE (Cam^R)</i>	Expression	MERCK
Lemo (DE3)	<i>fhuA2 [lon] ompT gal (λ DE3) [dcm] ΔhsdS/ pLemo (Cam^R)</i>	Expression	New England Biolabs
BL21 (DE3) pLysS	<i>F-, ompT, hsdSB (rB-, mB-), dcm, gal, λ(DE3), pLysS (Cam^R)</i>	Expression	Promega

Table 2. 9 Compounds used for *in vitro* and *in vivo* work

Compound	Structure	Solvent	Source
CDCA		DMSO	NZP UK
OCA		DMSO	NZP UK
Jed441		DMSO	NZP UK
Jed561		DMSO	NZP UK
Jed692		DMSO	NZP UK
Tropifexor		DMSO	AbMole
α -Guggulsterone		DMSO	Sigma

2.2 Methods

2.2.1 The Farnesoid X Receptor Gene

In this study, a variant of the human FXR α gene (NR1H4, GenBank accession number U68233) that encodes isoform 1(+) was used, corresponding to amino acid (aa) sequence accession number Q96RI1-3 (UniProt) (Appendix 2). The gene was synthesised by GenScript (Piscataway, NJ, USA) and codon optimized for use in *E.coli* expression systems.

2.2.2 Structure Prediction and Construct Design

The aa sequence of the full length FXR protein (1-486 aa) was submitted to the Pfam Protein families database (Finn *et al.*, 2016) to identify conserved domain boundaries. The amino acids for the full length FXR and individual domains of FXR also were submitted to the structure prediction servers IntFOLD (McGuffin *et al.*, 2015) and Phyre2 (Kelly *et al.*, 2015). The structures generated were visualised using the PyMOL Molecular Graphics System, Version 2.0 (Schrödinger, LLC). The predicted domain architecture and tertiary structure were used to design appropriate constructs encoding these regions. The aa sequences of the proposed domains were submitted to the ExPASy Protparam server (Gasteiger *et al.*, 2005) to predict several parameters of the final protein including the theoretical isoelectric point (pI), extinction coefficient and stability index.

2.2.3 Analysis of FXR Crystal Structures

FXR crystal structures were visualized using PyMOL Molecular Graphics System. Analysis was carried out using features of the PyMOL software to align multiple structures and to highlight individual amino acid residues, for comparison with the domain and structure prediction boundaries. Additional analysis of the selected FXR crystal structure, PDB ID:4QE6, was carried out using FLAP (Baroni *et al.*, 2007), which was used to probe the ligand binding pocket with small molecule groups representing hydrophobic, hydrogen bond donor/acceptor and shape interactions to calculate the corresponding GRID molecular interaction fields (MIFs).

2.3. Molecular Biology Methods

2.3.1 Primer Design

Forward and reverse primers used throughout this study are listed in the relevant chapters' methods section. The most important consideration for primer design was to have forward and reverse primers with a melting temperature (T_m) between 58-65°C and with less than 4°C difference between the two primers. Primers were designed to be between 18-25 bp in length,

with a GC content between 40-60 % and with the 3' end concluding with a GC clamp. Primers were checked using the BLASTp tool for intra- and inter- primer homology, and to check for non-specific products in the target species genome.

For the ligation independent method, In-Fusion™ cloning (Clontech), primers were designed to facilitate subsequent cloning. In addition to the sequence complimentary to the target DNA, the primers also had a 5'-, 15 bp complimentary sequence to the open ends of the respective destination vector, invalidating the requirement for restriction digestion and minimising the number of steps involved in cloning. The primers were automatically generated by submitting the target protein sequence to the OPTIC database (as implemented by OPPF-UK, Harwell Research Complex) and by selecting the relevant destination vector. A table of the primers used for this experiment can be found in Appendix 1.

2.3.2 Polymerase Chain Reaction (PCR)

Phusion Flash DNA polymerase was used to amplify the gene of interest (GOI) for subsequent insertion into the respective vector. A typical reaction consisted of Phusion Flash Master Mix (Thermofisher), 0.6 μ M of each forward and reverse primer, and 20-40 ng template DNA. Thermal cycling parameters are provided in Table 2.10, and annealing and elongation steps were repeated for 30 cycles.

Table 2. 10 PCR cycling parameters for use with Phusion Flash Master Mix

Step	Temperature (°C)	Time (seconds)
Initial denaturation	98	10
Denaturation	98	1
Annealing	60	5
Extension	72	15/Kb
Extension	72	120

2.3.3 Agarose Gel Electrophoresis

The PCR products were analysed by gel electrophoresis on a 1.6 % (w/v) agarose gel. Agarose was dissolved in 1x TAE by heating briefly before the addition of SYBRsafe DNA gel stain (1:10,000 dilution). PCR products (5 μ L) were mixed with 6x gel loading dye (1 μ L) and 5 μ L loaded into each well. The molecular weight marker, Hyperladder 1 Kb, was run alongside PCR products for comparison. DNA fragments were separated by running the gel at 100 V for 30-40 minutes and the gel visualised using the T:Genius imaging system.

2.3.4 Purification of PCR Products

As all PCR products were relatively small (less than 4 kb) and of good quality (observed by agarose gel electrophoresis), fragments were purified by paramagnetic bead purification. An

initial purification was carried out by incubating PCR products with the restriction enzyme *DpnI* at 37°C for 30-60 minutes. *DpnI* has specificity for methylated DNA and was used to digest the template DNA without interfering with the PCR product.

Purification from excess primers, dNTPs, enzymes and salts was carried out using the Agencourt AMPure XP system, which utilizes solid-phase reversible immobilization technology. AMPure XP reagent containing the paramagnetic beads, was mixed and incubated with PCR reactions to allow the binding of DNA fragments larger than 100 bp. Reactions were placed on a super magnet plate to separate the bead-bound PCR products from unbound reagents. Following multiple ethanol wash steps, DNA was eluted in 10 mM Tris pH 8.0.

2.3.5 Estimation of DNA Concentration

The concentration of DNA was estimated by measuring the absorbance of UV light at 260 nm on a Nanodrop spectrophotometer.

2.3.6 IN-Fusion Cloning

The insertion of the gene of interest (GOI) into the linearized vector was carried out by the sequence independent, ligase-free In-Fusion™ method, which enables the efficient cloning of PCR products into any linearized vector at any locus. The single-step procedure, whereby the purified PCR products were incubated with the linearized vector in the presence of the IN-Fusion™ enzyme at 42°C for 30 minutes, relies on the overlapping homology between the ends of the DNA fragments and significantly streamlines workflow, allowing cloning to be carried out in a high throughput manner.

2.3.7. Heat Shock Transformation of *E. coli* Cells

Aliquots of chemically competent *E. coli* cells (50 µL volume) were thawed on ice and 1 µL DNA (50-100 ng) was added and mixed gently by flicking the microfuge tubes. After a 30 minute incubation on ice, the cell/plasmid mixture was heat shocked for 45 seconds at 42°C and then immediately recovered on ice for 2 minutes. Pre-warmed SOC medium (250 µL) was added to the transformed cells, which were then incubated at 37°C for 1 hour. Following their recovery, 20-100 µL of transformed cells were spread on LB plates supplemented with the appropriate antibiotics, and plates were incubated overnight (approximately 16 hours) at 37°C. For transformations into One Shot® OmniMAX™ 2 T1 *E. coli* cells, LB agar was supplemented with 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) and 0.2% (v/v) 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (X-gal) to allow for blue/white colony screening. The cloned constructs were verified by PCR, using both a vector and an insert specific primer. Again, the PCR products

were run on an agarose gel to confirm the identity of the recombinant plasmid according to size.

2.3.8. Purification of Plasmid DNA

Overnight cultures were prepared by inoculating one single colony into 10 mL LB broth supplemented with appropriate antibiotics, and incubated overnight at 37°C, with aeration by shaking at 250 rotations per minute (rpm). Cells were harvested by centrifugation at 5,000 x g (relative centrifugal force) for 10 minutes. The cell pellet was used for plasmid DNA purification using the GeneJet Plasmid Miniprep kit (Thermofisher) following manufacturers' instructions. Plasmid DNA concentration was measured using a Nanodrop spectrophotometer at 260 nm.

2.3.9 Preparation of Glycerol Stocks

Glycerol stocks, of positively transformed colonies, were prepared from overnight bacterial cultures with a final concentration of 15% (v/v) glycerol. Stocks were mixed in sterile cryovials and stored at -80°C.

2.3.10 DNA Sequencing

The plasmid DNA of selected transformed clones was purified by plasmid miniprep as described in Chapter 2, section 2.3.8, and sequenced by Eurofins, using multiple primers that amplify the regions of the construct where insertions or mutations took place.

2.4. Production of FXR LBD

A more detailed description of the methods involved in the optimization of the expression and purification of FXR LBD can be found in Chapter 4. Methods described here are a generalised procedure.

2.4.1 Expression of FXR LBD

The protein expression protocol used in this study utilized the Rosetta (DE3) *pLysS* strain, designed for optimal expression of eukaryotic proteins, or parent strain BL21 (DE3) *pLysS*. Starting cultures were prepared for each construct in LB medium supplemented with chloramphenicol (35 µg/mL) and carbenicillin (50 µg/mL) and incubated overnight at 37°C with aeration by shaking at 250 rpm. After approximately 16 hours incubation, seed cultures were then used for the inoculation of LB medium, again supplemented with antibiotics as before, and cells initially grown at 37°C. Induction was initiated once the cells had reached a mid-exponential growth phase, indicated by an OD_{600nm} ~0.5, by adding IPTG to a final concentration 0.7 mM. The cells were grown for a further 20 hours at ~15°C, after which the cells were

harvested by centrifugation ($5,000 \times g$ for 10 minutes) and stored at -80°C until ready for use in protein purification.

2.4.2 Protein Purification by Immobilized Metal Affinity Chromatography

The expression of a His-tag on the N'- terminal of FXR LBD facilitated purification of this target protein by Immobilized Metal Affinity Chromatography (IMAC). Purification was carried out using an automated AKTA FPLC system with a 5 mL pre-packed, nickel-nitrilotriacetic acid (Ni-NTA)- based HisTrap or cobalt-based HiTrap TALON® column (both GE Healthcare Life Sciences). Pellets of induced cells were thawed on ice and resuspended in lysis buffer (50 mM tris pH 7.8, 500 mM sodium chloride, 30 mM imidazole, 0.2% (v/v) tween20, 10% (v/v) glycerol) supplemented with cComplete™ EDTA-free protease inhibitor cocktail tablets (Roche), lysozyme (0.5 mg/mL) and DNase I (400 U/mL) (both Sigma). Cells were lysed by sonication at 30% amplitude for 5 minutes with 5 second pulses, or by using a pressure cell homogenizer at 30 kPsi (Stansted Fluid Power) and the crude lysate clarified by centrifugation ($30,000 \times g$ for 30 minutes at 4°C). The supernatant containing soluble proteins was separated and loaded directly onto the HiTrap column. Non-specific, unbound proteins were removed by washing the column with a low imidazole content wash buffer (50 mM tris pH 7.8, 500 mM sodium chloride, 30 mM imidazole), and His-tagged proteins were finally removed from the column by elution in a high imidazole, high salt concentration buffer (50 mM tris pH 7.8, 500 mM sodium chloride, 500 mM imidazole, 10% glycerol).

2.4.3 Additional Purification by Size Exclusion Chromatography

Following IMAC purification, the proteins underwent additional purification from salts, other small molecules and non-specific proteins by size exclusion chromatography (SEC). IMAC purified proteins were dialysed overnight against gel filtration buffer to remove some of the imidazole prior to SEC. This second purification step also was carried out using an AKTA FPLC system with a pre-packed HiLoad 26/600 Superdex 200 prep grade column (GE Healthcare Life Sciences). The column was calibrated using standards of 6 proteins with known, diverse molecular weights, and elution times of these proteins were used to create a standard curve. The elution profile of target proteins was validated with the standard curve to identify the correct peak containing the protein of interest.

2.5 Protein Analysis and Quantification

2.5.1 Analysis of Proteins by Sodium Dodecylsulfate – Polyacrylamide Gel Electrophoresis (SDS-PAGE)

Analysis of expression levels and of the purity of proteins, after each purification step, was determined by SDS-PAGE. Proteins were run on a NuPAGE 4-12% Bis-Tris precast gel (Novex, Invitrogen). Samples were mixed with 4× Laemmli buffer and heated at 95°C for 10 minutes prior to loading. The gel was run for 50 minutes at 200 V in NuPAGE 1×MOPS SDS running buffer (Invitrogen). Proteins were compared to the Novex sharp prestained protein ladder (Invitrogen). The gel was either stained to visualise the proteins with Coomassie-based dye Quick Coomassie (Generon) or used to transfer separated proteins onto a nitrocellulose membrane via Western Blot (WB) analysis. Gels stained by Quick Coomassie were visualised using an Image Scanner III gel scanner (GE Healthcare).

2.5.2 Analysis of Proteins by Western Blot

Proteins were denatured and separated by SDS-PAGE as described (Chapter 2, section 2.5.1). Proteins were then transferred to a nitrocellulose membrane, using a Transblot SD semi dry transfer cell (Biorad), which allows efficient, economical blotting without the need for large volumes of buffer or a gel cassette. Three sheets of Whatman filter paper were pre-soaked in transfer buffer containing either 15% methanol (for the anode) or 0.1% SDS (for the cathode) and the membrane was also briefly pre-soaked in anode transfer buffer. A blotting sandwich was created with the filter paper, membrane and gel, ensuring that the nitrocellulose membrane was closest to the anode plate, whilst the gel was laid directly on top, closest to the cathode plate. Transfer was carried out at 15 V for 90 minutes, after which the membrane was washed briefly in deionised water. The nitrocellulose membrane was blocked for 1 hour at room temperature in TBST containing 5% (w/v) non-fat milk powder. The blocked membrane was then washed with TBST and incubated with the Horse Radish Peroxidase-conjugated Anti- 6x His tag antibody (Abcam). His tagged proteins were detected by chemiluminescence, using Clarity ECL substrate (Biorad) and imaged using the Image Quant LAS4000 (GE Healthcare).

2.5.3 Estimation of Protein Concentration

Protein concentration was estimated by measuring the absorbance at 280 nm, using a Nanodrop Spectrophotometer (Thermofisher). The molar extinction coefficient ($M^{-1}cm^{-1}$) and molecular weight (Da) of the target protein were calculated from its sequence, using the ExPASY ProtParam web-based tool (Gasteiger *et al.*, 2005). Protein concentration was determined by applying the Beer-Lambert law: $Absorbance = molar\ absorption\ coefficient \times concentration \times cell\ path\ length$.

Chapter 3.

Use of Computational Approaches for the Structural Analysis of FXR and Molecular Docking of Novel Ligands

3.1 Introduction

The use of computational methods in modern drug design has increased significantly in recent years, and it is now considered a *de facto* standard tool in medicinal chemistry, complementary to experimental approaches. With traditional drug discovery methods, involving the synthesis and screening of compounds, only 1 in 40,000 compounds tested will be brought to market, taking an estimated 7-12 years and \$1.2 billion for the development of a new drug (Shankar *et al.*, 2006). Furthermore, compounds identified in this way, may not have the optimal structure required for biological activity, meaning that larger doses may have to be administered, increasing the potential risk of undesired side effects. The advancement of the proteomics field and the burgeoning availability of structural information, in addition to high performance software, has meant that computer-aided drug design (CADD) can streamline drug discovery. A focussed, computer-assisted approach, often coined *in silico*, can rapidly identify the most promising compounds to be progressed, whilst eliminating poor candidates at an early stage. This limits chemical synthesis and biological evaluation, minimizing research resources and costs whilst accelerating the process of successful lead identification and optimization (Kapetanovic, 2008).

The two main applications of CADD are structure-based drug design and ligand-based drug design. The former relies on knowledge of the three-dimensional (3D) structure of the receptor and its active site in order to evaluate potential molecular interactions, binding energies and steric relationships between the protein and proposed ligands; whereas the latter, employs statistical analyses to identify relationships between ligands and their specific biological actions on target proteins. Both approaches can be used in the iterative process of lead compound identification, which can serve as a basis for further modifications to optimize the potency, stability, solubility or pharmacokinetics of the compound.

3.1.1 Structure- Based Drug Design

A pre-requisite of the structure- based process is the knowledge or structure determination of the target protein either by experimental methods, such as X-ray crystallography (MX) and Nuclear Magnetic Resonance Spectroscopy (NMR), or by theoretical bioinformatics methods, such as homology modelling. To date, structures obtained by MX have been the most widely used for the purposes of drug design. This is due to both the high resolution of the models generated and also due to its advantage over other techniques in that it can determine structural information of proteins varying in size from a few amino acids to 900kDa. Moreover,

the presence of ordered water molecules in experimental data often proves advantageous in observing key binding mechanisms for some drug targets (Anderson, 2003). NMR spectroscopy has also been valuable in observing the intramolecular dynamics of the target protein, due to its use of proteins in solution. Moreover, even if a structure for a target protein cannot be experimentally determined, advances in protein prediction algorithms and software have meant that homology modelling can provide sufficient spatial information to predict both ligand binding and protein-protein interactions.

An important consideration in structure-based drug design, is identification of the ligand binding sites. Whilst X-ray crystallography provides the advantage that the binding site is usually obvious and frequently deemed to be the position of the ligand within the crystal lattice, non-crystallographic methods require alternative methods, often small fragment probing techniques, to identify the pocket or protuberance with the potential to make various polar, electrostatic and hydrophobic interactions.

The three-dimensional structure of the receptor and the specific spatial arrangement of its interacting amino acids within its binding pocket are pivotal to the identification of putative binding modes of potential ligands; the practical application of this knowledge underlies molecular docking.

3.1.1.1 Molecular Docking

A detailed understanding of the principles that govern molecular interactions and mechanisms that most influence ligand binding, provide a framework for the design of new drugs. The application of these principles to computational approaches, allows for a search space to be explored and ligands to be virtually placed into its protein receptor and potential interactions to be predicted, enabling prospective drugs to be evaluated and accordingly ranked, prior to their synthesis (Mohan et al., 2005). Computational docking aims to generate an optimal ligand-protein conformation and explores different binding poses (conformations and orientations) of a ligand in its specified receptor (Mobley and Dill, 2009). The procedure utilizes molecular mechanics modelling programs, which use different force field parameter-based algorithms to both search for ligand poses, and to score the predictions based on the thermodynamics of the interaction. These force field parameters collectively describe the total potential energy of a given molecule; taking into account the bond energies due to changes in interatomic distance, changes in bond angles, changes in bond conformation, as well as energy contributions due to Van der Waals (VdW) forces, and the electrostatic attractive and repulsive forces acting in a molecule between atoms with a full or partial charge; and subsequently, help to simulate plausible geometry and interactions between receptor and ligand (Thomas, 2003). Many

docking procedures use a precalculated grid map to represent the receptor when defining interaction energies with each ligand. This grid can be reused for each ligand and significantly accelerates the scoring calculations.

Docking approaches and virtual screening are becoming increasingly automated, allowing multiple compounds to be processed with little manual intervention. The use of large chemical compound libraries for virtual screening is advantageous, since the majority of the compounds used are already commercially available and can easily be screened for biological activity. The *de novo* design of novel synthetic ligand scaffolds, however, allows for the bespoke tailoring of compounds to specifically target certain receptors based on the geometry and biochemical microenvironment of the binding site. However, the challenge often occurs in actually synthesizing the intended compounds.

Another major limitation to the procedure, is the fact that most simplistic docking approaches utilize a rigid receptor. Most ligand and receptor interactions are flexible in solution, and the receptor may adopt a range of dynamic conformations dependent on its bound or unbound states. Whilst newer algorithms can account for receptor and ligand flexibility, these are often more computationally expensive and can dramatically increase the time required for docking pose predictions (Anderson, 2003).

Although some examples have shown significant differences between docked and experimental ligand binding poses (Fritz *et al.*, 2001), most cases display very minor differences between the predicted and actual conformations, with deviations between the atomic Cartesian coordinates of less than 2Å (Shoichet *et al.*, 2002); and several success stories have been documented where drugs identified by *in silico* approaches have reached the market (Phillips *et al.*, 2018). As such, due to its speed, reasonable accuracy, and low cost, docking approaches are methods of choice for eliminating non-binding compounds and for identifying or scrutinizing potential leads, and for ultimately enriching hit rates in the initial stages of drug discovery research (Leach *et al.*, 2006).

3.1.2 Ligand-Based Drug Design

Ligand-based drug design is typically used when the target of certain ligands is unknown, or in the absence of a 3D structural model or ligand binding information of the target protein. It relies on the prior identification of ligands that are known to bind to the target, where the desired biological activity of interest previously has been established. The most widely used tools in the field of ligand based drug design are 3D quantitative structure activity relationships (3D-QSAR)

and pharmacophore modelling, and both can be used to understand the structural or physico-chemical properties that correlate with the desired activity of the ligand.

The premise underlying 3D-QSAR is that similar structural features will yield similar activity, and this computational approach aims to quantify the relationship between the two. The experimentally derived biological activity data serves as a dependent variable, whereas a set of molecular descriptors describing the 3D chemical features or ‘fingerprint’ of each ligand can be considered as the independent determining factor. The molecular descriptors, or interaction fields, include the various physical and geometric characteristics of the molecule, and its quantum chemical properties, and these can be used to generate a pharmacophore model which attempts to explain the variability of the biological activity in a set of ‘training’ ligands. This model can then be used for statistical analyses and to determine a final 3D-QSAR model, which, in turn, can be used to predict the activity of new, analogous ligands (Veselovsky and Ivanov, 2003).

One widely used, successful method of 3D-QSAR is comparative molecular field analysis (CoMFA) and the associated method, comparative molecular similarity indices (CoMSIA). Both methods assume that the interactions between a ligand and its receptor are non-covalent and are used to characterize molecular interaction fields around the molecules. By aligning the molecules against a template compound or common pharmacophore on a 3D grid lattice, the steric and electrostatic potential energies, at each grid point, can be simultaneously calculated. While CoMFA does not account for hydrophobic or polar interactions, CoMSIA includes these terms in its energy function.

Both methods use partial least squares (PLS) analysis to determine an ideal model, which correlates the molecular descriptors to activity. PLS combines the statistical techniques of multivariable linear regression and principal component analysis to perform a systematic search of molecular descriptors. By extracting information from these molecular descriptors and the biological data, the method can efficiently reduce the number of independent variables, thus improving the correlation and generating an ideal model (Acharya *et al.*, 2011). The reliability and quality of the model can be assessed by two statistically significant values obtained from PLS analysis, the conventional correlation coefficient (R^2), and the cross-validated coefficient (Q^2), (Höskuldsson, 1988). The model also can be validated by leaving out known compounds when generating the model, then using the model to predict the activity of said compound, and then comparing the predicted versus experimental values.

For 3D-QSAR to be successful, it requires the activity of the compounds to be determined for the same target, ideally by the same standard protocol, in order for the data to be comparable. Likewise, the models with the highest predictive power are derived from molecules that are diverse in both their chemical structure and biological activity, with the difference between the highest and lowest biological property recommended as being at least 4 orders of magnitude (Dearden *et al.*, 2009). Nonetheless, the feasible computational time and the fast generation of models make this indirect modelling approach advantageous (Melo-filho *et al.*, 2014). To date, 3D-QSAR has been shown to be adequately capable in the prediction of close analogues of known compounds, and the information obtained from this approach, regarding certain molecular regions and their ability to positively or negatively affect the activity of the compound, can be used by medicinal chemists to devise alternative functional groups, which may be introduced to specific locations on a molecule's scaffold to enhance biological efficacy or affinity.

3.2 Aims and Objectives

Conventional drug discovery methods were once dominated by the 'brute-force' approach of high throughput screening of synthetic and natural compound libraries to identify potential lead compounds. However, this technique was associated with a low hit-rate, and hits that were identified often had poor efficacy and little scope for further optimization (Sliwoski *et al.*, 2013). Moreover, the cost and time taken to synthesize compounds, test them and obtain results, drove the impetus for more rational procedures. The nascent field of structural biology brought with it structural information for thousands of previously uncharacterized proteins; and due to the availability of large protein and compound databases, together with advances in computer software and technology, the use of computational methods in drug design emerged and quickly became an attractive tool to accelerate the drug discovery pipeline. The use of computer-aided methods in the field of drug design can significantly reduce the time and cost required to bring a drug to market, as a wider, more diverse chemical space can be explored without the need to synthesize every compound to be tested *in vitro*. With virtual screening, inactive compounds can be quickly discarded, and potential active compounds can be prioritized, streamlining the process and resulting in a much higher hit-rate. Furthermore, computational tools provide the scope for the prediction of pharmacokinetic properties based on the structure of the ligand, or the prediction of toxic and carcinogenic side effects of compounds and may be used at other stages in the preclinical development process.

The overall aim of the work in this chapter, is to use computational tools to aid the understanding of the structure of the FXR LBD, and to complement other approaches in the design of novel FXR ligands. To date, several high-resolution structures of the FXR LBD have been published in the Protein Data Bank (PDB, www.wwpdb.org). These structures, which have been produced by co-crystallisation with diverse sets of ligands, provide additional insight into this receptor and the unique binding modes required for either receptor activation or inhibition. By analysing and comparing these experimentally derived structures, it may be possible to deduce the primary chemical features of the ligand binding pocket and key residues involved in small molecule binding. The advantage of readily available structures also means that molecular docking methods can be employed to screen potential ligands. An aim of this chapter is to dock potential ligands and to observe their mode of interactions with FXR, to drive the development of a novel set of compounds to specifically target this receptor. Furthermore, parallel to the structure-based approach, these novel compounds will be synthesized by NZP UK and used in FXR-specific reporter assays carried out by DiscoverX (Eurofins). The data provided by DiscoverX will be used in this lab for subsequent applications in ligand-based QSAR models, and will help to define the distinct chemical signatures on the ligand backbone that may lead to improved affinity and efficacy of the novel compounds.

3.3 Materials and Methods

3.3.1 Homology Modelling

For the most part of this work, a structure of the protein or domain of interest was available in the Protein Data Bank. However, in the case of the full length FXR protein, or TGR5, where no structures were available, a homology model was generated. To do this, the amino acid sequence of the target protein was initially used to perform a protein BLAST search (BLASTp) and homologous receptors aligned and compared, using the multiple sequence alignment server Clustalw2 (ebi.ac.uk/Tools/msa/clustalw2). The aa sequence was then also submitted to the online protein structure and function prediction server, IntFOLD (www.reading.ac.uk/bioinf/IntFOLD). The model used was the top ranked structure with the highest global modality score and confidence values, and template structures used for this model were cross-referenced in the PDB and evaluated for their rational use as a relevant template protein.

3.3.2 Initial Analysis of FXR LBD – Ligand Bound Conformations

Prior to *in silico* ligand docking and screening, the FXR LBD structure was scrutinised. The FXR LBD adopts different structural conformations according to the nature of the ligand bound, as shown by the diversity of the X-ray crystallographic structures of FXR deposited in the PDB. Structures were visualised and compared, using PyMOL Molecular Graphics System, (Version 2.0.4 Schrödinger, LLC). Structures were superimposed to facilitate comparison and the align script used in the command line to generate all atom root mean square deviation (RMSD) scores. Reported RMSD values are the output after several refinement cycles. One FXR LBD structure (PDB ID: 4QE6), from a human homologue of the gene, gave a high-resolution structure co-crystallised with the classic BA agonist, CDCA. This structure, based on its canonical active conformation, was chosen as a reference for all subsequent alignments. Furthermore, due to the similarity of the bound ligand to the compounds of interest herein and the high quality of the structure, PDB ID: 4QE6 was chosen for the initial analysis and docking studies.

3.3.3 Analysis of the FXR Ligand Binding Pocket

Analysis of the FXR ligand binding pocket was carried out using FLAP software (Fingerprints for Ligands and Proteins, version 2.2, Molecular Discovery, UK). The software combines GRID molecular interaction fields (Goodford, 1985), and pharmacophoric fingerprints. Molecular interaction fields (MIFs) can be described as the distribution of potentials of certain physical interactions, such as hydrogen bonding, hydrophobic, electrostatic or VdW interactions (Cross and Cruciani, 2010), and the use of the GRID force field which is specifically designed for the

characterization of ligand binding to biological macromolecules, improves the accuracy of the method. The PDB coordinate file 4QE6 was uploaded into the FLAP software for analysis. The ligand binding pocket was automatically defined, using the CDCA ligand as a guide, but the radius parameter for the cavity search was extended to 6Å to ensure that all possible sub-pockets were included in the analysis. The resolution of the GRID cage was left at 0.75Å, as per the automated settings. The GRID method places a ‘cage’ over the cavity and performs an iterative search of interactions at each position on the grid. The method routinely utilizes over 60 chemical probes, but only 6 of the main probes were used to describe the MIFs within the FXR LBD and on the CDCA ligand structure. The 6 probes used were; the carbonyl oxygen ‘O’ probe for hydrogen bond donors, amide ‘N1’ probe for hydrogen bond acceptors, ‘DRY’ probe for steric or hydrophobic entropy, ‘CRY’ probe for identifying both lipophilic and hydrophobic characteristics, ‘OH2’ probe for identifying key water molecule interactions and the ‘H’ probe for describing the shape of the interaction. MIFs were then merged to form discrete pharmacophoric fingerprints, representing favourable or unfavourable interactions. These MIFs can then be used for comparisons of potential ligands for their biological targets and can overcome the speed-accuracy constraints associated with other methods of *in silico* docking and virtual screening. In this case, the FLAP software was used solely for the initial characterization of the binding pocket, and later for attempts at conducting a 3D-QSAR study.

3.3.4 Molecular Modelling and Docking

In order to explore and rationalise the interactions and potential binding mechanisms of novel compounds, molecular modelling was performed using the SYBYL 7.2 software package (Tripos, St. Louis, USA).

3.3.4.1 Compound Library

Compounds were designed by colleagues at NZP UK. Most compounds maintained their 5-ring steroidal nucleus, with additions and substitutions made to the A and B rings, and additions made to the side chain (Figure 3. 1). Some compounds had an open A-ring structure. The 3D structure of each novel compound was built using ChemDraw 16 and Chem3D software (PerkinElmer Informatics). Energy minimisation of the compounds also were performed in Chem3D, using the MM2 force field. Some compounds were extracted from their published PDB coordinate files using the structure preparation tool in SYBYL, where the ligand was removed and saved as its own object. All compounds were imported into SYBYL and hydrogen atoms were added in idealised geometries, where necessary. Ideal bond lengths, bond angles and torsion angles were generated in SYBYL. An appropriate force field for BAs or steroidal compounds has not been previously described in the literature, and so charges were assigned

to each atom of each molecule using the Merck Molecular Force Field (MMFF94) available in the SYBYL software suite, which provides good accuracy for a broad range of organic molecules, and includes common atom types such as carbon, nitrogen, sulphur and fluorine. Localised energy minimisations of the conformation were performed using the Powell conjugate gradient minimization method with the default convergence criterion, 0.05 kcal/mol and the number of optimization iterations set to 1000. The final structure for each ligand of interest was taken in its lowest energy conformation and these minimised conformers were used for subsequent docking experiments. The resulting 3D coordinate files were converted to a mol2format for use in Surflex-Dock experiments, using SYBYL. A list of all compounds built and used for molecular docking can be found in Tables 3. 1-3. 5.

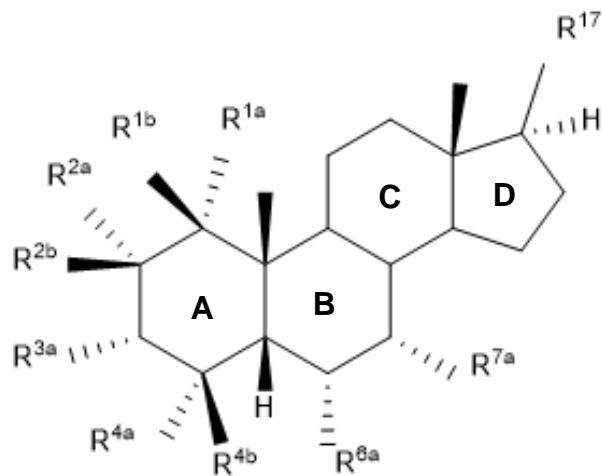


Figure 3. 1 General formula of novel compounds based on bile acid backbone. 'R' groups represent positions where additions or substitutions occur, with numbers representing the carbon position, and a/b differentiating the configuration of the group compared to the carbon stereocentre.

Table 3. 1 CDCA-derived compounds with alterations to A/B ring or side chain

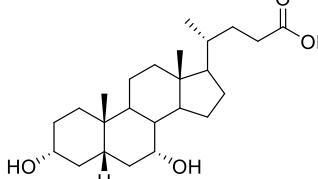
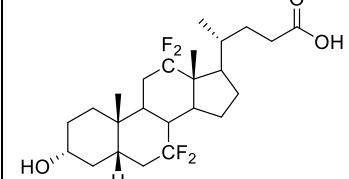
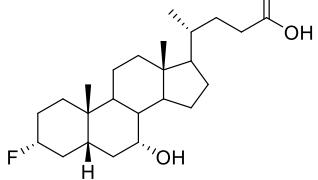
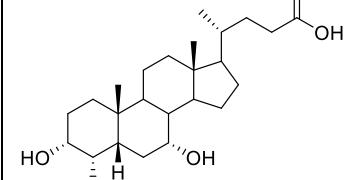
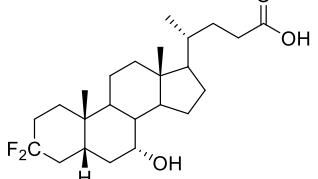
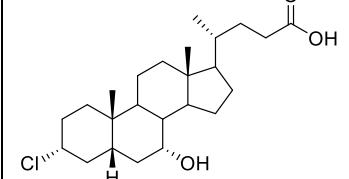
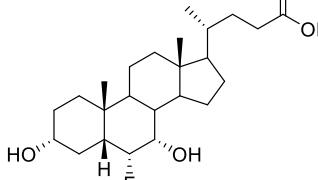
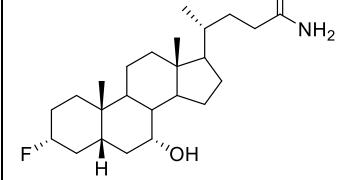
Compound ID (synonym)	Structure	Compound ID (synonym)	Structure
Jed18 (CDCA)		Jed104	
Jed79		Jed106	
Jed85		Jed130	
Jed86		Jed379	

Table 3. 2 OCA-derived compounds with substitutions at the tail end (R^{17}) only

Compound ID (synonym)	Structure	Compound ID (synonym)	Structure
Jed20 (OCA)		Jed584	
Jed150		Jed585	
Jed120		Jed586	
Jed243		Jed645	
Jed567		Jed588	
Jed572		Jed589	

Jed573		Jed590	
Jed575		Jed636	
Jed576		Jed637	
Jed577		Jed638	
Jed578		Jed639	
Jed579		Jed640	
Jed580		Jed641	

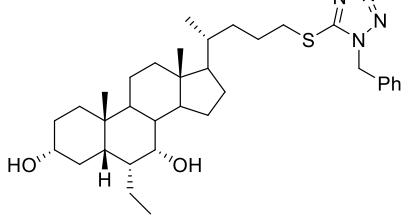
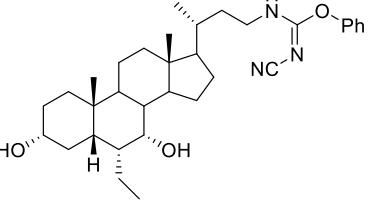
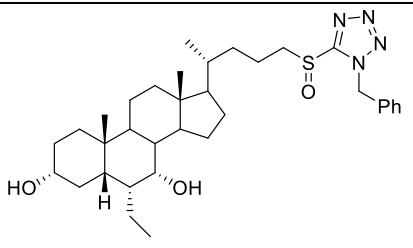
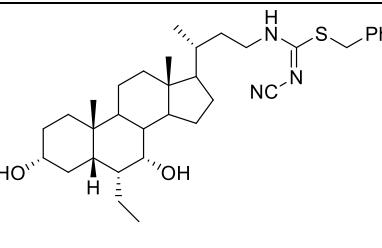
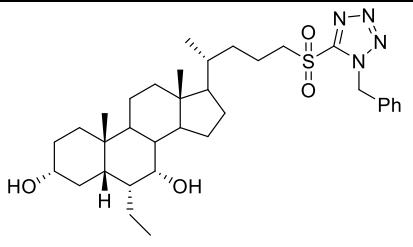
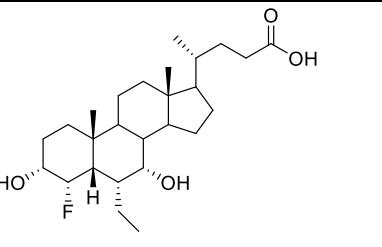
Jed581		Jed642	
Jed582		Jed643	
Jed583		Jed644	

Table 3. 3 OCA-derived compounds with extensions at 3C position and altered tail ends

Compound ID (synonym)	Structure
Jed181	
Jed183	
Jed184	
Jed303	
Jed401	
Jed402	
Jed403	

Jed404	
Jed420	
Jed431	
Jed433	
Jed434	

Table 3. 4 OCA-derived compounds with fluorine substitutions on A ring and additions to tail end

Compound ID (synonym)	Structure	Compound ID (synonym)	Structure
Jed678 (1 β -F-OCA)		Jed385	
Jed665 (2 α -F-OCA)		Jed406	
Jed397 (2 β -F-OCA)		Jed407	
Jed664 (4 α -F-OCA)		Jed441	
Jed432 (4 β -F-OCA)		Jed442	
Jed556		Jed443	

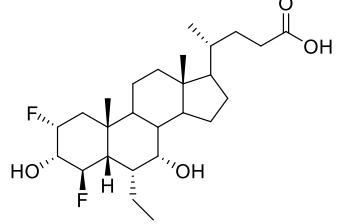
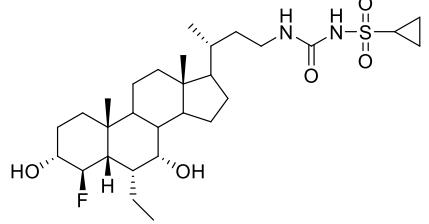
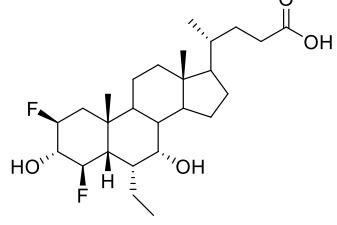
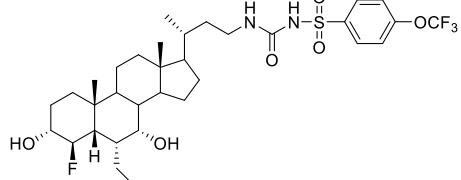
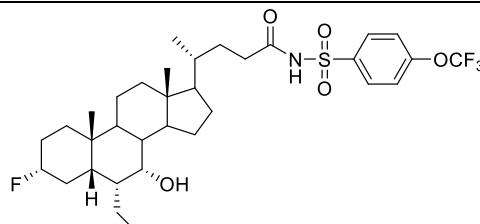
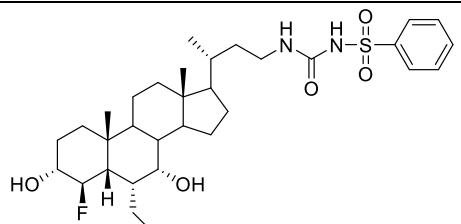
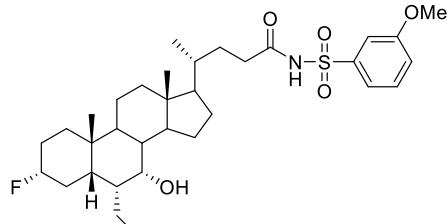
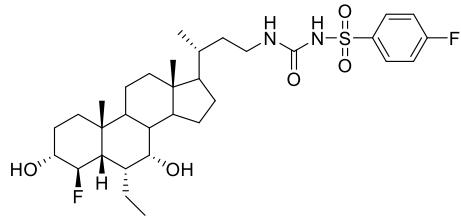
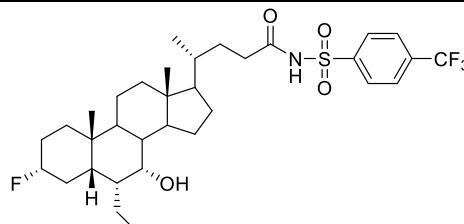
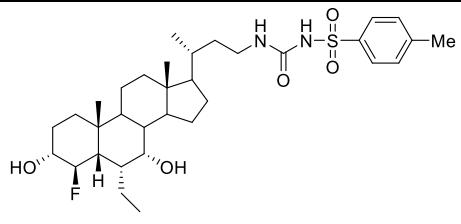
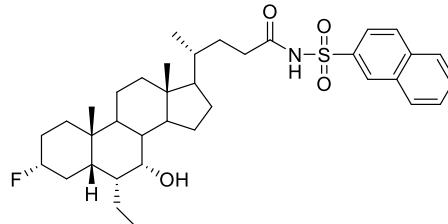
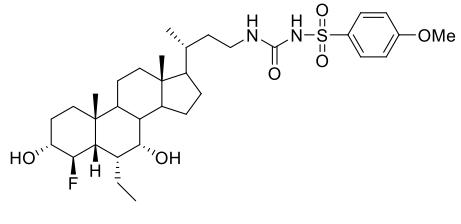
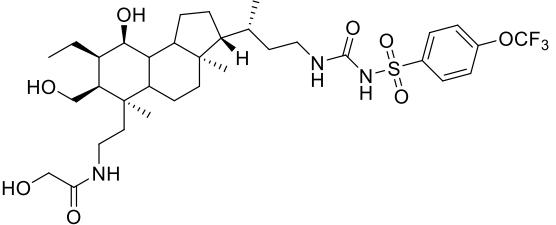
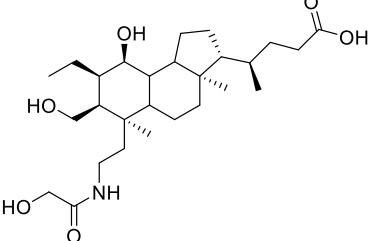
Jed557		Jed559	
Jed558		Jed560	
Jed381		Jed561	
Jed382		Jed562	
Jed383		Jed563	
Jed384		Jed564	

Table 3. 5 Compounds with open A ring structure

Compound ID (synonym)	Structure
Jed388	
Jed389	
Jed390	
Jed391	
Jed392	
Jed393	
Jed394	

Jed395	
Jed444	

3.3.4.2 Receptor Structure Optimization

The experimental crystal structures of FXR LBD co-crystallised with CDCA, N-Benzyl-N-(3-(tert-butyl)-4-hydroxyphenyl)-2,6-dichloro-4-(dimethylamino) Benzamide (NDB), or without a ligand (PDB IDs: 4QE6, 4OIV and 5Q0K, respectively) were obtained from the Protein Data Bank and optimized in SYBYL for molecular docking. To prepare the structure for docking experiments, the Biopolymer Structure Preparation tool, within the SYBYL software suite, was used to remove all non-protein atoms from the coordinate file, these included: the co-crystallised ligand, water molecules and any bound ions present. The receptor (protein only) then was prepared by repairing side chains, adding hydrogen atoms in idealised geometries, adding AMBER7 partial charges and fixing side chain amides and side chain bumps. The resulting optimised protein coordinate file was saved as a mol2 file for use in ligand docking, as described below.

3.3.4.3 Docking Procedure

The ligand docking experiments utilized Surflex-Dock, a fully automatic flexible algorithm in SYBYL that can rapidly generate suitable poses for molecular fragments. Surflex-Dock combines the empirical scoring function of the classic algorithm, Hammerhead, with a surface-based morphological similarity method. It employs a protomol-based method to generate a putative binding site, based on the protein structure provided. The automatic, site-based approach, similar to the GRID method, utilizes a probe-based system and associated scoring function, whereby H-bond donor probes (N-H), H-bond acceptor probes (C=O) and hydrophobic probes (CH₄) are placed into the ligand binding site in different, discrete positions and conformations, to assess the entire binding site for its potential to interact with the ligand. High scoring probes are merged to form a ‘pocket’, or protomol, which represents an assumed, standard pseudo-ligand or set of molecular fragments that characterize the binding site. The protomol directs the alignment of ligands in the active site and the conformation of each ligand is compared to the

protomol. To do this, each ligand is fragmented at positions of rotatable bonds (excluding rigid ring structures), and each fragment is conformationally aligned with the protomol. The ligand is then incrementally reconstructed based on the maximal similarity of the fragment to the protomol and using the constraints of the fragment position on the ligand scaffold.

Ten poses were generated for each compound and these were ranked based on molecular similarity to the protomol. All parameters within the docking suite were left at the default values for the docking experiments. The Surflex-Dock scoring function is based on empirical data from published protein-ligand X-ray structures, and taking into consideration VdW surfaces, favourable polar or non-polar atom pairs, and entropic terms, a score is generated to represent a potential value for binding affinity.

Docked ligands were visualised and interactions analysed using PyMOL Molecular Graphics System, Version 2.0.4 (Schrödinger, LLC) and BIOVIA Discovery Studio (Dassault Systèmes, San Diego, USA). Potential hydrogen bonds (H bonds) were assigned when the distance between two electronegative atoms was less than 3.3Å, whereas any separation greater than 3.3Å, but less than 4Å was considered as a VdW interaction. Binding affinities were disregarded in this work but results of binding mechanisms were documented.

3.3.5 Three-Dimensional Quantitative Structure Activity Relationship (3D-QSAR)

Three dimensional QSAR is an approach using a set of ligands with known activity against a certain receptor target molecule, and a variety of chemical descriptors. The relationship between these descriptors and the activity is defined using statistical analysis. The results may help to identify specific chemical descriptors that correlate most with improved activity or that are detrimental to activity, and as such the technique can help with the optimization of drug design.

3.3.5.1 3D-QSAR Using FLAP

A 3D-QSAR was attempted in FLAP (Molecular Discovery). Compounds, that had been built in Chem3D (PerkinElmer), and that associated FXR activation EC₅₀ data had been acquired, were imported into FLAP as a database and formed what is termed the 'training set'. The compounds were initially aligned, using the structure of OCA as a reference template. The alignment uses a fuzzy subgraph matching algorithm to identify common sub-structural features, and due to the complexity of the ring structures of the compounds, the parameters were set to align the molecules based on bond type only. Activity data was transformed to a logarithmic scale for small differences to be amplified, making it easier to discriminate between compounds with similar activities.

The structural components of each compound were analysed using the GRID MIFs as probes, as described previously (Chapter 3, section 3.3.3), and the 3D-QSAR procedure was run using the standard implementation in FLAP. A statistical model was then created, using PLS regression which correlates the independent variables, the structural parameters as determined by the GRID MIFs, with the dependent, EC₅₀ activity data. The PLS technique is useful for when the independent variables outnumber the data points, and the model generated summarises the data in the most simplistic format, using the least number of different combinations of the independent variables to produce an optimum model. The robustness of the model was then validated by a process known as ‘Leave One Out’, where one compound is excluded, and its activity predicted by the resulting model that is yielded. The process was repeated for all compounds (50), and the cross-validation correlation coefficient (Q^2), which serves as a quantitative measure of the predictive power of the model, as well as the correlation coefficient of determination (R^2), which serves as a measure of variance, was determined. Likewise, the optimum number of components needed to create a regression model also was computed. A plot of the ‘predicted’ versus ‘experimental’ activity values was produced, which also can be used to assess the validity and accuracy of the QSAR model that has been generated. Finally, the PLS loadings plot showed the contribution of each MIF to the activity of each compound.

3.3.4.2 3D-QSAR Using Volsurf+

A 3D-QSAR also was attempted using Volsurf+ software (Molecular Discovery), which similarly is based on the GRID MIFs. Unlike other comparative tools, in which the superimposition of the compounds often poses limitations on the usefulness of the model, Volsurf+ extracts information from 3D molecular fields, converts it to simpler descriptors, and is independent of conformational sampling and the need for compound alignment. Volsurf+ also allows patterns in advantageous physicochemical properties to be uncovered (Cruciani et al., 2000). The overall process for carrying out 3D-QSAR analysis in Volsurf+ was the same as for the procedure used in FLAP.

3.4 Results

3.4.1 The Structure of FXR and Analysis of the LBD

Computational approaches were employed to gain a deeper understanding of the FXR architecture, specifically regarding the LBD, with the aim that this knowledge would be beneficial in the design and development of new ligands for this receptor. To date, there have been no published structures of FXR in its entirety. As such, the amino acid sequence of the canonical isoform 1(+) of FXR (Uniprot Q96RI1-3) was submitted to the online integrated protein structure and function prediction server, IntFOLD (McGuffin *et al.*, 2015). The server generated several models of the full-length FXR protein, and the top-scoring models all used the closely related protein, human Liver X Receptor (LXR β), as a template. The server identified this protein as an evolutionary homologous protein and used it as a template due to its top ranking structure quality estimates. The top-ranked homology model, shown in Figure 3. 2 was based on a heterodimer structure of the DBDs and LBDs of LXR β and RXR (PDB ID: 4NQA). The FXR DBD was predicted to exist as a zinc finger fold, whereas the FXR LBD was predicted to be a 12 α -helix sandwich, typical of nuclear receptors. The *N*-terminal region (1-145aa) did not match any folding motifs of currently known proteins and its structure was not predicted. Furthermore, the IntFOLD DISOclust results predicted the *N*-terminal residues to have a higher probability of disorder. To determine any minor folding assemblies, the full length FXR sequence was split and submitted in three parts comprising the *N*-terminal region, DBD and LBD, respectively. The structures generated again used human NR structures as a template for the DBD and LBD (PDB IDs: 1KB6 and 1PQ6 respectively). However, no high-quality templates, with defined structures, were found to match the sequence of the *N*-terminal region (1-145aa).

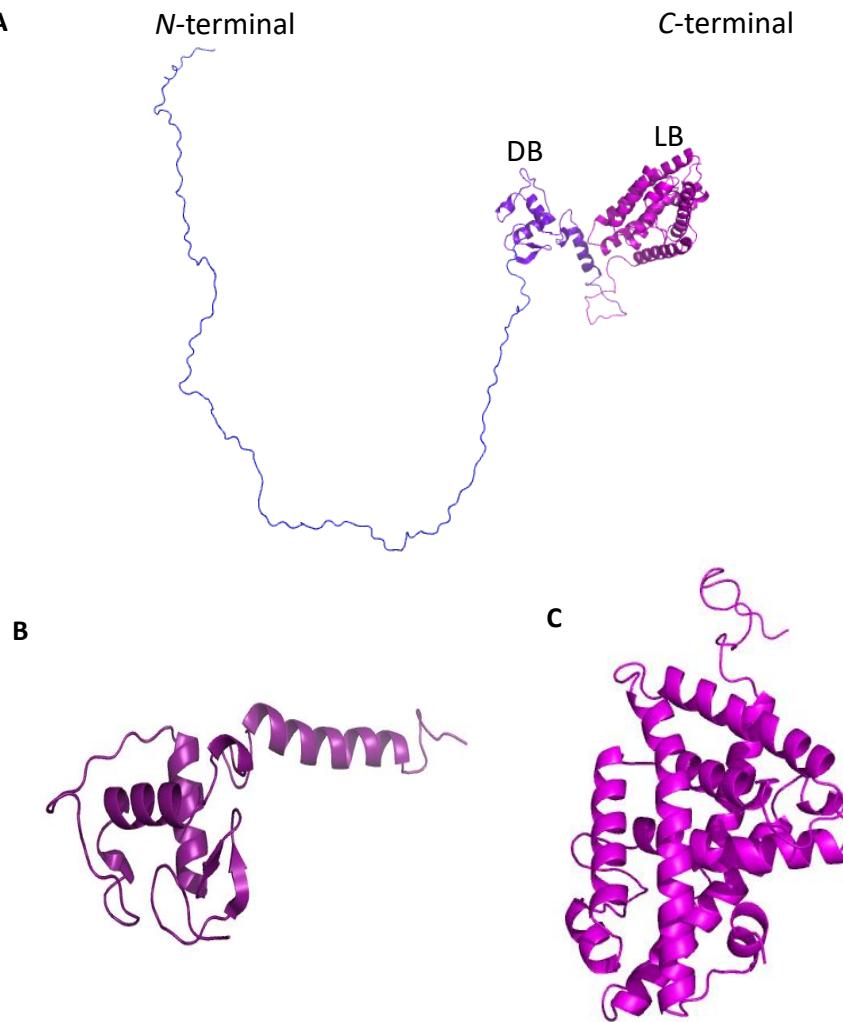


Figure 3.2 Full length and individual domain FXR structures predicted by the integrated modelling server IntFOLD. (A) Full length FXR based on PDB ID: 4NQA, (B) The DNA binding domain (DB) based on PDB ID: 1KB6, (C) The ligand binding domain (LBD) based on PDB ID: 1PQ6. Structures in B and C were submitted to the server independently of one another.

As the site of therapeutic interest, attention was focussed mainly on the FXR LBD. When the work was initiated, there were 26 structures of the FXR LBD deposited in the PDB, but over the duration of 3 years, a further 50 LBD structures were added to the database. These structures have been determined from the co-crystallisation with natural, semi-synthetic and synthetic agonists and antagonists alike. The diversity of these resultant structures highlights the plasticity of the receptor when bound to different ligands. In order to gain a better understanding of the structural flexibility of FXR and some of its possible binding modes, deposited structures of FXR were superimposed and compared. The structures of FXR with different ligand classes were compared to 4QE6, representing the structure that had been co-crystallised with the cognate ligand, CDCA.

Superimposition and comparison of the structures showed that there are no major conformational changes that occur, but rather, slight minor changes in individual helices and loops that are responsible for differential activation of FXR. Comparisons between the conformation adopted when CDCA was bound, 4QE6, and the apo conformation, 5Q0K, showed a slight minor change in most of the helices, but overall relatively low RMSD score, especially in comparison to some of the RMSD scores derived from conformations adopted in the presence of selective agonists (Table 3. 6). The lowest RMSD scores were observed in structures liganded with semi-synthetic BA-derived compounds, OCA and Iso-UDCA, and accordingly, there were only small movements in a few of the loops and helices.

Other steroidal agonists induced slightly more varied conformations, and one agonist 9LI (PDB 5Q0M) had the highest RMSD score, as well as the highest number of helices affected, of all the steroidal agonist-bound structures. The nonsteroidal full and partial agonists induced more varied conformations of the FXR LBD, and displayed higher RMSD scores, and movement in several different helices and loops. Of note, helix 2 and helix 6 displayed the most flexibility (Figure 3. 3A and B), and measurements within PyMOL, indicated that the former could move by as much as 10Å, and the latter by as much as 9.5Å, when FXR was liganded with XL355 (PDB 3FLI) as opposed to CDCA. Similarly, the loop between helix 11 and 12, and helix 12 itself, also displayed largely varying conformations in all the structures compared.

The FXR LBD structure adopted by binding with the antagonist NDB (PDB 4OIV), interestingly, does not display the biggest overall RMSD score, however, significant changes are seen in helix 11 and 12 (Figure 3. 3C). Helix 11 appears to be extended and takes on the structure of a β -sheet, protruding into the space that helix 12 occupies in agonist activated structures, whereas helix 12 is completely extended away from the main LBD core.

Table 3.6 Comparison and characterization of helix movement and ligand binding sites of FXR structures to classic agonist structure 4QE6

PDB Entry	Ligand	Resolution (Å)	RMSD (Å) compared to 4QE6	Helix movement compared to 4QE6	Ligand binding residues
4QE6/ 6HL1 (revised structure)	CDCA	1.6	-	-	M328, R331, S332, Y361, Y369, H447 (M290, A291, H294, I352, I357, W454)
5Q0K	Apo form	1.8	0.936	H2, H5, L5/6, H6, H7, H9, H10, H11, L11/12, H12	-
1OSV (<i>mus musculus</i>)	OCA (steroidal, agonist)	2.5	0.712	L1/2, L5/6, H9, L9/10	R331, S332, Y361, Y369, H447 (A291, M328, I332)
1OT7 (<i>rattus norvegicus</i>)	Iso-UDCA (steroidal)	2.9	0.717	L5/6	H294, R331, S332, Y369 (M290, M328, L348)
3BEJ	MUF (steroidal, agonist)	1.9	0.984	L1/2, H2, L5/6, H10/11, L11/12, H12	T288, R331 (A291, M328, F329, W454)
5Q0M	9LI (steroidal, agonist)	2.2	1.122	H2, L2/3, H3, L5/6, H6, H7, H10, H11, L11/12, H12	M265, T270, H294, R331 (M265, L287, F288, I335, I352, I357, Y361)
1OSH	Fexaramine (methylcinnate, cyclohexyl amide, nonsteroidal, agonist)	1.8	0.876	H2 (not resolved), H3, H6, H11, H12	H294 (L287, M290, I335, F336, L348, I352, M365, W469)
3DCT	GW4064 (isoxazole, nonsteroidal, agonist)	2.5	0.990	L5/6, H6, H7, H9, H11, H12	M265, R331 (T270, L287, M290, A291,)

					<i>I335</i> (<i>F329ππ</i> ,)
3RUU	GSK237 (GW4064 analogue, nonsteroidal, agonist)	2.5	0.905	H2, L5/6, H9, H11, L11/12, H12	R331, H447 (<i>M265, T270</i> , <i>L287, M290</i> , <i>A291, H294</i> , <i>F329</i>) (<i>F329ππ</i>)
3L1B	635 (Tetrahydro- azepinoindole, nonsteroidal, agonist)	1.9	0.754	H2 (not resolved), H3, L5/6, H6 (shifted by ~5Å), H11, L11/12	Y369 (<i>L287, M290</i> , <i>M328, I335</i> , <i>I337, L340</i> , <i>L348 M365</i>)
5Q0I	OXO (Tetrahydropyr- azolopyridines, nonsteroidal, agonist)	1.7	1.195	H2 (completely shifted by ~7Å), H3, L5/6, H6 (completely shifted by ~4Å), H7, L11/12, H12	N283, H294 , R331, S332 , Y369 (<i>M268, I273</i> , <i>L287, I335</i> , <i>F336, L340</i> , <i>I352, F366</i> , <i>W469</i>)
3OKI	OKI (benzimidole, nonsteroidal, agonist)	2.0	0.952	L1/2, H2 (completely shifted by ~9Å), L5/6, H6 (completely shifted by ~10Å), H11, L11/12, H12	Y369 (<i>I273, L287</i> , <i>M290, M328</i> , <i>I352, I335</i>)
3P88	P88 (quinolone, nonsteroidal, agonist)	2.95	1.01	H2, L5/6, H7, L11/12, H12	H447 (<i>M265, T270</i> , <i>L287, M290</i> , <i>A291, H294</i> , <i>M328, I335</i>) (<i>F329ππ</i>)
5WZX	Hedragonic acid (Triterpene, nonsteroidal agonist)	2.95	1.273	H2 (not resolved), H5, L5/6, H6 (not resolved), H7, H11, L11/12, H12	H447 (<i>L287, M328</i> , <i>F329, F366</i> , <i>F461, W469</i>) (<i>W454ππ</i>)
3FLI	XL335 (Azepino[4,5-b] indole, nonsteroidal,	2.0	0.773	H2 (completely shifted by ~10Å), L2/3, L5/6, H6	Y369 (<i>L287, I352</i>)

	selective agonist)			(completely shifted by ~9.5Å), H7, H12	
4QE8	31D (nonsteroidal, partial agonist)	2.62	0.888	H2, L5/6, H7, L11/12, H12	(<i>F</i> 284, <i>L</i> 287, <i>I</i> 357, <i>M</i> 450)
4WVD	Ivermectin (macrocyclic lactone, nonsteroidal, selective agonist)	2.9	0.874	H2 (completely shifted by ~6Å), L5/6, H6 (completely shifted by ~10Å), L6/7, H7, L11/12, H12 (not resolved)	N283 (<i>L</i> 287, <i>M</i> 290, <i>I</i> 335)
5IAW	T73 (Terpenoid, nonsteroidal, selective agonist)	2.58	1.104	H2 (completely shifted by ~8Å), L5/6, H6 (completely shifted by ~6Å), H7, H9, H11, L11/12, H12	T288, H447 (<i>L</i> 287, <i>A</i> 291)
4OIV	NDB (Benzamide, nonsteroidal, antagonist)	1.7	1.098	H2, L5/6, H7, H11, H12 (replaced with β-sheet)	H294 (<i>A</i> 291, <i>M</i> 328, <i>S</i> 332, <i>I</i> 352, <i>M</i> 365, <i>L</i> 451)

*RMSD and helix movements defined by alignment and measurement tool with structure 4QE6 in PyMOL. For ligand binding residues, residues involved in hydrogen bonding are shown in bold, residues contributing to hydrophobic interactions are shown in italics, and residues involved in π-π interactions are denoted by (ππ).

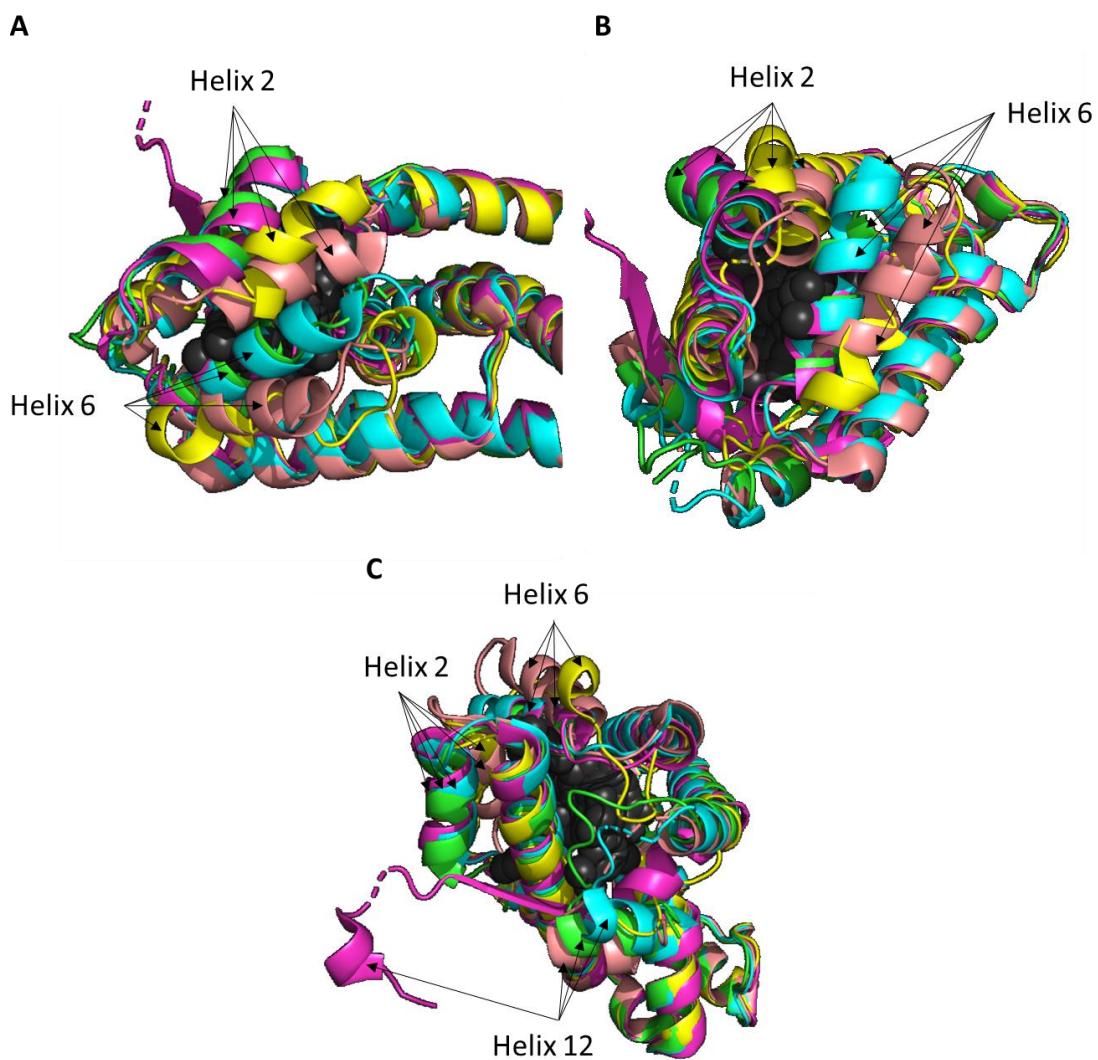


Figure 3.3 Observations of FXR LBD structural flexibility. Superimposition of structures 4QE6 liganded with natural agonist CDCA (green), 5Q0K apo structure (cyan), 3FL1 liganded with selective agonist XL355 (peach), 4WVD liganded with partial agonist ivermectin (yellow), 4OIV liganded with antagonist NDB (magenta). Bound ligands represented by black spheres. Structures are shown from different angles to aid visibility of helix movement (A, B, C).

The structures show that the ligand binding cavity is created between helix 3, helix 5, helix 6, helix 7 and helix 10/11 (Figure 3.4). The cognate ligand, CDCA binds with its A- ring in the corner of the pocket formed by helix 7 and 10/11, where it interacts with Y361, Y369 and H447, on their respective helices. The opening to the cavity is created by helix 5, helix 1 and loop 1/2, and the carboxyl tail extends out of the pocket, interacting with R331 on helix 5, near the cavity entrance. The hydroxyl group at the C7 position on ring B, interacts with S332, also on helix 5, for additional stabilisation of the receptor conformation. Likewise, several hydrophobic interactions were observed along the ring structure of CDCA, with a number of different residues lining the ligand binding pocket, namely M290, A291 and H294 on helix 3, I354 and I357 on helix 6, and W454 on helix 11. Comparisons of the ligand binding modes of XL355 (PDB 3FLI), ivermectin (PDB 4WVD) and NDB (PDB4OIV), show that whilst CDCA lies horizontal and relatively flat across the binding pocket, the other ligands extend into the spaces above and below the canonical ligand, protruding into helix 3 and 6 (Figure 3. 4).

Closer inspection of the exact ligand binding residues (Table 3. 6), shows that OCA hydrogen bonds with the same residues as CDCA. The other steroid compounds share a common interaction with residue R331, however, ligands in 3BEJ and 5Q0M interact with additional residues along helix 3, T288 and H294, respectively. The other non-steroidal agonists create hydrogen bonds with a variety of different residues, some known to interact with CDCA, and others which are novel, such as N283 on helix 3, which makes contacts with agonist OXO in structure 5Q0I, and the partial agonist ivermectin in structure 4WVD. Nonetheless, for all structures observed, the most common polar interaction residues are H447 and R331. In addition to these polar contacts, significant hydrophobic interactions were made by all ligands, with residues on helix 3 and 5. Furthermore, various aromatic ring-based ligands, such as GW4064 and its derivatives, were also shown to make pi-pi ($\pi\pi$) stacking interactions particularly with F329 on helix 5.

In order to discern which interactions were considered as most important for molecular recognition and ligand binding, the FLAP software was used to carry out analysis of the FXR ligand binding domain in the 4QE6 structure. GRID molecular probes were used to identify the molecular interaction fields of the cavity, and results were displayed as a MIF contour map, as shown in Figure 3. 5A. Three main types of MIFs were identified within the FXR LBD; hydrophobic hotspots, hydrogen bond-donating regions, and hydrogen bond-accepting regions. The regions of hydrophobic potential were confined to the top of the cavity, lining helix 3 and 6. Conversely, the regions with high potential for polar interactions, were identified at the base of the cavity and extending into an unoccupied distal sub-pocket between helix 3 and loop 1/2.

FLAP also was used to probe CDCA in its classic binding position (Figure 3. 5B). The contour maps highlighted that CDCA had complementary MIFs to the FXR ligand binding pocket, whereby the top (β) side of the molecule was preferential to hydrophobic interactions, and the bottom (α) face was most suited to polar contacts. The amphipathic nature of CDCA was exploited for the design and creation of novel BA-derived compounds, which could protrude into sub-pockets not currently occupied by the cognate ligands.

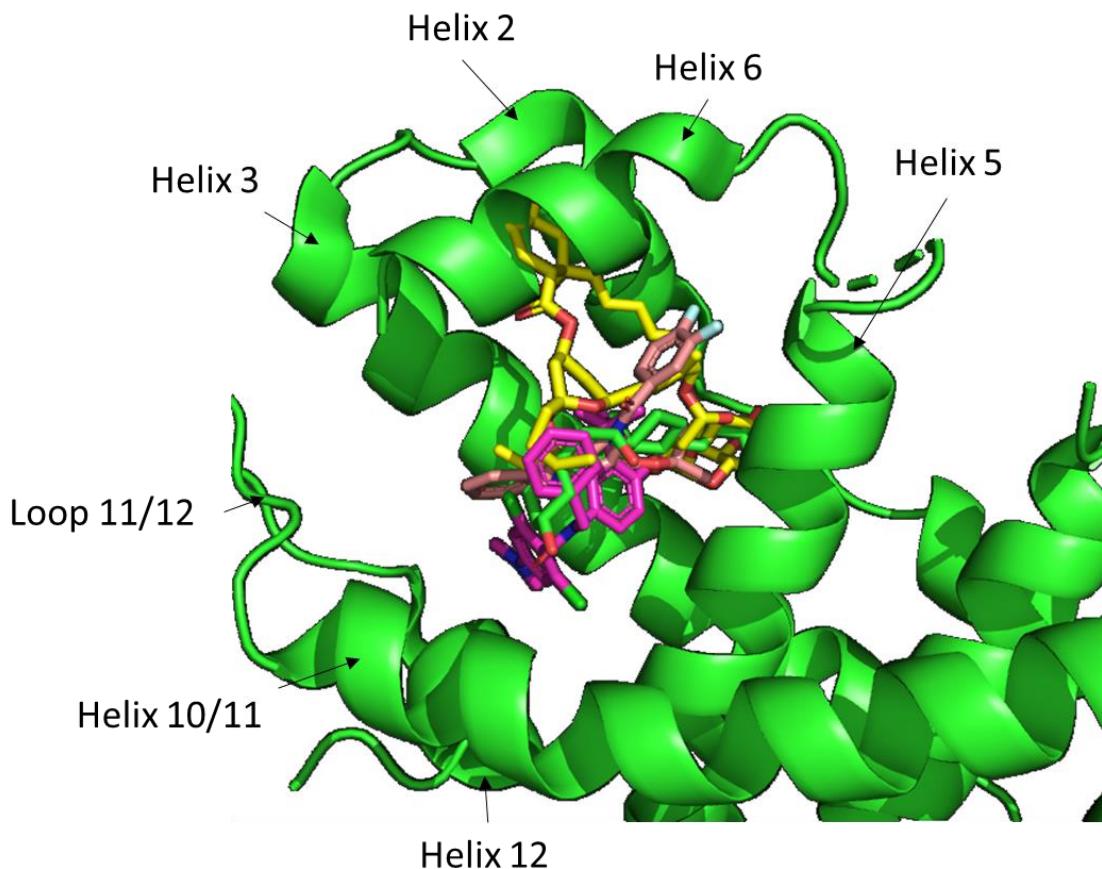


Figure 3. 4 Observations of ligand binding modes of compounds with different degrees of agonist behaviour. Ligands are displayed with structure 4QE6 for comparison. Natural agonist CDCA (green), selective agonist XL355 (peach), partial agonist ivermectin (yellow) and antagonist NDB (magenta). (Helix 7 is not shown to improve visibility of ligands).

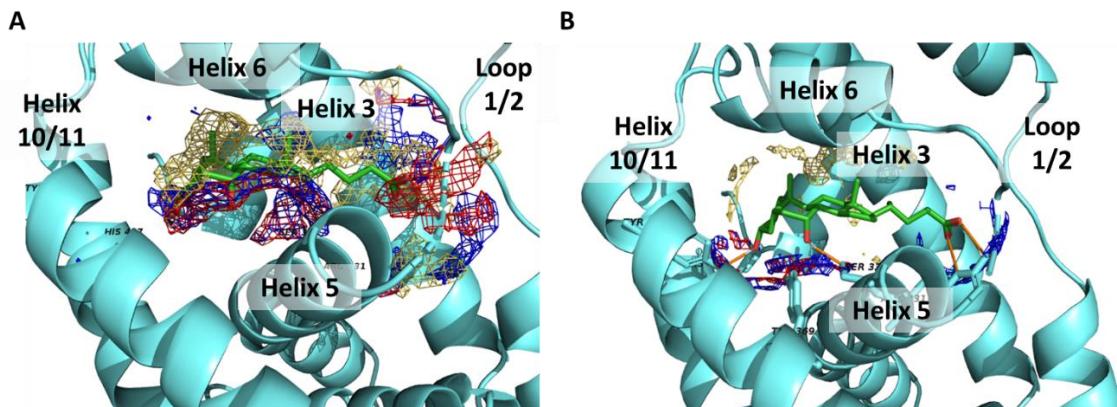


Figure 3.5 Comparisons of GRID molecular interaction fields for the receptor (A) and for the CDCA ligand (B). Crystal structure of FXR LBD (PDB 4QE6, cyan) with CDCA (lime) was used to generate hydrophobic (yellow mesh), hydrogen bond donor (blue mesh) and hydrogen bond acceptor (red mesh) molecular interaction fields, which were used to identify key hotspots in the LBD and potential binding mechanisms of the different ligands.

3.4.2 *In Silico* Docking of Compounds into FXR LBD

When molecular docking studies were initiated, the FXR LBD structure 4QE6, was one of the few structures published for the human homologue of FXR at high resolution. Due to its classic conformation with a BA ligand, this structure was used for docking the novel BA analogues. Molecular docking was carried out using the Surflex Dock algorithm in SYBYL. The CDCA structure was removed from the ligand binding pocket of 4QE6 and a protomol was generated automatically by probing the cavity with different probes representing either hydrophobic or hydrogen bond donor/acceptor interactions.

3.4.2.1 Validation of the Docking Method

To validate the docking procedure, the extracted CDCA molecule from the original 4QE6 coordinate file was converted to a mol2 file and redocked into the receptor following the structure preparation procedures implemented in SYBYL and using the newly defined protomol as a guide for placement. The results for the top-ranked solution showed that CDCA was docked back into the same site and in the same orientation that it occupied in the original crystal structure, however, there were slight differences in the bond torsion angles, particularly along the steroid ring's puckered structure (Figure 3.6). The slight difference in puckering and pose of the redocked CDCA, resulted in the observation of hydrogen bond interactions with Y361 as per the original structure, but yielded an additional interaction with S342. Nonetheless, the RMSD score was calculated in PYMOL without performing an outlier rejection or any fitting in the alignment, and the difference between all atoms of the two molecules was 2.1 Å, which was deemed to be sufficiently accurate (Ramírez and Caballero, 2018).

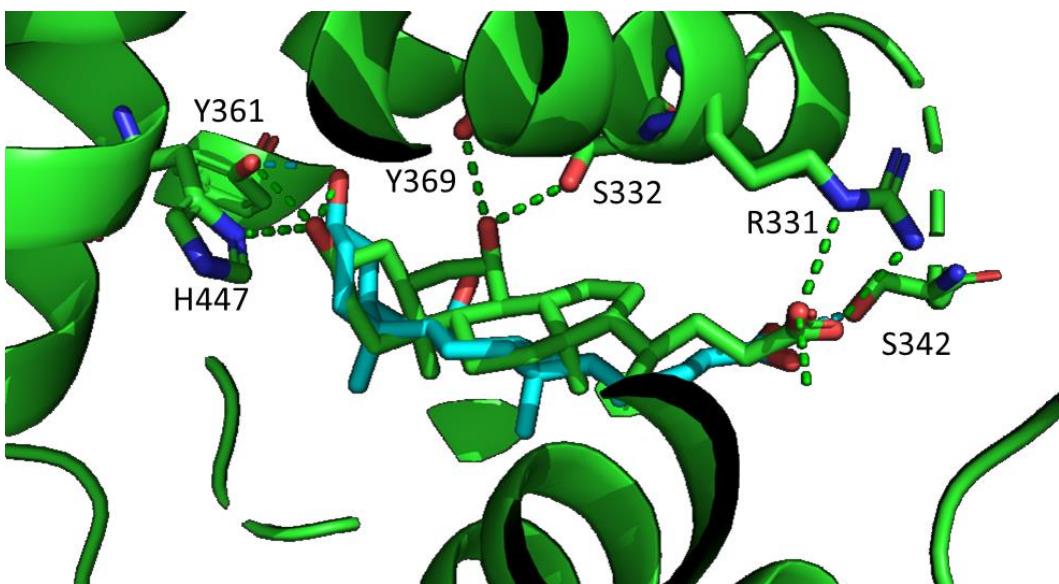


Figure 3. 6 Validation of the docking procedure. Re-docked CDCA (cyan) compared with the CDCA from the original structure (green). Hydrogen bonds are shown by dashed lines.

3.4.2.2 Docking Novel BA Derivatives into FXR LBD

The novel hypothetical compounds were designed by colleagues at NZP UK and built in ChemDraw 16 and Chem3D software (PerkinElmer Informatics). Following validation, compounds were docked using the automated method in SYBYL and resulting binding interactions were viewed in both PyMOL Molecular Graphics System, Version 2.0.4 (Schrödinger, LLC) and BIOVIA Discovery Studio (Dassault Systèmes). The dataset was grouped according to the minimum scaffold of the ligands, and Tables 3. 7-3. 11, summarise the potential polar contacts and VdW interactions between the ligands' functional groups and residues within the binding pocket.

The results showed that all compounds were docked into the same site as the classic BAs, and in general adopted the same orientation. The novel compounds all displayed polar contacts with at least one of the common residues, H447, Y361, Y369, S332, and R331, and many employed VdW interactions with various other residues, particularly those along the upper surface of the ligand binding pocket on helix 3. Whilst the majority of VdW interactions occur around the 'ABCD' ring with hydrophobic residues of the pocket, the addition of electronegative atoms to either end of the steroid nucleus resulted in novel hydrogen and halogen bonds taking place. Likewise, the addition of bulky aromatic groups, such as benzene, resulted in the additional noncovalent, π -stacking interactions with other residues.

3.4.2.2.1 The Effect of Simple Substitutions to CDCA Backbone

When comparing the docking results of the simplistic, CDCA-derived compounds (Table 3. 7), it was observed that all the compounds were oriented in the same position as CDCA, with the exception of Jed86, Jed104. Despite the relatively simple change to these structures, with the addition of electronegative atoms to various positions along the A/B rings, it was observed that there were some differences to the torsion angles of the docked compounds. This gave rise to the slightly different poses for these compounds as observed in Figure 3. 7. Nonetheless, the first set of these compounds all engaged in polar contacts with residues along the lower side of the pocket, and their carboxylic acid side chains sat by the entrance of the cavity, as observed in their precursor CDCA. Compounds Jed86 and Jed104, both having a fluorine (F) atom added to the C6 position on the B ring, were positioned back-to-front, with their carboxylic acid side chains in the crevice of the pocket, towards S355, whilst their steroid 'heads' were positioned at the opening of the pocket.

Table 3. 7 Table summarising polar contacts and VdW interactions between CDCA-derived compounds and FXR.

Compound (synonym)	Hydrogen Bonds	Van der Waals interactions
Jed18 (CDCA)	3 α OH – S352 7 α OH – S332 COOH – R331	L287, M290, A291, H294, M328, I335, I352, W454
Jed79	7 α OH – S332 COOH – R331	L287, M290, A291, H294, M328, I335, I352
Jed85	7 α OH – S332 COOH – R331	M265, L287, M290, A291, H294, M328, I352
Jed86	6 α F – M290 (halogen bond) COOH – S355	M265, L287, M290, A291, H294, M328, R331, I335, L348, I352, I362
Jed104	7 α F – S332, R331 (halogen bond), M328 (halogen bond) COOH – S355	M265, M290, H294, I335, F336, L348, I352
Jed106	7 α OH – L287 COOH – R331	M265, M290, H294, M328, I335, I352
Jed130	7 α OH – M328, S332, Y369 COOH – S342	L287, M290, M328, I335, L348, I352, I357 I362, Y361, M365
Jed379	7 α OH – S332 Side chain NH ₂ – Y361, H447	L287, M290, H294, I335, L348, I352, I362, M365, F366, Y369

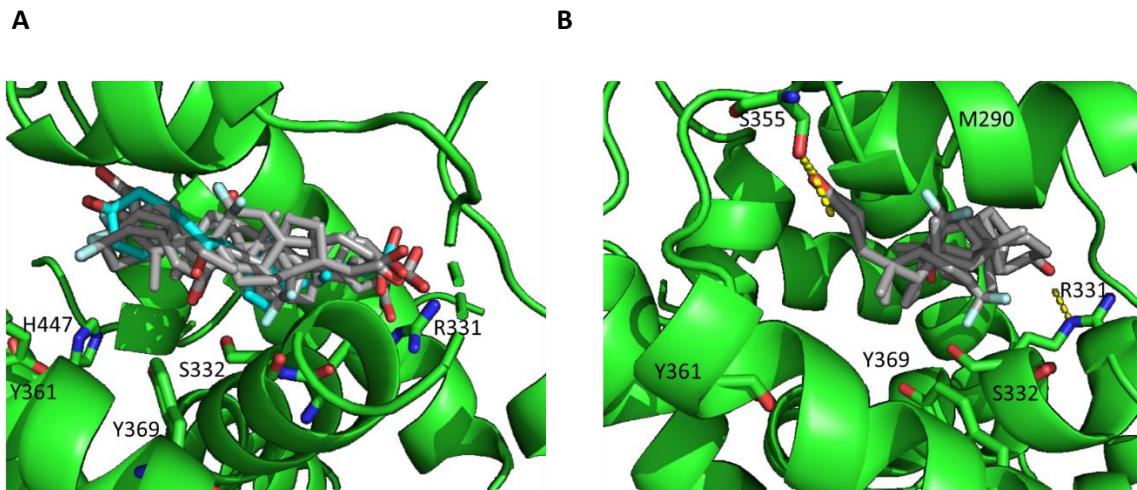


Figure 3.7 Docking results of CDCA-derived compounds. Comparisons of novel compounds (grey) with CDCA (cyan). Compounds that (A) share classic CDCA orientation and (B) docked in tail-first orientation. Common binding residues are shown as sticks. H bonds are shown as yellow dashes.

3.4.2.2.2 The Effect of Tail End Extensions

The rest of the compounds in the dataset utilised OCA as a starting scaffold, due to its enhanced potency compared to CDCA. Several compounds were designed to retain the steroid rings of OCA, but to substitute the carboxyl tail for longer, bulkier extensions with higher electrostatic potential. The docking results, shown in Table 3.8, revealed that as expected, the hydroxyl groups on the A and B rings engaged in polar interactions with the same key residues shown to interact with OCA, H447, Y361 and Y369. The elongated side chains, however, extended out of the pocket, or occupied smaller sub pockets created near to the flexible loop between helix 1 and 2. The results showed that the addition of more electronegative atoms to the side chain, resulted in very subtle structural changes observed by different polar and π -stacking interactions. The related compound series, Jed575, Jed576 and Jed577, which all possessed similar benzene containing side chains, all displayed very similar 'ABCD' ring poses, but varying tail end structures (Figure 3.8A). Jed575 and Jed576 both exhibited very similar contacts within the binding pocket, however, although it makes no changes to the potential hydrogen bonds made, the addition of 2 oxygen atoms to the sulphur atom in Jed577 adds an additional π -sigma interaction with H294. Similarly, the addition of 2 oxygen atoms to the sulphur on Jed581 results in a π -sulphur interaction with H294, orienting the tail towards the top of the pocket, and allowing the end carboxylic acid group to participate in hydrogen bonds with R265 on loop 1/2 (Figure 3.8).

Furthermore, the addition of other highly electronegative atoms such as fluorine to the end of the benzene, again results in the observation of slight structural differences. When comparing

Jed575 and its fluorinated derivative Jed585, the steroid rings of both compounds are docked in identical positions and poses (Figure 3. 9). However, their tail ends differ, whereby the addition of the fluorine atom(s) promotes halogen interactions with R331 and Q263, at the entrance of the binding pocket.

Designs also explored the use of other chemical groups preceding the benzene. Additions of other cyclic groups such as a cyclobutane causes a shift in the torsion angles of the compound, forcing the benzene group to sit in sub-pocket between helix 3 and loop 1/2, rather than protruding out of the pocket. An example of this can be seen in Figure 3. 10A, with Jed589 (yellow) compared to Jed590 (olive), an analogue lacking the cyclobutane group. Furthermore, the addition of a trifluoromethyl group to the end of the benzene ring, causes further distortions to the ligand pose, with the side chain of Jed637 extending up towards T270 on helix 2, which will likely move to accommodate this elongated compound (Figure 3. 10B).

Likewise, a compound series with the addition of a nitrile group before the benzene, displayed complete rotations of the nitrile and benzene groups dependent on the atom (nitrogen, oxygen or sulphur) linking the two functional groups. This is shown in comparisons of Jed641 (nitrogen linker), Jed642 (oxygen linker) and Jed643 (sulphur linker) in Figure 3. 11, where the nitrile groups adopt completely different angles (arrowhead), despite the steroid rings aligning perfectly.

Table 3. 8 Table summarising polar contacts and VdW interactions between OCA-derived compounds with extensions of the carboxyl tail only and FXR

Compound (synonym)	Hydrogen Bonds	Van der Waals Interactions
Jed20 (OCA)	3 α OH – H447 7 α OH – Y369 COOH – S342	L287, M290, I335, F336, L348, I352, I357, I362, M365, F366,
Jed150	7 α OH – S332 COOH (1) – R331 COOH (2) – S342	L287, M290, A291, H294, M328, F336, I352, Y369
Jed120	NH – A291 SO ₃ – W469	L287, M290, H294 (π -stacking), I335, F336, L348, I352, Y369, H447, W454
Jed243	7 α OH – M328, S332 Sulphonamide NH ₂ – R331 CF ₃ – N261, Q263 (halogen bond)	Y260, M265, L287, M290, A291, H294, V297, I335, F336, L348, I352, I355, F336, I359, Y361, I362, M365, F366, Y369
Jed567	7 α OH – S332	M265, L287, M290, A291, H294, V297, I335, L348, I355, Y361, W454
Jed572	7 α OH – S332 Triazole N – R331	M265, L287, M290, A291, H294, V297, M328, F329, I335, L348, I355, M365

Jed573	7 α OH – M328, S332 SO ₂ – S342 End OH – S345	L287, M290, A291, H294, I335, F336, L348, I357, F366, Y369, W454
Jed575	3 α OH – Y361 7 α OH – Y369	M265, L287, H294, V297, I335, F336, L348, I352, I357, M365, F366
Jed576	3 α OH – Y361 7 α OH – Y369	R264, M265, L287, N293, H294, V297, I335, F336, L348, I352, I357, I362
Jed577	3 α OH – Y361, H447 7 α OH – Y369	L287, M290, A291, H294 (π -sigma), R331, L348, M350, I352, I357, I362, F366, W454
Jed578	7 α OH – Y369 COOH – R331	L287, H294, M328, I335, L348, I352, I357, I362, M365, W454
Jed579	3 α OH – Y361 7 α OH – Y369	L287, H294, M328, I335, F336, L348, I352, I357, I362, M365, F366
Jed580	3 α OH – Y361 7 α OH – Y369 COOH – R264	L287, M290, A291, H294 (π -sulphur), L348, I352, I357, I362, M365, F366
Jed581	3 α OH – Y361 7 α OH – Y369	M265, P266, I269, L287, H294 (π -sulphur), V297, I335, F336, L348, I352, I357, I362, M365, F366, W454
Jed582	3 α OH – Y361 7 α OH – Y369	M265, L287, V297, I335, F336, L348, I352, I357, I362, M365
Jed583	3 α OH – Y361, H447 7 α OH – Y369 SO ₂ – R331	M265, L287, M290, A291, H294, V297, M328, L348, I352, I357, I362, M365, F366, W454
Jed584	3 α OH – Y361 7 α OH – Y369 CF ₃ – R331, Q263 (halogen bond)	M265, L287, V297, L348, I335, F336, I352, I357, I362, M365, F366
Jed585	3 α OH – Y361 7 α OH – Y369 CF ₃ – R331, Q263 (halogen bond), R264 (halogen bond)	M265 (π -sigma) L287, M290, A291, L348, I335, I352, I357, I362, M365, F366
Jed586	3 α OH – Y361 7 α OH – S332, Y369 CF ₃ – M265,	M265 (π -sigma), L287, M290, A291, H294 (π -sulphur), L348, I352, I357, I362, M365, F366
Jed645	3 α OH – Y361 7 α OH – Y369 Sulphonamide NH – R331 CF ₃ – M265,	M265 (π -sigma), L287, M290, A291, H294 M328, L348, I352, I357, I362, M365, F366, W454
Jed588	3 α OH – S352 Sulphonamide NH ₂ – R331	M265, L287, M290, A291, H294, Y369, H447, M450, W454
Jed589	Sulphonamide NH ₂ – R331 Sulphonamide SO ₂ – M265	L287, M290, A291, H294, Y369, H447, M450,
Jed590	7 α OH – L287 Sulphonamide NH ₂ – R331	R264, M290, A291, H294, M328, W454, W469

Jed636	3 α OH – Y361 7 α OH – Y369 Cyclobutene carbonyl – R331	R264 (π -stacking), M265, L287, M290, M328, L348, I352, I357, I362, M365, F366
Jed637	7 α OH – M328, S332 NH – R331 CF ₃ – T270	M265 (π -sigma), L287, M290, A291, H294, V297, F336, I352, F366, Y369
Jed638	3 α OH – S355 CF ₃ – V297 (halogen bond)	M265, M290, A291, H294 (π -sigma), M328, R331, I335, H447, M450, I452, W454, W470
Jed639	3 α OH – H447 7 α OH – Y369 SO ₂ – R331, S342 CF ₃ – R264, M290, N293, H294	M265, L287, M328, I335, L348, I352, I357, W454
Jed640	3 α OH – S355 7 α OH – L287 Sulphonamide NH – R331	M265, M290, A291, H294, M328, L348, I352, L466, W470
Jed641	3 α OH – H447 7 α OH – Y369 NH – M290 Nitrile N – R331	M265, L287, A291, V297, L348, I352, I357, I362, M365, F366
Jed642	3 α OH – H447 7 α OH – Y369	L287, M290, A291, H294, R331, I335, L348, I352, I357, I362, M365, F366
Jed643	3 α OH – H447 7 α OH – Y369	R264, L287, M290, A291, I335, L348, I352, I357, I362, M365, F366, W454
Jed644	3 α OH – H447 7 α OH – Y369 NH – M290	M265, L287, A291, V297, L348, I352, I357, I362, M365, F366, M451, W454

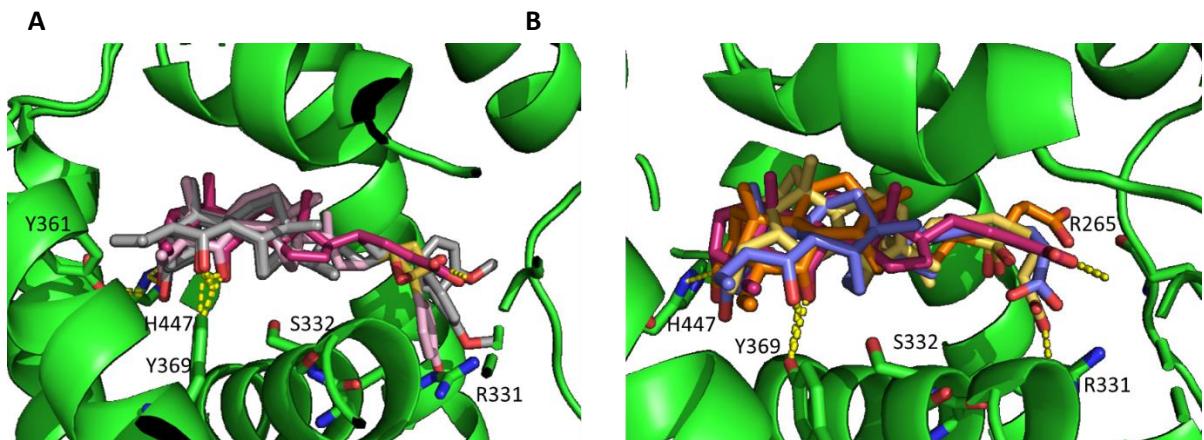


Figure 3.8 Docking results of OCA-derived compounds with extensions of their carboxyl 'tails'. (A) Comparisons of Jed575 (grey), Jed576 (grey), Jed577 (light pink) and OCA (magenta). **(B)** Comparisons of Jed578 (yellow), Jed579 (purple), Jed580 (orange) and OCA (magenta). H bonds are shown as yellow dashes, VdWs and other interactions are not explicitly shown.

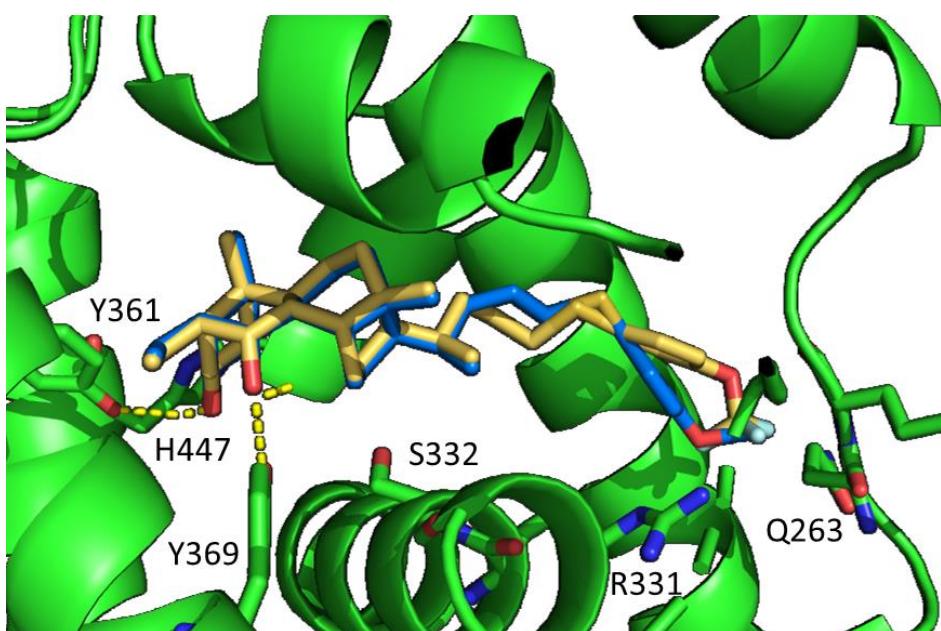


Figure 3.9 Comparisons of Jed575 (blue) and its fluorinated derivative, Jed584 (yellow). H bonds are shown as yellow dashes, VdWs and halogen bonds are not shown.

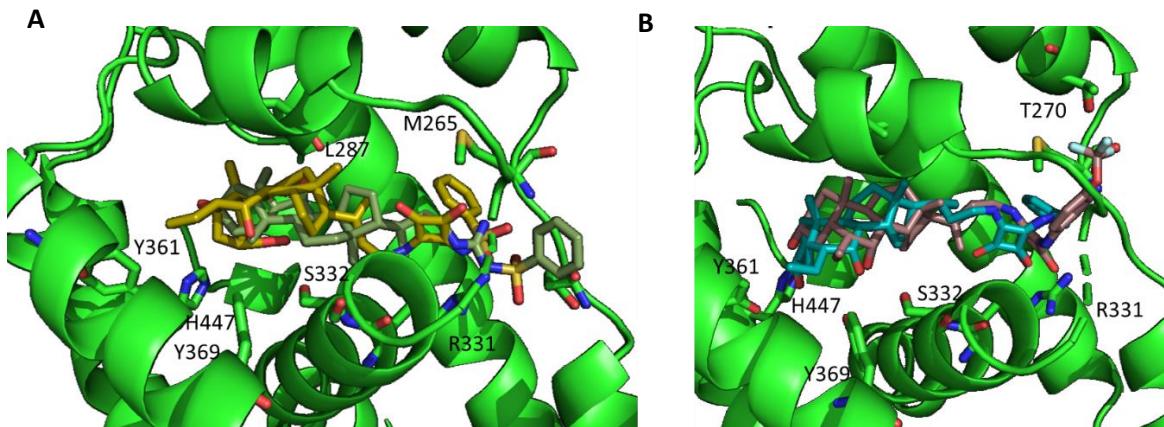


Figure 3. 10 Addition of cyclobutane to BA side chain. (A) Comparisons of compounds with and without cyclobutane Jed589 (yellow) and Jed590 (olive), respectively. (B) Addition of extra fluorine atoms to end of benzene in Jed637 (pink) vs compound without fluorination, Jed636 (teal).

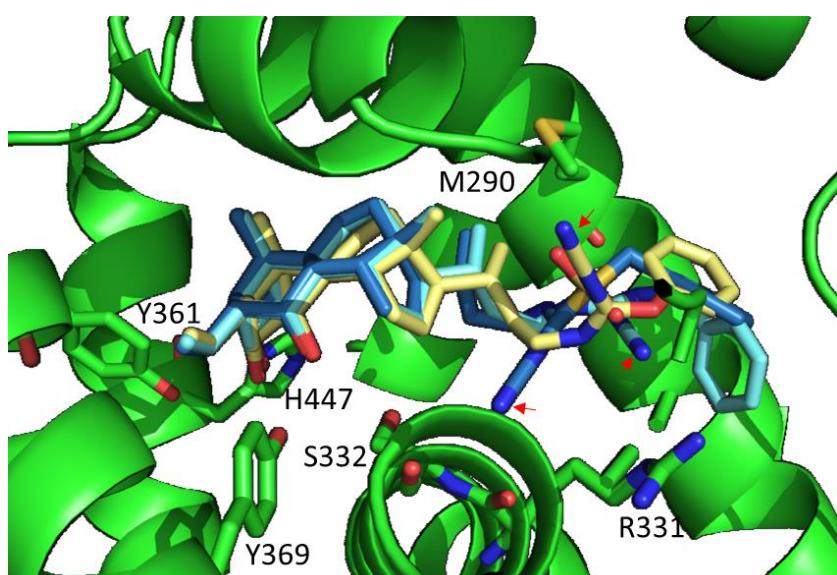


Figure 3. 11 Comparisons of nitrile containing side chains. The movement of the nitrile group (arrowhead) and benzene rings in Jed641 (blue), Jed642 (yellow) and Jed643 (cyan) are dependent on the linker atom.

3.4.2.2.3 The Effect of Extensions at the C3 Position

The addition of functional groups and extended chains at the C3 position resulted in the shift of compounds out of the pocket, as would be expected (Figure 3. 12A). Furthermore, the replacement of the C3 with a nitrogen heteroatom (Jed420 and Jed431), interestingly, also resulted in the displacement of the compounds further out of the cavity (Figure 3. 12B). The side chains of the compounds were seen to sit outside of the ligand binding pocket or occupy sub-pockets next to loop 1/2. While the novel functional groups were able to uphold

interactions with many of the common binding residues, many of these compounds made fewer VdWs interactions (Table 3. 9).

Table 3. 9 Table summarising polar contacts and VdW interactions between OCA-derived compounds with extensions at their C3 position and FXR

Compound (synonym)	Hydrogen Bonds	Van der Waals Interactions
Jed181	3- Pyrrole NH – Y369 7 α OH – M328, S332 Sulphate – M265	L287, M290, A291, H294, R331, I335, L348, I352, Y361, M365
Jed183	3- COOH – Y369 7 α OH – M328 Sulphate – R331	M265, M290, A291, H294, V297, I335, F336, L348
Jed184	3- end NH ₂ – Y361, H447 Sulphate – R331	M265, M290, A291, H294, V297, M328, S332, I335, F336, L348
Jed303	3- end OH – S355 7 α OH – S332 Sulphonamide NH – R331 CF ₃ – R331, E334 (halogen bond)	M265, L287, M290, A291, H294, I335, F336, L348, Y361
Jed401	3- Benzene OH – Y369 7 α OH – S332 Sulphonamide NH – R331	M265, L287, M290, A291, H294, V297, F336, Y361, M362
Jed402	7 α OH – L287 Sulphonamide NH – R331 CF ₃ – S259, Q263, F301, V297 (halogen bond), E300 (halogen bond)	M265, L287, M290, A291, H294, M328, F336, L348, I352, W470
Jed403	3- benzene OH – H447 Tail end carbonyl – R331 CF ₃ – E334, Q380	M265, M290, A291, H294, I335, F336, L348
Jed404	7 α OH – L287 Sulphonamide NH & SO ₂ – R331	M290, A291, H294, M328, F329, I335, F336, L348, I352, M365
Jed420	COOH – R331	M265, M290, A291, H294, V297, I335, F336, L348
Jed431	Sulphonamide NH – R331 CF ₃ – S262 (halogen bond)	L287, M290, A291, H294, M328, I335
Jed433	Sulphonamide NH – R331 CF ₃ – Y260 (halogen bond) V297 (halogen bond),	M265, L287, M290, A291, M328, I335, L348, I352, W454, W470
Jed434	3- end OH – I352, S355 COOH – R264	L287, M290, A291, H294, M328, R331, I335, F336

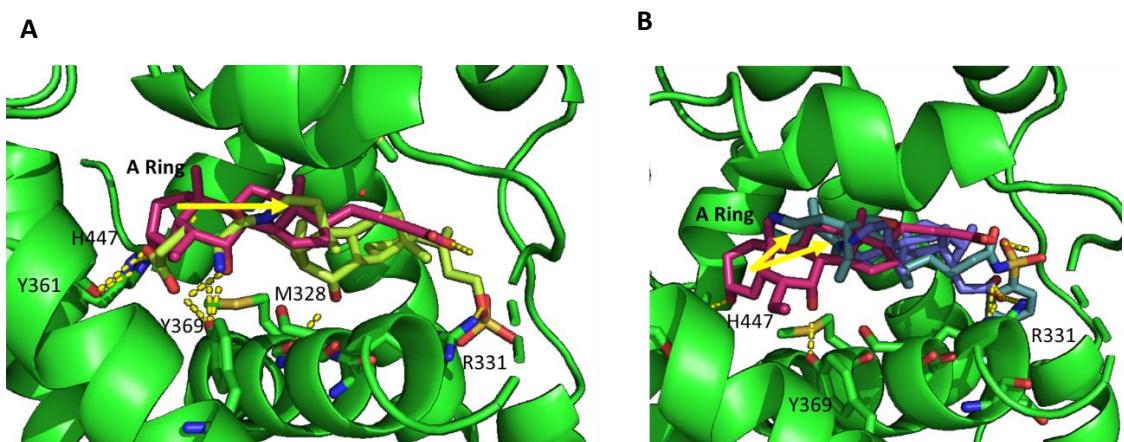


Figure 3. 12 The effect of extensions or substitutions at the C3 position. (A) Comparisons of OCA (magenta) and Jed183 (lime) show a shift (arrow) in the ABCD ring when the C3 is extended. (B) Comparisons between OCA (magenta) and Jed420 (purple) and Jed431 (cyan) show a shift (arrow) in the ABCD ring when the C3 is substituted. H bonds are shown as yellow dashes, VdWs and halogen bonds not shown.

3.4.2.2.4 The Effect of Simple Fluorine Substitutions on the A ring and Extended Carboxyl Side Chains

The exploration into the fluorination of the A ring resulted in several simple OCA derivatives being designed with fluorine atoms added to each position along the hexane ring in either the α or β configuration. The docking solutions showed that these derivatives, 1 β F- (Jed678), 2 α F- (Jed665), 2 β F- (Jed397), 4 α F- (Jed664), and 4 β F- (Jed432), all docked into the same position as OCA (Figure 3. 13A). A graphical representation of the results shown in Figure 3. 13A, displays all compounds in perfect alignment of their steroid rings. However, observation at a 90° rotation shows that the carboxyl side chains of 1 β F-OCA and 2 α F-OCA had bent back, making potential contacts with M328 rather than S342 or R331 (Figure 3. 13B). The visualisation of the docking results in PyMOL was unable to distinguish any differences between the position of the fluorine group and the consequential effect on binding modes of the compounds. However, inspection of the ligand interactions in BIOVIA discovery studio identified potential halogen bonds between fluorine atoms in either configuration on the C4 position and Y369, and the distance between the tyrosine residue and both fluorine atoms were measured to be 3.5 Å (Table 3. 10).

Designs also considered the addition of 2 fluorine atoms at different positions along the A ring. When 2 fluorines were added in both the α and β orientations, to the same C4 atom (Jed556), the potential for a halogen interaction with Y369 was lost. Adding a second fluorine atom to the 2 α or β positions whilst retaining the 4 β F (Jed557 and Jed558, respectively) resulted in a slight change in the steroid ring position in the pocket (Figure 3. 14). Jed558 occupied a similar

position and pose as OCA, but the fluorine atom on the α side of the molecule rotated the compound slightly and shifted it further out of the pocket.

When compounds retained a steroid ring structure, with the simple addition of one fluorine group on the A ring, but an extended side chain tail, the majority still retained interactions with the essential residues. Whilst interactions with H447, Y361, Y369 in the pocket corner were maintained by the C3/ C7 hydroxyl groups and by the additional fluorines, the extended tail chains, most containing sulphonamide moieties, were still able to interact with R331 via their sulfonyl group or amine group (Table 3. 10). The extension of the side chains beyond the sulphonamide, with the addition of a benzene ring resulted in more residues being engaged by VdWs and other interactions including π -stacking.

Interestingly, some of the compounds made non-canonical interactions with residues along helix 3. One example is Jed441, which, when visualised in PyMOL, was observed to adopt a 'flipped' orientation with its 6α -ethyl group sitting at the top, distal corner of the pocket and the steroid rings facing helix 3 where it can contribute to interactions with L287 (Figure 3. 15). Furthermore, the tail end of the compound also lies in close association with Helix 3 and the addition of the trifluoromethyl group at the end results in halogen interactions with V297, which also lies on helix 3. These binding mechanisms are not observed with closely related compounds without the benzene and fluorine groups, such as Jed442.

Another series of compounds also were seen to contribute to potential hydrogen bonds with residues along helix 3. Jed559 to Jed564 represent a series of structurally related compounds. Interestingly, these all made potential polar contacts with classic residues from their 3- and 7-hydroxyl groups (Table 3. 10). These compounds also all appeared to make halogen bonds from their 4β F atoms, again with classic residue(s) Y369, and Y361 in the case of Jed561. When visualised in PyMOL, these compounds align at their 'ABCD' rings, however, their tail ends are completely variable (Figure 3. 16). Some compounds, such as Jed560, interact with classic residue R331 via their sulphonamide group, but interact with other residues on helix3 and loop 1/2 via halogen bonds from their trifluoromethyl substituent. Others engage in hydrogen bonds between their sulphonamides and M290 on helix 3. One compound, Jed561, despite not having an extension beyond the benzene moiety, even participates in hydrogen bonding with 2 residues along helix 3, M290 and H294. To help rationalize which features of Jed561 were important for binding, whether the fluorinated A ring, or the extended side chain, an identical compound without the 4β F atom, was made, termed Jed692. Docking results of Jed692 and comparisons with Jed561, showed that the non-fluorinated derivative was pushed further out

the pocket, with its side chain protruding slightly out of the entrance in a more solvent exposed region (Figure 3. 17).

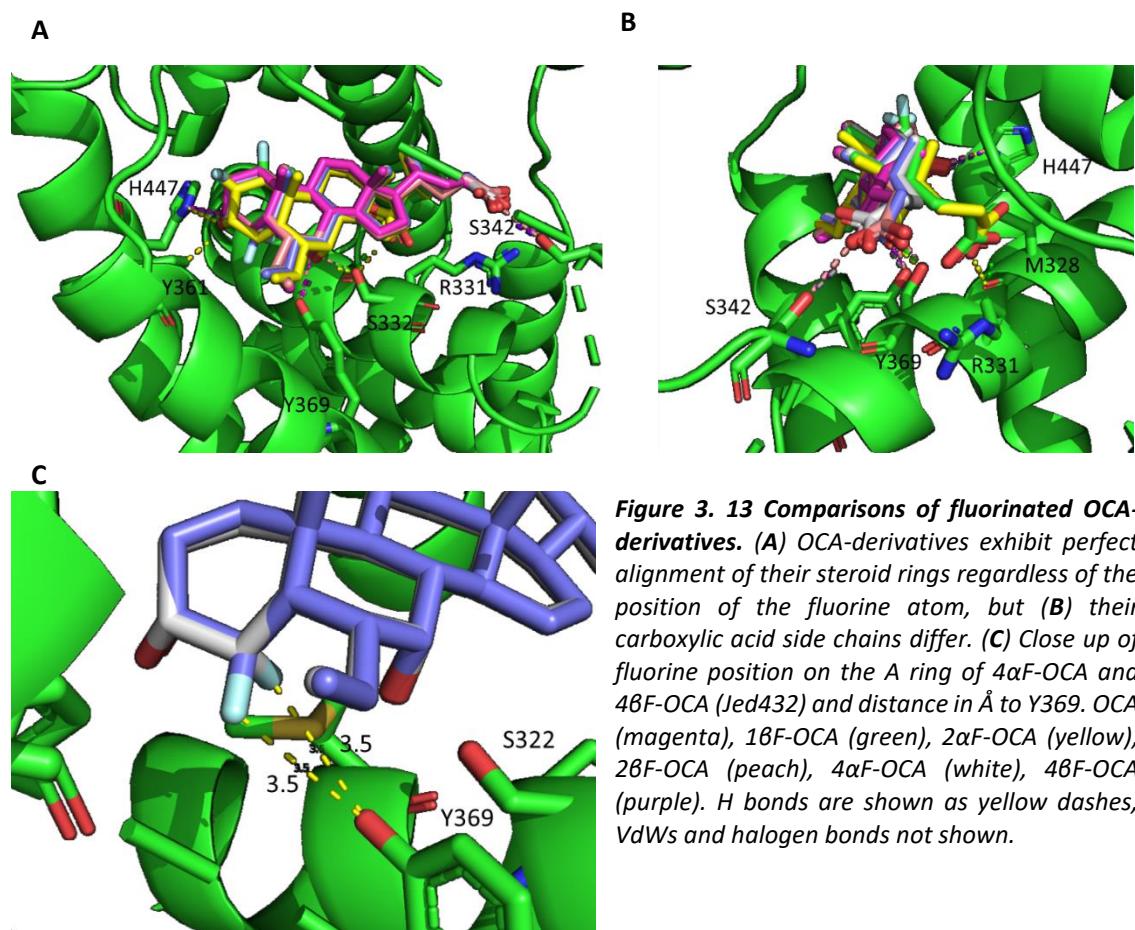


Figure 3. 13 Comparisons of fluorinated OCA-derivatives. (A) OCA-derivatives exhibit perfect alignment of their steroid rings regardless of the position of the fluorine atom, but (B) their carboxylic acid side chains differ. (C) Close up of fluorine position on the A ring of 4 α F-OCA and 4F-OCA (Jed432) and distance in Å to Y369. OCA (magenta), 18F-OCA (green), 2 α F-OCA (yellow), 28F-OCA (peach), 4 α F-OCA (white), 4F-OCA (purple). H bonds are shown as yellow dashes, VdWs and halogen bonds not shown.

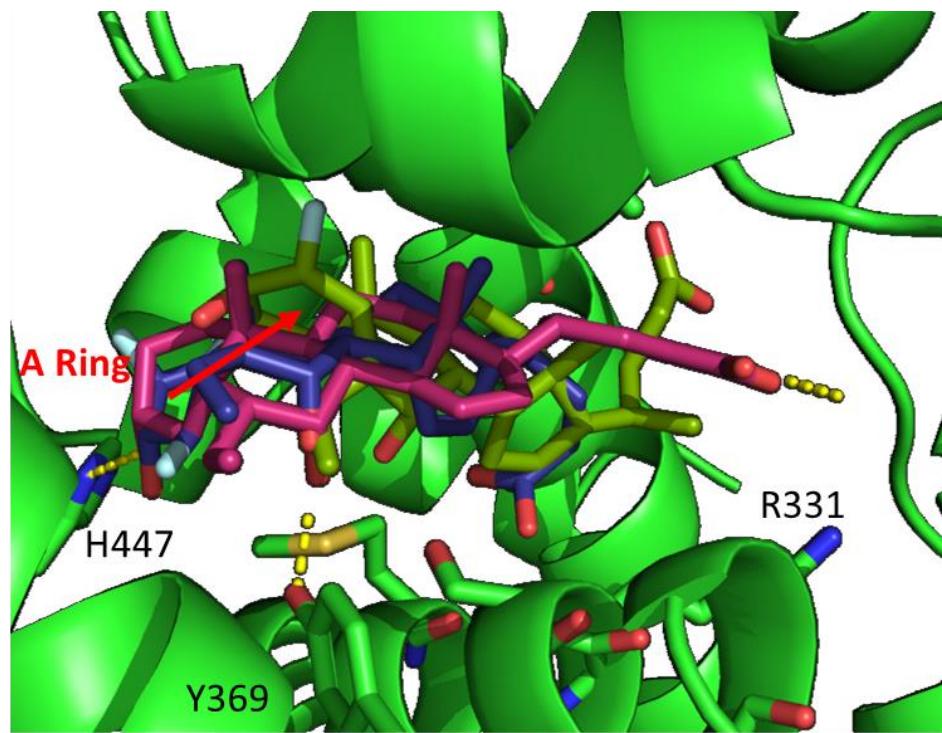


Figure 3. 14 Addition of multiple fluorine atoms to the A ring and their effect on ligand binding position.
 OCA (magenta), $2\alpha,4\beta$ diFluoro-OCA (Jed557, olive), $2\beta,4\beta$ diFluoro-OCA (Jed558, purple). Shift in A ring denoted by arrow. H bonds shown as yellow dashes, VdWs and halogen bonds not shown.

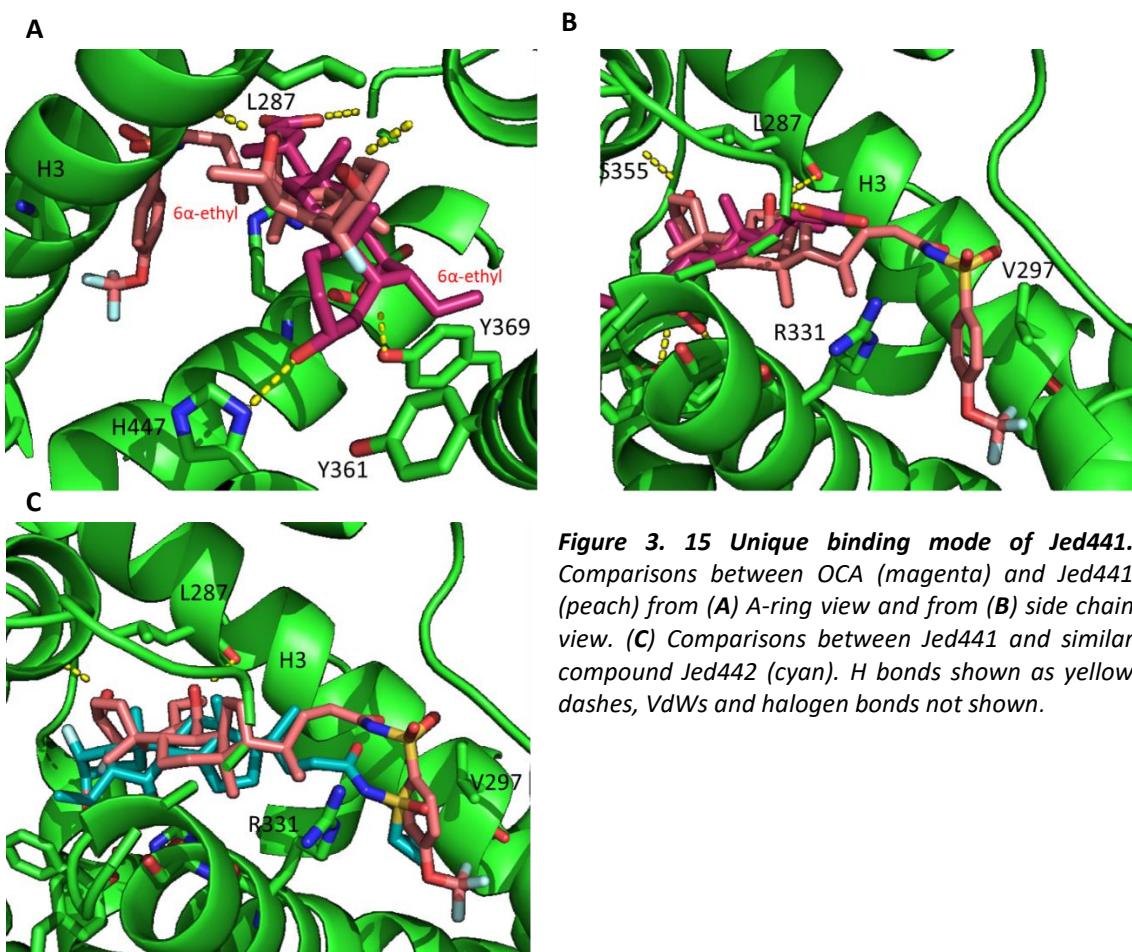


Figure 3. 15 Unique binding mode of Jed441.
 Comparisons between OCA (magenta) and Jed441 (peach) from (A) A-ring view and from (B) side chain view. (C) Comparisons between Jed441 and similar compound Jed442 (cyan). H bonds shown as yellow dashes, VdWs and halogen bonds not shown.

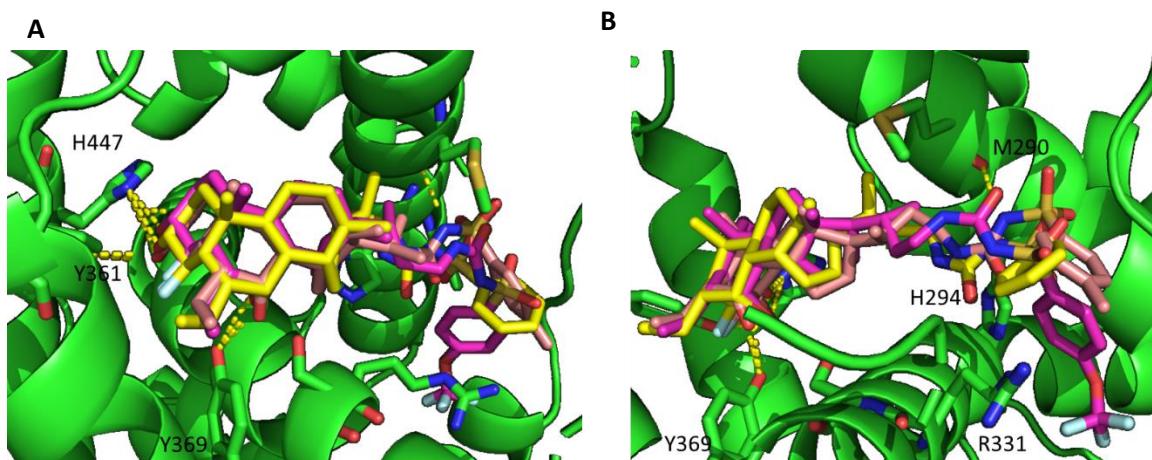


Figure 3. 16 Compounds which make polar contacts with helix 3 and the effect of different side chain functional groups on binding modes. Comparisons of Jed560 (magenta), Jed561 (yellow) and Jed563 (peach) from (A) angles viewing the A ring and (B) angles viewing the side chain. H bonds shown as yellow dashes, VdWs and halogen bonds not shown.

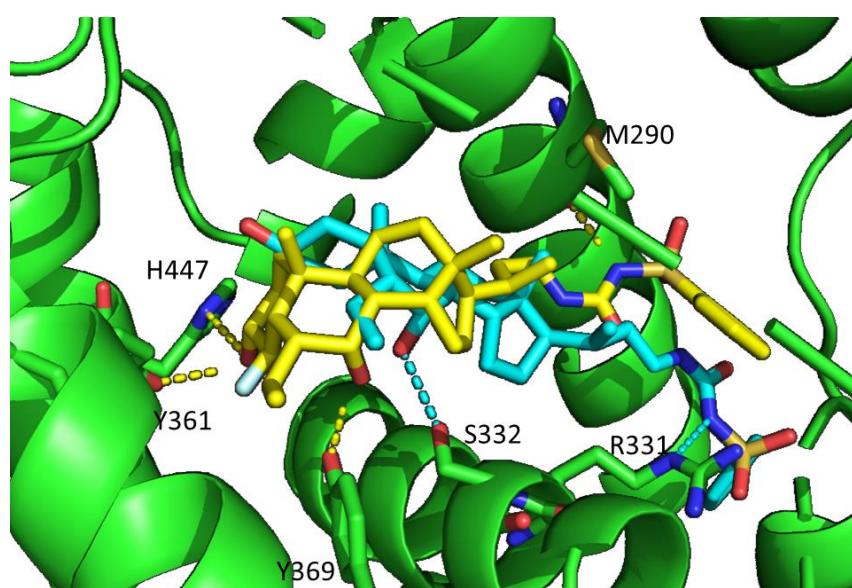


Figure 3. 17 Effect of fluorination at the C4 position. Comparisons of Jed561 (yellow) and the non-fluorinated derivative Jed692 (cyan). H bonds are shown as dashed lines.

Table 3. 10 Table summarising polar contacts and VdW interactions between OCA-derived compounds with substitutions on their A rings and FXR

Compound (synonym)	Hydrogen Bonds	Van der Waals Interactions
Jed678 (1 β F-OCA)	3 α OH – H447 7 α OH – S332	L287, M290, H294, V297, I335, F336, L348, I352, I357, I362, M365, F366
2 α F-OCA	3 α OH – Y361 7 α OH – S332 COOH- M328	L287, M290, H294, I335, F336, L348, I352, I357, I362, M365, F366
2 β F-OCA	3 α OH – H447 7 α OH – Y369 COOH- S342	L287, M290, I335, F336, L348, I352, I357, I362, M365, F366
4 α F-OCA	3 α OH – H447 4 α F – Y369 (halogen bond) 7 α OH – Y369 COOH- S342	L287, M290, I335, F336, L348, I352, I357, I362, M365, F366
Jed432 (4 β F-OCA)	3 α OH – H447 4 β F – Y369 (halogen bond) 7 α OH – Y369 COOH- S342	L287, M290, I335, F336, L348, I352, I357, I362, M365, F366
Jed556	3 α OH – H447 COOH – R331	M265, L287, M290, A291, H294, M328, R331, I335, L348, Y369
Jed557	7 α OH – S332	M265, L287, M290, A291, H294, M328, R331, I335, L348, Y369
Jed558	3 α OH – Y361 COOH – M328	L287, M290, A291, I352, I357, I362, M365, F366
Jed381	7 α OH – M328, S332 Sulphonamide SO ₂ – R331 CF ₃ – F301	L287, M290, A291, H294, V297, I335, W454, W470
Jed382	7 α OH – M328, S332 Sulphonamide SO ₂ – R331	M265, L287, M290, A291, H294, V297, I335, F336, L348, I352, I357, I362, M365, F366, Y369, W454
Jed383	7 α OH – M328, S332 Sulphonamide SO ₂ – R331	M265, L287, M290, A291, H294, V297, I335, F336, L348, I352, I357, I362, M365, F366, Y369
Jed384	7 α OH – S332 Sulphonamide NH – R331	L287, M290, A291, H294 (π -sulphur), V297(π -sigma), L298, I335, F336, L348, I352, I357, I362, M365, F366, Y369, W454
Jed385	3 α F – H447 7 α OH – Y369 Sulphonamide NH – R331	Y260, L287, M290, A291, H294, V297, L298, I335, F336, L348, I352, I357, I362, M365, F366, Y369, W454
Jed406	7 α OH – Y369 Sulphonamide NH – R331	Y260, L287, M290, A291, H294, V297, L298, I335, F336, L348, I352, I357, I362, M365, F366, Y369, W454

Jed407	7 α OH – Y369 Sulphonamide NH – R331	L287, M290, A291, H294, V297, L298, M328, I335, F336, Y369
Jed441	3 α OH – S355 7 α OH – L287 CF ₃ – V297 (halogen bond)	M290, A291, H294, I335, L348, I352, W454
Jed442	7 α OH – M328, S332 Sulphonamide NH – R331	M265, L287, M290, A291, H294, I335, F336, L348, I352, W454
Jed443	7 α OH – S332 Sulphonamide NH – R331	L287, M290, A291, H294, Y361
Jed559	3 α OH – Y361, H447 4 β F – Y369 (halogen bond) 7 α OH – Y369 Sulphonamide SO ₂ – S342, S345	L287, M290, A291, L348, I352, I357, I362, M365, F366
Jed560	3 α OH – Y361, H447 4 β F – Y369 (halogen bond) 7 α OH – Y369 Sulphonamide SO ₂ – R331 CF ₃ – R264, N293	M265, L287, M290, A291, V297, L348, I352, I357, I362, M365, F366
Jed561	3 α OH – Y361, H447 4 β F – Y361 (halogen bond), Y369 (halogen bond) 7 α OH – S332, Y369 Sulphonamide NH – M290, H294	R264, M265, L287, L348, I352, I357, I362, M365, F366, W454
Jed562	3 α OH – Y361 4 β F – Y369 (halogen bond) 7 α OH – Y369 Sulphonamide NH – M290	R264, M265, L287, M328, L348, I352, I357, I362, M365, F366
Jed563	3 α OH – Y361 4 β F – Y369 (halogen bond) 7 α OH – Y369 Sulphonamide NH – M290	R264, M265, L287, M328, L348, I352, I357, I362, M365, F366
Jed564	3 α OH – Y361 4 β F – Y369 (halogen bond) 7 α OH – S332, Y369	R264, M265, L287, M290, V297, M328, R331 (π -cation), L348, I352, I357, I362, M365, F366

3.4.2.2.5 The Effect of Opening the A Ring of the Bile Acid Backbone

When the A ring structure of the classic BA backbone was completely opened, the resulting compounds still participated in hydrogen bond interactions with essential H447 and Y361 residues via their various electronegative atoms on the substituent open chains (Table 3. 11). Furthermore, many of the common VdW interactions that were seen previously with classic BA compounds, were maintained. Despite the absence of the hexane A ring, these compounds were still observed to occupy a very similar binding position and pose as OCA, however, the BCD

rings were shifted slightly along in the pocket. Compounds like Jed395, which also has an additional side chain extension, including the sulphonamide, benzene and trifluoromethyl groups seen previously in Jed441 and Jed560, was also seen to be oriented towards helix 3, although its side chain bent downwards allowing the sulphonamide group to make polar contacts with R331 (Figure 3. 18).

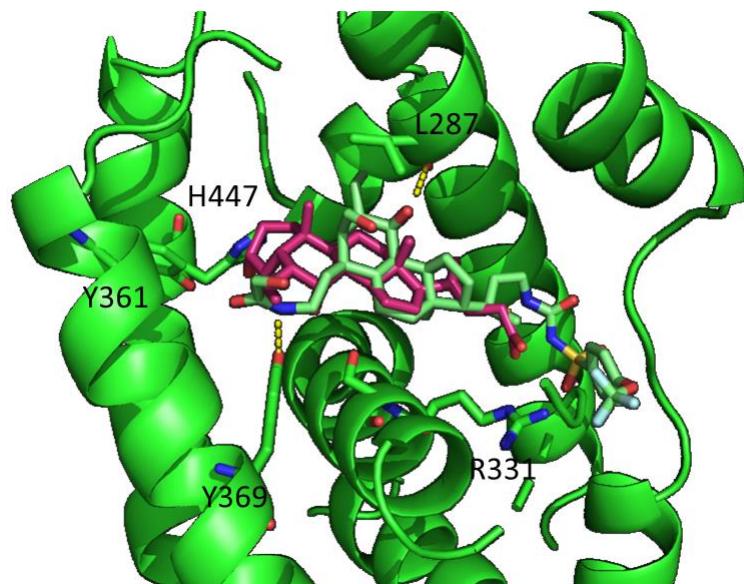


Figure 3. 18 The effect of open A rings and extended side chains on binding position. Comparisons between OCA (magenta) and Jed396 (mint) and their interactions. H bonds are shown by dashed lines.

Table 3. 11 Table summarising polar contacts and VdW interactions between BA-derived compounds with an open A ring structure and FXR

Compound (synonym)	Hydrogen Bonds	Van der Waals Interactions
Jed388	3- end Benzene OH – H294 7 α OH – S332 CF ₃ - A327	L287, M290, H294, A291, V297, M328, R331, I335, F336, L348, I352, F366, Y361, Y369, H447, M450, W454
Jed389	1- end NH – S355 7 α OH – S332 Sulphonamide NH – R331	M265, L287, M290, A291, L348, I352
Jed390	1- end NH – Y361, H447 4OH – Y369 COOH – S342	M290, A291, M328, W454
Jed391	1- end NH – Y361, H447 4OH – Y369 COOH – S342	L287, M290, A291, H294, M328, I352, I357, W454
Jed392	2- end NH – Y361, H447 4OH – S332, Y369 COOH – S342	L287, M290, A291, H294, M328, I352, I357, I362, F366

Jed393	1- end OH – Y369 CF ₃ – H294	L287, M290, A291, H294, M328, L348, I352, I357, I362, F366, W454
Jed394	3- end Benzene OH – Y369, H447 7 α OH – S332	L287, M290, A291, H294, V297, M328, I335, F336, L348, I352, I357, I362, F366, M450, W454
Jed395	1 OH – L287 4- end OH – Y369 Sulphonamide NH – R331	A291, H294, V297, F301, M328, W454, W460
Jed444	1- end OH – Y369 1- end NH – Y369 4 OH – L287 COOH – R331, S342	A291, H294, M328, L348, W454, W460

3.4.2.3 Docking Compounds into Other Structures of FXR LBD

As identified in Chapter 3, section 3.4.1, the FXR LBD structure has been shown to exhibit quite substantial plasticity and a flexible structure dependent on occupancy of an agonist or antagonist. In order to account for the different conformations that the receptor can adopt, the compounds were docked into other FXR LBD structures to identify whether these could potentially sustain an antagonist or inactivated conformation. Compounds were docked into the apo structure (PDB ID: 5Q0K) and an antagonist-bound structure (PDB ID: 4OIV).

The apo docking results showed the compounds occupied completely different binding orientations and poses and were placed in different sub pockets within the binding cavity. In the apo structure, OCA was docked in a back to front manner, with its carboxylic acid side chain tail in first and closest to the top, distal corner of the pocket (Figure 3. 19A). Similarly to OCA, other compounds like Jed561, whilst docking in the classic agonist sub-pocket, were found to be in a flipped or upside down orientation, with their 6 α -ethyl groups situated at the top of the pocket (Figure 3. 19B). Whereas compounds such as Jed441, although these dock in the standard BA orientation, are not positioned close enough to any of the polar residues to make any potential interactions (Figure 3. 19C).

When the compounds were docked into a structure originally obtained with an antagonist, the significant conformational changes, also lead to noteworthy changes in the predicted docking poses of the compounds. OCA again was docked with its carboxyl side chain near the top of the pocket, interacting with residues on helix 3 (Figure 3. 20A). Whilst the key interaction with H447 was recovered, the altered binding pose of OCA, meant that the interaction was made by the C7 hydroxyl rather than by the group at the C3 position. Moreover, the steroid ring of OCA also appeared to be bent back on itself, possibly as an attempt to fit the compound in a smaller,

unnatural binding cavity. Again Jed561 was positioned with its 6α -ethyl oriented towards the top of the pocket, and the 'ABCD' ring was located in the sub-pocket made by helix 3 and loop 11/12 (Figure 3. 20B). Interestingly, the side chain of Jed561 was positioned towards the base of the pocket, where it exhibits a steric clash, and thus inserts itself between the extended H11 loop and helix 7. Meanwhile, compounds such as Jed441, at its lowest binding energy, were not even predicted to dock inside the ligand binding pocket of the antagonist conformation, and instead were placed outside in the solvent exposed region (Figure 3. 20C).

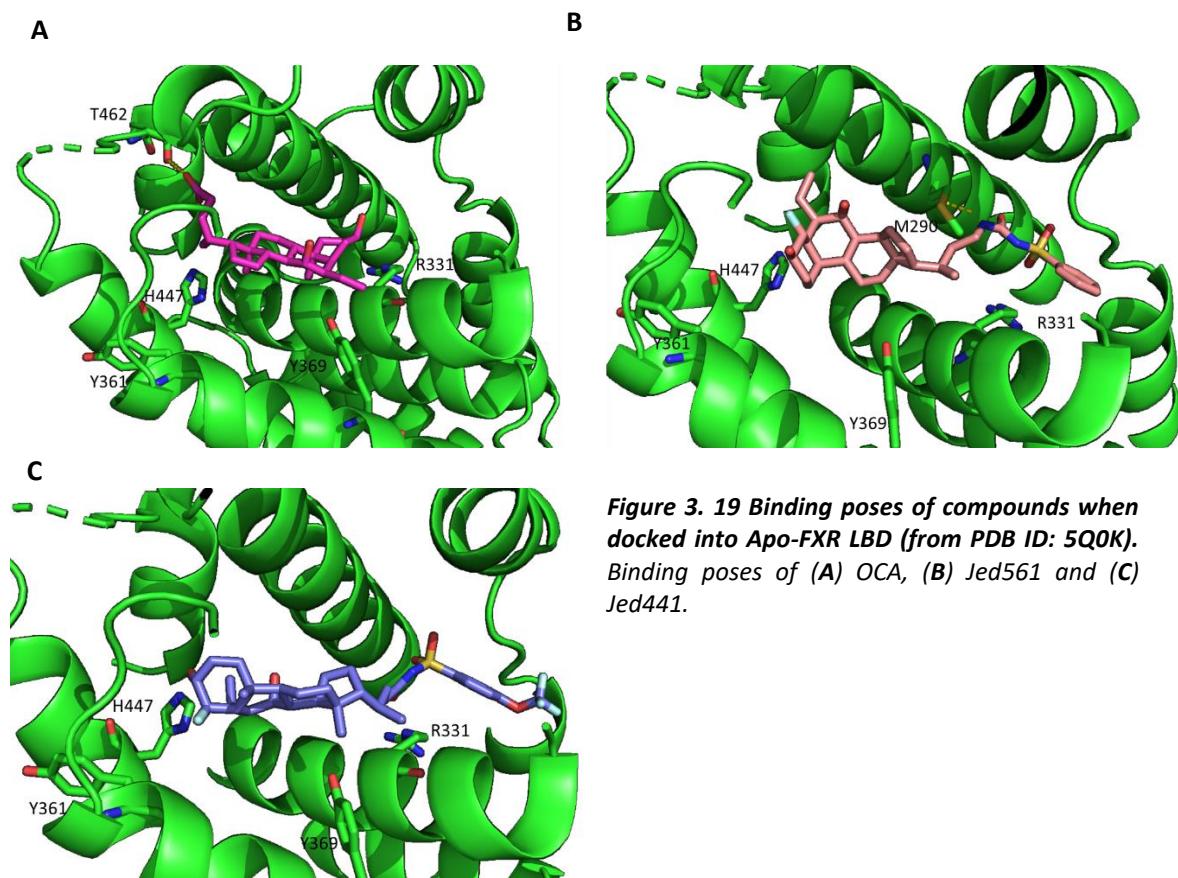


Figure 3. 19 Binding poses of compounds when docked into Apo-FXR LBD (from PDB ID: 5Q0K). Binding poses of (A) OCA, (B) Jed561 and (C) Jed441.

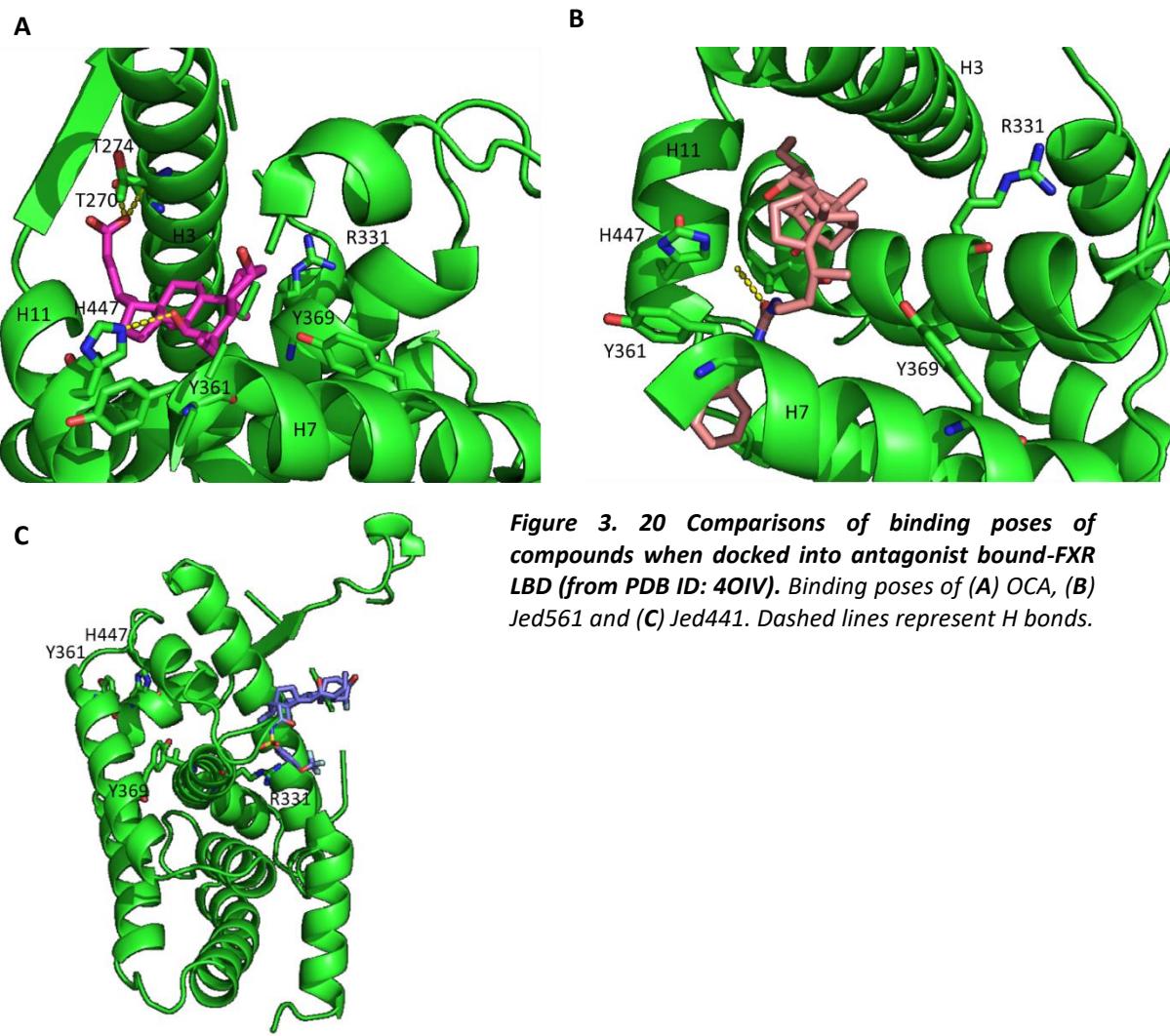


Figure 3. 20 Comparisons of binding poses of compounds when docked into antagonist bound-FXR LBD (from PDB ID: 4OIV). Binding poses of (A) OCA, (B) Jed561 and (C) Jed441. Dashed lines represent H bonds.

3.4.3 3D-QSAR of Novel FXR agonists

Compounds were synthesized by colleagues at NZP UK, and 50 were tested externally (by DiscoverX, Eurofins, CA, USA) in cell-based chemiluminescence assays for their ability to activate FXR. These 50 compounds were tested at a range of concentrations and dose-response curves used to determine and provide corresponding EC₅₀ values for the 3D-QSAR training set used in this work. The EC₅₀ data generated were normalised to OCA, which was set at 100 nM, and data ranged from 9 nM to 400 nM. For use in the 3D-QSAR model the EC₅₀ concentrations were converted to logarithmic values. In an attempt to identify the specific chemical features of the novel compounds which were responsible for the enhanced potency, a 3D-QSAR approach was employed. The first step of the process is to align the 3D structures of the compounds, to easily distinguish structural differences. The 3D-QSAR was carried out using FLAP software and utilised the same ligand structure files that had been created previously for the molecular docking experiments in SYBYL. As the minimal common scaffold, OCA was used as a reference template,

using the bond type as the basis for alignment. The software, however, was unable to accurately align every compound, and once outliers had been removed, the final training set, with which the model would be generated, only contained 21 compounds (Figure 3. 21A and B). A PLS analysis with these compounds and their logarithm EC₅₀ values, minimized the molecular interaction field data into just a few combinations known as latent variables, with the R² and Q² plot showing that up to 3 latent variables were required to produce a model with the maximum statistical value for R² (Figure 3. 21C). The cross-validated Q² values, however, resulted in a decline when more latent variables were added to the model, with values declining below zero. This suggested that the model had been overfitted, and the addition of more latent variables cannot justify the data, reducing its predictive ability. Furthermore, when validating the resulting model, by comparing the actual versus predicted activity of these compounds, the graph generated did not show a linear relationship between the experimental and predicted values (Figure 3. 21D). Moreover, compounds which had been shown to have very similar experimental activity, often showed quite varied predicted values, further indicating an unsatisfactory model with low predictive power. The final results for this 3D-QSAR model identified the shape and hydrogen bond donor molecular descriptors as the two variables with the most influence on activity. However, both were predicted to have a positive and a negative correlation with activity. In fact, all descriptors observed, had both a positive and negative correlation with activity, apart from the aromatic descriptor, which was predicted to have a small positive influence on activity.

It was initially thought that the unsatisfactory model may have resulted from the poor alignment of the compounds. To overcome this, Volsurf+ software was used to carry out a QSAR without the need for aligning the compounds. Volsurf+ compresses the information in 3D interaction energy grid maps into a few 2D descriptors that also can be used to describe the pharmacokinetic properties of the compound in question. Once 2D descriptors are generated, the procedure, much like the 3D-QSAR method in FLAP, uses principal component analysis and PLS to generate a model to correlate descriptors with activity. Following PLS analysis, again the results of the R² and Q² validation plots showed that the R² coefficient increased with an increasing number of latent variables, however, the Q² value started to decline after 2 variables (Figure 3. 21E). Again, this suggested that the inclusion of multiple descriptors could not add any more predictive power to the model without including some noise or irrelevant information. Further validation, by plotting the predicted versus experimental activity data, indicated that the model had poor predictive ability, as no linear relationship was observed between the two values (Figure 3. 21F).

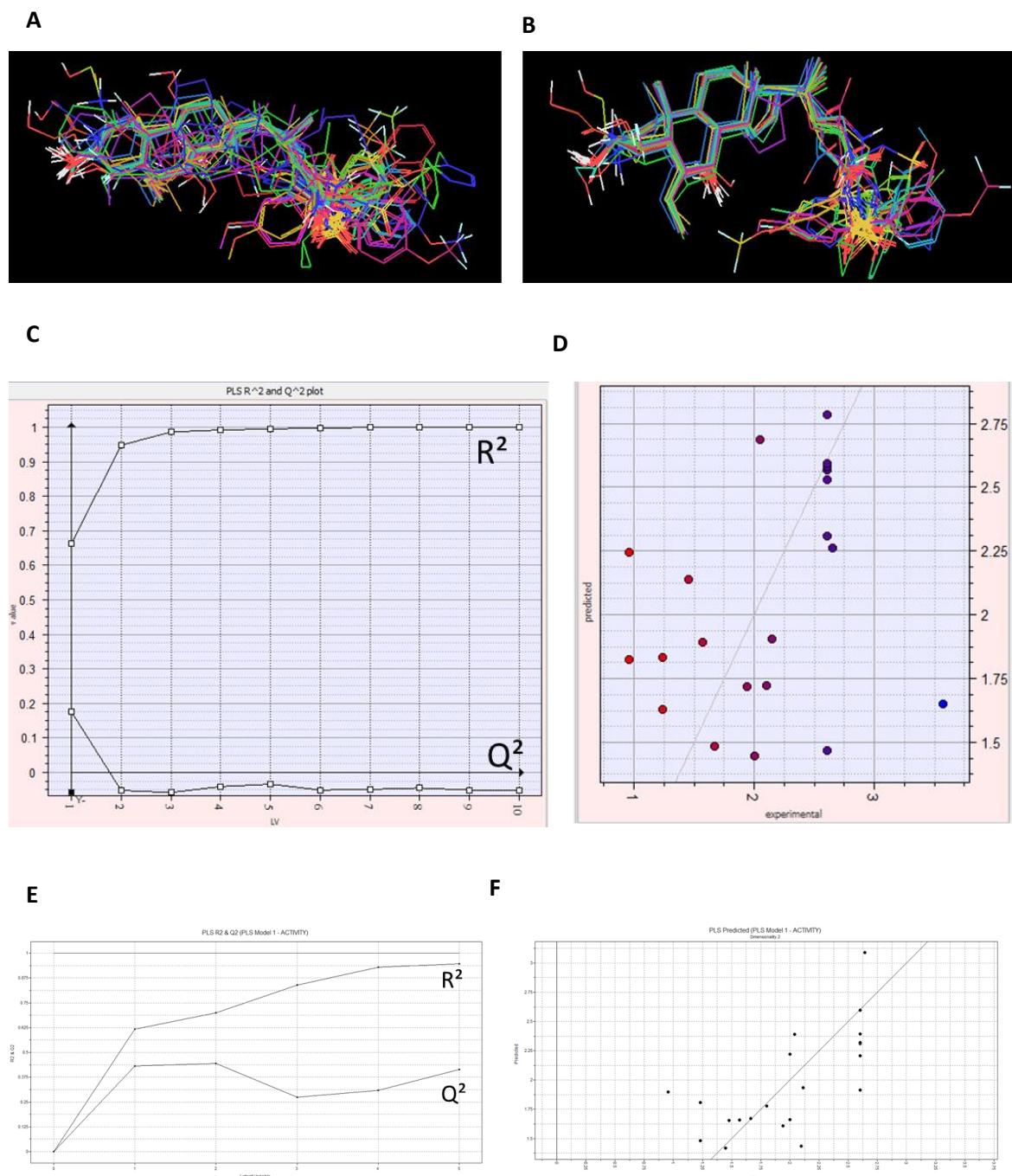


Figure 3.21 Correlating molecular descriptors with biological activity of compounds. Alignments of (A) novel compounds in the training set using OCA as a template and (B) with outliers removed. These alignments were used to generate a QSAR model correlating structural features with experimental EC₅₀ data. PLS validation plots to check the accuracy of the 3D QSAR model (C and E), and Volsurf+ model (D and F). In (C and E) R² / Q² plots, the Q² value should be close to the R², but in the models here the Q² values do not increase with more variables added, suggesting that the model is only accurate and specific for the training set. In (D and F) plot showing predicted vs experimental values, there is little correlation between predicted log EC₅₀ values and those experimentally determined, further displaying the model's inability to accurately predict EC₅₀ values of other compounds outside of the training set used to build it. In D, each compound is represented by a data point, and red to blue data points represent lowest to highest logarithmic EC₅₀ values. EC₅₀ data determined by dose-response curves from chemiluminescent reporter assays were provided by DiscoverX, but QSAR modelling and validation shown here performed by the author.

3.4.4 *In Silico* Docking of Compounds into TGR5

One of the aims of this work was to create compounds that are specific to FXR and will not activate off target receptors. One receptor of particular importance was TGR5 (G-protein BA receptor, GPBAR), whose activation had been associated with the adverse effect of pruritus in early clinical trials (Neuschwander-Tetri *et al.*, 2015). In order to identify whether these novel compounds could potentially activate TGR5, *in silico* experiments were also carried out to dock the compounds into this receptor.

3.4.4.1 Homology Modelling of TGR5

As to date, no structure of the TGR5 receptor has been published, docking studies utilised homology models of this receptor. An initial homology model was created by performing a BLAST search within the Protein Data Bank, to identify proteins with a high sequence similarity which also had structural data available. Results identified a structure of the lysophosphatidic acid receptor (PDB ID: 4Z36) as the top-rated homolog. Using the predicted binding residues as theorized by Gertzen *et al.* (2015) for interactions with the most potent BA agonist of TGR5, taurocholate, the sequence of top ranking homologs were aligned with TGR5 to see if key binding residues were conserved amongst these proposed substitute receptors. Gertzen and colleagues proposed that BAs bind in the top intramembrane region of the receptor close to the extracellular loops. It is thought that the C3 hydroxyl group, shared by human BAs, can interact via hydrogen bonds with a glutamate and tyrosine residue within the receptor, and via hydrophobic interactions with a leucine and tyrosine residue along the steroid ring (Figure 3. 22A). Multiple sequence alignment results from ClustalW2, identified the lysophosphatidic acid receptor as the closest related sequence, however, the only residues that were conserved in the proposed ligand binding region were Y240 and L244.

Another homology model also was generated, this time by submitting the TGR5 amino acid sequence to the secondary structure prediction server, IntFOLD. The structure generated was based on the structures of the turkey β -adrenergic receptor (PDB 2Y00), human β 2-adrenergic receptor (PDB 2RHI), and substance P receptor (PDB 2KSB). The IntFOLD results also listed R79, Y69, E169, Y240 and L244 as some of the potential ligand binding residues for this structure.

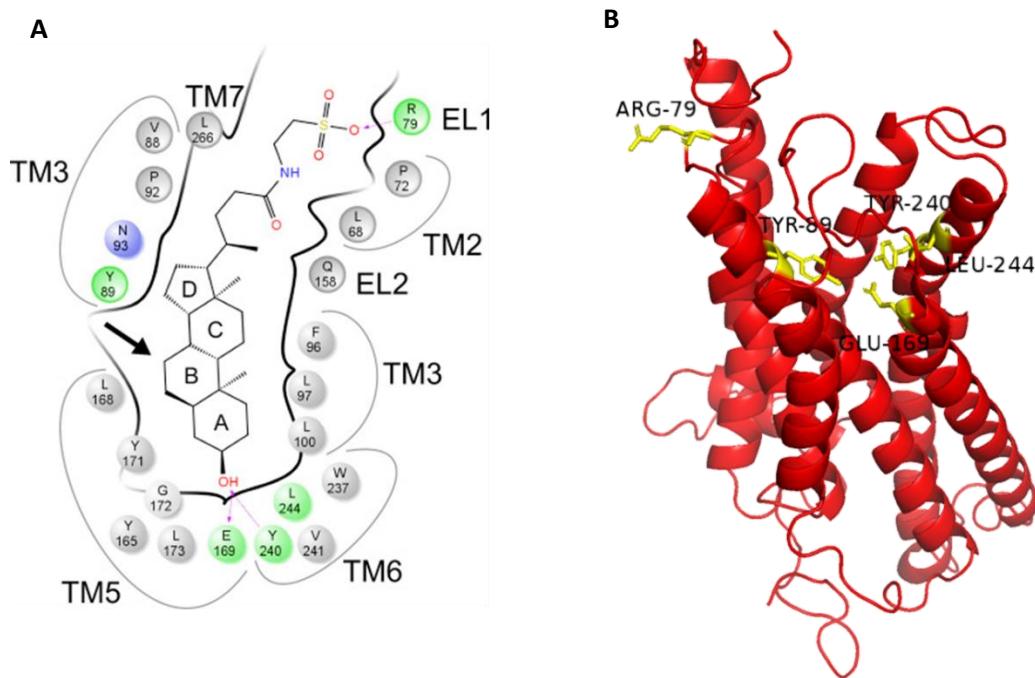


Figure 3.22 Proposed binding mechanism of BAs in TGR5 and key binding residues. (A) Taurocholate interactions with TGR5 taken from (Gertzen *et al.*, 2015), (B) Key residues mapped on to the homology model of TGR5 generated by the IntFOLD protein prediction server, in this work. 'TM' signifies transmembrane helices and 'EL' extracellular loops.

3.4.4.2 Docking Novel Compounds into TGR5

The structure for the TGR5 homology model was prepared in SYBYL, using the protein preparation tool as previously described (Chapter 3, section 3.3.4.2). Initially, a protomol was generated using the automated searching algorithm in SYBYL. However, this resulted in the placement of the compounds further down into the intramembrane region and not near the proposed extracellular binding site (Figure 3.23A). The protein was then re-probed this time using the residue directed procedure, whereby the key binding residues (as identified by Gertzen *et al.* 2015) were used for guiding the placement of the protomol. When this approach was used, only OCA was placed directly in the central binding cavity (Figure 3.23B). Results showed that OCA was positioned in the same orientation as the theorized mechanism, with the A ring headfirst. However, the compound was placed lower into the central region than the hypothesized site. Nonetheless, OCA was predicted to make hydrogen bonds with the key residue Y240, albeit with its carboxylic acid tail, rather than C3 hydroxyl group. Jed441, on the other hand, was placed in the opposite orientation, with its A ring facing the extracellular apex of the receptor. Furthermore, Jed441 was docked to the side of the intracellular domain and not directly inside the binding site. Likewise, Jed561, did not appear to fit inside the binding site,

and at its lowest binding energy was predicted to bind in the extracellular loop region of the receptor, outside of the binding cavity.

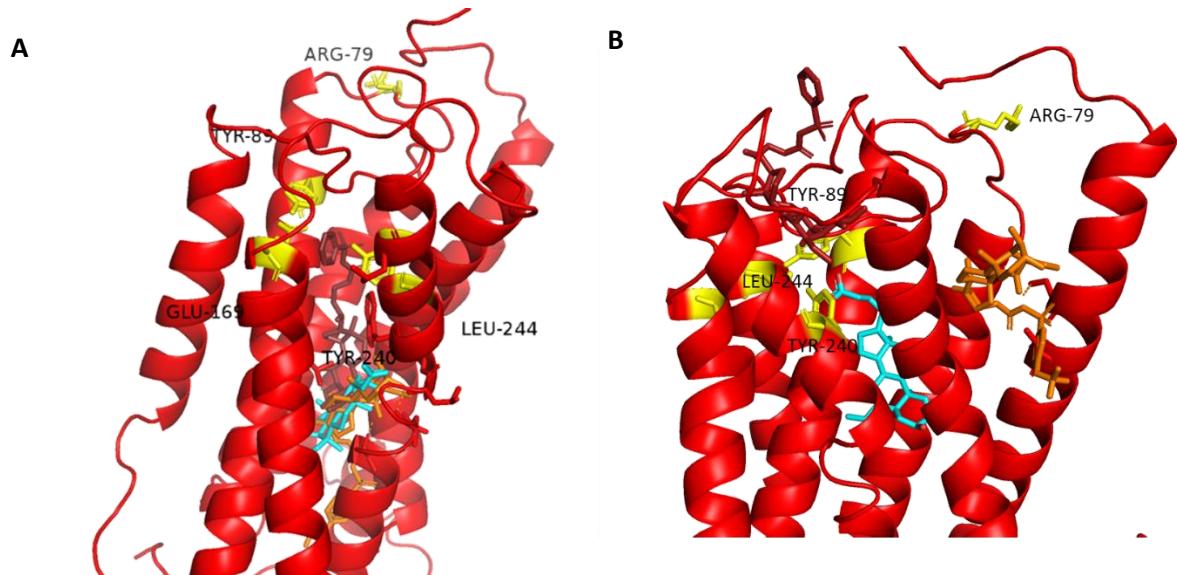


Figure 3.23 Compounds docked into TGR5 homology model using (A) automated protomol generation and (B) residue directed protomol generation. Key ligand binding residues are highlighted in yellow and labelled. OCA (cyan), Jed441 (orange), Jed561 (burgundy).

3.5 Discussion

3.5.1 The Structural Plasticity of FXR

The use of online bioinformatics servers and computational software has been instrumental in understanding structural features of FXR and its mechanisms of ligand binding and activation. The homology model determined by the protein structure and function prediction server, IntFOLD, supported the idea that FXR shares the common nuclear receptor modular architecture, with an intrinsically disordered *N*-terminal region, a zinc finger-containing DBD, and an α -helical barrel LBD (Rastinejad *et al.*, 2015).

The vast crystallographic data available for the FXR LBD, reinforces the idea that, like other nuclear receptors, FXR's LBD appears to be completely flexible and various helices can move to accommodate diverse ligand scaffolds. Superimposition of the available FXR LBD structures and comparison to the structure obtained with cognate CDCA ligand, PDB ID: 4QE6, highlighted the relatively subtle changes in individual helices that were required for the spectrum of receptor activation.

Overall structural changes were quantified by the RMSD score (Table 3. 6). The RMSD describes the average distance between identical atoms on the superimposed proteins, and larger scores designate larger distances and larger variation between the structures. Interestingly, the RMSD of the apo structure (5Q0K) compared to 4QE6 was not the highest of all the structures compared. Hence, highlighting the very minor global conformational changes needed for FXR to be activated by CDCA. Despite this, these aligned structures displayed slight differences in almost every helix, except for helices 1, 3 and 8, and the reduced ligand binding cavity observed in 5Q0K, altered the binding poses of docked compounds (Figure 3. 19A).

Of all the alignments, the structure of OCA-bound FXR, 1OSV, had the lowest RMSD score, and therefore the highest structural similarity to 4QE6. Furthermore, the semi-synthetic ligand was observed to make hydrogen bonds and hydrophobic interactions with the same residues as the classic ligand CDCA. OCA has the same common structural motif as CDCA, with the only difference being the addition of an ethyl group at the C6 position, but OCA has been determined to be 100 times more potent than its derivative (Pellicciari *et al.*, 2004). The structural data here further confirms that relatively small conformational changes in the FXR LBD can result in significant functional consequences.

Considerable deviation was observed in helix 2 and helix 6, which both adopted varying positions dependent on the ligand that was bound. In some cases, these helices were not

resolved, perhaps suggesting that their flexibility prevented the well-ordered, discrete structure required for accurate electron density maps produced by X-ray crystallography. Both helix 2 and helix 6 are positioned on either side of helix 3, and both have been proposed to act as a molecular spring to accommodate large ligand scaffolds and support the conformational movements of helix 3 upon ligand binding (Downes *et al.*, 2003).

Nonetheless, the most significant flexibility was observed for helix 12, which differed depending on whether the LBD was unliganded, bound with a full or partial agonist, or bound with an antagonist. In the apo structure of FXR, the loop between helix 11 and 12 is considerably disordered, and helix 12 lies slightly apart from the main LBD core. However, in the agonist bound structures, helix 12 lies slightly closer to helix 3, but at a more perpendicular angle. Although helix 12 was not resolved in the structure co-crystallised with partial agonist, ivermectin, the holo structure with the antagonist, NDB, displayed significant conformational changes in this end region of the LBD. The structure obtained with NDB consisted of a homodimer of two FXR LBD molecules, in which only one domain was bound with the antagonist. The structure observed included a β -strand immediately preceding helix 12, which, in turn, adopted an extended remote conformation far from the rest of the LBD helices. This dynamic extendibility of helix 12 underlies the mechanism of ligand-dependant transcriptional activation of FXR, as with other nuclear receptors. Helix 12 is also known as the activation function 2 (AF2) domain. Structural studies into various other nuclear receptors have proposed that in an apo or antagonist position, helix 12 is completely flexible and lies in a remote position, far from the rest of the domain, where it precludes protein-protein interactions with a coactivator protein (Mackinnon *et al.*, 2014). However, upon ligand binding, helices 2, 3, 6 and 11 change conformation, which ultimately alters the position of helix 12 to a more stable one. Moreover, molecular dynamic simulations have also proposed that loop 11/12 is also fundamental in the rearrangement of helix 12 in full agonist conformations (Costantino *et al.*, 2005). In the agonist position, helix 12 localises between helix 3 and helix 11, creating a hydrophobic binding groove on the surface of the LBD, which promotes the recruitment and interaction with a coactivator protein (Rastinejad *et al.*, 2013). This structural molecular switch mechanism is not unique to FXR, however, the mechanism by which it exclusively recognises and binds its cognate BA ligands is unique.

3.5.2 Ligand Binding Mechanisms of FXR

As previously discussed, the FXR LBD can assume various dynamic conformations, and several ligands can initiate an induced-fit mechanism whereby helices can move to accommodate varied ligand scaffolds. Within the ligand binding pocket, several polar and charged residues

engage in hydrogen bonds with ligands, and the presence of multiple hydrophobic residues promotes additional stabilisation through hydrophobic effects. Unlike other steroid hormones, BAs, bind ‘headfirst’ in the FXR binding pocket, with their A- ring in the corner of the pocket, and their side chain extending out of the opening. BAs like CDCA, are well suited to this unique binding cavity, in that their hydroxyl groups along their α face are well positioned to interact via hydrogen bonds with S332, Y361, Y369 and H447, along helix 5, 7 and 11. Furthermore, the positioning of methyl groups along the β face of their 4 ring nucleus, promotes supplementary hydrophobic interactions with residues on helix 3, 5 and 7. The addition of their polar carboxylic acids, mean that hydrogen bonds with R331 on helix 5 can further stabilise their position within the pocket. Other BAs, such as DCA and LCA lack the hydroxyl group at the C7 and, therefore, are expected to only interact with H447 and Y361, which is thought to account for their partial agonism (Makishima, 1999; Downes *et al.*, 2003). H447 and Y361 are thought to act as an activation trigger, whereby polar contacts with these residues place the protonated histidine residue in a position where it can form a cation- π interaction with W469 on helix 12, levering the AF2 into its stable, active conformation (Mi *et al.*, 2003). Conversely, the compilation of the ligand binding interactions observed in different agonist-bound structures (Table 3.6), suggests that whilst hydrogen bonds with H447 are common, the interaction is not always essential for potent agonists such as GW4064 (in structure 3DCT). Likewise, previous molecular dynamic simulations have suggested that the interaction between H447 and W469 is not sufficient to stabilize helix 12 in an active conformation (Costantino *et al.*, 2005). Furthermore, fundamental differences are observed in other structures where ligands engage in polar contacts with residues on helix 3 or probe alternative sub-pockets. This further suggests that receptor activation isn’t restricted to direct interactions with H447 and Y361, and that the coordinated effects of interactions with residues along other helices may be able to support an active LBD conformation.

Results of molecular interaction field analysis of the FXR ligand binding pocket, identified the potential for polar interactions in a sub-pocket between helix 3 and loop 1/2, which could be exploited in the design of novel ligands. This is further supported by previous docking studies, which identified the previously undescribed potential for small molecule placement in this region (Meyer *et al.*, 2005; Pellicciari *et al.*, 2006). The pocket, characterized by H294, V297, L298, F301 on helix 3, Y260 on helix 1, and M265 on loop 1/2, has been described as the non-canonical or ‘S2’ binding site; due to the disposition of guggulsterone in this alternative site during docking studies, it has been proposed as a site for the selective modulation of the receptor (Meyer *et al.*, 2005). Indeed, interactions with helix 3 in this sub-pocket, may be

sufficient to alter its conformation and perhaps create a stabilised, but suboptimal coactivator binding groove with helix 12.

3.5.3 Docking Novel Compounds into the FXR LBD

Previous literature recommends performing docking of novel ligands in a receptor structure that is complexed with a similar ligand (Sutherland *et al.*, 2007). As such, the 4QE6 structure was chosen for docking due to its crystallisation with a BA structural analogue, CDCA. Validation of the docking mechanism, whereby the original CDCA ligand was extracted from the FXR LBD structure, processed as per the standard ligand preparation procedure, and docked back into the same receptor, showed marginally different bond angles of the CDCA compound. These may have arisen from the energy minimization of the ligand and resulted in a slightly altered placement of the re-docked ligand compared to the crystallised ligand. Nevertheless, in the top scoring solution, the self-docked ligand occupied the same binding site as the original and most closely resembled the original crystal structure; although it did not reproduce all of the same potential polar contacts as the original co-crystallised ligand, the overall RMSD score was within the threshold limit deemed acceptable to good, 2.0 to 3.0Å (Ramírez and Caballero, 2018).

Close to 100 novel compounds were designed and docked for work presented herein. By using the automated protomol generation procedure, ligands were docked in their most energetically favourable loci, without the unintentional bias arising from placing the compounds in the classic steroid binding site. In general, novel compounds occupied the same site as CDCA, and were orientated with their A rings closest to the pocket corner. Novel compounds also engaged in hydrogen bond interactions with one or more of the common binding residues, but bulky additions and substituent groups also meant that several non-canonical interactions were formed, and side chains could occupy the alternative sub-pockets.

Substitutions or additions of side chains at the C3 position on the BA 'A' ring caused a shift in the steroid nucleus further along in the pocket. However, many of the polar interactions with classic residues in the corner of the pocket, such as H447, Y361 and Y369, were maintained by substituent electronegative atoms and polar functional groups on the novel compound scaffolds. Likewise, extensions of the compounds at their carboxyl side chains, were also able to maintain many of the common interactions with R331 and other residues at the opening of the cavity. However, small changes in the composition of the side chains, and the inclusion of new functional groups in their extensions, often resulted in changes in the torsion angles and conformations around the compound's rotatable bonds, ultimately resulting in the engagement of different potential residues; while the pose and position adopted by the steroid nuclei of

these compounds were often very similar, their distal side chain ends often displayed considerable conformational differences. Furthermore, whilst some of the extended side chains protruded out of the cavity and into the solvent exposed space, possibly making the most distal functional groups redundant, other chains formed electrostatic interactions with residues in the auxiliary binding pocket between helix 3 and loop 1/2. Of particular note was the addition of a highly electronegative, polarizable trifluoromethyl group to the end of the compound side chain. Trifluoromethyl groups are often added to compounds as a substitute for the more easily oxidisable hydrocarbon methyl groups, resulting in a more metabolically stable, lipophilic compounds (Esterhuysen *et al.*, 2017). Results shown here, suggest that inclusion of this fluorinated functional group to the compounds, increased the potential for the side chain to occupy the secondary binding pocket, where it could form favourable halogen interactions with a range of residues on helix 3 and loop 1/2.

Although, the results shown here provide insight into the unique binding modes of the novel compounds and their occupancy of alternative sub-pockets, a major limitation of the docking method is that, by itself, it cannot sufficiently predict the functional activation of the receptor, nor the biological effect of said interactions. A challenge of this receptor, as noted previously, is that there are no residues solely responsible for the binding of ligands or for the complete, direct activation of FXR. Furthermore, by virtue of FXR's activation mechanism, there is no obligate link between the binding energies of docked ligands and their efficacy (Pellicciari *et al.*, 2006) and although binding affinities weren't considered in this case, using docking studies alone, it is impossible to distinguish full and partial agonists from competitive inhibitors.

3.5.4 Proposed Binding Mechanism of Lead Compounds

In order to identify which compounds were biologically active and which ligand scaffolds were most associated with the improved efficacy, 50 of the compounds were synthesised and screened in an *in vitro* chemiluminescent FXR activation assay. The resultant EC₅₀ data were normalised to OCA (at a value of 100 nM), which is regarded as the gold standard. Results of EC₅₀ values (third party data not shown) showed that compounds with either an extension or nitrogen heteroatom replacement at the C3 position (e.g. Jed303, Jed420, Jed431, Jed433), had the lowest affinity for FXR LBD, with values similar to CDCA. On the other hand, compounds which retained the C3 hydroxyl group, but had an additional fluorine at the C4 β position (Jed441-443, Jed559-564), all appeared to have a higher potency than other compounds measured. Most members of this series of fluorinated compounds had the potential to form polar interactions with classic residues H447 and Y361 from their C3 hydroxyl groups, with the additional potential to form supplementary halogen bonds with Y361 and Y369 from their 4 β F

atoms. Furthermore, comparisons of the docking results of Jed561 with Jed692, an identical derivative without the fluorine at this position, suggests that Jed692 cannot be anchored into the corner of the pocket by interactions with H447 and Y361, and instead lies further along the pocket with its side chain protruding out of the cavity. In addition to their effect on metabolic stability and lipophilicity, fluorine functional groups are often added to compounds to affect the reactivity of neighbouring functional groups (Graton *et al.*, 2012). Whilst it's possible that the fluorine atom at this position on the A ring affords the compound some enhanced potency, possibly due to the engagement of classic residues by halogen bonds, the fluorine may also be affecting the hydrogen bond potential of the adjacent hydroxyl group on the C3. However, whether this impact on neighbouring alcohol groups is positive or negative, is debated and may be dependent on the chemical environment of the atoms (Graton *et al.*, 2012; Linclau *et al.*, 2016); and the improved activity of these compounds cannot be solely defined by this simple addition.

Comparisons of OCA and Jed432, its 4 β F-derivative, suggest that both share identical docking poses and participate in interactions with the same classic residues. However, both share very similar EC₅₀ values, suggesting that the 4 β F atom alone is not enough to significantly affect the potency of the compound. Extensions of the tail ends of these compounds must also be contributing to their activity. Accordingly, the top two compounds, Jed441 and Jed243, which both had an EC₅₀ value of 9 nM, had identical structures with the exception of the 4 β F, which was not present in Jed243, thus, highlighting that the side chain of the compound may be more important for increased potency.

Docking studies were able to identify the unique binding mode of Jed441, which was placed in a flipped orientation, facing helix 3, with its 6 α -ethyl group oriented towards the top of the pocket. In this orientation, Jed441 had the potential to interact via hydrogen bonds with L287, and via halogen bonds with V297, both on helix 3. Similarly, the next most active compound, Jed561, which had an EC₅₀ value of 26 nM, also made potential hydrogen bond interactions from its side chain to residues on helix 3. Moreover, several other related compounds in this series, within similar range of EC₅₀ values (e.g. Jed560 and Jed563), also were able to make polar contacts or halogen interactions with residues along helix 3. These data imply that by interacting with helix 3, these compounds can promote activation of FXR, presumably by affecting the hydrophobic coactivator binding groove, which is made between helix 3 and helix 12. Interactions with helix 3 have been observed with several other potent agonists such as Fexaramine (in structure 1OSH) and GW4064 (in structure 3DCT) where these compounds have been noted to stabilise the hydrophobic core of the receptor and help to sustain a protein

conformer required for coactivator recruitment (Downes *et al.*, 2003; Akwabi-Ameyaw *et al.*, 2008). Whilst docking results certainly suggest that this helix 3 interaction mechanism may be responsible for the improved activity of novel compounds presented here, the challenges of a rigid or even semi-rigid (as used herein) docking procedures in a highly plastic receptor, emphasizes the need to confirm these ligand binding modes by X-ray crystallography or other experimental methods.

3.5.5 Limitations of a Static Docking Procedure

The long-established methods of compound screening by virtual docking, typically use a high-resolution static structure derived from X-ray crystallography. However, the holo-receptor structures determined by this method, often only represent an extreme atomic 'snapshot' of conformational states. In the case of the FXR LBD, ligand binding does not lead to one particular conformation, and instead results in the inherently, dynamic conformational changes of helix 12. As such, the static picture provided by docking results may not necessarily, sufficiently explain the molecular recognition and receptor transactivation upon ligand binding, and other techniques which can sample the receptor's numerous conformational microstates are required.

The use of computational methods to accurately account for structural plasticity in ligand screening has recently been established, and the nascent field of flexible docking is becoming more frequent. Older methods such as that of induced fit docking, involve the formation of a loose ligand-receptor complex that induces a conformational change in the receptor leading to a conformation that supports a tighter-bound complex (Koshland, 1958). Induced fit docking was used previously by Fu *et al.* to identify a representative FXR state which was then used as a model for the subsequent virtual screening of compounds (Fu *et al.*, 2012). However, the use of induced fit docking, may not be able to sample the entire spectrum of binding scenarios and its use to identify binding modes of a large compound library will be significantly computationally expensive and time consuming (Feixas *et al.*, 2014). An alternative approach is to use ensemble docking, which utilises numerous conformational states to account for ligand interactions with targets in different conformations. This approach has been previously used to identify potential oestrogen receptor (ER) ligands, and furthermore, to discriminate full agonists and selective allosteric modulators (Mackinnon *et al.*, 2014). This ensemble approach was found to yield consistently higher enrichments and diverse actives than when a single structure was used for docking.

To overcome the ambiguity associated with docking into a single structure of the FXR LBD, and to indirectly account for different receptor conformations, lead compounds were also docked into the apo and antagonist-derived structures. Compounds displayed markedly different binding orientations in the apo structure, and within the antagonist structure, they either caused a steric clash or were not even deposited inside the receptor. These results further suggest that the compounds are indeed agonists rather than antagonists, although the extent of agonism cannot be determined by the docking method used here.

Additional methods, such as molecular dynamics simulations, can be used to explore a full continuum of structural states, generating novel receptor conformations not exhibited in existing crystal structures; and it can even be used to characterize allosteric binding sites (Durrant and McCammon, 2010). Moreover, molecular dynamics can overcome limitations of traditional docking, by accounting for solvation effects whereby key, ordered water molecules in the binding site can be exploited. Extensive molecular dynamics simulations were used previously, subsequent to rigid docking into a multitude of FXR LBD crystal structures with the anticipation that it would help to refine predicted ligand binding poses (Bhakat *et al.*, 2018). However, refinement did not significantly improve predictions, and simulations were unable to accurately discriminate kinetic stability between correct and mis-docked poses (Bhakat *et al.*, 2018). Furthermore, due to the low microsecond timescale of simulations, it is thought that molecular dynamics may unintentionally exclude the capture of important conformational changes (Feixas *et al.*, 2014). Again, although the simulations may provide insight into structural rearrangements made when a ligand binds, it may not necessarily reflect the transcriptional activity of the ligand. Consequently, for the identification of lead compounds, molecular dynamic simulations are often used in conjunction with other virtual screening methods, as well as other experimental techniques.

3.5.6 Ligand-based Approaches

The structural flexibility observed for the FXR LBD presents a considerable challenge for binding predictions obtained by docking methods. To account for this, work also focussed on the use of ligand-based approaches to identify the specific chemical features that were responsible for the improved activity of the novel compounds. Three-dimensional QSAR is a common technique used for the identification of molecular descriptors that correlate with activity, and to date numerous groups have used this approach for fine-tuning the development of non-steroidal FXR agonists (Honório *et al.*, 2007; Zhang *et al.*, 2007). Unfortunately, the procedure employed here was associated with several problems. The first being the inadequate alignment of the compounds. Molecular alignment is one of the crucial steps in conducting a 3D-QSAR, and the

FLAP software aligns the compounds as to distinguish the common ligand scaffolds from the unique chemical characteristics. However, an inherent assumption of the method is that all compounds bind in a common manner (Kim *et al.*, 1998). Although most of the compounds shared a very similar binding mode, some compounds such as Jed441 were seen to be flipped upside down, and the considerable rotation in the side chain, often resulted in the compounds engaging in interactions with alternative residues. Alignment based on structural similarity of the compounds, therefore, may not necessarily represent the appropriate orientations or conformations in the target receptor that is associated with their activity, resulting in a weaker predictive ability of the model.

Nevertheless, use of the alignment independent Volsurf+ technique also failed to produce a reliable statistical model, which was able to accurately predict the activity or pharmacokinetic properties of the compounds. This suggests additional or alternative problems with the data. For example, a fundamental requirement to obtaining a predictive QSAR model is a structurally and functionally diverse dataset, with responses ranging over several orders of magnitude (Shi *et al.*, 2001; Dearden *et al.*, 2009). The FXR activity data for these compounds, obtained from DiscoverX (Eurofins), appears to have an upper limit possibly due to the detection limit of the assay, and many compounds near this threshold have the exact same EC₅₀ value. To obtain more accurate results, a higher range and more intermediate ligand concentrations should have been used. Nonetheless this lack of diverse data undoubtedly skews the results and prevents accurate validation of the model, thus preventing it from accurately predicting important structural or physico-chemical features correlated with activity.

3.5.7 Conclusion

The work presented in this chapter highlights how computational methods can be applied to aid the identification of novel agonists of FXR. Bioinformatics and structural visualisation tools have provided insight into the full length structure of FXR, as well as the intrinsically flexible LBD. Computational docking methods have been able to characterize the potential binding mechanisms of novel ligands, and also have been applied as an approximate indicator of off-target TGR5 activation. It is realised that docking results cannot be naively taken as is, since many factors, including structural plasticity and solvent effects, play an important role in the *in vivo* binding of ligands, and on its own, docking poses and binding affinities cannot accurately predict receptor transactivation. However, in conjunction with *in vitro* screens, such complementary computational methods have provided a rationalisation of the interactions with helix 3 that may account for the improved activity of lead compounds Jed441 and Jed561. Although the 3D-QSAR executed here was unable to produce meaningful results, X-ray

crystallographic analysis of FXR complexes with the lead compounds may help to elucidate specific chemical characteristics of the compounds that participate in specific interactions in the ligand binding pocket, which may be crucial for activation of the receptor.

Chapter 4. Structure Determination of FXR

4.1 Introduction

FXR belongs to a large superfamily of receptors that underlie several important physiological processes. Accordingly, they are of substantial interest in modern biomedical research and drug discovery. Despite the growing progress in technology and computational approaches in the field of drug discovery, structural biology has been instrumental in the adoption of some orphan receptors, in providing atomic-level insights into the nuclear receptor binding domains, and in understanding the unique events that determine ligand binding (Ottow and Weinmann, 2008).

Previously published structures of FXR have helped to identify binding modes of cognate BA ligands, in addition to novel means in which non-steroidal ligands bind in the LBD. Furthermore, the structure of the FXR LBD has aided the design of novel agonists and enabled predictions to be made as to the specific molecular recognition mechanisms associating ligand binding and receptor transactivation (Chapter 3). Nonetheless, due to the flexible nature of the FXR LBD, and the limitations of a semi-rigid docking procedure, it is imperative to identify exactly how FXR encompasses the novel ligand scaffolds, and to assess the reliability of the docking process. Several strategies can be applied to evaluate molecular docking procedures, including the calculation of docking accuracies, and the use of enrichment factor analysis (Huang *et al.*, 2006). However, by obtaining structural data specific to the FXR LBD complexed with these novel ligands, it will not only allow docking solutions to be correlated with actual observed poses, but it may provide the potential for novel ligands to be chemically modified and improved further. In addition, the new, more accurate receptor structure can be used for future rounds of molecular docking. Moreover, the structural determination of FXR's currently uncharacterized DBD may help to advance the understanding of DNA binding mechanisms, or even intra-domain signalling and allostery.

4.1.1 Methods of Structure Determination

There are several methods for determining the intricate three-dimensional structure of biological macromolecules. These methods gave rise to the field of structural biology and revolutionized the understanding of concepts surrounding the structure and function of proteins. Since its inception in the 1950's, MX has made the largest contribution to the understanding of proteins and, to date, has been the most popular method for determining the spatial relationships and locations of individual atoms within a protein. The advent of other methods such as NMR spectroscopy and the more recent advancement of Nobel prize-winning, breakthrough technologies in the method of cryo-electron microscopy (cryo-EM), has made it

possible for scientists to determine the high resolution structure of over 140,000 proteins (www.rcsb.org/stats, as of August 2019).

4.1.1.1 Macromolecular X-ray Crystallography (MX)

Of the 140,000 proteins deposited in the Brookhaven Protein Data Bank (PDB), approximately 130,000 of these have been determined by X-ray crystallography; with the burgeoning use and success of this technique owing to significant advances in automated purification and crystallization methods, increasingly powerful synchrotron X-ray sources, and the development of sophisticated computer software for data collection, structure solution and refinement (Wlodawer *et al.*, 2013). Resolutions of up to 0.2nm, or 2Å, are routinely obtained with MX, which is sufficient to distinguish peptides from the protein's main chain. Yet, structures with resolutions of 1Å or less, are becoming increasingly available, whereby individual atoms can be clearly defined and even hydrogens may be resolved.

The fundamental principle underlying MX is based on Bragg's law, which identified that when a crystal is placed in an X-ray beam, similarly to the phenomenon of reflection, X-ray scattering will be observed. This scattering arises from the highly ordered, regular arrangement of homogenous molecules in the crystal, whereby the electrons in the investigated sample, can diffract the incident X-ray beam at many discrete angles and intensities, depending on the spacing and planes of the crystal, its orientation relative to the X-ray beam, and the wavelength of the X-rays (Bragg and Bragg, 1913). The resulting diffraction pattern (recorded either by film or digital detectors) can be used to deduce the positions of the atoms that gave rise to the diffraction spots in the original structure. In addition to the measured intensities of diffraction spots, additional information concerning the 'phase' of the spots is required, and this can be estimated by other experimental or indirect computational methods. Together, data regarding the intensity and phase of the diffraction spots, can be transformed and used to reconstruct the crystal structure. The use of MX for the structure determination of proteins is now a mature technique, in which there have been improvements to nearly all of the time-consuming, critical steps (Dauter and Wlodawer, 2016). However, the overall quality of the structure is dependent on sharp diffraction spots, which, in turn, are determined by the size, quality and degree of internal order in the crystal. As such, a major bottleneck in the process of MX is in the production of highly ordered, diffraction quality, 3D crystals.

Protein crystals are formed by adopting a low energy, regular arrangement of protein molecules in 3D space. Regular packing involves repetition of the unit cell (simplest repeating unit) along different dimensions, and each unit cell is made up of a number of asymmetric units representing the rotated or translated macromolecule (Ilari and Savino, 2008). Despite being

well ordered, biological macromolecule crystals are usually loosely packed, allowing 40-60% of the volume to be occupied by solvent (Ilari and Savino, 2008). As a result, protein crystals are very fragile, but the fact that the protein molecules are surrounded by water, means that their structure and biochemical features, closely resemble that of their fully solvated variant (McPherson and Gavira, 2014).

For crystals to form, a highly pure (>95%), monodisperse sample of the protein is required. The initial protein solution is needed at the highest possible concentration, without the occurrence of aggregation or precipitation, and a concentration of 10 mg/mL is often recommended as an optimal starting point. It is critical for the concentration of the protein to gradually increase to a supersaturated level, providing a thermodynamic force that drives the proteins out of solution, allowing the formation of crystal growth centres or nuclei (McPherson, 1990). Once a stable nucleus is formed, the crystal can continue to grow, by non-covalent and intermolecular physical bonds that hold the molecules together in a stable, perfectly ordered, periodic crystal lattice (McPherson *et al.*, 1995). If the sample is not homogenous, contaminants can cause flaws or dislocations, and can even prevent the formation of crystal contacts, hindering the extension of the crystal lattice.

Macromolecular crystal growth is largely empirical, and an a difficult technique in practice; the objective of crystallization trials is to significantly reduce the solubility of the proteins in solution, either by the addition of precipitants directly to the sample (batch crystallization), or by allowing the gradual dehydration of the sample (vapour diffusion, Figure 4. 1). The method of vapour diffusion is typically the most common choice of crystallization. A droplet containing a mixture of the purified protein with a precipitant is positioned next to a larger reservoir containing the precipitant, either in a well (sitting drop) or suspended from a coverslip (hanging drop). As the water vapour diffuses out of the drop to achieve an equal osmolarity with the reservoir, the protein concentration is gradually increased (Dessau and Modis, 2011). In addition to precipitants, the process of crystallization can be extremely sensitive to the ionic strength of the solution, the pH and temperature. However, the method has been greatly advanced by readily available liquid handling robot systems, which can accurately aspirate submicrolitre drops in a matter of seconds. The use of multi-chamber plates and other consumables facilitates the high throughput (HTP) screening of vast matrices of crystallization conditions.

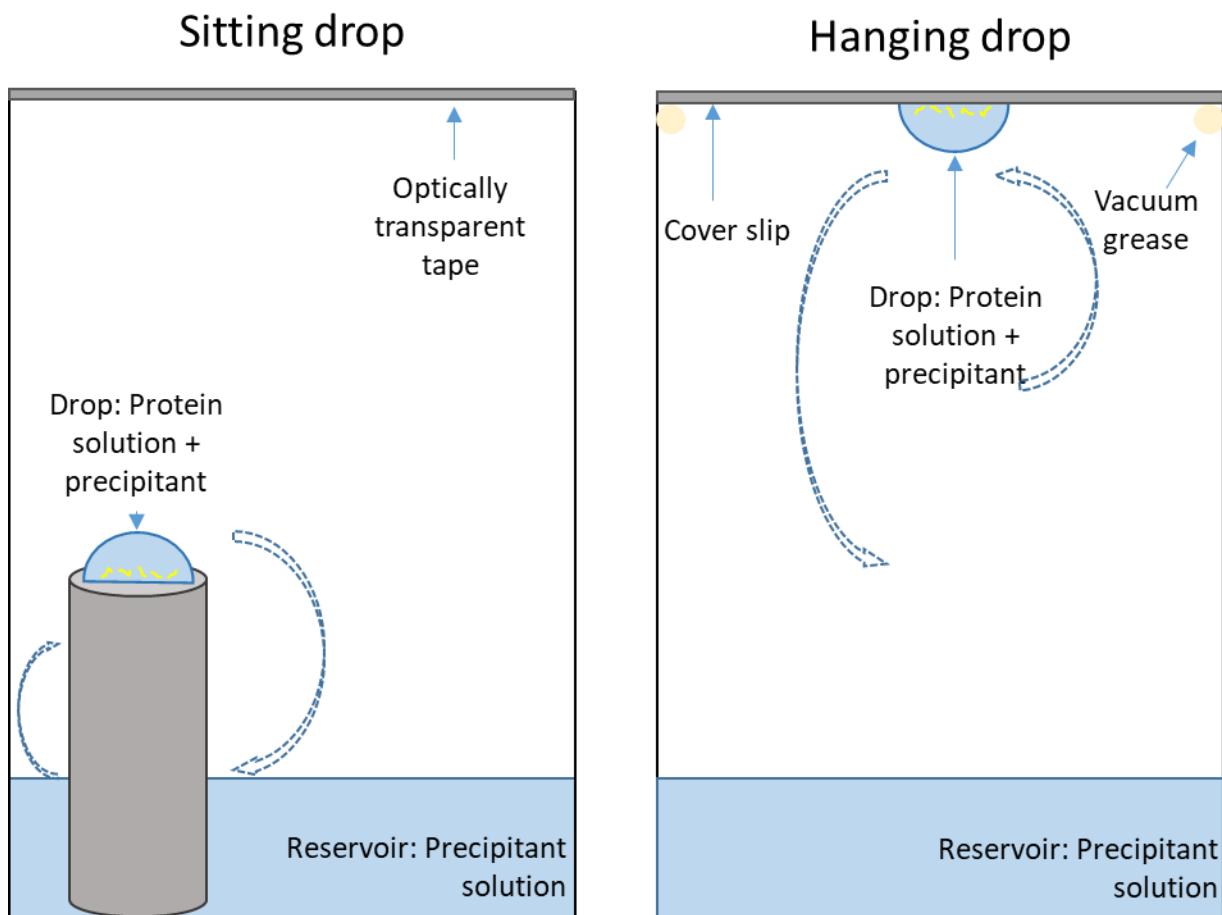


Figure 4. 1 Method of protein crystallization by vapour diffusion. In sitting drop vapour diffusion, the drop containing the protein and precipitant is elevated on a bridge above the reservoir, and the well is sealed with tape. In hanging drop vapour diffusion, the drop is suspended from a cover slip, which is sealed with vacuum grease to create an air-tight environment. The drop is allowed to equilibrate (blue dotted arrows) with the reservoir and as water diffuses out the concentration of both the protein and precipitate in the drop gradually increases promoting the growth of well-ordered crystal lattices under the appropriate conditions.

4.1.1.2 Nuclear Magnetic Resonance (NMR) Spectroscopy

NMR spectroscopy can also provide rich, detailed, atomic-level information about the 3D structure of biological macromolecules. Moreover, NMR provides the additional benefit of being able to use proteins in a solution, near physiological conditions, and can be applied to study time-dependent processes, such as intramolecular dynamics, molecular recognition and protein folding.

The phenomenon of NMR arises from the intrinsic 'spin' of atomic nuclei with odd mass numbers, which generates a magnetic field, or moment. When these spinning, magnetised nuclei are placed in an external magnetic field, the nuclei can adopt either a lower or higher energy spin state, and a resonance spectrum can be recorded following the application of electromagnetic radiation, which causes transition between a lower energy state to an excited

one (Lesk, 2016). These properties can be used to determine the chemical environment of the atom, where the flow of electrons in the vicinity of the nucleus can perturb the applied external magnetic field, to an extent dependent on the surrounding electron density. Consequently, nuclei in different surroundings will change states at different magnetic field strengths or radiation frequencies, and each chemical group will appear in different regions of the spectrum (Berg *et al.*, 2002). By using this information, it is possible to examine the effect of different conditions, such as a disordered or folded protein structure, on particular chemical groups or individual amino acids. Likewise, due to the fact that the spin of one atom's nuclei can have an effect on the spin of neighbouring nuclei, proportional to the distance between the two nuclei, the technique can be applied to detect the relative proximity of atoms to one another in a 3D protein structure. Furthermore, due to the use of molecules in solution and the potential conformational heterogeneity of protein structures, NMR analysis can detect a range of different conformations that the protein may adopt.

One limitation of the technique is that it requires NMR active atoms, with an odd atomic number, giving rise to nuclear spin. Naturally occurring carbon and nitrogen atoms usually have an even number of neutrons and protons (^{12}C and ^{14}N), and ^{13}C and ^{15}N isotopes are usually only present in organic material in trace amounts. To overcome this, isotopic labelling is routinely used where the protein of interest is recombinantly expressed in *E. coli* grown on media sources containing $^{15}\text{NH}_4\text{Cl}$ or $(^{15}\text{NH}_4)_2\text{SO}_4$ and ^{13}C -glucose (Cai *et al.*, 1998). A significant limitation of NMR is its limited use for proteins above 35kDa, which are subject to slower tumbling rates, reduced sensitivity and more complexity of the spectrum (Yu, 1999). Moreover, although the technique does not require crystals, which can often be tedious, time consuming or problematic to achieve, NMR still requires high concentrations and high purities of the sample to obtain a good signal to noise ratio (Berg *et al.*, 2002). Additionally, it requires a highly stable protein that can withstand unfolding or precipitation, during the days to weeks that the NMR experiment takes place. Finally, the spectral data analysis also can be extremely exhaustive, and the large number of structures compiled can require complex, time consuming experiments and lengthy analysis.

4.1.1.3 Cryo-Electron Microscopy (cryo-EM)

Cryo-EM is a method that can resolve structures from a variety of biological macromolecules in their near native environments, ranging in size from a several tens of kilodaltons (kDa), to megadaltons (mDa) and even to whole cells (Murata and Wolf, 2018). It uses a high energy electron beam to image very thin specimens in a transmission electron microscope. By using a magnetic objective lens, it can produce both the diffraction pattern of the sample, and the

magnified image, which already contains all of the structural information (Wang and Wang, 2017).

The procedure arose out of developments of earlier methods of transmission electron microscopy (TEM) and electron crystallography, which were initially deemed unsuitable for biological specimens, due to their sensitivity to damage from the electron beam radiation, and due to the difficulties in obtaining diffraction quality two dimensional crystals. Limitations of these methods were overcome by the introduction of an approach, now known as 'single-particle EM', which allows structures to be determined by the computational alignment of multiple particle images and subsequent reconstruction of a 3D structure. This process overcame the need for crystallization, however, it was still restricted by the necessity to maintain a native hydrated sample within the ultrahigh vacuum environment of the microscope column (Shen, 2018). Attempts to address this challenge, lead to the discovery that flash freezing the samples under native conditions, could not only prevent the formation of crystalline ice, preserving the material in a vitrified, amorphous state, but also protected the samples from radiation damage. Nonetheless, cryo-EM was widely critiqued due to the need for large symmetrical molecules, and due to the low resolution of structures obtained (typically above 5Å, and up to 80Å, (Lawson *et al.*, 2011; Egelman, 2016)), where even the main chain of the protein was hard to trace (Shen, 2018). It wasn't until 2012, after the innovation of direct electron detectors, that the use of cryo-EM gained momentum. The new detectors were more sensitive than traditional generations of detectors, and the high speed acquisition of data in the form of movies, allowed radiation-damaged components to be filtered out, and the blurred images to be sharpened, overall enabling high resolution structures to be achieved (typically in the region of ~3 Å, but as low as 1.8 Å; (Merk, *et al.*, 2016; Herzik Jr *et al.*, 2017; Khoshouei *et al.*, 2017)).

In addition to its bypassing crystallization, cryo-EM offers the advantage that it can resolve multiple conformational states from a conformationally heterogenous sample. Furthermore, it requires significantly less biological sample to use, compared to either MX or NMR, which often does not require rigorous molecular manipulation (Wang and Wang, 2017). However, although to date the highest resolution achieved by cryo-EM currently stands at 1.8Å (Merk, *et al.*, 2016), only a handful of structures have been determined at a side-chain level resolution, and the majority do not surpass 3Å. Moreover, the observable object size of the specimen also limits the use of cryo-EM to resolve certain structures; where very large macromolecules may be limited by the thickness of the sample that can be penetrated by the electron beam, and very small samples may be restricted by their lower threshold for radiation damage. Currently, the

lower molecular weight limit for cryo-EM stands at 38 kDa (Murata and Wolf, 2018). Furthermore, the low throughput, manual sample preparation and analysis procedures, which often require substantial expertise and time, often make other techniques more favourable, particularly for drug discovery pipelines.

4.1.2 Structural Determination of Nuclear Receptors

Nuclear receptors pose unique challenges with regards to their structure determination, arising from their highly disordered N-terminal regions, highly dynamic domain architecture, and largely hydrophobic ligand binding pockets. To date full length or individual domain structures of NRs have been determined by MX, NMR and cryo-EM, and structural data has been supplemented with the use of additional techniques, such as small angle X-ray scattering (SAXS).

A structure of the full length vitamin D receptor in complex with the DBD and LBD domains of partner receptor RXR, also bound to response element DNA, was achieved by cryo-EM (Orlov *et al.*, 2012). However, the structure determined could only be interpreted in the range of 12-13 Å. Such low resolution meant that it could only provide general information regarding the overall shape of the molecule, and principal characteristics of NRs, such as their DBD-DNA interactions, LBD-ligand interactions, or LBD-coactivator interactions could not be validated. The VDR structure resolved by cryo-EM was susceptible to interpretative errors, and there were inconsistencies between this structure and decades worth of previous literature based on high resolution crystallographic structures (Rastinejad *et al.*, 2015). Whilst cryo-EM certainly holds the potential for study of the entire FXR protein, the relatively low resolution limits its applicability for investigating ligand binding mechanisms in detail.

The majority of NR structures have been elucidated using NMR and MX, and most have focussed on the individual DBD or LBD domains. The size limitations of NMR have meant that it can only be applied to a single, truncated domain, as larger molecular weight complexes display overlapping signals and slower tumbling dynamics (Rastinejad *et al.*, 2013). Early studies were able to elucidate secondary structural information and identify a common fold of NR DBDs of the oestrogen receptor, glucocorticoid receptor and retinoic acid receptor (Hard *et al.*, 1990; Schwabe *et al.*, 1990; Knegtel *et al.*, 1993). However, more comprehensive studies were later carried out by MX, using these domains bound to DNA (Luisi *et al.*, 1991; Schwabe *et al.*, 1993). Few studies have used this technique to investigate the dynamic properties, and obtain a structure of a NR LBD (Lu *et al.*, 2006, 2008; Michiels *et al.*, 2010). Both the apo and holo forms of the LBD have been determined by NMR, and the technique has been successfully used in conjunction with transverse relaxation-optimized spectroscopy to understand the structural

dynamics of the RXR LBD upon ligand binding (Lu *et al.*, 2006). These studies were able to provide some information regarding open helix 3 and 12 conformations in RXR, where crystallographic studies were deficient and limited by the packing forces in the crystal lattice. The structure also was able to demonstrate changes in receptor dimerisation interfaces upon ligand binding. However, the data was limited by overlapping cross peaks, thought to be due to the exchanges between monomeric and multimeric forms of the protein. Additionally, the model contained several unassigned regions, including the loop between helix 6 and 7, and between helix 10 and 11, despite these regions showing contacts with the ligand in crystal structures (Lu *et al.*, 2006). Regardless of the advantages of using NMR, NR models derived from this technique have always been complemented with and confirmed by models derived by MX.

MX has been the primary method of choice for resolving structures of NRs. The technique has been able to successfully determine the multi-domain structures of PPAR γ , HNF4 α and LXR DBD-LBDs in complex with partner protein, RXR, and their respective DNA response elements, ligands and coactivator peptides, at resolutions between 2.8-3.2 Å (Chandra *et al.*, 2008, 2013; Lou *et al.*, 2014). At this resolution, sufficient detail has been provided to unambiguously trace main chain helices and strands, and interaction interfaces between the DBD and DNA, LBD and coregulators have been clearly established. Furthermore, models derived by MX have provided a common template of domain arrangement and architecture for all NRs (Rastinejad *et al.*, 2015).

As common drug targets, there are hundreds of NR LBD structures that have been delivered by MX, including 83 for the FXR LBD. Despite their static nature, data from different crystallographic models served as early indicators of structural plasticity, and comparisons between structures have provided an insight into active and inactive receptor conformations. LBD structures determined by MX have adequate resolution to identify, at an atomic level, the precise residues involved in ligand recognition and binding, and this technique has been widely applied in the drug discovery process.

A significant limitation with NR LBDs is their instability and their inability to be crystallized in a apo state (Rastinejad *et al.*, 2013). Furthermore, their propensity to aggregate, due to their leucine-rich, hydrophobic binding pockets, can hinder the soluble expression of the recombinant protein in *E. coli*, and can cause problems during the purification process (Mossakowska, 1998). Unfortunately, it is not possible to determine which NRs will pose challenges, but the fact that all previously published structures of FXR have been resolved by

MX, using protein that had been heterologously expressed in *E. coli*, suggests that structure determination by this method is attainable.

4.2 Aims and Objectives

The numerous methods of structure determination can help to further the understanding of FXR architecture, structural dynamics and receptor activation. While cryo-EM may help to establish an overall scaffold of the entire FXR protein, the elucidation of specific mechanisms driving the molecular recognition of DNA and ligands, requires atomic-level detail. Size constraints would limit the use of NMR to individual domains, but the technique is well suited to provide dynamic information regarding the conformational rearrangements that the 29kDa LBD may adopt upon binding different ligands and/or coregulators. Nonetheless, MX is thought to be more reliable at producing high-resolution structures, which is of particular importance when considering the detail needed to accurately determine the binding mechanisms of novel ligands. Furthermore, crystallization trials can be set up and screened for diffraction, in a high throughput format, allowing multiple ligand/coactivator combinations, and several crystallization conditions to be explored. MX can also theoretically determine the structure of the full-length protein, and it may certainly be possible to generate a model of FXR's DBD and LBD in complex with DNA response elements, novel ligands and coactivator peptides.

To this end, the aim of the work presented in this chapter is to further the understanding of FXR function and activation, by obtaining a model of its structure. In order to validate the proposed binding mechanisms of novel BA ligands (Chapter 3, section 3.5.4), work will primarily focus on achieving a high resolution structure of the FXR LBD complexed with the top candidate ligands, as well as with poorly performing analogues, in an attempt to identify structural features that may account for differences in activity. In order to obtain these structures, work will attempt to address the challenges associated with the expression, purification and crystallization of NR LBDs, and to produce a sample of protein that is highly concentrated and homogenous, adequate for crystallization. A secondary objective will be to determine the likelihood of expressing other domains of the FXR, either individually or in tandem with another domain.

4.3 Methods

For final optimised methods, refer to Chapter 2.

4.3.1 Sequence of FXR

The 486 amino acid sequence of the full length FXR canonical isoform 1 can be found in Appendix 2. Initial constructs encoding the FXR LBD included parts of the sequence of the hinge region between the DBD and LBD and started at residue 240 (constructs C2-F2). The new construct later designed for the expression of the FXR LBD, started at residue 258, immediately before helix 1. The new construct also contained the mutations E291A and E364A (established through private communication with Markus Rudolph, Roche Pharmaceutical Research and Early Development, Basel, Switzerland). The codon-optimized gene encoding the full length FXR was synthesized by Genscript (Piscataway, NJ, USA) for use in high throughput cloning and expression in pOPIN vectors (kindly provided by the Oxford Protein Production Facility, OPPF, Harwell Research Complex). The sequence encoding the mutant FXR LBD (herein referred to as FXR LBDmut1) was synthesized and subcloned directly into the pET15b vector by Genscript.

4.3.2 High Throughput Cloning and Expression Screens

The HTP cloning and expression screening of 23 different constructs, each encoding a different region of the full length FXR protein in combination with a different *N*- or *C*-terminal tag was carried out at OPPF. The highly specialised robotic equipment and optimised HTP set up at OPPF facilitated the efficient screening of potential constructs and conditions, which could easily produce the target protein. The molecular biology methods related to the cloning of these constructs is outlined in Chapter 2, sections 2.3.1 to 2.3.10.

Screening was carried out using two different *E.coli* expression strains, Lemo 21 (DE3) and Rosetta 2 (DE3) *placI*, and two induction regimes, addition of IPTG and autoinduction. Screening was carried out in a 96-well format for parallel processing. Purified constructs were transformed into both host strains for each construct. A starting seed culture was prepared for each construct in LB medium supplemented with carbenicillin (50 µg/mL) and chloramphenicol (35 µg/mL). Seed cultures were diluted by a factor of 20 for inoculation into 3mL media, Power Broth (Molecular Dimensions) for IPTG induction or Overnight Express™ Instant TB medium (TBONEX, Merck Millipore) for autoinduction, both supplemented with antibiotics as before. Cells were initially grown at 37°C and induction was initiated once the cells had reached an OD_{600nm} ~0.5. Power broth cultures were supplemented with IPTG (to a final concentration 1 mM) and grown for a further 20 hours at 20°C. For autoinduced cultures, protein expression was initiated by reduction of temperature to 25°C and grown for a further 20 hours.

Aliquots (1 mL) of the cultures were made and these cells harvested by centrifugation at 6000 $\times g$ for 10 minutes (Beckman Avanti centrifuge, rotor JS5.3). The supernatant was removed and cell pellets stored at -80°C for 20 minutes or until required. The cell pellets were resuspended completely in lysis buffer (50 mM sodium phosphate, 300 mM sodium chloride, 10 mM Imidazole, 1% Tween 20, pH 8.0, Qiagen) supplemented with 1 mg/mL lysozyme and 400 U/mL DNase I, using a micro titre orbital plate shaker for 30 minutes at 1000 rpm. The crude lysate was clarified by centrifugation at 6000 $\times g$ for 30 minutes at 4°C, and supernatants were transferred to a new 96 well plate where they were mixed with Ni-NTA magnetic agarose bead suspension (Qiagen) by shaking at 1000 rpm for 30 minutes at room temperature. The plate was transferred to a 96-well magnet to remove unbound proteins. The beads were washed twice with wash buffer (50 mM sodium phosphate, 300 mM sodium chloride, 20 mM imidazole, 0.05% Tween 20, pH 8.0, Qiagen) and proteins bound to the Ni-NTA beads eluted in elution buffer (50 mM sodium phosphate, 300 mM sodium chloride, 250 mM imidazole, 0.05% Tween 20, pH 8.0, Qiagen).

Purified soluble proteins were analysed by SDS-PAGE as outlined in Chapter 2, section 2.5.2, and run alongside a wide molecular weight range SigmaMarker in the first lane, and a low molecular weight range SigmaMarker, in the last lane.

4.3.3 Optimization and Scale up of His₆-FXR LBD Expression

Selected constructs were transformed by the heat shock method into Rosetta (DE3) pLysS or BL21 (DE3) pLysS, as described in Chapter 2, section 2.3.7. Single colonies were used to create starter cultures, which were in turn used for inoculation into larger scale expression cultures or stored as glycerol stocks. In cases where glycerol stocks were used for starter cultures, a loopful of frozen cells were scraped from the stock and immersed into 10-50 mL LB broth. Starter cultures were incubated overnight at 37°C, shaking at 220 rpm. Cultures were scaled up to 1 L, 4 L or 8 L simply by adding more flasks to the batch. Cultures were grown in 2.5 L baffled flasks with a 500 mL volume of LB supplemented with carbenicillin (50 µg/mL) and chloramphenicol (35 µg/mL). A 1 in 50 dilution was used to inoculate flasks, and all flasks were seeded with the same starter cultures. Expression was tested following induction with 0.1 mM, 0.5 mM, 0.7 mM and 1 mM IPTG. Post induction temperatures were tested at 25°C, 20°C and 15°C. After overnight incubation, cells were harvested by centrifugation at (5,000 $\times g$ for 10 minutes), and cell pellets from each flask were pooled. For cultures larger than 4 L, cells were grown in 4 L batches and pooled during purification.

4.3.4 Additional Purification by Anion Exchange Chromatography (AEC)

Another purification step was added in order to separate non-specific proteins from the similar sized target protein when separation could not be achieved through size exclusion chromatography alone. At pH 8, the His₆-tagged FXR-LBD is negatively charged, so a positively charged Sepharose Q based, HiTrapQ column (GE Healthcare Life Sciences) was used to separate proteins based on charge. Prior to loading on the column, the protein solution was diluted in a low salt concentration buffer (50 mM Tris pH8, 50 mM sodium chloride, 10% glycerol) by dialysis. As the protein is loaded on to the column, the negatively charged target protein will interact electrostatically to the positively charged resin, binding this protein to the column for longer. The target protein is then eluted by gradually increasing the salt concentration on the column, whereby the protein is displaced by the increasing number of chloride (Cl⁻) ions. This step was not included in the final purification protocol.

4.3.5 Degradation Tests

A sample of His₆-FXR LBD was incubated at room temperature 1 day post IMAC purification for a further 3 days. Aliquots were taken at various time intervals and immediately mixed with 4× Laemmli buffer and denatured. Samples were stored at -21°C until all aliquots had been collected. Samples were then analysed on an SDS-PAGE gel as outlined in Chapter 2, section 2.5.1.

4.3.6 Thrombin Cleavage

Initial thrombin cleavage tests were carried out on a sample of His₆-FXR LBD immediately following IMAC purification. The concentration of the protein in the sample was determined by Nanodrop spectrophotometry. Thrombin from human plasma (Sigma) was dissolved in cold phosphate buffer solution to a stock solution of 1 U/μL. Thrombin was added to the protein sample at a concentration of 10 U/mg, and the reaction was mixed gently by inversion. The sample was split into two 3 mL aliquots and each was added to a 3 mL GeBAflex 10 kDa molecular weight cut off (MWCO) dialysis tube and dialysed against 500 mL gel filtration buffer (20 mM tris pH 7.8, 200 mM sodium chloride, 1 mM DTT, 10% (v/v) glycerol). Dialysis was carried out overnight and aliquots were taken at various intervals and immediately mixed with 4× Laemmli buffer and denatured. After 22 hours, the cleavage reaction was inhibited by the addition of 100 mM phenylmethanesulfonyl fluoride (PMSF). The final thrombin cleavage protocol used reaction conditions at 4°C and for 22 hours.

Cleaved FXR-LBD was separated from contaminant proteins by a second IMAC purification, as outlined in Chapter 2, section 2.4.2, except the flow through fraction, containing the purified target protein was collected. The eluted proteins were collected for analysis only.

4.3.7 Concentration

Following a final purification by SEC (Chapter 2, section 2.4.3), ligands and coactivator peptides were added to the dilute protein. Ligand compounds had been dissolved in 100% DMSO to a stock concentration of 20 mM. Ligands were added to the FXR LBDmut1 at final 12× molar concentration, and final DMSO concentration in the sample was less than 1%. SRC2-2 coactivator peptide (KHKILHRLLQDSS) was custom synthesized by Lifetein (New Jersey, USA) and lyophilized peptide dissolved in water, due to the overall basic polarity of the sequence. SRC2-2 peptide also was added to a final 12× molar concentration of the FXR LBDmut1 protein. The protein-ligand-coactivator complex was incubated overnight at 4°C. Following incubation, the sample was divided between two 15 mL Amicon Ultra-15 centrifugal devices with a 10 kDa MWCO (Merck). The protein was concentration by centrifuging at 2,500 × *g*, at 4°C for 5 minutes at a time. Retained protein near the crevice of the membrane was resuspended by gently pipetting after each spin. When the volume had reduced to ~1500 µL, the retentate was transferred to two 0.5 mL Amicon Ultra-0.5 microcentrifugal devices (10 kDa MWCO) and concentrated further. Final retained proteins were removed by aspiration with a pipette and pooled together. The final concentration of the protein complex was estimated by absorbance at 280 nm, using a Nanodrop spectrophotometer.

4.3.8 Crystallization Screening

Preliminary screening for initial crystallization conditions was performed by sitting drop vapour diffusion method using the high-throughput robotic crystallization facility at OPPF-UK, Research Complex at Harwell. The Nuclear Receptor (Billas *et al.*, 2001), PACT Premier (Newman *et al.*, 2005), JCSG+ (Newman *et al.*, 2005) and Proplex screens (all Molecular Dimensions) were used to test several different buffers, salts and precipitants at different concentrations and pH values to identify potential conditions for crystal growth. Screening was carried out at 2 different ratios in 96 well Crystal Quick X plates, using 200 nL droplets consisting of, either 100 nL protein and 100 nL well solution, or 50 nL protein with 150 nL well solution. The reservoir solutions were dispensed by a Hydra 96 robot (Robbins Scientific) and small volume droplets, by a Mosquito LCP robot (TTP labtech) at 54% humidity; the plates were transported for droplet visualization at regular intervals using automated systems. All crystallizations were set up and stored at 295K.

Regions of interest were identified and plates were scanned in 2 µm grids over these regions at the VMXi beamline ($\lambda = 0.95\text{\AA}$), which collects data *in situ*, without the need to harvest crystals from the plate or manipulate the plates in any way.

4.3.9 Densitometry

Densitometry to quantify expression and purification of different bands was carried out using ImageJ software (NIH, Maryland, USA). Gel images scanned and saved using the Image Scanner III Gel Scanner (GE Healthcare) were uploaded into the Image J software. After selecting the relevant bands, peaks were plotted corresponding to the respective intensity of the band. Quantification was carried out by the automatic measurement of the area under each peak.

4.4 Results

4.4.1 High Throughput Cloning and Small-Scale Expression Screens

4.4.1.1 Rational Construct Design

The NR1H4 gene encoding the canonical isoform 1(+) of FXR (Uniprot Q96RI1-3) was synthesised by GenScript (Piscataway, NJ, USA) and codon optimized for expression in *E. coli* systems. Multiple constructs were designed that contained different domains of the full-length FXR protein in various vectors (pOPIN suite, OPPF-UK), each expressing a different fusion tag, maximising the chances of soluble expression (Table 4. 1, Appendix 3). To design these novel constructs, the Pfam protein families database was used alongside the results from the structure prediction tool, IntFOLD (Figure 3. 2), for the previously undescribed structures of the *N*-terminal domain and DBD of FXR. As expected, the Pfam results, which are generated by aligning the primary protein sequence with previously determined domain families, suggested a C-terminal nuclear receptor LBD (303-486aa, PF00104) connected via a hinge region to a more central zinc finger containing domain (150-219aa, PF00105). The *N*-terminal domain, again, did not match any common folding motifs and was predicted to be highly disordered. The amino acid sequence of the proposed domains also were submitted to the ExPASy ProtParam tool (Gasteiger *et al.*, 2005), for the prediction of chemical and physical parameters of each respective region, and to the SEquence-based CRystallizability EvaluaTor (SECRET) online tool (Smialowski *et al.*, 2006), to predict if the final protein of interest is likely to crystallize. Results suggested that the FXR LBD constructs should yield soluble protein that is able to crystallize, whereas the other domains were less likely to do so. The predicted domain boundaries of the *N*-terminal region and DBD, as well as the protein parameter prediction tools, allowed constructs to be designed to include the most appropriate amino acid residues, which could yield a folded, stable protein structure.

Table 4. 1 List of constructs designed for expression of FXR domains with different fusion tags.

Construct No. (Well position)	pOPIN Vector	Insert (amino acid residues)	Expression Product	Expected Molecular Weight (kDa)
A1	pOPINE	Full Length FXR (1-486)	Full FXR-KHis ₆	56
B1	pOPINF	Full Length FXR (1-486)	His ₆ -3C-Full FXR	56
C1	pOPINS3C	Full Length FXR (1-486)	His ₆ -SUMO-3C-Full FXR	67
D1	pOPINHALO7	Full Length FXR (1-486)	His ₆ -HALO7-3C-Full FXR	88

E1	pOPINJ	Full Length FXR (1-486)	His ₆ -GST-3C-Full FXR	81
F1	pOPINE	DNA and Ligand Binding Domain (134-483)	DBD+LBD-KHis ₆	41
G1	pOPINF	DNA and Ligand Binding Domain (134-483)	His ₆ -3C-DBD+LBD	41
H1	pOPINS3C	DNA and Ligand Binding Domain (134-483)	His ₆ -SUMO-3C-DBD+LBD	52
A2	pOPINHALO7	DNA and Ligand Binding Domain (134-483)	His ₆ -HALO7-3C-DBD+LBD	73
B2	pOPINJ	DNA and Ligand Binding Domain (134-483)	His ₆ -GST-3C-DBD+LBD	66
C2	pOPINE	Ligand Binding Domain (240-483)	LBD-KHis ₆	29
D2	pOPINF	Ligand Binding Domain (240-483)	His ₆ -3C-LBD	29
E2	pOPINS3C	Ligand Binding Domain (240-483)	His ₆ -SUMO-3C-LBD	40
F2	pOPINJ	Ligand Binding Domain (240-483)	His ₆ -GST-3C-LBD	54
G2	pOPINE	<i>N</i> -terminal region (1-136)	<i>N</i> region-KHis ₆	16
H2	pOPINF	<i>N</i> -terminal region (1-136)	His ₆ -3C- <i>N</i> region	16
A3	pOPINS3C	<i>N</i> -terminal region (1-136)	His ₆ -SUMO-3C- <i>N</i> region	27
B3	pOPINJ	<i>N</i> -terminal region (1-136)	His ₆ -GST-3C- <i>N</i> region	41
C3	pOPINE	<i>N</i> -terminal and DNA binding domain (1-239)	<i>N</i> region+DBD-KHis ₆	29
D3	pOPINF	<i>N</i> -terminal and DNA binding domain (1-239)	His ₆ -3C- <i>N</i> region+DBD	29
E3	pOPINS3C	<i>N</i> -terminal and DNA binding domain (1-239)	His ₆ -SUMO-3C- <i>N</i> region+DBD	40
F3	pOPINHALO7	<i>N</i> -terminal and DNA binding domain (1-239)	His ₆ -HALO7-3C- <i>N</i> region+DBD	61
G3	pOPINJ	<i>N</i> -terminal and DNA binding domain (1-239)	His ₆ -GST-3C- <i>N</i> region+DBD	54
H3	pOPINE-3C-eGFP	-	GFP (+ve control)	27

* Where His₆ is 6 histidine residue tag (1kDa), 3C is a 3C protease cleavage site, SUMO is a solubility tag (12kDa), HALO7 is a fluorophore labelling and purification tag (33kDa) and GST is a glutathione tag (26kDa).

4.4.1.2 Cloning

For each construct, insert DNA corresponding to the respective domain, was amplified from the full length NR1H4 gene that had been synthesised by Genscript (Piscataway, NJ, USA). The primers used (Appendix 1), were not only complementary to the insert gene, but also contained

a 15 bp overlap complementary to their respective vector, in order to facilitate cloning by the In-Fusion® method. Successfully amplified PCR products were purified from both template DNA and reaction reagents, followed by their insertion into their appropriate, linearized vectors. The In-Fusion reaction mixtures were subsequently used for transformations into a cloning-grade *E. coli* strain, OmniMaxII (Invitrogen) and tested for successful target gene insertion by blue/white screening. The α -complementation mechanism underlying blue/white screening, results in the presence of white colonies that contain a successfully rejoined vector and insert, due to the prevention of a functional β -galactosidase enzyme being formed; blue colonies contain a vector that was not efficiently linearised and therefore not recombinant. Around 90-100% of colonies were expected to be white (Berrow *et al.*, 2009), and due to the highly efficient In-Fusion method, white clones were indeed predominant for each construct.

Colonies for each construct were then used to inoculate LB and small scale (1 mL) cultures were grown in a 96-well format. After sufficient growth overnight, cultures were harvested, and the recombinant construct plasmids extracted from their host *E. coli* cells by plasmid mini preparation. Final constructs were re-verified by PCR screening, utilizing a vector specific and insert specific primer. Visualisation of the amplicons on an agarose gel revealed the presence of a single, dominant product in every lane except for those corresponding to constructs A2 and D2, which had multiple bands (Figure 4. 2). Colonies were re-picked for the A2 and D2 constructs and the verification process was repeated. All amplicon sizes were as expected, suggesting the successful insertion of the target DNA into the appropriate vector.

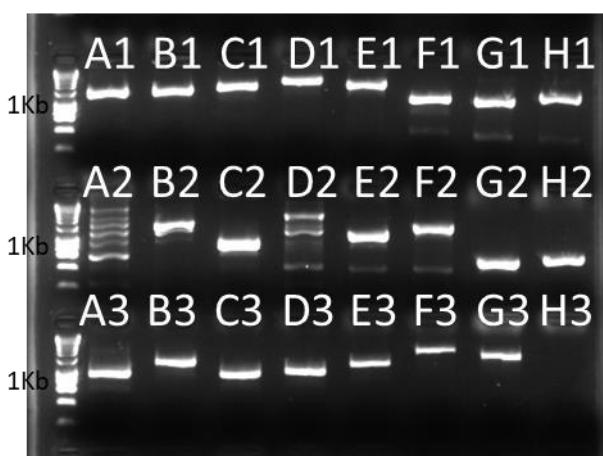


Figure 4. 2 High throughput PCR verification of cloned constructs. PCR products run on a 1.5% agarose, 1xTAE gel. Lane 1 of each row is the molecular weight marker, Hyperladder (1kb) and 1kb position indicated. Rest of lanes labelled with their construct number.

4.4.1.3 Protein Expression and Purification

The highly specialised and robotic technology at OPPF-UK (Research Complex at Harwell) allowed these constructs to be quickly and conveniently transformed into two different engineered *E. coli* strains, Lemo 21(DE3) and Rosetta 2(DE3) *placI*, and expression was tested

using two different regimes, either growth on Power Broth with induction by the addition of IPTG, or growth on auto-induction medium. Following expression, proteins were purified by Ni-NTA column purification, and analysis of which constructs and conditions favoured expression was facilitated by SDS-PAGE (Figure 4. 3). The results indicated that Rosetta (DE3) appeared to be the more successful strain at expressing the desired proteins (Figure 4. 3B and D). However, constructs A2 and C2-H2, appeared not to have expressed any proteins in Lemo cells for either induction regime, suggesting that there was an error in the original transformations of these constructs into the expression host, rather than reflecting failure of the expression conditions. Regardless of expression strain and induction conditions, the full length FXR, N-terminal region, and N-terminal domain linked to the DBD did not appear to express easily, as only non-specific and low molecular weight bands, suggesting proteolysis and cleaved fusion tags, are seen in these lanes (A1-E1, G2-G3, Figure 4. 3). It is possible that these protein constructs are expressed, but insoluble and further work would need to be carried out to determine this and to optimise expression. Meanwhile, the LBD appears to have been successfully expressed in Rosetta 2 (DE3) cells under both induction conditions, as designated by the red arrows, with the *N*- and *C*-terminally His₆-tagged variants (constructs D2 and C2) displaying highest levels of expression (Figure 4. 3B and D, lanes C2 and D2). The screen also suggested that the DBD, when linked to the LBD, may also be expressed, albeit at low levels. In both Lemo 21 (DE3) and Rosetta 2 (DE3) cells, various constructs containing the DNA-LBD insert, display a detectable band at the size expected for this construct, also designated by red arrows; these bands do not correspond to the sizes of those suspected to be degraded proteins or cleaved fusion tags (Figure 4. 3B, C and D, lanes F1-H1 and B2).

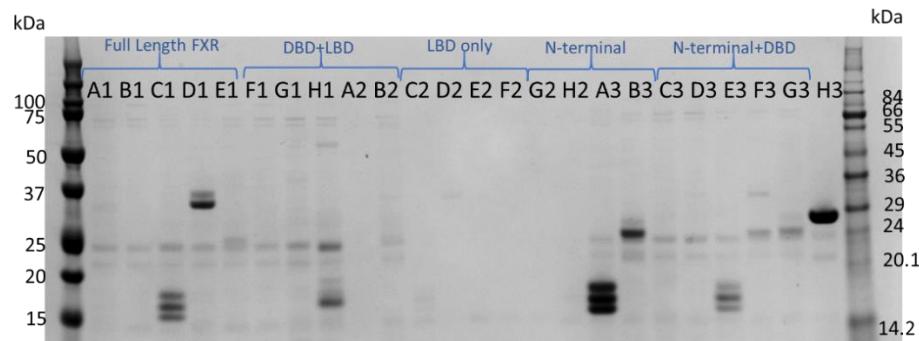
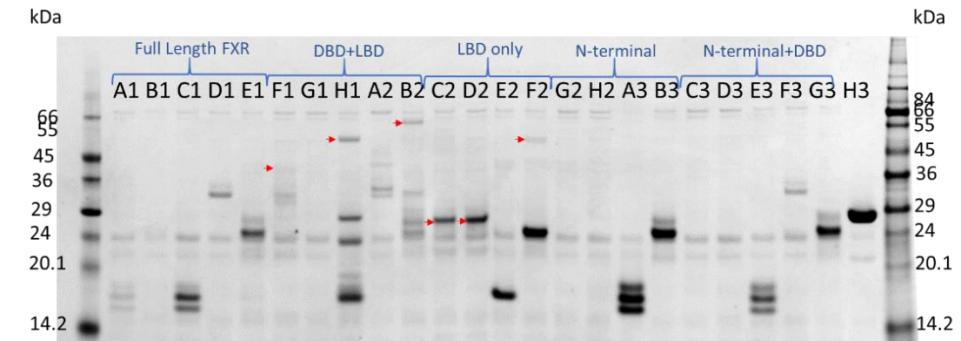
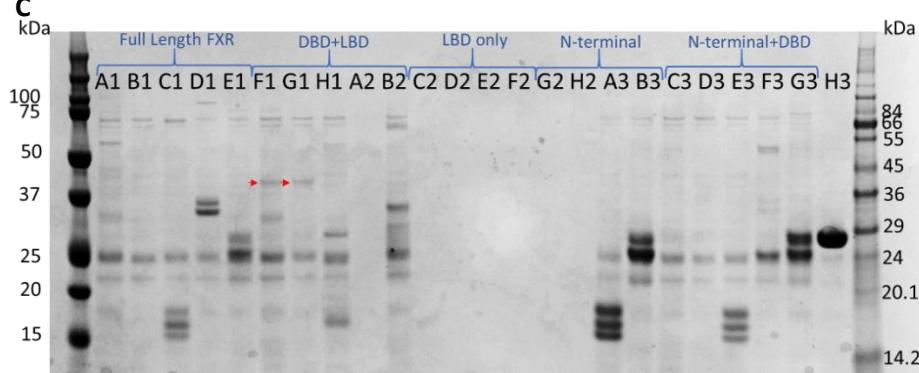
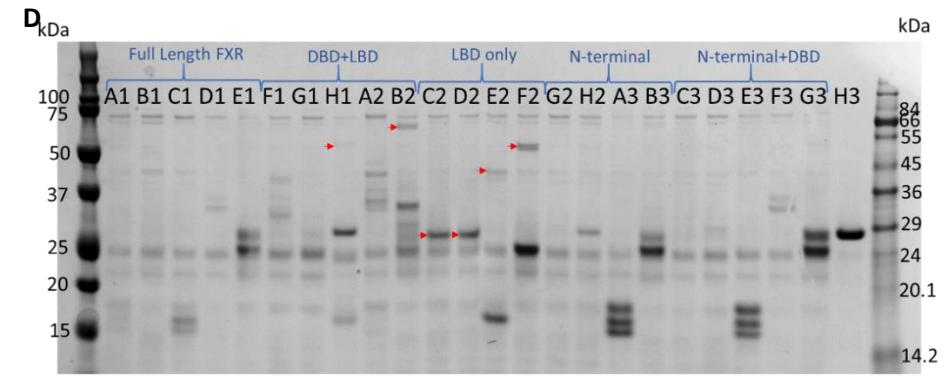
A**B****C****D**

Figure 4. 3 SDS-PAGE Analysis of Ni²⁺-NTA purified proteins from constructs expressed in **A)** Lemo 21 (DE3), grown in Power broth and induced with IPTG, **B)** Rosetta (DE3), grown in power broth and induced with IPTG, **C)** Lemo 21 (DE3), grown and induced in Autoinduction medium, **D)** Rosetta 2 (DE3), grown and induced in Autoinduction medium. Lane labels refer to construct number. Expected band sizes (kDa) for constructs: **A1, B1**- 56; **C1**- 67; **D1**- 88; **E1**- 81; **F1, G1**- 41; **H1**- 52; **A2**- 73; **B2**- 66; **C2, D2**- 29; **E2**- 40; **F2**- 54; **G2, H2**- 16; **A3**- 27; **B3**- 41; **C3, D3**- 29; **E3**- 40; **F3**- 61; **G3**- 54; **H3**- 27. Constructs **A1-E1** express full length FXR, **F1-B2** express the FXR DBD+LBD, **C2-F2** express the FXR LBD, **G2-B3** express the FXR N-terminal and **C3-G3** express the FXR N-terminal + DBD. Construct **H3** was the positive control, expressing a 27kDa protein. Small red arrows indicate potential target protein expression. Proteins separated on a 4-12% NuPAGE Bis-Tris gel (Invitrogen) and stained with Quick Coomassie (Generon).

4.4.2 Scale-up and Yield Optimization

With the aim of solving the structure of the FXR LBD in complex with lead compounds, Jed441 and Jed561, work was directed towards obtaining a highly concentrated, soluble, homogenous sample of FXR LBD sufficient for crystallization. Due to the highly flexible nature of the C-terminal helix 12, and its importance in determining receptor activation, the *N*-terminally His₆-tagged variant, construct D2 (pOPINF-FXRLBD), was used henceforth. Using the D2 construct that had been transformed into the Rosetta (DE3), cultures were initially scaled up to 500 mL volumes, and target protein expression was induced by the manual addition of 0.5 mM IPTG. Cells were harvested following ~16 hours incubation at 25°C post-induction. Rosetta cells were lysed by sonification and purified on a Ni-NTA column (as described in Chapter 2, section 2.4.2). During purification of the His₆-FXR LBD target, it was quickly identified that the amount of protein generated from this scale culture, would not be adequate (less than 1 mg of protein/L culture). The chromatogram produced during this purification did not contain a peak, as expected, for the elution of proteins from the column; instead plateaued at around 350 milli-absorbance units (mAU), which likely represented the presence of high concentrations of imidazole. Furthermore, although there was a band at the appropriate size (29kDa) on the SDS-PAGE gel, this protein was not significantly more distinct than the other non-specific bands (Figure 4. 4).

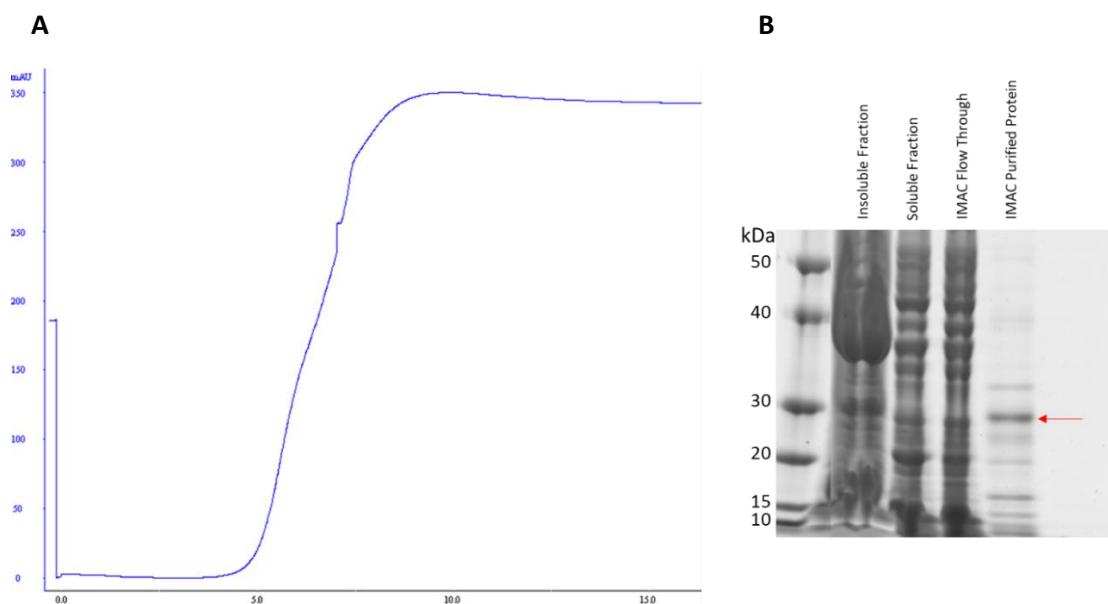


Figure 4. 4 Scale up and purification of FXR-LBD. Expression in 500mL Rosetta (DE3) cultures, induction by 0.5mM IPTG, lysis by sonication and purification by Ni-NTA column. (A) IMAC purification chromatogram displaying UV (280nm) absorbance (blue trace) in mAU. (B) SDS-PAGE of purification fractions separated on a 4-12% Bis-Tris NuPAGE gel (Invitrogen) and stained with Quick Coomassie (Generon). Target protein indicated by red arrow.

In addition to the simple scale up of the *E. coli* cultures, other factors affecting the yield of purified protein obtained also were investigated. These included the concentration of IPTG, the temperature and incubation time post-induction, and the lysis procedure. Following optimization, a final protocol was established whereby 4 L cultures were induced with 0.7 mM IPTG and then incubated at ~18°C for 20 hours. Cells were harvested and stored at -80°C for at least 2 hours. Cells were then thawed, initially lysed in buffers containing 0.5 mg lysozyme per mL lysate, and subsequently lysed using a cell homogenizer (Stanstead Fluid Power). This protocol generated significantly increased amounts of target protein, which was observed by the 2000 mAU peak on the IMAC purification chromatogram, and by the corresponding thick ~29 kDa band on the SDS PAGE gel (Figure 4. 5). The use of the nanodrop spectrometer to measure the concentration of this purified protein fraction (also accounting for the extinction coefficient estimated by the ProParam tool), determined that an approximate total of 56 mg of purified protein had been produced from 4 L of cell culture. This concentration should be sufficient for the setup of crystallization screens, however, the presence of non-specific proteins, imidazole and high salts in this fraction, meant that additional purification steps were required.

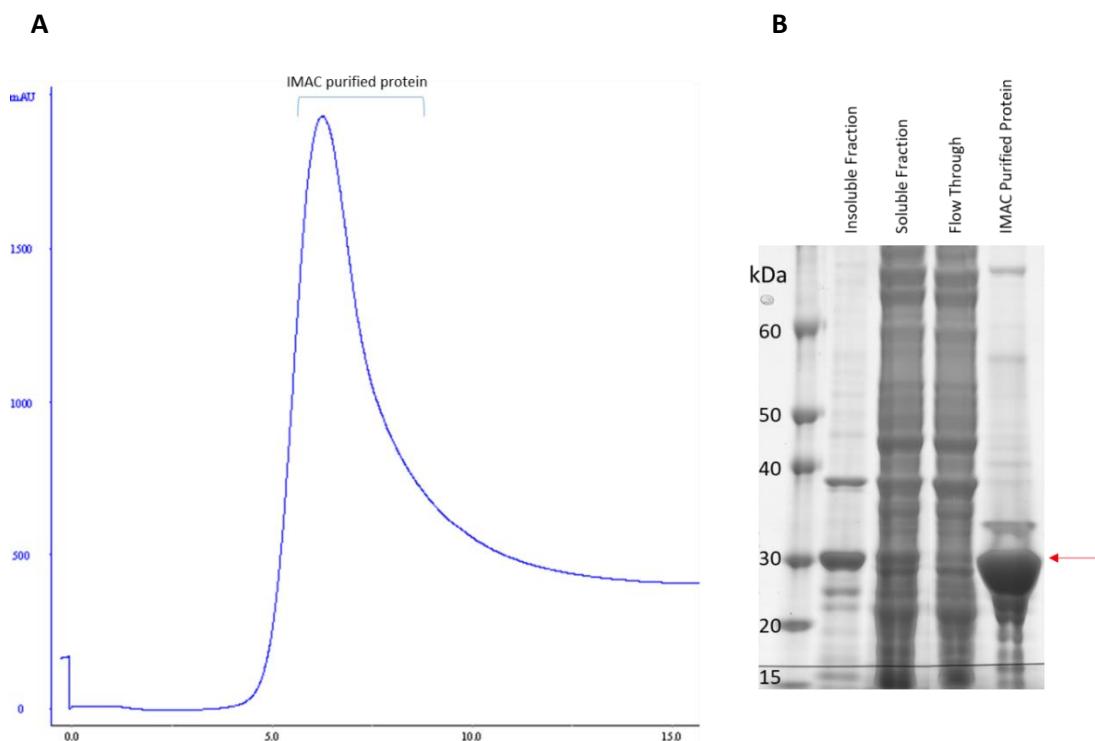


Figure 4. 5 Further scale up and purification of FXR-LBD. Expression in 4L Rosetta (DE3) cultures, induction by 0.7mM IPTG, lysis by cell disruption and purification by Ni-NTA column. (A) IMAC purification chromatogram displaying UV (280nm) absorbance (blue trace) in mAU. (B) SDS-PAGE of purification fractions separated on a 4-12% Bis-Tris NuPAGE gel (Invitrogen) and stained with Quick Coomassie (Generon). Target protein indicated by red arrow.

4.4.3 Removal of Contaminants

For the preparation of crystallization-grade protein samples, purification by size exclusion chromatography (SEC) was applied, following IMAC purification. The cross-linked agarose resin facilitates the separation of proteins according to their size. Pre-equilibration of the column, using standards of known molecular weights (Sigma), indicated that carbonic anhydrase, which is a 29kDa protein, eluted off the column at a retention volume of 235 mL. The application of the IMAC purified His₆-FXR LBD sample to the SEC column resulted in one wide, symmetrical peak between 210 and 230 mLs, and one shouldered peak at 320 mL. Whilst the shouldered peak likely corresponded to imidazole and other salts being eluted after one total column volume of buffer, verification of the fractions collected between 210 and 230 mLs by SDS PAGE, confirmed that this peak corresponded to the 29kDa His₆-FXR LBD (Figure 4. 6). However, it also showed that, due to their similar sizes, purification by SEC was unable to separate contaminant proteins from the target; although at a lower intensity, these bands appeared in every lane with the His₆-FXR LBD, indicating that the peak seen on the chromatogram was actually a conglomerate of multiple single peaks. Even when the flow rate was decreased, complete resolution of a single band was not achieved.

In a further endeavour to separate the co-purified bands from the suspected His₆-FXR LBD, purification by ion exchange chromatography was attempted. The isoelectric point (pI) of His₆-FXR LBD was estimated to be 5.9, and so at the working pH of 7.8, the protein was expected to have an overall net negative net charge, allowing it to bind to the positively charged anion column matrix, whilst other, less negative proteins were expected to flow through the column unbound. The addition of increasing concentrations of salt to the column causes the competitive binding of the negative salt ions present in the buffer, to the electrically charged exchangers covalently bound to the resin, displacing any bound proteins. However, separation of the non-specific proteins by anion exchange was unsuccessful. Dialysis of the His₆-FXR LBD into an appropriate starting buffer with a low salt concentration often resulted in the precipitation of the protein, and losses to the overall protein yield. However, when higher salt concentrations were used, although precipitation was prevented, the increased salt ions in the buffer interfered with the electrostatic interactions between the target protein and the column, resulting in insufficient binding of the protein and the subsequent presence of His₆-FXR LBD in the flow through (Figure 4. 7B), again resulting in losses to the overall protein yield. The salt-gradient elution of proteins from the column produced 4 merged peaks (Figure 4. 7A). SDS-PAGE analysis of these peaks revealed that all contained the target protein, but all fractions also contained the contaminant proteins.

It was initially thought that the non-specific bands seen in purified His₆-FXR LBD samples were products of the degradation or cleavage of the target protein. In order to determine this, the elution fraction following an IMAC purification of His₆-FXR LBD, was incubated at room temperature over a 4 day period, with samples taken at regular intervals. Analysis of these samples on a SDS-PAGE gel showed that, whilst some of the higher molecular weight polypeptide bands 'disappear' over time, indicative of degradation, the presence of the target protein was detected across the entire time course measured (Figure 4. 8). Moreover, the staining intensity of the band remained unchanged, suggesting that the protein was relatively stable. The low molecular weight doublets were observed in the very first sample, and no unique bands were introduced at later time points, suggesting that these non-specific bands in fact were not degradation products.

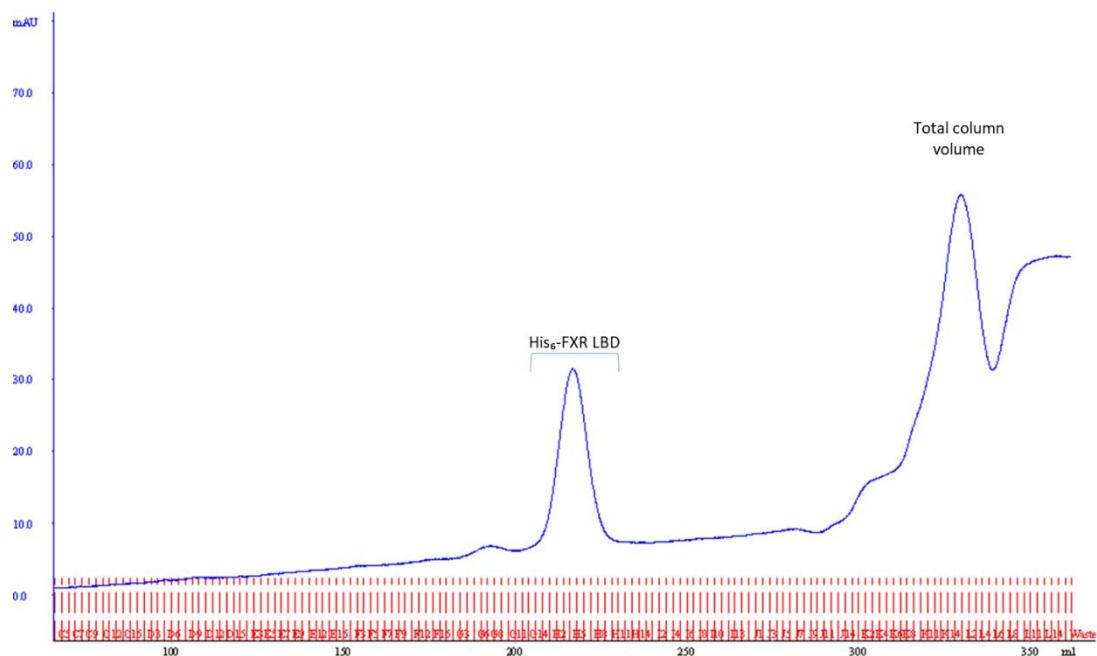
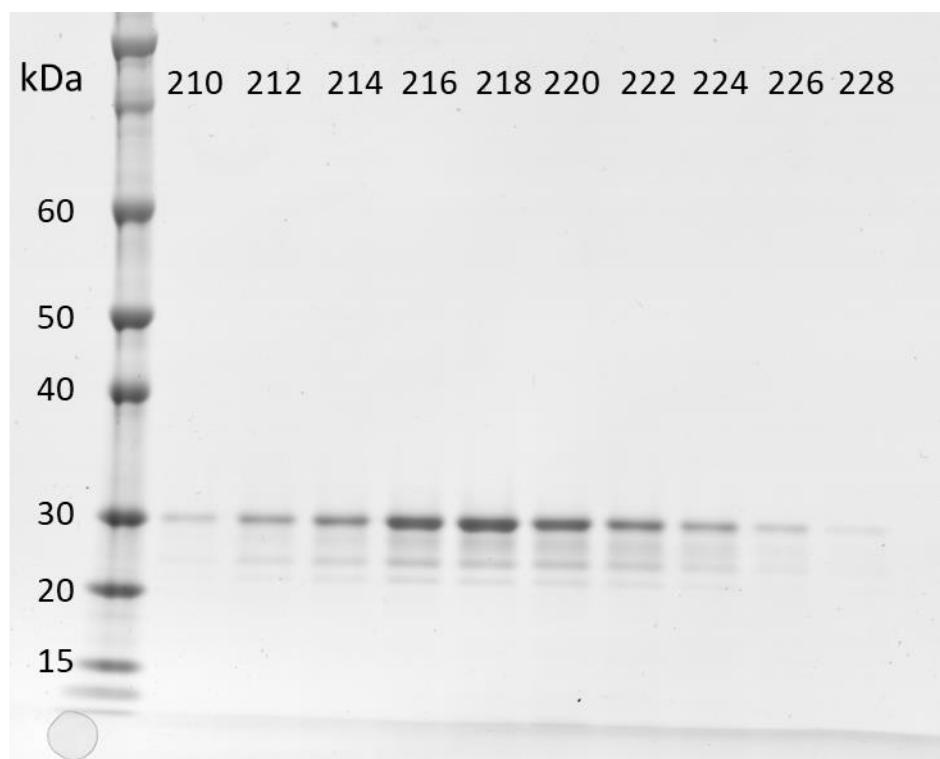
A**B**

Figure 4. 6 Size Exclusion Chromatography of His₆ -FXR LBD. Purifications were carried out on a HiLOAD 26/600 Superdex 200 prep grade column. Sample was loaded using a 5mL injection loop and flow rate set to 1mL/min for 1.5CV with 2mL fractions collected. (A) SEC purification chromatogram displaying UV (280nm) absorbance (blue trace) in mAU. (B) SDS-PAGE of SEC purification fractions (210-228mL) separated on a 4-12% Bis-Tris NuPAGE gel (Invitrogen) and stained with Quick Coomassie (Generon).

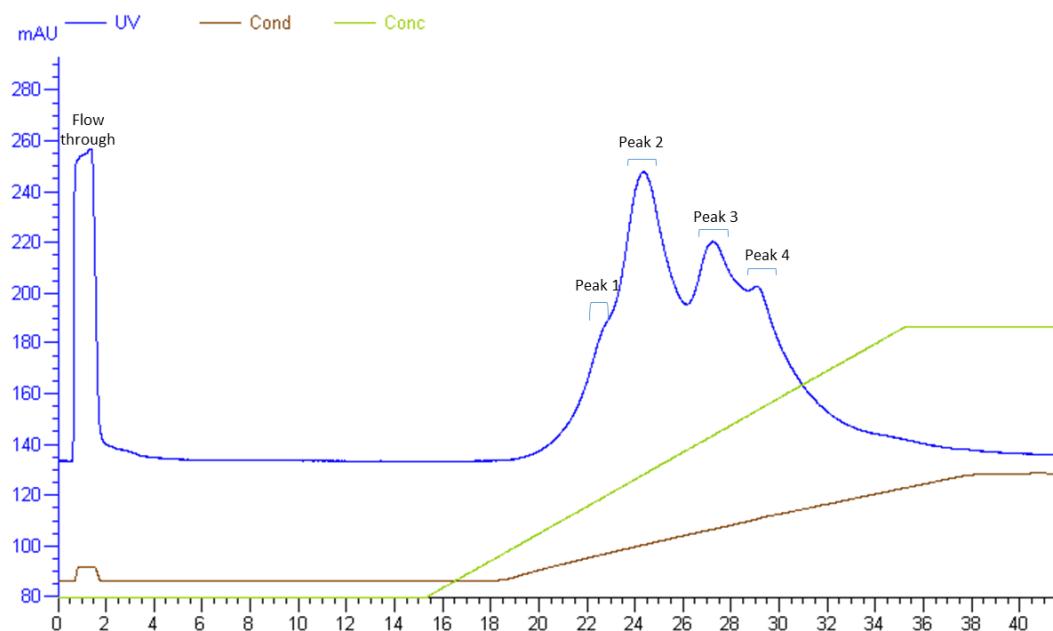
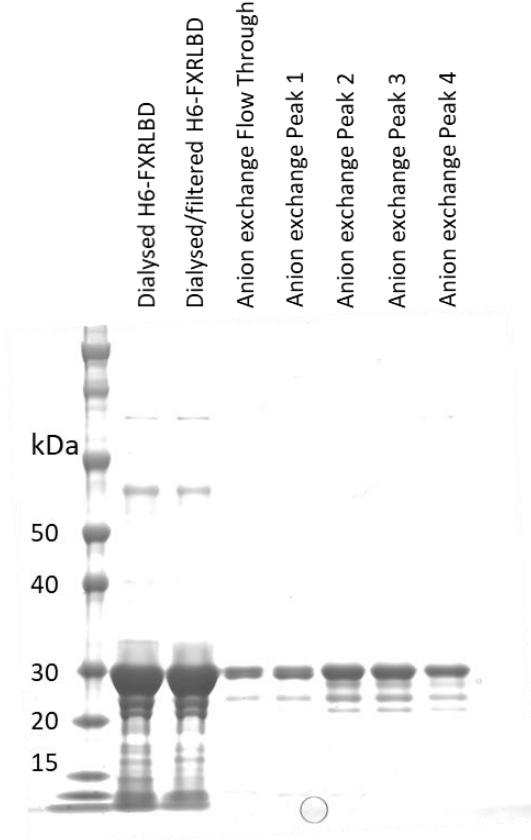
A**B**

Figure 4.7 Anion exchange chromatography of His₆-FXR LBD. Following dialysis into a low salt start buffer, His₆-FXR LBD was filtered and loaded on to a 1mL HiTrapQ column. Unbound proteins were washed off with low salt buffer and bound proteins were eluted with increasing concentrations of NaCl, up to 1M. (A) AEC purification chromatogram displaying UV (280nm) absorbance (blue trace) in mAU, programed salt gradient (green) and the actual changes in conductivity (brown). (B) SDS-PAGE of AEC purification fractions separated on a 4-12% Bis-Tris NuPAGE gel (Invitrogen) and stained with Quick Coomassie (Generon).

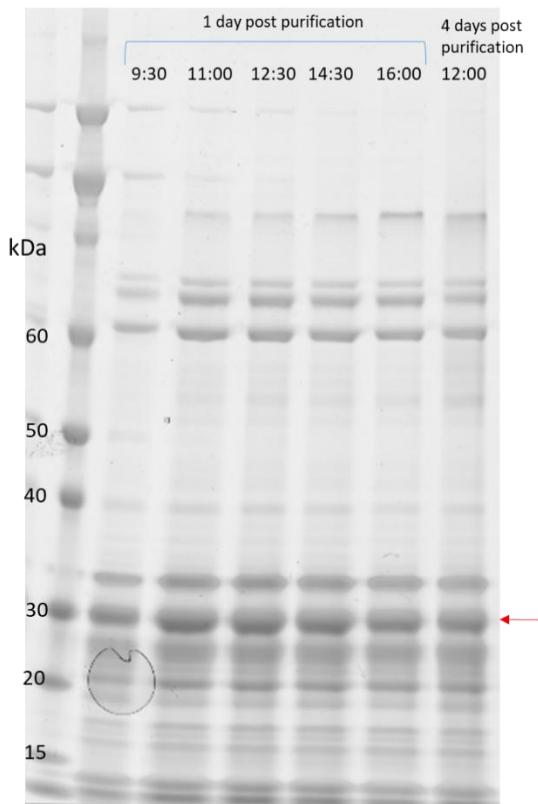


Figure 4. 8 SDS PAGE analysis of protein degradation. 1 day post-purification, an IMAC elution fraction was incubated at room temperature and samples taken at regular time intervals to observe any changes in protein profiles. Proteins were separated on a 4-12% Bis-Tris NuPAGE gel (Invitrogen) and stained with Quick Coomassie (Generon). Target protein indicated by red arrow.

4.4.4 Challenges with Protein Precipitation

A reoccurring challenge with the His₆-FXR LBD protein, was its tendency to precipitate. Time-dependent precipitation occurred when the protein was stored at room temperature or 4°C for more than 15 hours, despite the fact that all buffers used were at pH values deemed to be sufficiently different from the pI. The FXR LBD contained 2 surface exposed free cysteines, however, this precipitation phenomenon was not corrected even in the presence of 10% glycerol and 1 mM dithiothreitol (DTT), which acts to reduce free monothiols, preventing non-native disulphide bond formation. Procedures, such as dialysis into low salt concentration (less than 150 mM sodium chloride) buffers, exacerbated the precipitation. Likewise, when trying to concentrate the protein by itself in a centrifugal spin column, it was prone to precipitation, even in after small reductions in solvent volume, and in concentrations as low as 2.5 mg/mL protein. Several different centrifugal devices with different membranes and surface areas were tested, and the temperature at which the concentration was carried out was also varied. Although precipitation was still observed in all procedures, the percentage of total protein lost, was less when using the Amicon Ultra devices (Merck), and when the concentration procedure was

employed at 4°C. Furthermore, it was observed that the most effective way to stabilise the protein and prevent precipitation, was to add a ligand prior to its concentration.

4.4.5 Design of a New Construct

Although expression of the His₆-FXR LBD was initially easy and straightforward, a significant bottleneck occurred in the large scale production and purification of the protein, with several challenges being presented. At best, 5 mg/mL of His₆-FXR LBD complexed with OCA was achieved, and despite showing promising results in initial pre-crystallization trials, designed to indicate optimal protein concentrations for crystallization screening, screens with the Nuclear Receptor LBD (Billas *et al.*, 2001) and the PACT premier screen (Newman *et al.*, 2005), but did not yield any crystals.

The initial D2 construct was designed to express several residues of the flexible hinge region in addition to the defined LBD of FXR. It was later thought that the inclusion of these flexible residues in the recombinant protein may be inhibiting the formation of ordered protein molecules required for crystal formation. As such, a new construct was designed to only include the more rigid LBD, starting at residues directly before the first alpha helix. Similarly to the group at Roche Pharmaceutical Research Centre, who were able to publish 36 high-resolution structures of the FXR LBD (Gaieb *et al.*, 2018), the new construct was designed to also include two surface mutations, E291A and E364A to reduce surface entropy of the protein. The sequence encoding the doubly mutated FXR LBD (258-486aa) was synthesized and subcloned by GenScript (Piscataway, NJ, USA), and inserted into the pET15b vector between the *Nde*I and *Xho*I restriction sites, as to lie directly downstream of the His₆ tag. In addition to the production of an *N*-terminally His₆ tagged fusion protein, the pET15b vector was chosen due to the presence of a thrombin recognition sequence between the tag and the cloning site, which also provided the additional possibility of His₆ tag removal.

4.4.6 Protein Expression and Purification of the New His₆-FXR LBD Mutant

As with the previous construct, expression and purification of the new mutant protein (His₆-FXR LBDmut1), needed optimization. Again, large scale cultures (>4L) were required for high protein yields, and the optimal conditions established previously, were utilized for the expression of this new protein mutant. In addition to these variables, one of the biggest factors contributing to the amount of protein produced was the *E. coli* host strain being used for the expression. Comparisons between IMAC purifications of His₆-FXR LBDmut1, from either Rosetta (DE3) *pLysS* cells or the parental strain BL21 (DE3) *pLysS*, revealed that despite identical culture conditions, significantly more target protein was expressed in the Rosetta cells (Figure 4. 9). Chromatogram traces of the IMAC purification of His₆-FXR LBDmut1 from these strains revealed that, although

the peak intensities of the unbound proteins (flow through) were very similar (\sim 3000 mAU), purification from Rosetta cells produced an elution peak of approximately 1000 mAU, whereas the elution peak from purifications of BL21 was only 500 mAU (Figure 4. 9A and B). These differences were confirmed spectrophotometrically, where absorbance measurements on a Nanodrop and subsequent concentration determination indicated that 4 L of Rosetta cells produced an approximate total of 2.5 g of total proteins, whilst 4 L of BL21 cells produced 2.48 g of total proteins. However, measurements of the purified target protein showed that overexpression in Rosetta cells produced approximately 78 mg, whereas BL21 cells only yielded 28 mg. Comparison of the purification fractions on a SDS-PAGE gel showed a dense, thick \sim 29 kDa band in all of the fractions from the Rosetta cells (Figure 4. 9C). And although it indicates that some of the protein was not sufficiently released during the lysis step, and that some of the target protein may have been lost in the flow through, the His₆-FXR LBDmut1 is being overexpressed in this strain, and is the predominant species in the eluted purified fraction. BL21 cells on the other hand, did not significantly overexpress the target protein and the band in the elution fraction was less intense. Furthermore, although both gels revealed the presence of non-specific bands in the purified product, proteins eluted from lysed BL21 cells contain certain 20 kDa and 25 kDa non-specific proteins (green and purple arrows), which appear to be more dominant than the target His₆-FXR LBDmut1 (Figure 4. 9C).

Even with the new construct, the problems of non-specific protein aggregation persisted. Despite the higher yields of target protein, IMAC purification was not able to separate the target protein from other non-specific binding proteins, even when a cobalt-based TALON column was used (results not shown). Once again, SEC was used in an attempt to separate the associated proteins. Even when setting the flow rate to 0.5 mL/min (a 5-fold reduction in the flow rate recommended by the manufacturer (GE Healthcare)), resolution was not improved. The chromatogram from the SEC purification displayed a peak near the void volume of the column (100 mL) (Figure 4. 10). Analysis of this peak on an SDS-PAGE gel identified the presence of some target protein in addition to some other protein bands, suggesting that some of the protein had likely aggregated and, having no access to the resin, caused it to flow straight through the column. Analysis of the second peak eluted showed that, in addition to the target protein, co-purified, contaminant proteins were present on the in each fraction, with the unknown 20 kDa and 25 kDa species being most prevalent.

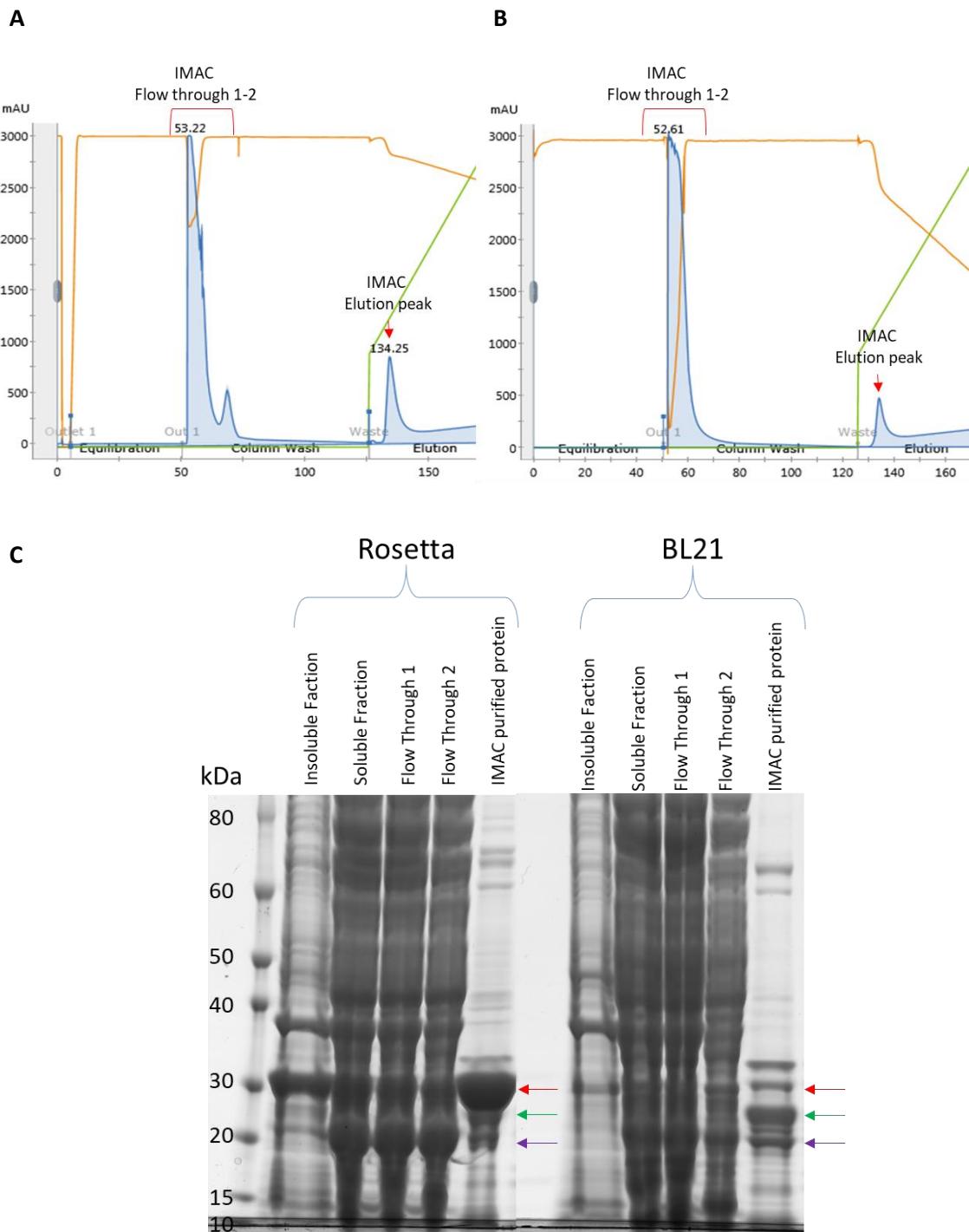


Figure 4. 9 Comparisons between expression and purification of His -FXR LBDmut1 in Rosetta (DE3) pLysS and BL21 (DE3) pLysS. IMAC purification chromatogram of cells purified from (A) Rosetta (DE3) pLysS cells and (B) BL21 (DE3) pLysS cells, displaying UV (280nm) absorbance (blue trace) in mAU, conductivity (orange trace) and programmed 30-100% imidazole gradient (green trace). (C) SDS-PAGE of purification fractions separated on a 4-12% Bis-Tris NuPAGE gel (Invitrogen) and stained with Quick Coomassie (Generon). Target protein indicated by red arrow, and predominant non-specific proteins indicated by green and purple arrows.

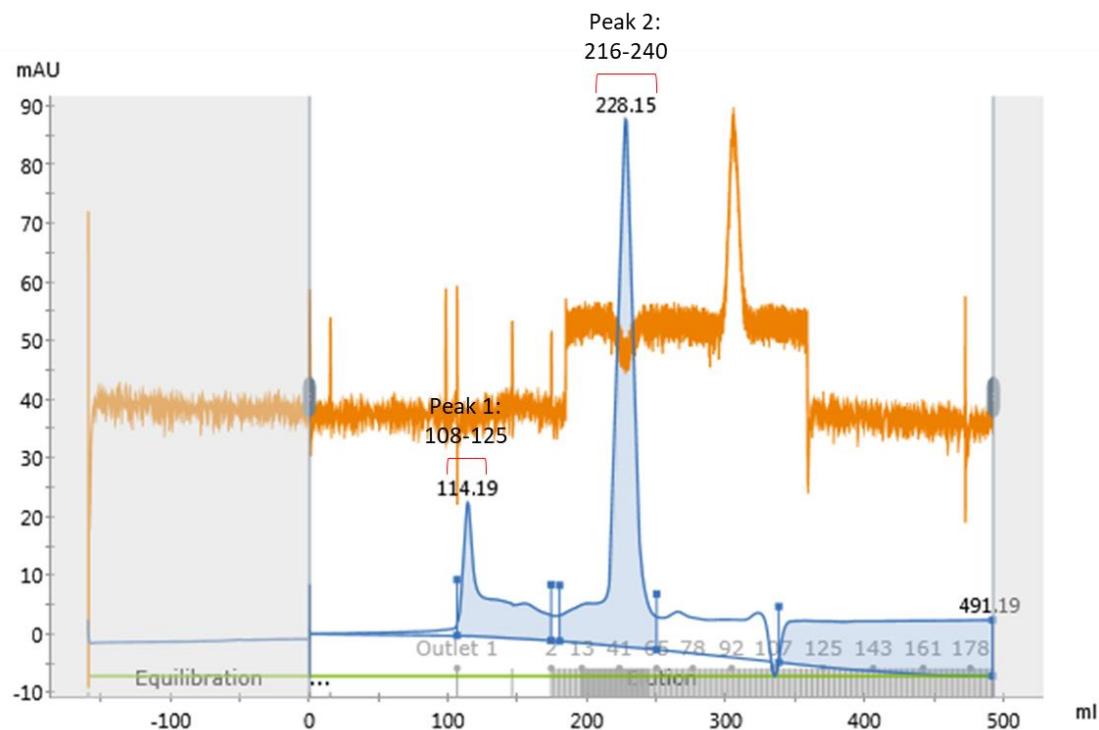
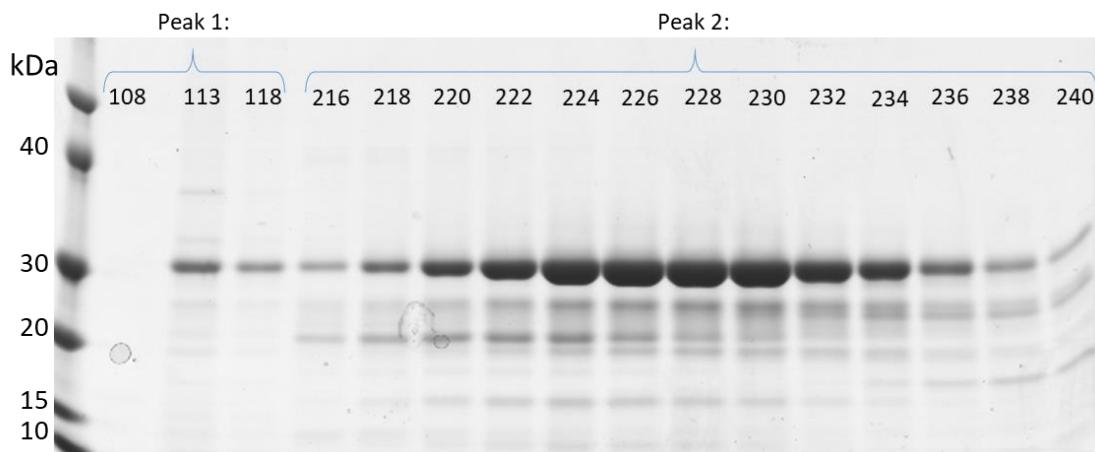
A**B**

Figure 4. 10 Size Exclusion Chromatography of His6 -FXR LBDmut1. Purifications were carried out on a HiLOAD 26/600 Superdex 200 prep grade column. Sample was loaded using a 5mL injection loop and flow rate set to 0.5mL/min for 1.5CV with 2mL fractions collected. (A) SEC purification chromatogram displaying UV (280nm) absorbance (blue trace) in mAU, and conductivity (orange trace) (B) SDS-PAGE of SEC purification fractions separated on a 4-12% Bis-Tris NuPAGE gel (Invitrogen) and stained with Quick Coomassie (Generon).

Due to the consistent presence of these unknown proteins, and the difficulty in removing them, it was imperative that these proteins be identified. The polypeptide bands were extracted from the polyacrylamide gel and sent for sequence analysis by nano-scale liquid chromatographic

tandem and matrix-assisted laser desorption/ionization mass spectrometry (Mass Spectrometry and Proteomics Facility, St. Andrews University, UK). Results identified peaks in the 20 kDa protein that matched with the native *E.coli* protein, FKBP-type peptidyl-prolyl cis-trans isomerase, SlyD, whilst the 25kDa band was determined to be an *E.coli* variant of cAMP regulatory protein, CRP.

SlyD and CRP are both common contaminants in IMAC purifications of recombinant proteins heterologously expressed in *E. coli*, and both have been reported to interact with metal ion affinity columns (Bolanos-Garcia and Davies, 2006). Consequently, an extra thrombin cleavage step was added subsequent to initial IMAC purification, with the hopes that a secondary IMAC purification with the cleaved FXR LBDmut1, would result in the unbound target protein flowing through the column, leaving the contaminant proteins attached.

To test the conditions that were most optimal for thrombin cleavage and complete His₆-tag removal, a sample of IMAC purified His₆-FXR LBDmut1 was used to set up reactions with high grade purity, human thrombin at approximately 10 units per mg protein, and incubated at either room temperature or 4°C. Aliquots were taken at various time intervals and proteins were immediately mixed with 4×Laemmli sample buffer and denatured at 95°C, to stop the cleavage reaction. After 22 hours, thrombin was inactivated by the addition of 100 mM PMSF. SDS-PAGE analysis of the time point aliquots demonstrated a small, band shift in the target protein, from just above, to just below the 30 kDa marker, indicating cleavage of the ~1 kDa His₆ tag. Results showed that cleavage had occurred by 2 hours into the reaction, although the sample at this time point contained a mixture of both His₆ tag-bound and unbound FXR LBDmut1 (Figure 4. 11). The most efficient cleavage was observed after 22 hours, however, the sample was not completely homogeneous, and 2 low molecular weight proteins between 15 and 20 kDa had been introduced. Nonetheless, results suggested that the cleavage reaction was slightly more successful when carried out at room temperature as opposed to 4°C.

A final purification protocol was established, whereby following lysis by cell pressure homogenization, His₆-FXR LBDmut1 was purified from the total protein content of the lysate by an initial IMAC. The purified His₆-FXR LBDmut1 was then incubated with thrombin for 22 hours at 4°C, and then subjected to an additional purification by IMAC. When this procedure was applied, and the secondary IMAC fractions run on an SDS-PAGE gel, the flow through fractions, which were now expected to contain the purified, cleaved FXR LBDmut1, showed that multiple proteins were still being co-eluted with the target protein (Figure 4. 12B). The fourth fraction of the flow through contained the least contaminating bands, however, the amount of FXR

LBDmut1 was also reduced in this fraction. Meanwhile, fractions containing the proteins that had bound to the IMAC column, and were eluted by the addition of imidazole, displayed a modest amount of target protein that had not been cleaved and had adhered to the column, in addition to some FXR LBDmut1, which appeared to have had its His₆ tag successfully removed. Nonetheless, the prevalent bands in these IMAC elution fractions corresponded to the non-specific proteins, suggesting that a significant proportion had been purified from the cleaved FXR LBDmut1.

A final SEC purification was employed to separate the FXR LBDmut1 from high salt concentrations, residual thrombin, and remaining contaminant proteins. The size exclusion chromatogram again identified a peak shortly after the void volume at ~115 mL (Figure 4. 13A). Again, this peak contained a small amount of target protein, and because its retention volume was exactly half that of the monomeric target protein (~230 mL), this again suggested that some of the protein was perhaps dimerizing and flowing through the column much easier and at a faster rate. The second peak, containing the target protein, showed that the FXR LBDmut1 was not completely free of contamination (Figure 4. 13B). However, densitometry analysis, using Image J software, identified that the ratio of target protein to contaminants was slightly better subsequent to thrombin cleavage (Figure 4. 13B), compared to when the procedure was not included (Figure 4. 10B). Consequently, as the target protein was indeed the major component, these samples were progressed to concentration and crystallization screening procedures.

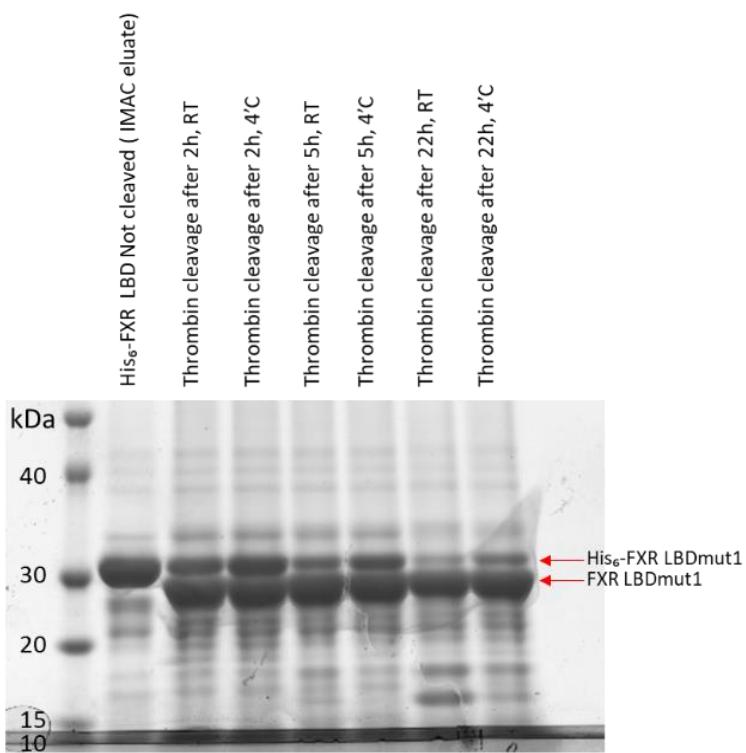


Figure 4. 11 Optimisation of conditions for thrombin cleavage reactions. His -FXR LBDmut1 initially purified by IMAC was incubated with thrombin (10U/mg protein) at either room temperature or 4° C. Samples were taken after 2, 5 and 22 hours and proteins analysed for changes to molecular weight by SDS-PAGE. Tagged and cleaved FXR LBDmut1 indicated by red arrows.

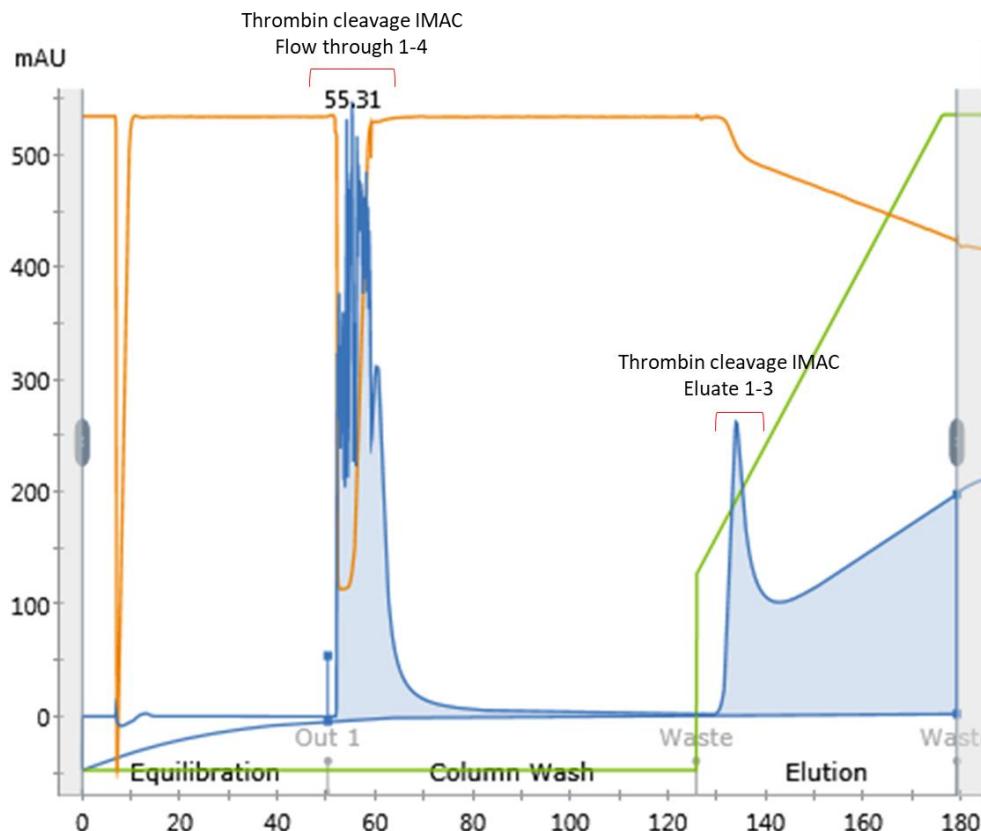
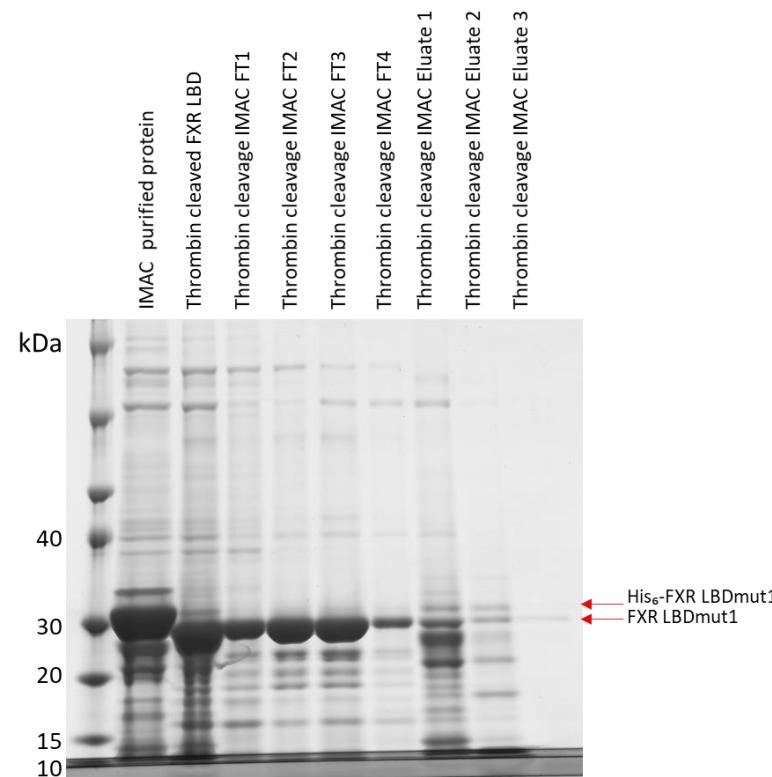
A**B**

Figure 4. 12 Thrombin cleavage and secondary purification of FXR LBDmut1. IMAC purified His₆-FXR was incubated with thrombin (10U/mg protein) for 22 hours at 4°C. Cleaved FXR LBDmut1 was separated from contaminants by an additional IMAC purification. (A) IMAC purification chromatogram displaying UV (280nm) absorbance (blue trace) in mAU, conductivity (orange trace) and programmed salt concentration gradient for elution (green) (B) SDS-PAGE of initial IMAC purified protein sample (not cleaved), thrombin cleavage reaction, and purification fractions separated on a 4-12% Bis-Tris NuPAGE gel (Invitrogen) and stained with Quick Coomassie (Generon). Cleaved target protein was expected in the flow through (FT) fractions. Tagged and cleaved FXR LBDmut1 indicated by red arrows.

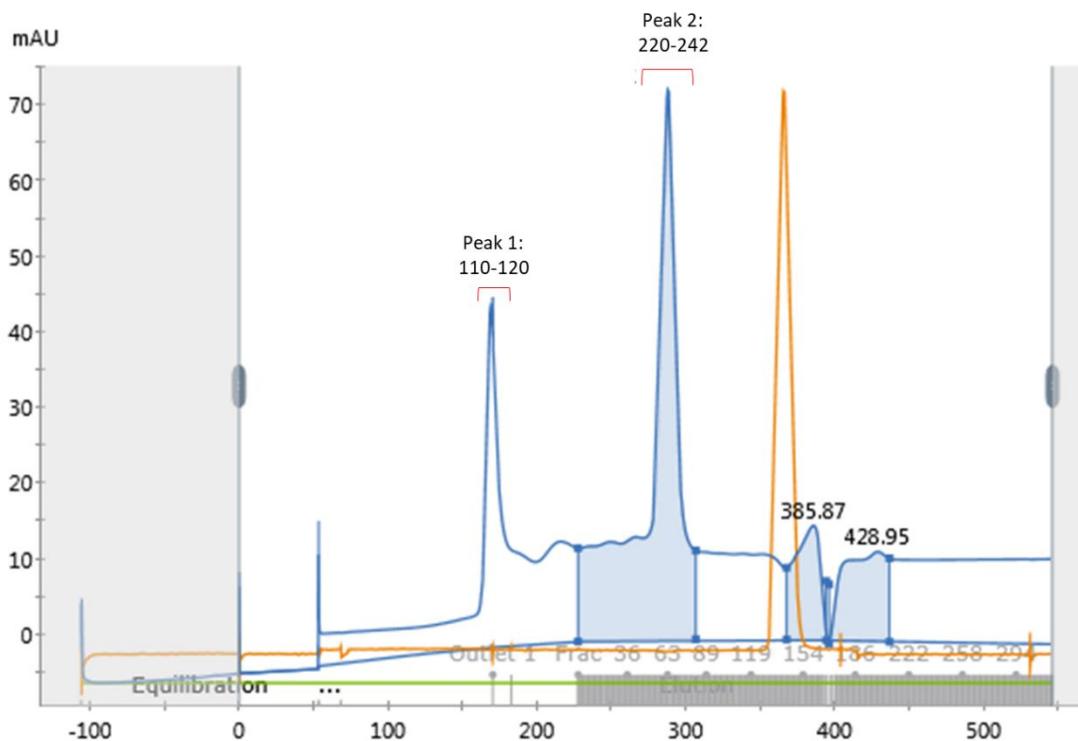
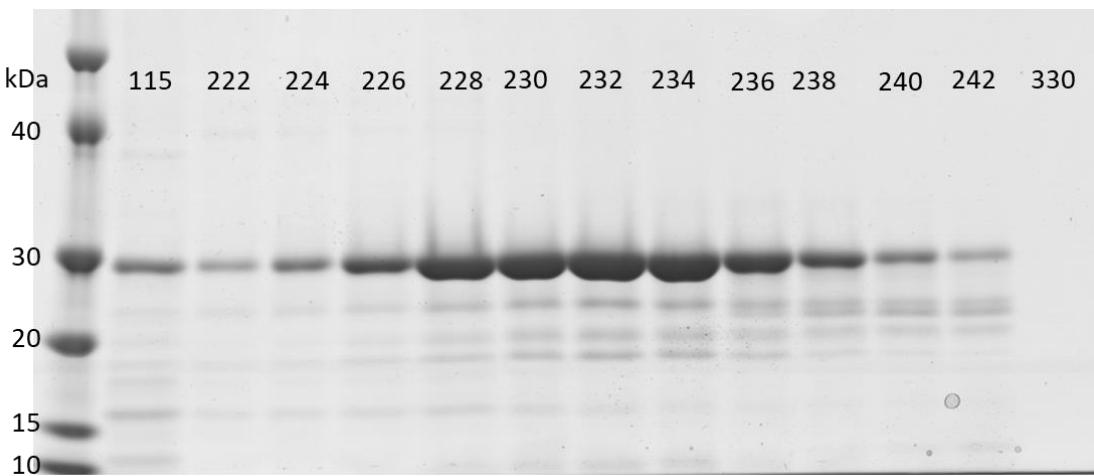
A**B**

Figure 4. 13 Final SEC purification of cleaved FXR LBDmut1. Following the second IMAC purification to remove the majority of non-specific proteins, SEC was applied as a final step before concentration. (A) SEC purification chromatogram displaying UV (280nm) absorbance (blue trace) in mAU, and conductivity (orange trace). (B) SDS-PAGE of purification fractions separated on a 4-12% Bis-Tris NuPAGE gel (Invitrogen) and stained with Quick Coomassie (Generon).

4.4.7 Concentration and Crystallization of FXR LBDmut1

Following the optimized purification of FXR LBDmut1, SEC fractions were pooled and the concentration of pooled protein determined to be approximately 0.22 mg/mL (4.4 mg in 20 mL

total). In order to stabilise the protein, ligand was added to the dilute protein. Initially, lead compound, Jed441 was added alone, but after trials with this complex did not yield any crystals, further protein was mixed with Jed561, as well as the coactivator peptide SRC2-2. In order to fully saturate the protein, ligand compounds dissolved in 100% DMSO were added at a 12× molar concentration to the FXR LBD, and the SRC2-2 peptide, which had been dissolved in water, also was added at a 12× molar concentration. The protein-ligand-coactivator mixture was incubated at 4°C overnight to allow complex formation. Following incubation, the 20 mL protein sample was split between two 15 mL centrifugal concentration devices with a 10 kDa MWCO point. As a precautionary measure, concentration was carried out by centrifugation for 5 minutes at a time, with intermittent pipetting of the retentate, to prevent an extreme concentration gradient, which could lead to precipitation. When the volume of the solution had reduced to ~1500 µL, the protein was removed from the filter and the concentration checked by nanodrop. The concentration was approximately 2.5 mg/mL, which with a 13-fold reduction in solvent volume, indicating that very little protein had been lost. The protein solution was transferred to two 500 µL centrifugal devices, again with a 10 kDa MWCO. The solution was added 250 µL at a time to each column and spun with intermittent pipetting. Concentration was halted when the volume in each device was approximately 75 µL and the retentate had started to turn a yellowish colour. The solutions were pooled and total concentration measured to be approximately 24 mg/mL (in 150 µL).

Following concentration, crystallization trials were set up for the FXR LBDmut1-Jed561-SRC2-2 complex using a sitting drop format in Crystal Quick X2 plates. The NR LBD, JCSG+, PACT premier and Proplex screens were used to maximise the chances of identifying optimal crystallization conditions. Drops were also set up using two different protein:buffer ratios, 1:1 and 1:3, and plates incubated at 20°C. Objects started to appear in crystallization screens by 8 days post set up. One condition (0.1 M Bis-Tris propane pH 8.5, 0.2 M sodium acetate, 20% PEG3350) produced a slightly birefringent object seen in Figure 4. 14A, which appeared by day 5 and grew larger in appearance by day 13. Another condition (0.1 M PIPES pH 7, 0.2M sodium thiocyanate, 1.6 M ammonium sulphate) produced several little spherulite-like droplets (Figure 4. 14B). Whereas several other conditions produced a non-amorphous precipitate, which eventually grew into small clusters of microcrystalline precipitate like that seen in Figure 4. 14C and D. Overall, conditions that were most promising or favoured microcrystals included a buffer between pH 6 and 8, a sodium or magnesium salt, and a mid-high molecular weight polyethylene glycol (PEG, 3350-8000).

Data was collected at the variable, microfocus VMXi beamline (Diamond Light Source) which is dedicated to the automatic, *in situ* screening of plates and was designed for the rapid assessment of potential microcrystalline hits. Grid scanning of regions of interest, however, did not identify any diffraction quality crystals in any of the wells, and no discernible protein diffraction spots were seen (Figure 4. 15).

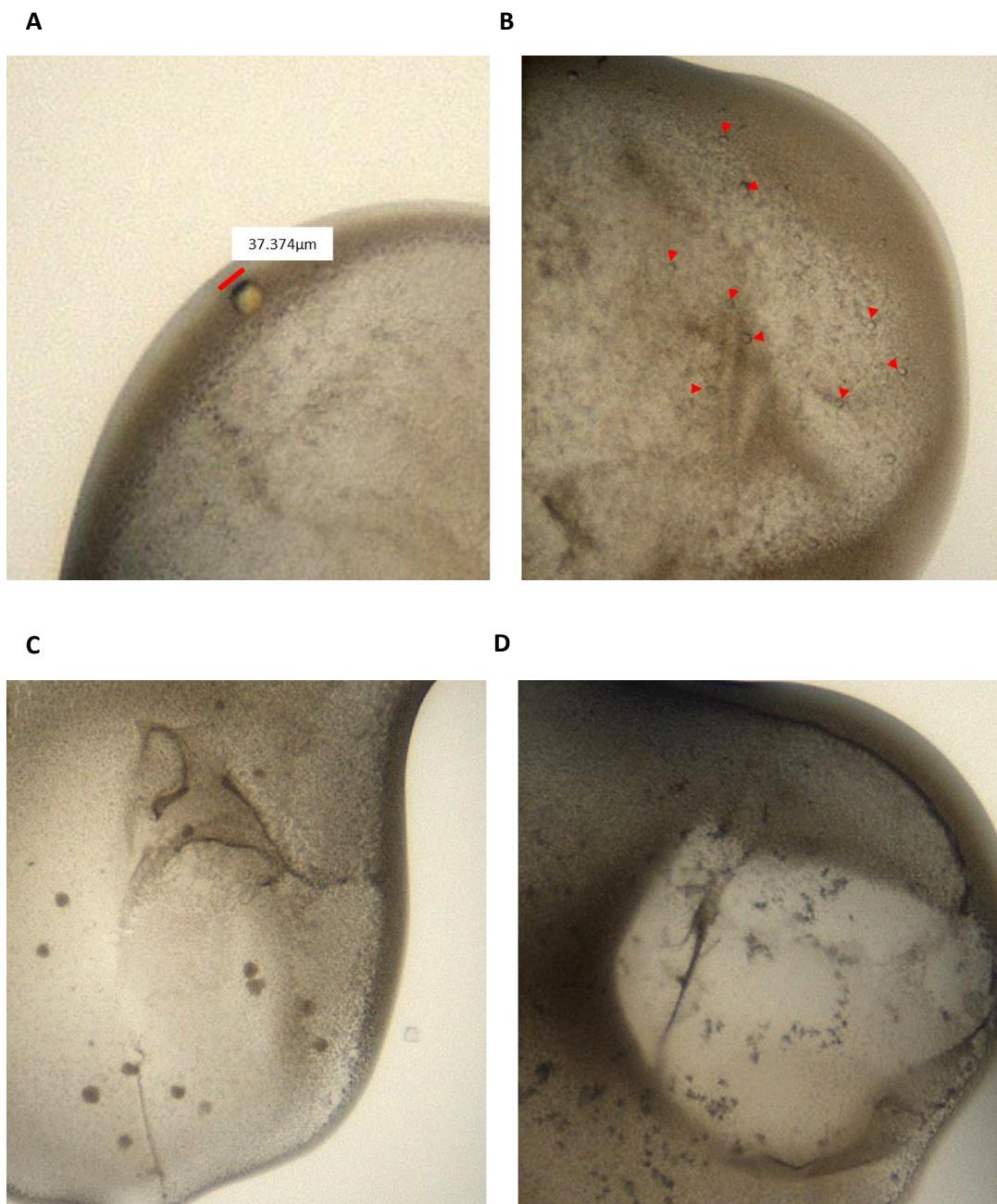


Figure 4. 14 Objects produced in crystallization screens of FXR LBD complexed with Jed561 and SRC2-2.
(A) Birefringent globular object seen in well conditions 0.1 M Bis-Tris propane pH8.5, 0.2 M sodium acetate and 20% PEG3350. **(B)** Spherulite-like objects (indicated by red arrows) seen in well conditions 0.1M PIPES pH7, 0.2M sodium thiocynate, 1.6M ammonium sulphate. **(C)** Precipitate clusters seen in well condition 0.1M PIPES pH7, 0.2M magnesium chloride and 12% PEG8000. **(D)** Crystalline precipitate clusters seen in well conditions 0.1M MOPs pH7.5, 0.1M magnesium acetate, 12% PEG8000.

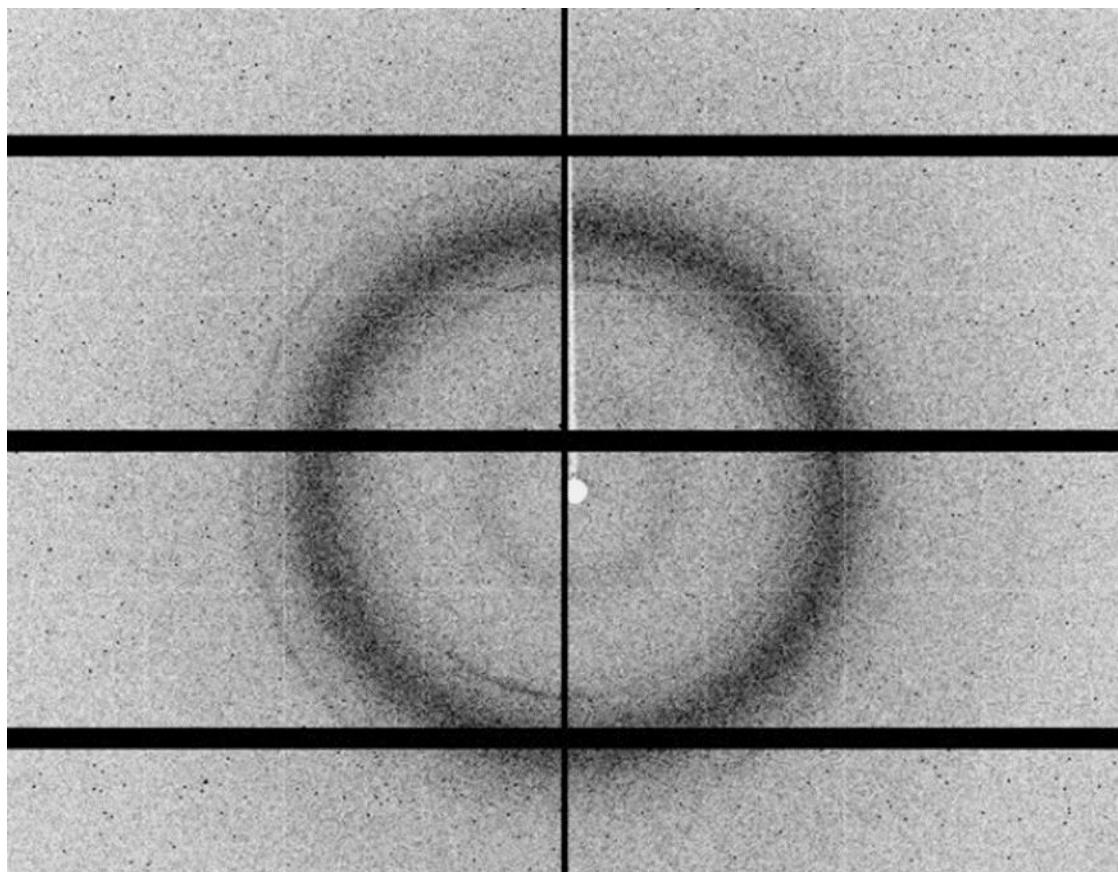


Figure 4. 15 Example of diffraction image observed from *in situ* plate screening on VMXi beamline. Any spots seen are present in all images and are likely to be background noise.

4.5 Discussion

4.5.1 High Throughput Expression Screening of FXR Domains

A high throughput screening approach was employed to find out the likelihood of recombinant FXR expression in *E. coli* cells. Large amounts of purified protein are required for crystallization, and due to its fast growth kinetics, high cell density, and well established transformation and culturing conditions, *E. coli* is the preferential host organism for the simple and easy production of target protein (Rosano and Ceccarelli, 2014). In all previous NR domain structural studies, including the FXR LBD, the heterologous expression of recombinant protein was carried out using *E. coli* systems (Mi et al. 2003; Bass et al. 2011; Xu et al. 2013), suggesting that it is an appropriate starting platform to use for current work. Nevertheless, a contingency strategy was employed by using the pOPIN vector suite, which allows for parallel screening in multiple expression hosts (Rada, 2017), in the event that expression was not achieved in bacterial cells.

Presently, there is nothing in the literature to suggest that the full-length FXR protein has been expressed or its 3D structure determined. However, to further the understanding of the full-length receptor, including its DNA binding function, the DBD also needed to be cloned and expressed in addition to the LBD. As such, multiple constructs were designed encoding different domain combinations fused to different affinity tags, and the high throughput expression screens allowed the quick and easy identification of which domains could be readily expressed in *E. coli*, in a soluble form, and without the need for alternative expression hosts or significant optimization.

Results identified that the full-length FXR protein could not be easily expressed in this host, although the possibility that it forms inclusion bodies was not explored. This aligns with the traditional view of full length NRs, that they are hard to express in the soluble form, and require truncation for improved yields (Mossakowska, 1998). Previous attempts to express full length VDR and androgen receptor (AR), resulted in low yields, and most of the protein was identified in the insoluble fraction (Rennie *et al.*, 1993; Hsieh *et al.*, 1995). Although Chandra and colleagues did not report their ability to express full length PPAR γ and HNF4 α , attempts to crystallize these full length proteins were futile (Chandra *et al.*, 2008, 2013). Inclusion bodies are often formed due to the extreme workload and simplistic machinery in *E. coli* cells, which, as a result of the lack of eukaryotic chaperones or post-translational modifications, lead to unfolded or partially folded proteins. It is possible that to produce soluble forms of this full length eukaryotic FXR protein, much 'gentler' induction regimes, or expression in a eukaryotic host is needed.

Nonetheless, screening identified that the DBD-LBD could be expressed in *E. coli*, when fused to a SUMO or GST tag. Although not carried out here, optimization of the induction conditions, may help to produce improved yields of this protein. Currently, there are no published structures of the FXR DBD-LBD, and by determining the structure of these linked domains, it will not only help to elucidate ligand binding mechanisms, but it will also help to determine important DNA binding interactions too. Structural studies of the FXR DBD-LBD may even help to uncover inter-domain cross talk and allosteric signalling, where observations can be made regarding any changes in DNA binding interactions in response to different ligands occupying the binding pocket or coactivators interacting with the LBD surface, as previously proposed (Meijer *et al.*, 2019).

High throughput expression screens identified that the FXR LBD could be expressed, with the best yields arising from constructs that encoded the domain fused to a simple 6xHistidine

residue tag, at either the *N*- or *C*-terminus (construct C2 and D2). The His₆ tag is ideal because it can facilitate the purification of the protein, yet its small size is unlikely to significantly interfere with the tertiary structure of the protein and, in most cases, it does not prohibit crystal formation. Construct D2, encoding the *N*-terminally tagged variant was chosen for future work due to the fact that the *C*-terminus of the domain comprises the AF2, whose conformation determines ligand specific transactivation by changing the accessibility of coactivator binding sites. Although for X-ray crystallography, dynamic changes of the AF2 will not be seen, the conformation of helix 12 will be important for the structure determined, particularly when the receptor is co-crystallized with a ligand and coregulator. Therefore, it was postulated that the *C*-terminal His-tag may interfere or block the coactivator binding site.

4.5.2 Yield Optimization

A major bottleneck in obtaining the 10 mg/mL concentrations of FXRLBD required for crystallization was in gaining a sufficient amount of target protein from *E. coli* cells, right from the start. Visual inspection of the SDS-PAGE gels of proteins purified during the initial high throughput expression screens, indicated that Rosetta (DE3) cells produced a slightly higher yield of overall proteins (Figure 4. 3). Later studies, comparing expression of the new construct, FXR LBDmut1, in Rosetta (DE3) and the parent BL21(DE3) strain, indicated that significantly higher yields of the target protein could be achieved in Rosetta cells (Figure 4. 9). Rosetta cells are derived from BL21 (DE3) but contain an additional plasmid that can express rare codons, which are not normally used in *E. coli*, although may be important for the translation of eukaryotic proteins. The original gene sequences were codon optimized before synthesis to minimize codon usage bias, but expression in the Rosetta strain may help to supplement the tRNA pool, allowing complete synthesis of the target sequence, and the availability of rare codons has been described as a major determinant of protein yields (Rosano and Ceccarelli, 2014).

Initial 500 mL Rosetta (DE3) cultures did not yield a large amount of protein, resulting in the lack of a chromatogram peak following the elution of His₆-FXR LBD during the Ni-NTA purification (Figure 4. 4). Although analysis by SDS-PAGE identified a 29 kDa band in the purified fraction, indicative of the His₆-FXR LBD, this band was not significantly thicker or more dense than the non-specific proteins, suggesting that the culture conditions did not favour its overexpression compared to other proteins. In addition, the presence of multiple, smudged bands in the insoluble cell pellet suggested the potential insufficient lysis of the cells.

His₆-FXR LBD yields were improved by changing several factors. One of the main changes that lead to improved yields was the simple scale up of culture volumes. By merely increasing the size of the cultures, there is an increase in biomass, which ultimately means that there are proportionally more cells producing the His₆-FXR LBD, and previous groups working on the FXR LBD have reported using large scale cultures in batches in excess of 3-6 L (Akwabi-Ameyaw *et al.*, 2011; Lu *et al.*, 2018).

Other parameters that were optimized were the concentration of IPTG and the conditions following induction of protein expression. IPTG is a lactose-mimic that triggers the downstream expression of T7 RNA polymerase, which ultimately is responsible for the overexpression of recombinant proteins (Rosano and Ceccarelli, 2014). By fine-tuning the IPTG concentration, it is possible to maximise the transcription of the target protein, whilst minimising the extreme burden on the cells, which could lead to inclusion body formation. Likewise, by lowering the temperature after induction with IPTG, protein production is slowed, allowing time for the proteins to be properly folded, increasing the likelihood of soluble expression.

Although important factors, changes to IPTG concentration and culture conditions did not significantly influence His₆-FXR LBD yield. However, the lysis method employed was observed to have a notable effect on the amount of protein obtained in the soluble fraction, and thus, the amount of target protein that was purified. Optimal conditions aim to maximise cell lysis and the amount of recombinant protein that is obtained, whilst minimizing the oxidation and degradation of that protein. The most common methods include lysis by freeze-thawing cycles and the addition of lysozyme, or mechanical lysis by high-pressure homogenization or sonication. Purification methods from cells lysed by sonication, lead to suboptimal yields of protein. Sonication itself isn't detrimental to proteins, but the heat generated by sonication can denature proteins. Although all samples were sonicated on ice, it is possible that the proteins nearest the probe may have been subject to overheating and degradation, therefore, resulting in losses. Furthermore, for large cell cultures (>4 L), the difficulty in maintaining temperatures and the lengthy time period required for sonication to achieve adequate lysis, made other methods such as cell disruption far more favourable. The final method of freeze-thawing the cell pellet, adding lysozyme to the lysis buffers, and passing the protein through high pressure cell homogenizer lead to significantly increased target protein yields (Figure 4. 5).

Although improved yields of soluble FXR LBD were achieved here, NRs have been associated with solubility issues (Mossakowska, 1998). A key requirement in obtaining enough protein for the structural studies of the androgen receptor, progesterone receptor, and glucocorticoid

receptor, was the co-expression of these proteins with their cognate ligands (Hassell *et al.*, 2006).

4.5.3 Purity Optimization

In addition to a high concentration, a high purity (>90%) and homogeneity of target proteins is also required for crystallization (McPherson, 1991). Another challenge faced, was in substantially purifying the His₆-FXR LBD, so that contaminant proteins were efficiently removed. The 20-25 kDa proteins that were consistently seen to co-elute with His₆-FXR LBD in purification fractions, were not due to the degradation of this protein, as SDS-PAGE gels showed that there were no changes to the His₆-FXR LBD band size or intensity, and there were no proteins that were 'introduced' at any later time point. Thus, the His₆-FXR LBD appeared to be stable over 4 days at room temperature (Figure 4. 8).

The non-specific proteins were later identified as *E. coli* contaminants, SlyD (20 kDa) and CRP (25 kDa). IMAC purification columns utilize a chelator to bind and immobilize divalent metal ions to the column resin. These metal ions, in turn, form coordination bonds with histidine residues, allowing target proteins with multiple histidines to be easily separated from the rest. However, proteins such as SlyD are frequently found to contaminate IMAC purified samples due to their short clusters of histidine and cysteine residues, which have been observed to have a high affinities for nickel and cobalt ions in particular (Parsy *et al.*, 2007). Whilst the CRP protein is not significantly histidine rich, nor a known metal binding protein, it contains 3 surface exposed histidine residues, which may also act to sequester metal ions; CRP is another common recurring contaminant identified in previous IMAC purification studies (Bolanos-Garcia and Davies, 2006). The His- tag like nature of these proteins explain why these particular proteins could not be completely separated from the His₆-FXR LBD on either a nickel nor cobalt based column. Furthermore, it explains why the addition of high salt concentrations and glycerol (which should prevent the electrostatic and hydrophobic interactions between non specific proteins and the column) was not sufficient to prevent these proteins from binding. To circumvent this contamination with native *E. coli* proteins, it is possible to use a fusion protein with a longer polyhistidine tag. The addition of His₈ or His₁₀ to the FXR LBD may increase the binding efficiency of the target protein to a Ni²⁺-NTA column, but longer histidine tags have been associated with solubility issues in some classes of proteins (Woestenenk *et al.*, 2004). Another solution is to use an *E. coli* expression host that has been engineered to express background proteins at a lower level or that expresses histidine-null forms of common contaminants (Andersen *et al.*, 2013). His₆-FXR LBD expression initiated in the low background

strain *E. coli* (LOBSTR), but this host produced low yields of His₆-FXR LBD (not shown), and the specific requirements for culturing this strain were harder to produce on a large scale.

Instead, an approach was taken to try to remove the contaminant proteins by other forms of chromatography. Size exclusion chromatography was carried out, using a HiLoad 16/600 Superdex 200 prep grade column. Equilibration of the column, using the gel filtration protein standards (Sigma), determined that a 29 kDa protein would have a retention volume of approximately 235 mL. Application of the His₆-FXR LBD to the SEC column was consistently recovered as a wide, ~20 mL peak with an apex between 220 and 230 mLs. Despite the symmetrical peak acquired on the chromatogram, which usually indicates a homogenous sample, analysis by SDS-PAGE showed that SEC was unable to fully separate the target from contaminant proteins, as non-specific bands were present in every fraction at an extent proportional to that of His₆-FXR LBD. Size exclusion separates proteins and molecules according to their molecular weight, by passing them through a packed bed column, where a porous matrix allows smaller molecules to diffuse into the pores, delaying their travel through the column, whereas bigger molecules, which cannot enter the pores, instead flow straight through. The isocratic elution procedure separates molecules in one column volume of buffer. Even after attempts to improve the resolution of the target protein, by using an appropriate sample volume less than 2% of the total column volume, and by significantly reducing the flow rate to allow complete partitioning between the stationary and mobile phases, complete separation of the proteins could not be achieved. The HiLOAD column used, contains a Superdex agarose-dextran medium, which is deemed suitable for the fractionation of proteins between 10 kDa and 600 kDa (Ó'Fágáin *et al.*, 2017). Although this particle size can separate monomers from aggregated proteins at a high resolution, it is possible that the 29 kDa His₆-FXR LBD and 20-25 kDa contaminant proteins are too similar in size to be separated by this medium. As such, a column packed with the Superdex 75 prep grade medium, which can fractionate small proteins between 3 kDa and 70 kDa, may be more appropriate.

Removal of the contaminant proteins also was attempted by separation on an anion exchange column, which has a higher resolving power than SEC (Gräslund *et al.*, 2008). Ion exchange chromatography is capable of separating proteins that have very minor differences in their overall surface charge. Despite their molecular weights, the His₆-FXR LBD and contaminant proteins, would have different amino acid compositions, which result in different pI values. By exploiting their differences in affinity to a charged resin column, ion exchange chromatography should be able to separate different proteins. His₆-FXR LBD has a pI of 5.9, hence an anion exchange column was used, at a buffer pH of 7.8, which allows the negatively charged His₆-FXR

LBD molecules to interact with the positively charged, ammonium ion HiTrapQ resin. However, a pre-requisite of the method is that the start buffer contains a very low concentration of salt, which sees to maximise the chances of protein-column interactions. When the IMAC purified His₆-FXR LBD was exchanged to a low salt concentration, the protein began to precipitate. Precipitation in low salt concentrations has been well observed, and, although poorly rationalised, high ionic strengths are recognised to maintain the solubility and stability of recombinant proteins (Gräslund *et al.*, 2008). A challenge arose in trying to determine a salt concentration low enough to prevent interactions with the anion exchange column, but high enough to maintain soluble forms of the His₆-FXR LBD protein. Eventually, when AEC was applied, it produced poorly resolved, overlapping peaks. When the fractionated peaks were separated on an SDS-PAGE gel, it was seen that a substantial amount of protein was lost in the flow through, likely due to insufficient initial binding to the column, due to interferences from the numerous Cl⁻ ions present in the start buffer. Furthermore, each peak again contained a mixture of both the target protein and non-specific proteins (Figure 4.7).

After the discovery that the contaminating bands were due to the co-elution of target protein with histidine rich *E.coli* proteins, it was anticipated that His₆ tag cleavage and subsequent additional IMAC purification would facilitate the separation of these associated proteins. The initial thrombin cleavage tests (Figure 4.11) sought to identify the most optimal conditions for maximum cleavage conditions, while minimising the non-specific cleavage and degradation of the protein. The thrombin used was from human plasma and was the highest purity grade available from Sigma. Although less abundant, and therefore more expensive than thrombin from bovine plasma, the human variant was used as it is deemed to be the most site-specific and free from secondary proteases. As recommended, 10 units of the endoprotease were used per mg protein, and although the optimum incubation temperature for thrombin cleavage has been determined as 45°C (Waugh, 2011), reactions were compared at room temperature (22°C) and 4°C, to identify which conditions could result in efficient cleavage, whilst maintaining a stable target protein. Results showed that complete cleavage could not be obtained in any of the reaction conditions tested. This may have resulted from suboptimal salt concentrations, (1 M is recommended (Waugh, 2011)), the inclusion of the reducing agent, DTT, in the buffer (it has been shown to inhibit thrombin activity (Waugh, 2011)), or due to the reaction not being given sufficient time to complete (some methods suggest a 2-4 day incubation (Hefti *et al.*, 2001)). However, cleavage was most efficient after incubation for 22 hours at room temperature, as this lane displayed the slight band shift taken to indicate His₆ tag removal. This fraction also contained the faintest band representing the higher molecular weight tagged FXR

LBD. However, separated proteins from the reaction incubated for 22 hours at room temperature, also contained 2 bands between 15 and 20kDa. These bands were more prominent than those seen in the other conditions, and with the possibility that these might represent non-specific cleavage, future methods were carried out at 4°C. Furthermore, although not completely efficient, cleavage reactions were not extended beyond 22 hours due to concerns of non-specific cleavage and FXR LBD stability.

Despite the measures taken with His₆ tag removal, a secondary IMAC purification subsequent to thrombin incubation, was still not able to fully separate contaminant proteins from the FXR LBD (Figure 4. 12). Although cleavage was incomplete, very little target protein remained bound to the Ni²⁺-NTA column and the majority of FXR LBD was recovered in the flow through. Nonetheless, subsequent purification by SEC and analysis by SDS-PAGE, suggested that some of the FXR-LBD could be dimerizing (Figure 4. 13). An initial peak was seen at approximately 115 mL on the chromatogram, exactly half the retention volume for the monomeric FXR LBD. When run on a gel, this fraction certainly contained small amounts of target protein. Furthermore, when fractions pertaining to the target protein peak were analysed by SDS-PAGE, again it was seen that contamination by the low molecular weight non-specific proteins persisted. In a natural cellular context FXR has been shown to homodimerize, or heterodimerize with partner receptor RXR (Forman *et al.*, 1995), and recent structural data have shown that heterodimerization is achieved through interaction interfaces on helices 7, 9 and 11 of the RXR and FXR LBD (Wang *et al.*, 2018). The RXR LBD was predicted to be 26 kDa by the ProtParam tool (Expasy server). It is possible that in the absence of RXR, the FXR-LBD is dimerizing with small, non-specific proteins as a substitute. While this sample of FXR LBD was used to see if crystals would form, regardless of its heterogenous nature, future studies may need to consider the approach of co-expressing the FXR LBD along with the RXR LBD, as achieved by Wang *et al.* (2018). Although the majority of groups have successfully expressed and purified the FXR LBD by itself, the cistronic co-expression of the RXR and other NR LBDs, under a single promoter, has reportedly improved the solubility of certain NR LBDs, and RXR has been proposed to act similarly to a chaperone protein, assisting the correct folding of the target receptor (Mossakowska, 1998). Furthermore, an added benefit of determining the FXR LBD/ RXR LBD dimeric structure, would be in furthering the understanding of the structural basis for interactions between the two receptors, and identifying any potential structural changes as a result of ligand binding.

4.5.4 Preventing Protein Precipitation

Another important proviso for crystallization of proteins is being able to achieve a high concentration of macromolecules, without the formation of aggregates or precipitate

(McPherson and Gavira, 2014). Unfortunately, initial constructs of the His₆-FXR LBD, were particularly associated with time-dependant precipitation, which, as previously discussed, was accelerated in buffers with low salt concentrations. This time-dependant precipitation was noted in previous work with the FXR LBD (Soisson *et al.*, 2008), and the propensity of NRs to aggregate has been attributed to their largely hydrophobic surfaces (Mossakowska, 1998). It is possible that in this case, the FXR LBD may be slightly unfolded, exposing their interior hydrophobic pockets which then promote aggregation with exposed hydrophobic regions on other molecules. However, the addition of 10% (v/v) glycerol, a stabilizing osmolyte, which is thought to act as an amphiphilic interface between the hydrophobic surface and polar solvent molecules (Vagenende *et al.*, 2009), was not able to ameliorate the aggregation. Although other methods, such as the addition of polar amino acids, like arginine, to the solution have been suggested to inhibit precipitation (Gräslund *et al.*, 2008), these were not investigated.

Another potential cause of protein precipitation was the oxidation of thiol containing residues. The initial His₆-FXR LBD construct contained two free, surface exposed cysteine residues in the C-terminal of LBD and an additional cysteine in the extra N-terminal hinge region residues. It is possible that the oxidation of cysteine's sulphydryl group, and the subsequent formation of disulphide bonds with other cysteine molecules, is causing the proteins to aggregate. The reducing agent DTT can interact with disulphide bonds and forms a stable oxidise product. To prevent precipitation by thiol interactions, DTT was added to the protein solution shortly after purification from the Ni-NTA column. Initial studies included DTT in the gel filtration buffers, but due to the short half-life of DTT, later purification procedures added a freshly made aliquot directly to the protein solution, to a final 1 mM concentration. Although, the addition of reducing agents greatly minimized precipitation in work described here, a previous group used an additional alkylation step with iodoacetamide to covalently bind to the free thiol groups of the FXR LBD, preventing the reformation of disulphide bonds (Soisson *et al.*, 2008). Similarly, other groups chose to mitigate the problem entirely, by using a variant of the FXR LBD that had its free cysteine residues mutated to glutamate, glutamine or lysine residues (Wang *et al.*, 2018).

Although the new FXR LBD mutant used in later studies, did not include mutations of the free cysteine residues, removal of the hinge region residues, was sufficient to create a noticeably more stable protein, which was less prone to precipitation. It is possible that the flexible or partly folded hinge region residues promoted the amorphous aggregation of proteins. Together with the addition of DTT, the addition of the ligand, with or without the addition of the coactivator, significantly increased the stability and solubility of the FXR LBD. Occupancy of the

hydrophobic pocket may have prevented hydrophobic interactions with other proteins. And, by addressing the possibility of aggregation due to surface exposed cysteine residues, the FXR LBDmut1-ligand complexes were concentrated to levels not achieved with the initial FXR LBD. Furthermore, the method adopted for concentrating the complex by centrifuging for just 5 minutes at a time, with protein resuspension between each spin, allowed the protein to be concentrated gradually. Previous attempts at concentrating the FXR-LBD, using long centrifugation times, often resulted in precipitate building up in the crevice of the filter, on the surface of the membrane. It is thought that large reductions in the solvent volume relatively quickly, could create a concentration polarization gradient, whereby the protein molecules accumulate towards the bottom of the retentate and at the surface of the filter membrane, where they can interact with each other and aggregate. This occurrence was minimised by separating the sample between two devices, increasing the surface area of the filter membrane and reducing the likelihood of concentration gradient formation. Concentration in this manner typically led to an 80-90% recovery of protein.

4.5.5 Crystallization

Protein crystallization involves numerous combinations of variables, but can often result in a low yield of crystals with sufficient quality to solve structures (McPherson *et al.*, 1995). A single condition can include a combination of a precipitant, pH-dependent buffer and an additive; to date, there are hundreds of different crystallization reagents, making up millions of unique combinations. Several commercially available screens, using random or methodical combinations of predefined reagents have been developed to aid the identification of optimal conditions. For early crystallization trials, using the initial variant of the His₆-FXR LBD, the NR LBD screen, tailored specifically towards the crystallization of flexible LBDs, and the PACT premier screen, one of the most effective systematic screens available, were used (Billas *et al.*, 2001; Newman *et al.*, 2005). Despite screening these 192 different reagent combinations, no crystals were obtained. It is very likely that the formation of crystals was restricted due to the low protein concentration (~5 mg/mL) derived from the early purifications.

Whilst purification conditions were being optimized, a group at Roche Pharmaceutical Research centre, were able to publish a series high-resolution structures of the FXR LBD (Gaieb *et al.*, 2018). Inspection of these structures highlighted two mutations of surface glutamate residues to alanine residues. These mutations were reported to make the proteins more crystallisable (Gaieb *et al.*, 2018). By the principles of surface entropy reduction, it is thought that certain residues with high intrinsic conformational entropy, such as glutamate, can impede the nucleation and crystallization of proteins, by increasing the energy required to stabilise or order

their side chains at the point of crystal contacts (Cooper *et al.*, 2007). Accordingly, it is suggested that by engineering proteins to remove or reduce these surface residues, it may be possible to promote crystallization (Cooper *et al.*, 2007). The FXR LBD contains two regions of glutamate clusters, and due to the previous success of the Roche group's glutamate-alanine mutations, a new construct was designed to incorporate these mutations also. Furthermore, as discussed earlier, the addition of the flexible hinge region residues in the former His₆-FXR LBD, also may have contributed to the recalcitrant nature of this protein and its resistance to crystallization.

Following purification of the new FXR LBDmut1, a completely homogenous sample of target protein was not acquired. However, densitometry analysis identified a higher proportion of target protein compared to contaminant proteins, than previous purification runs, and although unlikely, it is still possible that a moderately heterogeneous sample can form crystals, when the protein of interest is the major component. Crystallization trials were set up from ~10 mg/mL FXR-LBDmut1 complexed with Jed441, using the aforementioned screens, and additionally the JCSG+ screen, which uses a more 'shotgun', random approach to its component matrix, and the Proplex screen, which has been formulated for crystallization of protein complexes. Crystallization drops were set up at two different ratios of protein:buffer condition for optimization purposes. However, screens with this FXR LBD-Jed441 complex did not produce any crystals and most of the drops remained clear, indicating a suboptimal protein concentration. Moreover, although many NR LBDs have been crystallised without a coactivator peptide, fluorescence anisotropy analysis has revealed that, in some cases, crystallization is negatively correlated with helix 12 flexibility, and it's thought that the addition of a coactivator peptide can help to stabilize this helix, thus promoting crystallization (Nahoum *et al.*, 2008).

A second round of crystallization screens was attempted, again using the FXR LBDmut1, this time complexed with lead compound, Jed561, and coactivator peptide SRC2-2, which was identified to have the highest affinity for the FXR LBD (shown in Chapter 5). Concentrations of 24 mg/mL were achieved with this complex, and again each condition was set up in two different ratios. Subsequent drops contained birefringent objects, microcrystals, and non-amorphous precipitates. A birefringent globular object was seen in a well with conditions 0.1 M Bis-Tris propane pH 8.5, 0.2 M sodium acetate and 20% PEG3350. Although crystals don't usually feature curved edges, several previous studies have documented crystallization of the FXR LBD using 20-25% PEG3350 as a precipitant (Soisson *et al.*, 2008; Akwabi-Ameyaw *et al.*, 2011; Bass *et al.*, 2011; Gaieb *et al.*, 2018). On the other hand, crystallization with PEG3350 may be coincidental as PEG3350 is one of the most commonly used reagents in crystallization studies, and the same protein displays very different reagent preferences based on the ligand

that it is complexed with (McPherson and Cudney, 2014; Kirkwood *et al.*, 2015). Some of the drops contained precipitated protein, suggesting that the concentration of the protein was potentially too high and would not favour nucleation. However, many of the drops contained non-amorphous precipitate, with random patterns and localized clusters, which are often precursors to spherulites or crystals (Bergfors, 2009). The appearance of spherulites in some of the drops, suggested that these conditions could be optimized, or the spherulites used as seeds in new trials.

A 'large' crystal is no longer a necessity to obtain high resolution diffraction data, and the development of microfocus beamlines, such as the VMXi at Diamond Light Source, allow for diffraction data to be collected *in situ* from microcrystals (Sanchez-Weatherby *et al.*, 2019). Regions of interest were screened at the VMXi beamline for diffraction. Unfortunately, none of the drops screened produced any reflection spots, suggesting that these protein molecules were not in a uniform, periodic arrangement, as indicative of the reinforced diffraction from multiple parallel planes of atoms in crystals.

4.5.6 Conclusion

Due to its superior ability to consistently produce high resolution structures, MX was the method of choice for determining the structure of FXR. The atomic level detail provided by crystallographic structures is sufficient for identifying specific residues involved in either DNA or ligand binding and, as such, the method provides an excellent means for the validation of proposed ligand binding mechanisms, identified by *in silico* molecular docking. In addition, the structures obtained could provide scope for further engagement of the ligand binding pocket, or insight into selective modulation of the receptor, ultimately leading to the expansion of the compound library.

The results presented herein highlighted that heterologous expression of the full length FXR protein would not be easily achieved using an *E. coli* host, and other methods may be needed to acquire enough protein for structure determination by crystallography. The results did raise the possibility of expression of the DBD in conjunction with the LBD in bacterial cells, although it is possible that these protein domains may also be associated with problems surrounding the procurement of a highly pure and soluble product, and co-expression with the respective RXR domains may be required, as seen with other NRs (Chandra *et al.*, 2013; Lou *et al.*, 2014).

Unfortunately, in this work, structure determination of the FXR LBD in complex with top candidate, novel ligands could not be achieved by MX. Work, however, identified and highlighted several challenges associated with the expression, purification and crystallization of

this protein, which were not documented in previous publications of the FXR LBD. Whilst the majority of FXR LBD – ligand complex structures have been produced by industrial, or pharmaceutical labs, the work presented here offers solutions for the large scale expression and purification of the protein in an academic lab setting, without the capacity for large batch fermentations, or automated purification processes. Whilst the FXR LBD is relatively easily expressed in *E. coli*, yields can be increased by careful consideration of the strain used to synthesize this eukaryotic protein, by significantly increasing the overall biomass of the host cells, and by using lysis procedures that can facilitate the complete and rapid disintegration of large scale cultures. The problem of contaminant proteins can be minimized or even removed by more stringent wash processes during IMAC purifications, by thrombin cleavage of His-tags and a second IMAC purification, or even by co-expression and purification with the partner receptor, RXR. Unfortunately, contaminant *E. coli* proteins in this work, were of a similar size to the target FXR LBD and the use of more sensitive SEC column resins may help to fully separate these in the future. It was realised that time-dependent precipitation could be minimized by the addition of reducing agents to the protein, and by stabilisation of the protein with the addition of a ligand and cofactors, however, this problem can only be completely mitigated by mutagenesis of surface-exposed cysteine residues. Together, these steps can help to achieve sufficient homogenous protein for crystallization, but screens are still needed to determine the unique conditions required for nucleation and crystal growth.

Although diffraction quality crystals were not obtained here, refinement of some of the conditions that yielded microcrystals, may be able to produce higher quality crystals, and should be investigated further. In the absence of a clear indication of which parameters are more important for the crystallization of this complex, a systematic, incremental exploration of a range of conditions may help. The limitation to this approach, is the substantial amount of protein this would require, however, such a trial and error process can be overcome by the use of some of the microcrystals and spherulites as 'seeds', which may help to accelerate the formation and growth of new crystals. Microseeding takes advantage of the fact that crystallization is a two-step process, involving nucleation and subsequent crystal growth. By separating these two processes, the introduction of a submicroscopic crystal seed into a new drop can bypass the kinetic energy barrier needed for spontaneous nucleation, and can allow new crystals to form in a more dilute sample (Bergfors, 2003). The presence of microcrystalline precipitate certainly allows for seed streaking to potentially be applied.

In the event that crystals of this complex are never obtained, other methods provide the potential for ligand binding to the FXR LBD to be investigated, although to a different level of

atomic detail. As mentioned previously, NMR provides the potential for dynamic movement of the LBD helices to be studied. Historically, NR LBDs in complex with partial agonists, and selective modulators, have been hard to crystallize, and NMR was the first method to provide a glimpse into the dynamic mechanisms of partial agonism (Kojetin and Burris, 2013). The conformational changes induced by Jed441 or Jed561, binding to the ligand binding pocket, may be observed by NMR and may help to provide information about how these ligands promote transactivation of the receptor. Furthermore, the use of hydrogen/deuterium exchange (HDX) also has been used to complement other structural studies in the observations of conformational changes upon ligand binding in the LBD (Kojetin and Burris, 2013). The exchange of amide hydrogens to deuterium in proteins can be used to indicate changes in hydrogen bonding capacities and can measure changes in solvent accessibility and conformational dynamics. The experiment has been extensively used to compare the apo and holo forms of several NRs (Kojetin and Burris, 2013), and it has been used in conjunction with mass spectrometry to confirm the binding of novel prenylflavonoid ligands of FXR and to study the mode of interaction of guggulsterone with the FXR LBD (Liping Yang *et al.*, 2014; Yang *et al.*, 2016). Furthermore, HDX analyses can discriminate between differently graded agonists, and HDX profiles have been used to predict tissue specificity in a number of selective oestrogen receptor modulators (Dai *et al.*, 2008).

MX can play a central role in understanding structure-function relationships, as well as the minutiae of ligand, DNA and protein binding mechanisms. However, the technique can be supplemented with additional information from other structural and biochemical techniques, which provide relevant insight into the conformational features of a dynamic receptor such as FXR. Whilst structure determination of the FXR LBD with top candidate ligands, Jed441 and Jed561, can help to validate docking predictions, it's important to remember that the accuracy of the pose predictions does not determine the overall rank of the ligand. Ultimately, structural biology must be used synergistically with a combination of other cellular and molecular approaches to provide a complete picture of the complex mechanisms that arise from ligand binding and subsequently lead to FXR activation and regulation of its target genes.

Chapter 5.

Activity and Specificity of Lead Compounds

5.1 Introduction

FXR is a typical NR in both its structure and activation. NRs constitute a large family of ligand-activated transcription factors that share a similar modular structure and are known to regulate a plethora of vital processes involved in growth and development, cell differentiation and proliferation, and the maintenance of homeostasis (Renaud and Moras, 2000). Their underlying roles in the development of certain cancers and metabolic disorders, in addition to their ligand-dependent activation, has lauded NRs as obvious, attractive drug targets; and it is estimated that NRs represent 10-20% of global pharmaceutical targets (Ottow and Weinmann, 2008).

5.1.1 Nuclear Receptor Interactions with Coregulators

As established previously in Chapter 3, NRs have a highly modular architecture. The variable, disordered *N*-terminal region is followed by a conserved DBD, responsible for the recognition and binding of specific response elements on its target genes. The DBD is connected, via a variable hinge region, to the LBD, which is also highly conserved. The LBD has been described as the most functionally important domain, not only due to its ability to bind structurally diverse small molecule ligands, but also due to the fact that it contains the ligand-dependant AF2, as well as interaction surfaces for partner receptors and accessory coregulator proteins, which tightly regulate the initiation or repression of gene expression (Meyer *et al.*, 1989).

As described previously (Chapter 3), the LBD of NRs is made up of a 12 α -helix bundle whereby ligand binding in the hydrophobic binding pocket triggers a mousetrap-like mechanism. It is thought that the AF2, found on helix 12, is relatively mobile and flexible when the receptor is in an unliganded state. Agonist binding inadvertently repositions the AF2 in a compact position that precludes corepressor protein complexes and exposes interaction surfaces that have a high affinity for coactivator proteins, facilitating their recruitment. Previous studies have determined that this conserved AF2 helix is essential for the ligand-dependent transactivation of the receptor; mutations along this helix do not alter ligand binding or receptor dimerization, but hinder coactivator recruitment and the consequential transcriptional activation of target genes (Durand *et al.*, 1994). In addition to the AF2, a highly conserved lysine residue on helix 3 also was shown to be important in coactivator recruitment, and mutagenesis of this residue ablates receptor activation (Feng *et al.*, 1998). Together, helix 3 and helix 12, make the hydrophobic binding groove that interacts with leucine rich pentapeptide LXXLL motifs (where 'L' denotes a leucine residue, and 'X' is used to represent any amino acid), christened 'NR boxes', residing in distinct stretches of conserved sequences on the coactivator protein (Heery *et al.*,

1997). Whilst these leucine residues arrange themselves to face inwards into the cleft, making VdW interactions with hydrophobic sub pockets in the cleft of the NR, the lysine on helix 3 and a conserved glutamate residue on helix 12, both form a charge clamp by making hydrogen bonds with residues on the *N'* and *C'* termini of the coactivator main chain, anchoring it to the surface of the LBD and stabilizing the interaction (Renaud and Moras, 2000). Figure 5. 1 depicts the position of a steroid receptor coactivator (SRC2) peptide in the groove formed by helix 12 and helix 3 in the structure of FXR LBD (PDB:4QE6).

In the absence of an agonist, helix 12 is mobile and the extended position it adopts distorts the AF2 moieties and does not promote the formation of the surface exposed LXXLL binding groove. Likewise, antagonist binding does not support the structural reorganisation of helix 12 that is required for coactivator recruitment. Instead, apo- and inhibitor-bound receptors repress gene transcription by binding to corepressor protein complexes via interactions between a hydrophobic binding site on the LBD surface, which overlaps the binding site responsible for coactivator recruitment, and an extended signature motif, LXX(I/H)IXXX(I/L) (where 'I' represents isoleucine residues, and 'H' histidine residues), on the corepressor protein (Rosenfeld *et al.*, 2006). This extended binding motif, however, fits poorly into the hydrophobic cleft created by agonist binding, and the inability of a charge clamp to form, forces the corepressor protein to dissociate. The association between NRs, such as FXR, with coactivator or corepressor proteins determines whether the epigenetic changes required for gene transcription takes place or not.

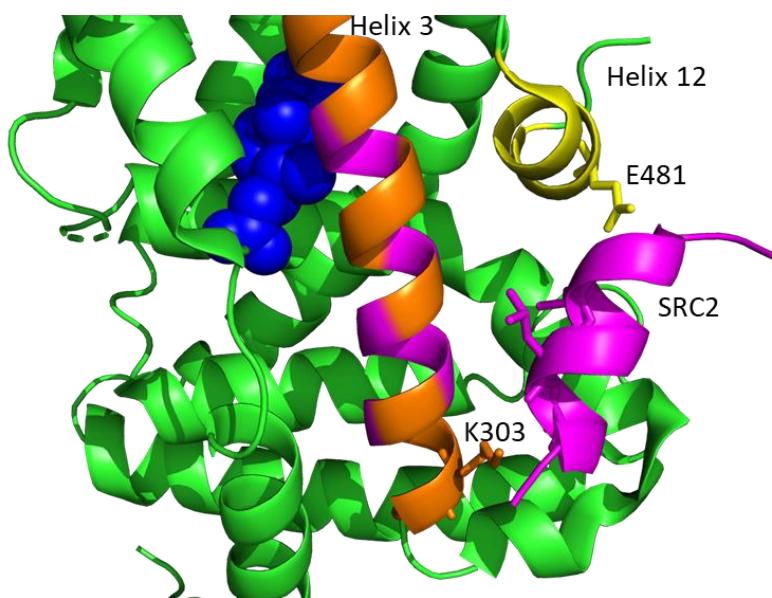


Figure 5. 1 Coactivator binding site on FXR LBD (PDB:4QE6). A peptide of motif 3 of the SRC2 coactivator (magenta) is recruited to the binding groove between helix 12 (yellow) and H3 (orange) and the Leucine residues (shown as sticks) interact with hydrophobic residues along this binding surface (magenta). Meanwhile Glutamate 467 on helix 12 and Lysine 303 on H3 form a charge clamp between the two ends of the coactivator chain.

5.1.2 Epigenetic Actions of Coregulators

Activation of gene transcription is dependent on the accessibility of target DNA, and the assembly of the preinitiation complex (RNA polymerase and general transcription factors). Under transcriptional silencing mechanisms, the highly ordered, condensed chromatin structure presents a physical obstacle to this core transcriptional machinery. However, coactivators induce the post-translational modifications responsible for chromatin remodelling and thus the binding of RNA polymerase to the promoter region of target DNA, and by mediating interactions with corepressor or coactivator complexes, NRs can act as a molecular switch to regulate the transcription of target genes.

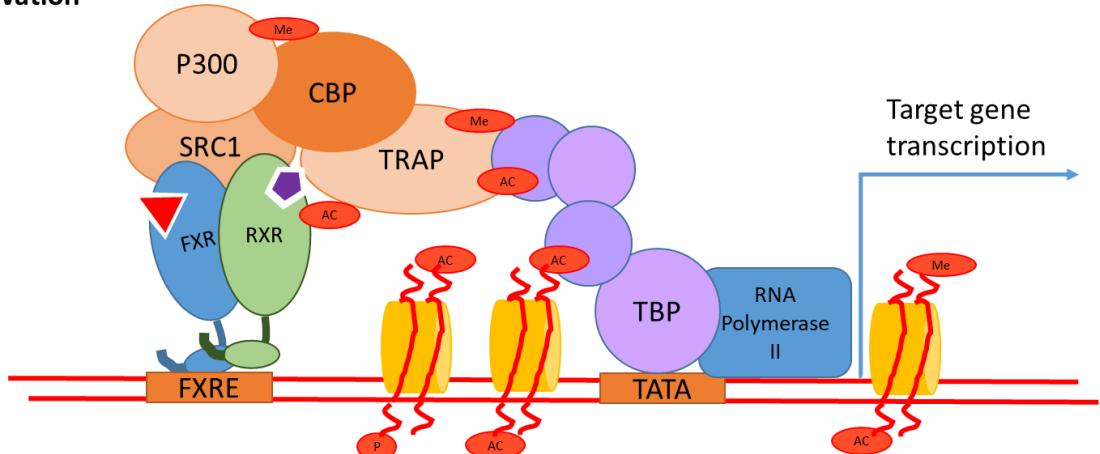
Coactivators can cause the post-translational modification of target proteins or undergo modification themselves, acting as either molecular chaperones, histone modification enzymes, chromatin remodellers or recruiters of other complexes that interact with basal transcriptional machinery (Wolf et al., 2008). One of the main classes of coactivator families is the p160 family, which includes Steroid Receptor Coactivator (SRC) 1, SRC2 and SRC3 (also known as Nuclear receptor coactivator, NCOA1, NCOA2, NCOA3, respectively). These possess intrinsic histone acetyl transferase activity, whereby the lysine-specific acetylation of histone tails causes a disruption to interactions between positively charged lysine side chains, and negatively charged DNA bases. This in turn disrupts nucleosome packing, leading to the relaxation of chromatin, thus making DNA more accessible to the preinitiation complex, resulting in more efficient gene expression. The CREB Binding Protein (CBP) and its homolog, p300, have also been shown to harbour histone acetyl transferase activity, however, it has been reported that it binds very weakly to certain nuclear receptors and requires other coactivators, such as SRC1, in order for recruitment to nuclear receptor-regulated promoters (Sheppard et al., 2001). Although it does not possess intrinsic histone modifying enzymatic activity, and is thought to have minimal transcriptional activity by itself, coactivators, such as peroxisome proliferator-activated receptor- γ coactivator alpha (PGC1 α), increase the affinity of certain transcription factors to additional acetyl transferase-displaying coactivators, such as SRC1 and CBP, increasing their transcriptional potential (Liang and Ward, 2006). As such, coactivators and other histone acetyl transferases have been shown to work in large, metastable, multi-protein complexes, whose composition is determined by the specific target DNA-binding sites and response elements within a genomic region, the particular composition of coregulators produced in a given cell type, and the physiological status of that cell type (Daramont et al., 1998).

Conversely, the repression of gene transcription, is generally mediated by chromatin condensation and the action of histone deacetyl transferases. It is the current view that corepressors, such as nuclear corepressor 1 (NCOR1) and silencing mediator of retinoic acid and thyroid hormone receptor (SMRT), act as both a platform to recruit multiple deacetyl transferases, and as a core component for the catalytic activity needed to modify both histone targets and other transcriptional regulators, such as p53 (Guenther *et al.*, 2002). Furthermore, it's thought that nuclear corepressors can directly bind to transcriptional activators and inhibit the assembly of transcriptionally active complexes (Aranda and Pascual, 2017).

In addition to the covalent modification of histones, other coactivators, such as thyroid hormone activating protein (TRAP) or vitamin D receptor interacting proteins (DRIP), by virtue of their LXXLL motifs, act to facilitate and enhance the function of other regulators, and are thought to be essential for the activation of NR mediated gene expression. Specific subunits of the TRAP complex, such as TRAP220, act as a bridge between ligand-activated NRs and their coactivators, and RNA polymerase II and other general members of the basal transcription machinery (Pandey *et al.*, 2005).

Similarly, nuclear receptor interacting protein (RIP140), has also been described to primarily act as a scaffold between NRs and chromatin remodelling enzymes. However, RIP140's regulatory function is thought to alternate between activating and repressing mechanisms dependent on the relative level of RIP140 in comparison with other coregulators, post translational modifications, and interactions with other transcription factors (Chung, 2013). RIP140 is thought to act as a corepressor in tissues involved in metabolism and energy expenditure, and as a coactivator in cells involved in innate inflammation, highlighting the shifting roles coregulators play in the tight epigenetic control of gene expression.

Activation



Repression

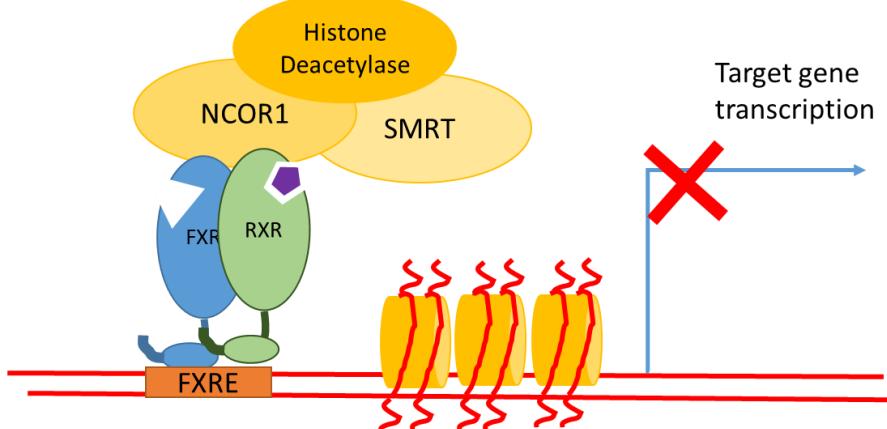


Figure 5. 2 General model of ligand activated NR regulation of target gene transcription. Nuclear receptors bind to response elements in the promoter of their target gene. Agonist mediated conformational changes facilitate the parallel or sequential recruitment of coactivator proteins (orange) which form large multi subunit complexes and cause the post translational modifications (red) of histone tails, other coactivators and other transcription factors (purple). This leads to chromatin remodelling and the assembly of the preinitiation complex at the promoter, ultimately leading to gene transcription. Apo receptor structures do not adopt a conformation which is able to displace the corepressor complex, and so histone deacetyltransferase enzymes covalently modify histone tails and recruit other transcriptional repressors to prevent target gene transcription.

5.1.3 Specificity of Nuclear Receptor-Coregulator Interactions

The assembly of coactivator proteins into large complexes, with the multi-functionary ability to affect various stages of the transcriptional process is, by definition, rate limiting in the activation of NRs. The assembly of these multi-subunit complexes is subject to the cellular concentration, localization, stability and protein-protein interactions of coactivators. These can often be regulated by post translational modifications of the coactivator proteins as a result of exogenous stimuli-derived signalling cascades (Aranda and Pascual, 2017). This strategy affords a degree of specificity to NRs, whereby their specific molecular actions are determined by the

cellular context and distributions of coactivator proteins; for example, a reduction in DRIP205 coactivator expression in HepG2 cells (by 50%) led to significant decreases in FXR-mediated regulation of the kininogen gene (Torra *et al.*, 2004). In addition, the physiological function of progesterone receptor has been shown to be differentially mediated by SRC1 in the uterus, and SRC3 in the breast, dependent on the protein stability of the coactivators (Han *et al.*, 2006); and the repressive actions of neuronal interacting factor X1 on NRs were limited to the brain by the highly restricted, spatiotemporally-regulated expression patterns of the protein (Greiner *et al.*, 2000).

Furthermore, despite the fact that coactivators have the potential to interact with several different NRs, it has been well documented that certain coactivators have relatively selective preferences for specific groups of NRs; for example, Androgen Receptor associated protein 70 (ARA70), which displays specificity for the Androgen receptor (Yeh and Chang, 1996). This specificity is thought to arise from differences in affinity of NRs for the residues immediately flanking the 'LXLL' motifs (Daramont *et al.*, 1998). Moreover, the affinity for particular coregulators may in part be defined by the structural conformation the receptor adopts as a result of binding to different classes of ligand (Ozers *et al.*, 2005; Han *et al.*, 2006). Vitamin D receptor was shown to differentially recruit either SRC1 or SRC2 depending of the ligand that was bound (Takeyama *et al.*, 1999). Similarly, different synthetic agonists of LXR were able to recruit coactivators distinctly from one another and were associated with differential gene expression (N. Li *et al.*, 2017).

This promiscuous mechanism of NRs, whereby different ligands can differentially recruit specific coactivators to exert its effects on particular groups of target genes does not exclude FXR. It has been described previously that FXR can activate target genes in a tissue-specific manner, thought to be mediated by tissue-specific patterns of coactivators or chromatin modifications (Thomas *et al.*, 2010). Moreover, a synthetic retinoid-derived compound was shown to exhibit FXR antagonist properties in *in vitro* reporter assays, but displayed differential responses dependent on the FXR-target gene *in vivo* (Dussault *et al.*, 2003). By harnessing this interplay between the receptor, ligand and coactivators, it may be possible to activate FXR in a gene-specific fashion, enhancing desired effects without promoting the negative effects associated with receptor activation.

5.1.4 Selective Modulation of Nuclear Receptors

Given the pleiotropic roles of NRs in a vast array of different tissues and organs, the clinical use of synthetic ligands for these proteins are usually associated with unwanted side effects. There

are countless examples where the clinical use of NR agonists is in decline due to their adverse effects. The use of Peroxisome Proliferator-Activated Receptor-gamma (PPAR γ) agonists, such as thiazolidinediones for the treatment for diabetes, has been reassessed due to their effects promoting weight gain, congestive heart failure and bladder cancer, and ligands have been developed for PPAR γ that exert post translational modifications of the receptor, independent of classic transcriptional agonism (Choi *et al.*, 2010; Cariou *et al.*, 2012). Accordingly, recent emphasis has shifted towards approaches based on selective receptor-coregulator interactions to target the specific transcriptional responses that are different to those seen by prototypical agonists.

5.2 Aims and Objectives

Whilst the use of *in silico* molecular docking and crystallographic structural studies provide essential insight into potential mechanisms of ligand binding, these do not provide conclusive information regarding the functional activation of the receptor. Published structures of the FXR LBD represent discrete conformations of the domain, and as such, do not identify subtle structural differences induced by different ligands, which may be sufficient to distinguish the full and partial agonists. Likewise, other methods are required to fully establish whether these novel ligands can induce the dynamic changes required for coactivator recruitment and thus receptor activation.

The aim of this work was to identify whether the lead compounds, Jed441 and Jed561, were bona fide ligands for FXR, in their ability to recruit LXXLL-containing coactivators. Perhaps more so than agonising FXR, it is imperative that these compounds act as selective modulators of the receptor and, as such, are able to mitigate some of the problems associated with global FXR activation. With the ability of these compounds to occupy the allosteric binding pocket, in addition to the canonical binding site (Chapter 3), a secondary aim was to determine whether Jed441 and Jed561 were sufficiently structurally distinct, enabling them to orchestrate the selective recruitment of coactivators to FXR, that they may potentially be able to propagate responses specific to a certain clusters of target genes.

5.3 Materials and Methods

5.3.1 LanthaScreen™ Assay (principle)

The LanthaScreen Time Resolved Fluorescence Resonance Energy Transfer (TR-FRET) FXR Coactivator assay kit (Invitrogen) was used with slight modifications to the manufacturer's protocol. The assay involves the energy transfer from a donor fluorophore to an acceptor fluorophore (Figure 5. 3). In this case, the fluorescent donor was the terbium-labelled anti GST antibody (which indirectly labels a recombinant, GST-tagged FXR LBD protein), and the acceptor, a fluorescein label on the coactivator peptide. When agonist bound FXR LBD changes conformation, the increase in affinity for a coactivator peptide in the helix 12 binding groove, and its subsequent recruitment, results in the close proximity between terbium and fluorescein. When terbium is excited at 340 nm, it is capable of transferring energy to the fluorescein acceptor, which in turn emits light at 520 nm, causing an emission shift from donor (495 nm) to acceptor (520 nm) emission. This energy transfer can be detected by the increase in fluorescence of the acceptor, fluorescein, and the decrease in fluorescence of the donor, terbium, and can be expressed as a ratio of intensities between the two fluorophores (Figure 5. 3). The use of a ratiometric measurement minimises interference and signal quenching by test samples. Furthermore, the use of time-resolved fluorescence utilizing the lanthanide element, terbium, which has a delayed decay in signal, ensures a prolonged emission of fluorescence beyond that of any autofluorescence, allowing the signal to be measured throughout a longer interval and without the interference of background signals.

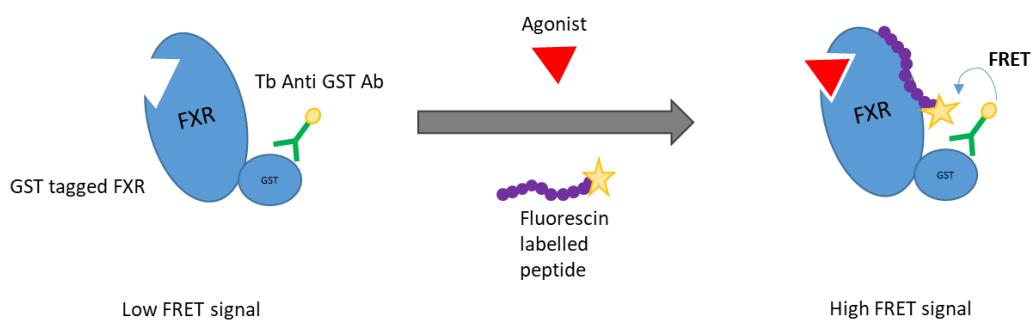


Figure 5. 3 Principle of the FXR agonist coactivator peptide recruitment assay. Terbium (tb)-anti GST antibody indirectly labels the FXR LBD by binding to the GST tag. Addition of an agonist causes the conformational change that results in the increased affinity for the fluorescein labelled coactivator peptide. Recruitment of the coactivator peptide to FXR LBD causes an increase in FRET signal due to the close proximity of the fluorescent donor (Terbium) and acceptor (fluorescein).

5.3.2 Compound Affinity Assays

The assay was performed in 384-well, black, round bottom plates (Corning) using 20 μ l reaction mixtures per well. Initial assays used reaction mixture components to the final concentrations

recommended in the manufacturer's protocol; varying concentrations of ligand in DMSO (1% (v/v) final concentration), 500 nM coregulatory peptide, 5 nM Terbium anti-GST antibody, 10 nM FXR LBD-GST. As recommended by the manufacturer, serial dilutions of each ligand were prepared in DMSO at 4x the final concentration needed for the assay. Initial assays used a 12-point, 3-fold serial dilution as per the example given in the manufacturer's instructions, whereas later assays utilized 5-point 10-fold dilutions to maximise the number of assays that could be carried out with each kit. Serially diluted ligands were then mixed with an equal volume of coregulator buffer G provided in the kit. A 10 μ l aliquot of ligand was added to each well first, however the coactivator/antibody/FXR LBD-GST solutions were premixed, also in coregulator buffer G, before a single 10 μ l aliquot was dispensed in each well. This was to minimise the time taken to set up each reaction and to minimise the number of additions needed for each well. Reactions were also carried out using DMSO alone (1% (v/v), without a ligand compound) to identify any ligand-independent coactivator recruitment, and also without FXR LBD-GST to provide an indication of autofluorescence and to provide an absolute bottom baseline for the assay. Each reaction was replicated in 4 wells to account for technical variation, and the average fluorescence readings were used. Each assay plate was run in triplicate with different batches of FXR LBD protein, to account for any biological variation between different purifications of the protein.

Plates were mixed briefly on a plate shaker and incubated at room temperature for 3 hours, protected from light. After incubation, TR-FRET fluorescence was measured, as an endpoint assay, using a FlexStation[®]3 Microplate reader (Molecular Devices) and an excitation wavelength of 332 nm and, detection of terbium and fluorescein emission signals at 488 nm and 518 nm, respectively. A 50 μ s delay time and 400 μ s integration was used, as recommended in the LanzaScreen module by the FlexStation[®]3 manufacturers.

5.3.3 Coactivator Affinity Assays

For assays assessing the affinity of the coactivator peptide for ligand-bound FXR LBD, a total of 28 different coregulator peptides was used. These peptides were derived from known coactivators and corepressors, and from similar sequences identified by random phage display (Invitrogen). Peptides were purchased from Invitrogen and a list of their sequences can be found in Table 5. 1. For these assays, a supersaturating final concentration, 10 μ M, of ligand was used. The FXR LBD-GST and anti-GST antibody concentrations remained the same as previous assays, 10 nM and 5 nM, respectively. A 10-point, 2-fold serial dilution of coactivator peptides was created by diluting the fluorescently labelled peptides in coregulator buffer G to concentrations 4x that of the final assay. Again, reagents were premixed to minimise the

number of additions to each well, and assays consisted of 4 replicate reactions, and 3 independent assays were used to determine average fluorescence. A 10 µl aliquot of ligand was added to each well, and a 10 µl aliquot of premixed coactivator/antibody/FXR LBD-GST was added. Plates were mixed by shaking and incubated at room temperature for 3 hours, protected from light and fluorescence measured as described above.

Table 5. 1 Sequences of fluorescein-labelled coregulator peptides

Peptide	Description	Fluorescein-labelled Sequence
Members of the p160 steroid coactivator family		
SRC1-1	Motif of SRC1 NR box 1	KYSQTSHKLVQLLTTAEQQL
SRC1-2	Motif of SRC1 NR box 2	LTARHKILHRLLQEGSPSD
SRC1-3	Motif of SRC1 NR box 3	ESKDQHQLLRYLLDKDEKDL
SRC1-4	Motif of SRC1 NR box 4	GPQTPQAQQKSLLQQQLTE
SRC2-1	Motif of SRC2 NR box 1	DSKGQTKLLQLLTTKSDQM
SRC2-2	Motif of SRC2 NR box 2	LKEKHKILHRLLQDSSSPV
SRC2-3	Motif of SRC2 NR box 3	KKKENALLRYLLDKDDTKD
SRC3-1	Motif of SRC3 NR box 1	ESKGHKKLLQLLTCSSSDR
SRC3-2	Motif of SRC3 NR box 2	LQEKHRLHKLLQNGNSPA
SRC3-3	Motif of SRC3 NR box 2	KKENNALLRYLLDRDDPSD
Other known coactivators		
CBP-1	Motif 1 of CREB-binding protein	AASKHKQLSELLRGGSQSS
TRAP220/DRIP-1	Motif 1 of Vitamin D Receptor interacting proteins	KVSQNPILTSLLQITGNGG
TRAP220/DRIP-2	Motif 2 of Vitamin D Receptor interacting proteins	NTKNHPMLMNLLKDNPAQD
RIP140 L6	Motif 6 of nuclear receptor interacting protein	SHQKVTLQLLLGHKNEEN
RIP140 L8	Motif 8 of nuclear receptor interacting protein	SFSKNGLLSRLLRQNQDSY
PPAR γ-specific coactivators		
PGC1a	PPAR gamma coactivator protein 1a	EAEEPSLLKKLLLAPANTQ
PRIP/RAP250	Motif of PPAR interacting protein	VTLTSPLLVNLLQSDISAG
Androgen Receptor preferring coactivators		
AR-N	Sequence from the N terminal of Androgen receptor involved in a ligand induced intradomain interaction with LBD	SKTYRGA FQNLFQSVREVI
ARA70	Motif of Androgen specific coactivator	SRETSEK FKLLFQSYNVND
Coactivator-like peptides from random phage display		
D22	Random phage display sequences (resembling RIP140, PGC1, DAX1, SHP)	LPYEGSLLL KLLRAPVEEV
C33	Random phage display sequences (resembling TRAP220, RIP140)	HVEMHPLL MGLLMESQWGA
EAB1	Random phage display sequences (interactions with ER)	SSNHQSSRL IELL SR

EA2	Random phage display sequences (interactions with ER)	SSKGVLWRMLAEPVSR
TB3	Random phage display sequences	SSVASREWVWVRELSR
Corepressor peptides		
SMRT ID1	Motif 1 of the Silencing Mediator for Retinoid and Thyroid Hormone Receptors (SMRT) corepressor	GHQRVVT LAQHISEVITQDYTRH
SMRT ID2	Motif 2 of the SMRT corepressor	HASTNMGLEAI IRKALMGKYDQW
NCOR1 ID1	Motif 1 of Nuclear corepressor 1	RTHRLIT LADHICQIITQDFARN
NCOR1 ID2	Motif 2 of Nuclear corepressor 1	NLGLEDIIRKALMG

* LXXLL motif shown in bold

5.3.4 TR-FRET Data Analysis

Data are shown as a TR-FRET ratio calculated by the signal at 518(520) nm/488(495) nm and graphs were plotted by subtracting background fluorescence (signal detected with no ligand, only DMSO in reaction mixture) from data. The data were fitted by non-linear regression, using the four-parameter log-logistic model equation for sigmoidal dose response (variable slope) in GraphPad Prism and EC₅₀ values also calculated automatically by GraphPad Prism (GraphPad Software Inc., San Diego, CA, USA). The screening window coefficient, 'the Z-factor', used to assess the precision and reliability of the assay performance, was calculated using data from assays with the highest agonist concentration and assays without any agonist present, according to Zhang et al (Zhang *et al.*, 1999). The Z-factor was above the 0.5 threshold for all assays and was taken to indicate that the assay was robust.

5.4 Results

5.4.1 The Affinity of Lead Compounds for FXR

Coactivator recruitment assays were used to examine and quantitate the agonist-dependent interactions between the GST-tagged FXR LBD and fluorescein-labelled coactivator peptides. Seven ligands were tested in coactivator assays. Natural ligand, CDCA, semisynthetic ligand OCA, non-steroidal competitor Tropifexor, and the compounds of interest, Jed441 and Jed561, were tested for their ability to recruit different coactivator peptides. The non-fluorinated derivative of Jed561, Jed692, also was included by way of comparing the compounds and the ability of the fluorine group at the 4 β position of the BA backbone to influence activity. The natural antagonist, guggulsterone also was used as a negative control and to test its ability to disrupt agonist-induced coactivator peptide recruitment. Initial assays were set up using the Lanthascreen FXR coactivator kit as recommended by the manufacturers; using the SRC2-2 peptide provided with the kit, carrying out a 12 point, 3-fold dilution series of each ligand. This was to obtain preliminary results regarding the affinity of each ligand and to determine a suitable range of ligand concentrations to use in subsequent assays. The results showed that all compounds were able to recruit the peptide SRC2-2, although the magnitude of the response was varied (Figure 5. 4). The sigmoidal dose-response equation was automatically used by the GraphPad software to determine the effective concentration (half the maximal response, EC₅₀) when the bottom of the curve was constrained to zero (Table 5. 2). The slope for Jed441 however, was an ambiguous fit, due to the absence of a sigmoidal response and improper curve, resulting in a very high confidence interval for the EC₅₀ value determined.

Disregarding Jed441 (due to the ambiguous response curve and dubious EC₅₀ value), Tropifexor was both the most potent compound, with an EC₅₀ value of 1.6 nM, and the most efficacious compound, with the highest maximal response at an emission ratio of approximately 0.4. This was followed by Jed692 and Jed561, which exhibited EC₅₀ values of 3.0 and 4.4 nM, respectively. OCA had a ten-fold higher EC₅₀ value at 36 nM, whereas CDCA had the lowest affinity for FXR LBD with an EC₅₀ value in the 10 μ M range. Interestingly, the steroid BA-derived compounds, OCA, Jed441, Jed561 and Jed692, had similar maximal responses, with emission ratios approximately 0.2, and although lower than the maximum emission seen with Tropifexor, these compounds are more efficacious than CDCA.

Assays using the antagonist guggulsterone did not produce the effect expected for an antagonist. Guggulsterone was unable to displace the 50 μ M CDCA from the ligand binding

pocket, as suggested by its inability to quench the fluorescent signal with an increasing concentration of compound. Instead, fluorescence emission ratio was observed to even rise slightly, although no sigmoidal response was seen.

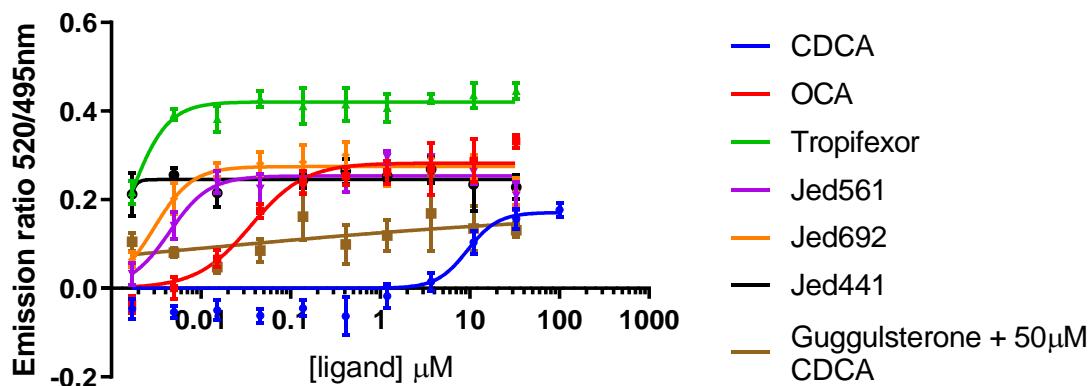


Figure 5.4 TR-FRET FXR Coactivator recruitment assay. Serial dilutions of various compounds (1% final DMSO concentration) were assayed for their ability to recruit SRC2-2 peptide to the FXR LBD. Curves were fitted using a sigmoidal dose-response equation (variable slope) in GraphPad Prism. (n=4 technical replicates, from one independent experiment, represented as mean \pm SEM).

Table 5.2 EC₅₀ values of compounds tested in FXR coactivator recruitment assays

Compound	EC ₅₀ \pm SEM (nM,)
CDCA	9,394 \pm 4,595
OCA	36 \pm 8.3
Tropifexor	1.6 \pm 0.15
Jed561	4.4 \pm 1.04
Jed692	3.0 \pm 0.53
Jed441	1.43 \pm 1,233

The ligands were then assayed for their ability to induce an association between FXR LBD and different fluorescein-labelled coregulator peptides in a dose dependent manner. Six different LXXLL-containing coactivators and one LXX(I/H)XXX(I/L)-containing corepressor were used for these assays, all at a final concentration of 500 nM. Increasing agonist concentrations are expected to increase the TR-FRET ratio seen with coactivator peptides, and decrease the ratio seen with corepressor peptides. As guggulsterone was unable to competitively inhibit FXR recruitment of SRC2-2 when 50 μ M CDCA was included in the reaction, assays here used guggulsterone alone to determine whether it behaved as an agonist or antagonist. Moreover, to further elucidate the effect of the fluorine atom, another compound, Jed678, a derivative of OCA with the single substitution of a fluorine at the 1 β position, also was assayed for its ability to activate FXR.

The maximum fluorescence ratios seen for each coactivator with each ligand are shown in Figure 5. 5. The results showed that none of the ligands were able to induce significant changes in fluorescence compared to DMSO-induced fluorescence, when using SRC1-1 and SRC3-1. Likewise, the addition of ligands to the reaction induced greater increases in maximal response in certain coactivator peptides, such as SRC2-2, compared to others, such as CBP-1, which displayed only mild increases. The maximum emission ratio in reactions containing DMSO only, was slightly higher when NCOR1 ID1 is used as the coregulator, and when the FXR agonists are included in the reaction, a very slight decrease in emission ratio was seen. The addition of guggulsterone to reactions was unable to significantly induce increases in fluorescence emission ratios with coactivator peptides, or significantly decrease the emission ratio seen with corepressor peptides.

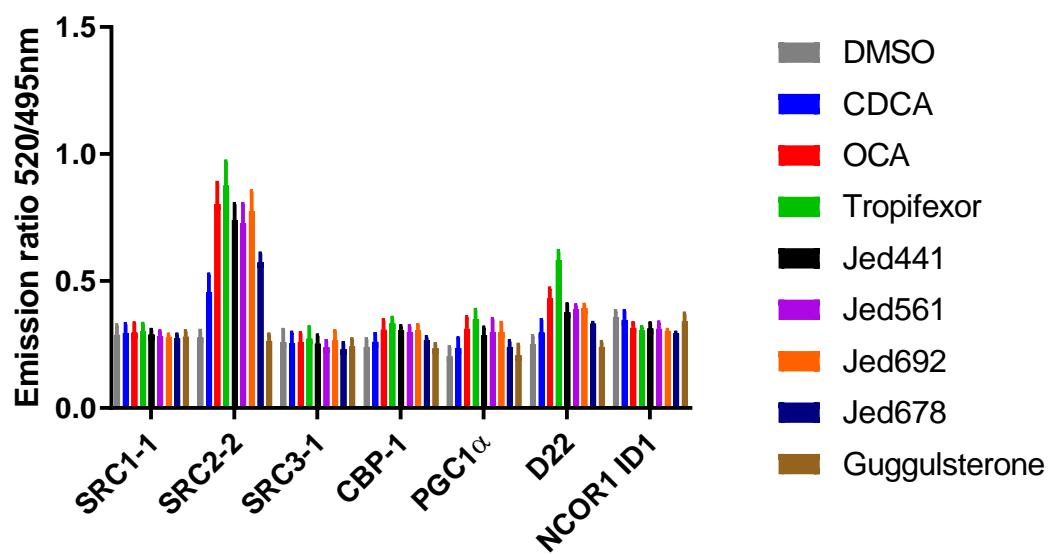


Figure 5. 5 Max response values determined by ligand-activated coactivator recruitment. Ligands were tested for their ability to recruit SRC1-1, SRC2-2, SRC3-1, CBP-1, PGC1 α and D22 coactivator peptides to FXR LBD resulting in an increase in fluorescence emission. Ligands also were tested for their ability to promote the dissociation of NcoR1 resulting in the decrease in fluorescence emission ratio ($n=3$ independent assays, represented as mean \pm SEM).

Dose-response curves were plotted without background fluorescence, and only depicting agonist-induced fluorescence changes (Figure 5. 6). Results of these dose-response curves again depicted that neither SRC1-1 nor SRC3-1 could be recruited to the FXR LBD in any of the assays with any ligand concentration. Again, the results reflect the max response data, whereby assays with NCOR1 ID1 do not display any increases in fluorescence and instead showed a slight decline in emission ratios with the increasing addition of agonists such as OCA and Tropifexor (Figure 5. 6B and C). Assays with CDCA confirmed the initial assays, showing that this natural agonist has

both a low efficacy and low potency, and was only able to considerably recruit SRC2-2 at the highest concentration tested 10 μ M (Figure 5. 6A). Interestingly, Jed678 displayed poor coactivator recruitment in comparison to its non-fluorinated original compound, OCA (Figure 5. 6G and B, respectively). Concentrations of Jed678, were needed at 1 μ M or more, to recruit SRC2-2 and D22 peptides, and the max emission ratios achieved were also fairly low in comparison to assays with OCA. As with the other ligands, Tropifexor was able to induce higher emission ratios when SRC2-2 was being recruited (Figure 5. 6C). For SRC2-2 recruitment, there was a very sharp increase in fluorescence ratio between 1 and 10 nM concentrations of Tropifexor. However, the increase in fluorescence emission, pertaining to D22 recruitment, is much more gradual between these concentrations of Tropifexor. Compounds Jed441, Jed561 and Jed692 displayed very similar recruitment profiles to one another (Figure 5. 6D, E and F, respectively). All compounds recruited SRC2-2 to produce the highest response in fluorescence, but whilst Jed441 and Jed692 displayed a gradual recruitment and steadily increasing sigmoidal response, Jed561 induced a sharper increase in SRC2-2 recruitment between 1 and 100 nM concentrations.

Further examination of the EC_{50} values, extrapolated from these dose-response curves (Table 5. 3), suggested that all compounds able to induce measurable increases in fluorescence emission, had relatively similar EC_{50} , regardless of which coactivator was being used in the assay. Whilst Tropifexor had the lowest values across all the assays, Jed441 and Jed561 appear to have a higher affinity for FXR than OCA, as seen by the 2-10-fold lower EC_{50} values for these compounds. Moreover, these assays suggested that Jed561 had an even higher affinity for FXR than non-fluorinated derivative, Jed692. The general rank order of EC_{50} values is Tropifexor> Jed561> Jed441> Jed692> OCA> CDCA> Jed678. With the exception of Jed678, novel BA-analogues were seen to perform better than classic and semi-synthetic BA agonists.

Table 5. 3 EC₅₀ values of compounds tested in FXR coactivator recruitment assays with different coactivator peptides

Compound	EC50 ±SEM (nM)			
	Recruitment of SRC2-2	Recruitment of CBP-1	Recruitment of PGC1α	Recruitment of D22
CDCA	1824 ± 127	1019 ± 336037	1360 ± 12522	1087 ± 257
OCA	92 ± 0.02	86 ± 0.03	123 ± 0.05	88 ± 0.02
Tropifexor	9 ± 5.4	8 ± 0.01	9 ± 4.1	7 ± 0.01
Jed441	19 ± 0.01	23 ± 0.01	27 ± 0.01	35 ± 0.02
Jed561	12 ± 5.7	12 ± 0.77	12 ± 14.9	11 ± 3.06
Jed692	22 ± 0.01	23 ± 0.01	33 ± 0.01	22 ± 0.00
Jed678	2603 ± 1.38	934 ± 417	1052 ± 45	1702 ± 319

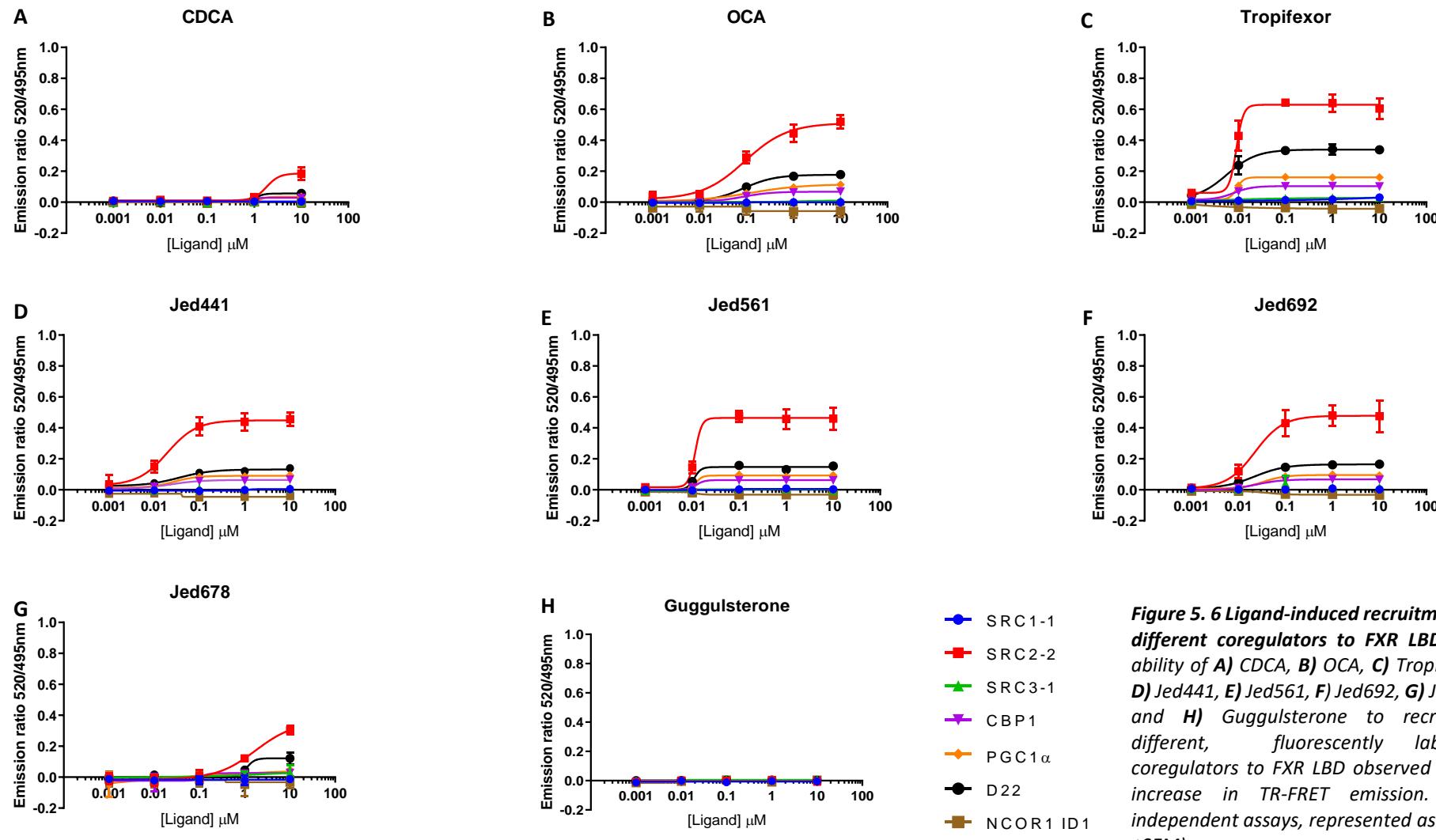


Figure 5.6 Ligand-induced recruitment of different coregulators to FXR LBD. The ability of **A) CDCA**, **B) OCA**, **C) Tropifexor**, **D) Jed441**, **E) Jed561**, **F) Jed692**, **G) Jed678** and **H) Guggulsterone** to recruit 7 different, fluorescently labelled-coregulators to FXR LBD observed by an increase in TR-FRET emission. ($n=3$ independent assays, represented as mean \pm SEM)

5.4.2 Jed561-Induced Interactions Between FXR LBD and Coregulators

To further elucidate the interactions made by FXR LBD with different coregulators when bound with the lead compound, Jed561, similar assays were used with variable ligand concentrations, but constant amounts FXR LBD, terbium labelled antibody, and with the addition of supersaturating concentrations of different fluorescently labelled coregulator peptides. The assays utilized twenty-eight different peptides containing the relevant coactivator or corepressor motifs, derived from naturally occurring coregulator proteins or from the random phage display of peptide sequences that resemble the LXXLL motif and flanking regions (Invitrogen). The results showed that Jed561 was able to promote a conformational change in FXR that allowed the recruitment of most of the coactivators (Figure 5.7). As expected, Jed561-bound FXR did not recruit any of the corepressor peptides, confirming that the compound is an agonist. Furthermore, it did not recruit any of the peptides thought to be specific for other nuclear receptors (Figure 5.7C).

Again using the non-linear fit function in GraphPad to determine EC₅₀ values, it was seen that regardless of which coactivator was used, the majority of EC₅₀ values were in the low nanomolar range, between 9 and 12 nM (Table 5.4). Comparing the agonist-specific fluorescence emission to DMSO control autofluorescence, it also was observed that the magnitude of fluorescence emission ratios was coactivator-dependent (Figure 5.8). In the presence of Jed561, assays with the peptide SRC2-2, as seen previously, displayed the highest change in emission ratios, with an increase of 0.45, whereas the other coactivators had agonist-specific emission ratios of less than 0.3. Other coactivator peptides recruited to high levels were SRC1-2, TRAP220/DRIP-1, RIP140 L8 and C33. For subsequent assays only SRC2-2, TRAP220/DRIP-1 and RIP140 L8 were used as these each represent a different class of coactivators and are all derived from naturally occurring coactivator proteins.

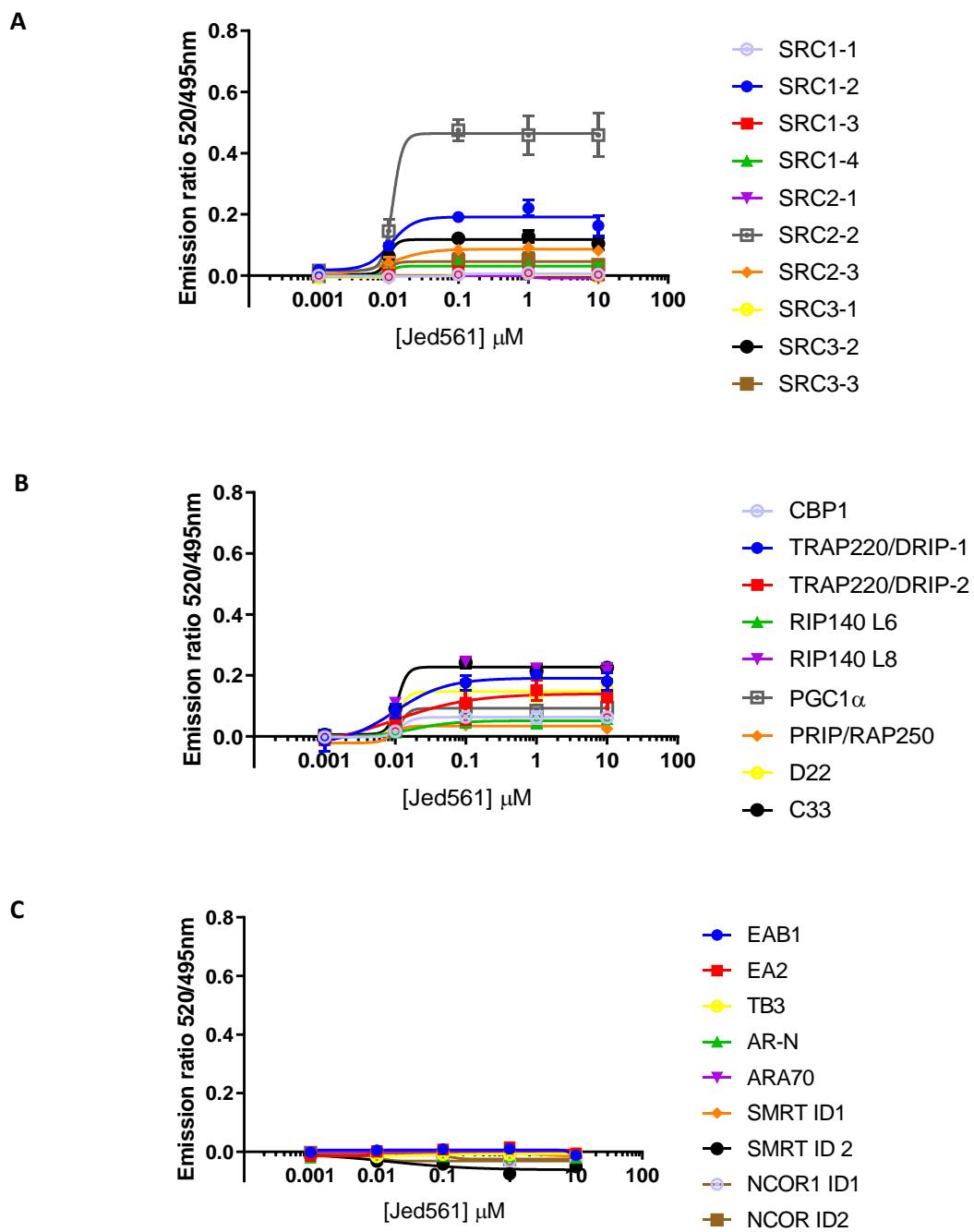


Figure 5. 7 The effect of different NR coregulator peptides on Jed561 induced TR-FRET. Jed561 was tested for its ability to stabilise FXR LBD in a conformation that could recruit **A)** coactivator peptides from the p160 family, **B)** coactivator peptides from other coactivator subclasses or random phage display, or **C)** coactivator peptides known to specifically bind to other NRs or known corepressor peptides. ($n=3$ independent assays, represented as mean \pm SEM).

Table 5. 4 EC₅₀ values for Jed561 in FXR coactivator recruitment assays with different LXLL-containing peptides

Coactivator Peptide	EC ₅₀ ± SEM (nM)
SRC1-2	10.7 ± 5.1
SRC1-4	10.5 ± 5.6
SRC2-2	11.5 ± 5.7
SRC2-3	11.2 ± 3.5
SRC3-1	9.5 ± 0.0
SRC3-2	9.9 ± 6.3
SRC3-3	9.6 ± 12.9
CBP-1	12.0 ± 0.0
TRAP220/DRIP-1	9.9 ± 3.6
TRAP220/DRIP-2	10.1 ± 18
RIP140 L6	17.4 ± 12.1
RIP140 L8	10.0 ± 3.7
PGC1 α	12.1 ± 15.0
PRIP/RAP250	9.0 ± 11.0
D22	11.0 ± 3.1
C33	10.8 ± 4.7

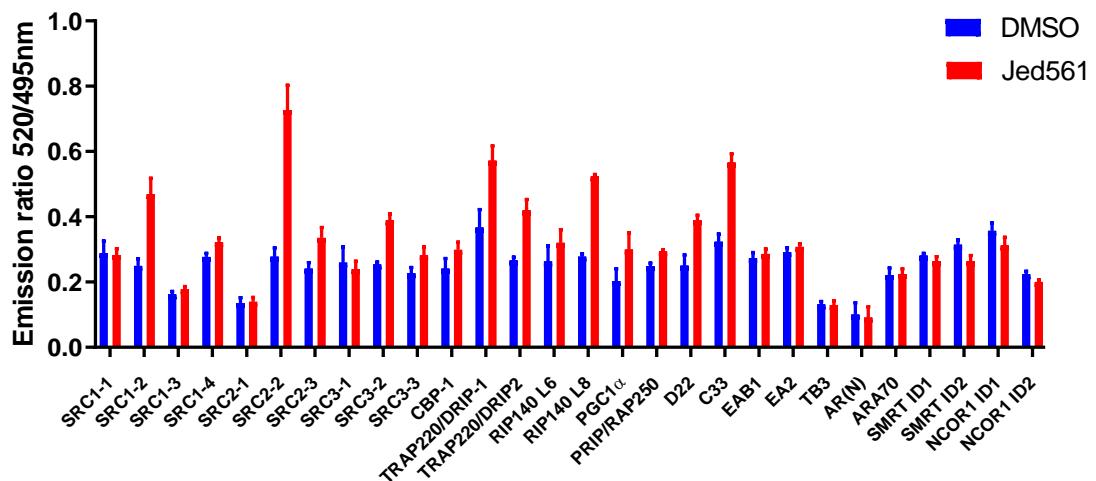


Figure 5. 8 Maximal fluorescence emission ratios in assays with different NR coregulator peptides in the presence or absence of Jed561. Comparisons of emission ratios seen in FXR LBD recruitment assays with unliganded (DMSO) or Jed561-bound receptor. Increases in fluorescence emission ratios beyond background DMSO levels indicate the recruitment and close proximity of fluorescently labelled coactivator peptides. (n=3 independent assays, represented as mean ± SEM).

5.4.3 The Affinity of Different Coactivators for FXR LBD

In order to assess the affinity of different coactivators for the FXR LBD, and to determine whether there are any differences dependent on the ligand that occupies the binding pocket, assays were set up, differing in the use of a titration of coactivator peptides and a saturating concentration of ligand. The emission ratios measured were plotted against the coactivator peptide concentration (Figure 5. 9). Unfortunately, the point of maximal activation, where the FXR LBD would be completely saturated with coactivator peptide and the fluorescent signal would be expected to plateau, was not achieved at the concentrations used. Furthermore, this meant that EC₅₀ values could not be accurately extrapolated from the data, and a quantitative descriptor of coactivator affinity could not be determined. The assays did, however, display slight differences in the concentration-response profiles dependent on the compound used in the assay.

When OCA occupied the LBD, coactivator recruitment emission ratio slopes for SRC2-2, RIP140 L8 and TRAP220/DRIP-1 almost appear to be parallel, suggesting that they may plateau at similar concentrations, which could suggest that, the affinities for FXR will be similar for all 3 peptides (Figure 5. 9A). Tropifexor on the other hand, shows, initially, very similar recruitment of the 3 coactivator peptides at most of the concentrations tested. However, at 2560nM, the emission ratios for RIP140 L8 and TRAP220/DRIP-1 begin to plateau, whereas SRC2-2 continues to rise. This suggests when complexed with Tropifexor, FXR LBD may have a slightly higher affinity for these two compounds, despite them being recruited to lower levels than that of SRC2-2 (Figure 5. 9B). Likewise, lead compounds, Jed441 and Jed561, also display their own unique coactivator recruitment profiles. Jed441-bound FXR was able to recruit SRC2-2 and RIP140 L8 to higher levels than TRAP220/DRIP-1 at the concentrations tested, and although the response for SRC2-2 was below than that seen for RIP140 L8 at lower concentrations, by 2560nM, the fluorescence emission for SRC2-2 starts to rise more dramatically (Figure 5.9C). In assays with Jed561, SRC2-2 recruitment began to reach a saturating level at 2560nM, whereas RIP140 L8 and TRAP220/DRIP-1 did not, again suggesting that FXR-Jed561 complexes may have a higher affinity for SRC2-2 than other coactivators (Figure 5.9D).

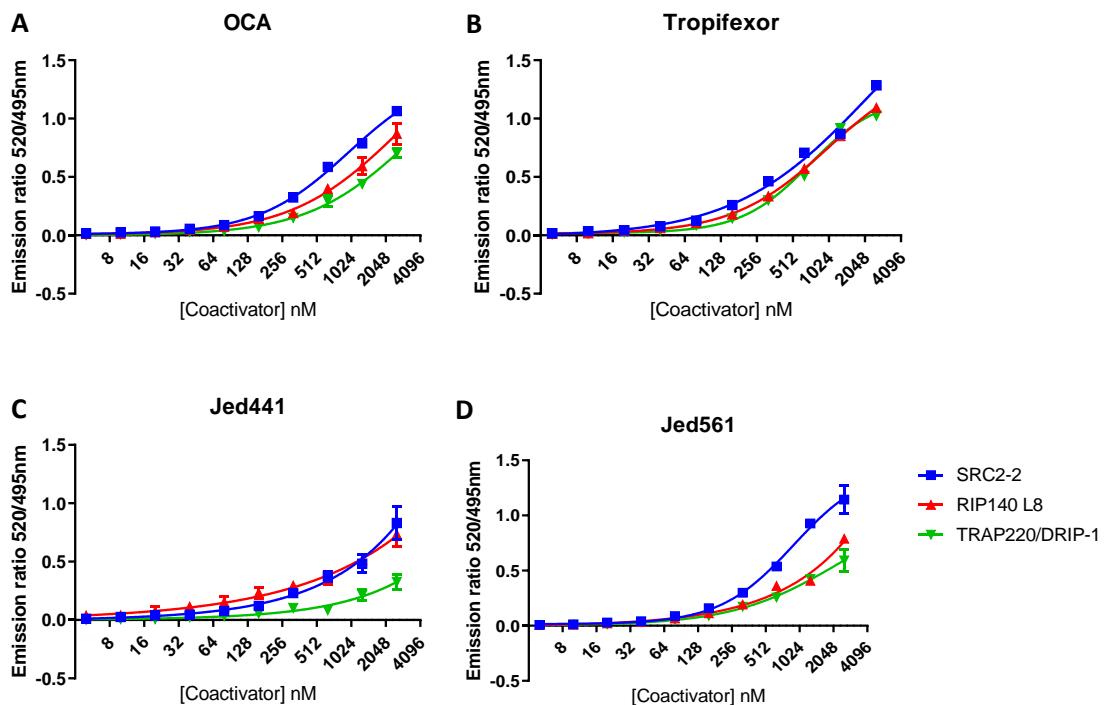


Figure 5.9 The effect of ligands on the affinity of FXR LBD for different LXLL-containing coactivator peptides. The affinity of FXR LBD for SRC2-2, RIP140 L8 and TRAP220/DRIP-1 was tested when **A**) OCA, **B**) Tropifexor, **C**) Jed441 and **D**) Jed561 occupied its binding site. ($n=3$ independent assays, represented as mean \pm SEM).

5.5 Discussion

5.5.1 Lead Compounds are Bona Fide Agonists for FXR

TR-FRET based coactivator assays were used to determine the ability of the proposed FXR agonists to activate the receptor by promoting the recruitment of coactivator peptides. The method, which could be used in a high throughput format, provided a sensitive and robust means to easily quantify the affinity of different compounds for FXR. Results from initial assays, which intended to simply determine EC₅₀ values, and gain an indication of an optimal range of ligand concentrations, supported data originally described in the literature, whilst also providing further insight about the novel compounds. Initial assays were based on the recruitment of a peptide that resembled the NR box 2 of the SRC2 coactivator protein (SRC2-2). The EC₅₀ value of 9.5 µM determined for CDCA was comparable to the 11.7 µM value previously reported by Bramlett and colleagues in similar assays recruiting SRC1 and using increasing concentrations of CDCA (Bramlett *et al.*, 2000). On the other hand, the EC₅₀ value for OCA at 36 nM was lower than the previously described value of 100 nM (Pellicciari *et al.*, 2002). However, these differences may have arisen from the use of alternative coactivator peptides, or from the use of different fluorescence labels. Moreover, the EC₅₀ values determined in this assay, are a composite value and are influenced by multiple equilibria, involving the ligand binding to the receptor as well as the peptide binding to the receptor-ligand complex. Therefore, values determined here may be different to EC₅₀ values determined by other means.

Nevertheless, the assays here determined that both Jed441 and Jed561 had 10-fold higher affinities for FXR than OCA. The initial EC₅₀ results, suggested that Jed441's affinity for FXR was akin to the non-steroidal agonist, Tropifexor, which previously has been determined to have EC₅₀ values in the sub-nanomolar range (Tully *et al.*, 2017); however as the basal plateau of the response curve was out of range, the EC₅₀ values estimated herein are not as accurate. Tropifexor, however, was indeed shown to induce a higher magnitude of response in comparison to the steroidal compounds, suggesting that BA-derived compounds may only display partial agonism of the receptor, unlike Tropifexor which may be a full agonist for the recruitment of SRC2-2.

Assays with different coregulator peptides, further confirmed the role of lead compounds Jed441 and Jed561 as agonists of FXR. Both compounds were able to stabilise FXR in a conformation whereby SRC2-2, CBP-1, PGC1α and D22 were recruited, resulting in an increase in fluorescence beyond the levels of DMSO control-induced background fluorescence. Neither

SRC1-1 nor SRC3-1 could be substantially recruited by Jed441, Jed561 or any of the other agonists currently available, implying that the FXR LBD does not have any affinity for these particular peptide sequences, regardless of the conformation it adopts. Coactivator NR box sequences have a sequence identity of around 50%, and the affinity of NRs for specific coactivator motifs is dependent on the recognition of sequences flanking the LXLL residues (Daramont *et al.*, 1998). Both peptides have several positively charged residues immediately preceding the LXXLL motif, and several polar residues immediately after this region, and it is possible that FXR has a low affinity for these neighbouring residues.

The fact that neither Jed441 nor Jed561 were able to promote increases in fluorescence when the corepressor NCOR1 was used, further confirms the idea that the conformation adopted by FXR when bound to these compounds, is one of an activated receptor. Moreover, there is a very slight decline in fluorescence in NCOR1 assays, with an increase in both Jed441 and Jed561 concentrations, suggesting that these compounds also are promoting the dissociation of corepressors from the LBD surface.

5.5.2 Effect of Fluorination on Activity of the Novel Compounds

The addition of fluorine atoms to the BA-scaffold was proposed to affect the activity and/or the stability of the novel compounds. Furthermore, the addition of a fluorine moiety to these compounds adds a unique chemical signature which can be patent protected. Despite showing slight disparities in docking results (Chapter 3, section 3.4.2.2.4), in which the binding pose and orientation of the ligand was dependent on the addition and position of fluorine atoms on the BA backbone, *in silico* studies could not discern whether fluorination had any effect on the activity of these compounds. As such, Jed678, an OCA derivative with a fluorine at the 1 β position, and Jed692, an analogue of Jed561, were both used in these assays for comparison.

Results from these assays suggested that the addition of a fluorine at the 4 β position may not be significantly advantageous for the activity of the compounds. Initial assays suggested that Jed561 and its non-fluorinated version, shared very similar affinities for FXR, and very similar efficacies in the recruitment of SRC2-2. Moreover, Jed692 even appeared to have a slightly higher maximal response and lower EC₅₀ values (Table 5. 2). Interestingly, subsequent assays recruiting different coactivator peptides, suggesting that Jed561 had a slightly higher affinity for FXR (Table 5. 3). However, as fewer ligand concentrations were used in these secondary assays, resulting in fewer data points, especially in the 10 to 100 nM range, may mean that the non-linear regression curves, and therefore the EC₅₀ values generated, are slightly less accurate than those values determined in the early assays.

The specific positioning of the fluorine along the BA backbone may have significant effects on the activity of the compounds. Comparisons between OCA and its 1β F derivative, showed that whilst OCA was able to promote the recruitment of various coactivator peptides with EC_{50} values between 86 and 128 nM, Jed678 was only able to recruit the peptides at much higher concentrations, and consequently had EC_{50} values between 1 and 2.5 μ M (Figure 5. 6, Table 5. 3). The substantially decreased affinity of Jed678, suggests that fluorination at the 1β position may in fact be inhibiting the binding of these compounds to the FXR ligand binding pocket. While results are inconclusive about the effect of fluorination on the A ring of the BA, it is still possible that the superior affinity of Jed441 and Jed561 may be a result of chemical groups in their modified carboxyl ‘tails’.

5.5.3 Different Compounds Display Varying Degrees of Efficacy

Initial assays with different coregulator peptides identified Jed561 as the most consistently potent novel compound in the recruitment of different coactivators. To assess the full extent of the ability of Jed561 to induce an active conformation of FXR, allowing the recruitment of different coactivators, 28 different coregulator peptides were used. The results further established the role of Jed561 as an agonist of FXR, with the ability to promote the interaction with several coactivators from different coactivator subclasses, and the ability to dislocate corepressor associations with the FXR LBD. Again, all of the EC_{50} values, determined with each of the coactivators, were in the low nanomolar concentrations, ranging from 9 to 17.5nM (Table 5. 4). With the format of this assay, whereby the ligand is titrated, EC_{50} values are indicative of the affinity of the FXR LBD for the ligand. Hence, the affinity of FXR for Jed561 is relatively consistent, regardless of the coactivator being recruited.

Nevertheless, it was observed that the magnitude of fluorescence was dependent on the coactivator being recruited. Results in Figure 5. 8, comparing the maximum fluorescence emission ratios arising from the apo receptor with the Jed561-bound LBD, shows innate differences in fluorescence depending on the coactivator used. The emission ratios seen with DMSO-control across all assays, show varying magnitudes with different coregulator peptides. These values describe the ligand-independent background fluorescence of these assays, and the differences seen may be, in part, due to slight differences in intensity of the fluorescence labels on the peptides themselves. On the other hand, previously reported structural and computational studies have implied an auxiliary coactivator binding groove directly adjacent to the canonical site (Nettles and Greene, 2003; Costantino *et al.*, 2005). Although in the normal cellular context, the role of this second binding site remains unknown, it has been proposed that coactivator motifs may be able to bind into this site prior to receptor activation. As such,

some of the background fluorescence observed may be attributed to the recruitment of the coactivators in the absence of a ligand.

The deduction of this autofluorescence from the graphs plotted in Figure 5. 7, indicate that there are differences in the ligand-specific response of each coactivator. Jed561 produces the highest response for the recruitment of SRC2-2, whereas its recruitment of the other coactivators only produced a partial response. It is possible that Jed561 is inducing a conformation of FXR LBD in which the coactivator binding groove formed is most complementary to the size, shape and electrostatic composition of the SRC2-2 peptide. However, as the other compounds also recruited SRC2-2 to a higher extent than the other peptides in previous assays (Figure 5. 6), it also is feasible that the FXR LBD itself has a higher affinity for the flanking regions of the SRC2-2 NR box. The SRC2-2 peptide is capped by hydrophobic residues at both ends, and there is a hydrophobic residue immediately preceding the NR box, which may enhance the affinity of the peptide for the hydrophobic binding groove. SRC2-2 previously has been reported to bind to other NRs with an affinity comparable to the full length SRC2 protein (Huang *et al.*, 2010), and it is possible that the higher magnitude of response seen with SRC2-2, is due to this intrinsic superior affinity of SRC2-2 for NRs in general.

In light of this, more assays were required to try to fully establish if there is any ligand dependency in the interactions between the FXR LBD and different coactivator peptides.

5.5.4 Coactivator Affinity for FXR may be Dependent on the Ligand Occupying the Binding Pocket

Titrations of coactivator peptides in assays with a saturating concentration of ligand, were used to determine the affinity of FXR LBD for each coactivator, when different ligands occupied its binding pocket. SRC2-2, RIP140 L8 and TRAP220/DRIP 1 were used as representatives of different coactivator subclasses. Although a quantitative measure of affinity could not be determined in these assays, the diverse recruitment profiles observed with the different compounds, suggests that the varying degrees of affinity of FXR for the coactivators, may be dependent on the compound used to activate the receptor. Interestingly, each compound displayed a unique recruitment profile, and there were no obvious distinctions between the responses of the steroid or non-steroidal nuclei of the ligand scaffolds. Likewise, the 3 BA-derived ligands had diverse coactivator recruitment profiles, suggesting that the conformations adopted by the FXR LBD, and thus the coactivators being recruited, after binding to Jed441 or Jed561, may be significantly different from the conformation induced by OCA. Although recruitment levels between the three coactivator peptides were very similar, again, it was seen that SRC2-2 generated a higher emission ratio, even at the same concentrations of RIP140 L8

and TRAP220/DRIP 1 peptides. This echoed the results seen in the previous assays where SRC2-2 peptides also produced the maximal fluorescent emissions, again supporting the idea that the SRC2-2 peptide is the most complementary coactivator peptide in terms of physical shape, stereochemistry, and biochemical affinity.

Although initially displaying a very gradual increase in fluorescence, the response of SRC2-2 in Jed441-occupied FXR assays, increases rapidly by the highest concentrations of the peptide, where it overtakes the response seen with RIP140 L8. This highly sensitive response may be consistent with a cooperative mechanism whereby coactivators are recruited in a synergistic, cumulative manner, and where the binding of one coactivator may enhance the affinity of a second. The presence of a second coactivator binding site may help to explain this observation. It has been previously observed that BAs can potentially interact with the second coactivator binding cleft through their carboxylic side chains (Pellicciari *et al.*, 2006), and with the extended side chains of our novel compounds such as Jed441, it is feasible that there is some impact on this ancillary coactivator site. The relationship and cooperativity between these two sites is yet to be determined, as some molecular dynamics simulations have suggested that the helix 12 agonist position obstructs the second site and causes the coactivator bound there to dissociate (Costantino *et al.*, 2005). Nonetheless, synergistic recruitment of SRC and other coactivators has been established previously, through the use of cell-free transcription systems (Liu *et al.*, 2001), and although the cooperative mechanism is thought to occur via interactions between protein interfaces on the coactivators themselves, the secondary recruitment of coactivators to the LBD surface may still be possible.

The assays showed that FXR bound with Tropifexor may have a higher affinity for RIP 140 L8 and TRAP220/DRIP-1, whereas Jed561 appeared to have a preference for SRC2-2 (Figure 5. 9). SRC2-2 has intrinsic histone acetylase activity, whereas RIP140 and TRAP220/DRIP are responsible for the recruitment of RNA polymerase, other chromatin remodelling enzymes and additional transcription factors. As such, these results suggest that there may be differences in the post-translational modifications of target gene promoters, and therefore the genes being transcribed, determined by the differential activation of FXR by Tropifexor, Jed441 or Jed561. Moreover, comparisons of mRNA expression in the Expression Atlas database (EMBL-EBI), indicate that there are higher levels of mRNA for SRC2 than the others in the liver (Appendix 4). This in turn, suggests that it may be possible to specify the actions of Jed561 to preferentially target FXR-mediated genes in the liver, as opposed to other tissues.

Further assays are required to fully clarify the extent to which lead compounds can preferentially recruit different coactivators and orchestrate specific actions of FXR signalling. Supplementary to the TR-FRET based assays used here, other techniques may help to overcome some of the experimental challenges involved in measuring NR-coactivator interplay. In addition to the static modelling of receptor-ligand interactions, computational approaches may be useful in accurate modelling simulations and analyses of dynamic conformational changes and subsequent interactions with coactivators. Nanosecond timescale molecular dynamics previously have been utilized to study different conformational states of the FXR LBD, unliganded, or in complex with OCA and different coactivator peptides (Costantino *et al.*, 2005), and although computationally expensive, may be a way to identify the exact mechanisms by which the FXR LBD under different conformations, recruits different coactivators. Alternatively, other *in vitro* biochemical experiments such as pull-down assays, or the previously mentioned, cell-free transcription assays, may help to identify the interactions between ligand-bound FXR and other coactivator proteins or transcription factors.

5.5.5 Conclusion

The results in this chapter confirm that lead compounds, Jed441 and Jed561, can activate FXR at concentrations lower than OCA. Whilst fluorination at the A ring may or may not be directly responsible for this enhanced potency, fluorine groups and other chemical moieties at the distal end may afford these compounds an improved selectivity for different coactivators compared to OCA. Whilst the compounds could recruit different coactivators to varying degrees, additional data is required to fully elucidate whether these compounds can preferentially promote the recruitment of certain coactivators over others, and to further determine whether this coactivator selectivity can proffer any tissue, cell or promoter specific activity to the novel compounds. Whilst these FRET-based assays were robust and simple to perform, their simplistic nature does not take into consideration several other factors affecting receptor activation and integration.

Firstly, these assays use a single LXLL-containing peptide, and do not consider the full-length protein of the coactivators, in which several pentapeptide motifs may exist. As such, full length coactivators may display a higher affinity for FXR than a single NR box, and the recruitment of one motif region, may affect the affinity of other motifs within the same coactivator protein (Wong *et al.*, 2001). Furthermore, as discussed briefly earlier, the full length coactivators have been shown to cooperatively interact with other coactivators to enhance binding and signal propagation (Liu *et al.*, 2001). Likewise, some coactivators are thought to utilise one of their NR box motifs for binding FXR, whilst using another motif further along the protein, for interactions

with RXR, which also has a high-affinity binding site for LXXLL sequences (Nettles and Greene, 2003; de Vera *et al.*, 2017)..

Although FXR interactions with its partner receptor, RXR, have not been considered here, RXR is thought to be an additional component in the regulation of FXR's transcriptional activity. FXR is considered a 'permissive' receptor, whose transcriptional activity can be stimulated by the binding of 9-*cis* retinoic acid to its cognate receptor, RXR, which can have an additive or synergistic effect. It is also thought that dimerization with RXR itself, can allosterically alter the FXR LBD, by stabilizing it in an active conformation and altering the coactivator binding site to promote coactivator binding (Zheng *et al.*, 2018). Accordingly, characterizing interactions and signal integration between FXR and RXR, may help to further fine tune FXR-mediated signalling.

Similarly, contributions from both the *N*-terminal domain and DBD have been excluded here. Although still poorly understood, members of the p160 coactivator family, have been shown to interact with the AF1 domain in the *N*-terminal region of several NRs. Coactivators have been shown to interact with the AF1 domain through glutamine-rich residues rather than their LXXLL motifs; and modifications by phosphorylation pathways have been shown to enhance the recruitment of coactivators to this site (Bevan *et al.*, 1999; Tremblay *et al.*, 1999). The AF1 domain is responsible for the ligand-independent, constitutive activity of the receptor, but it is thought that coactivators are involved in the cooperativity between both AF domains by forming a functional link between them (Aranda and Pascual, 2017). It has also been observed that phosphorylation of specific residues in the hinge region of FXR by AMPK signalling pathways, can inhibit the activation of the receptor (Lien *et al.*, 2014). Furthermore, several studies utilising NMR and hydrogen/deuterium exchange with mass spectrometry (HDX-MS), have shown that NR DBDs, in addition to the DNA response element sequences that these are bound to, can remotely impact the LBD, by altering the conformation of binding surfaces to a more energetically favourable configuration (Zhang *et al.*, 2011; de Vera *et al.*, 2017). This alters the interactions with both RXR and coactivator proteins, and consequently, can affect the affinity of the LBD for the coactivator, and the potency of the ligand (de Vera *et al.*, 2017). With regard to the selective modulation of NRs, the attention of pharmaceutical scientists is now being turned away from the LBD and towards other sites of receptor modulation. As an attempt to overcome problems such as limited selectivity and the emergence of resistance due to the blockage of signal transduction, especially in the treatment of NR-based cancers, some groups have focussed on the DBD-DNA interaction surfaces (Veras *et al.*, 2019). However, the relationship between different DNA motifs and FXR target gene selectivity is yet to be elucidated, and so it is unclear whether targeting FXR-DNA interactions will be a useful

therapeutic strategy (Massafra et al., 2018). Nevertheless, other sites along the receptor, in addition to the LBD, may hold the potential for strategically targeting the transactivation of FXR.

In conclusion, the entire FXR receptor has a complex modular nature, whereby cross talk between the individual domains, signalling derived from exogenous stimuli and the binding of DNA can affect the interactions between the receptor, coactivators and other transcription factors. This means that several other factors, in addition to ligand binding, can affect the transcriptional potential of FXR. Thus, it may be naïve to conclude that these novel compounds are superior agonists for FXR based on the data from these cell-free coactivator recruitment assays alone.

Chapter 6.

Function of Lead Compounds

6.1 Introduction

6.1.1 FXR Signalling in Health and Disease

Since its de-orphanisation in 1999, the role of FXR as a transcriptional regulator in several entero-hepatic metabolic pathways slowly has been unravelled. As such, with its involvement in the development of several pathophysiologies, FXR has emerged as an attractive pharmacotherapeutic target for the treatment of diseases including, but not limited to, fatty liver disease, primary biliary cholangitis and inflammatory bowel disease; providing the impetus for the discovery of novel agonists.

As with other nuclear receptors, FXR binds to its associated target genes through its DBD. The FXR DBD recognises and binds to a response element (FXRE), a consensus repeat hexanucleotide sequence, found on its target DNA. The most common FXRE with which FXR interacts, is an inverted, palindromic repeat of the six nucleotides 'AGGTCA', arranged as two copies separated by a single nucleotide. However, FXR has been shown to bind to a variety of response elements with diverse geometry, including negative FXREs, which have been associated with the transcriptional repression of genes upon FXR activation (Fiorucci et al., 2007). Chromatin immunoprecipitation followed by deep sequencing (ChIP-seq), in addition to gene ontology analysis, identified a large number of FXRE sites adjacent to numerous genes involved in cholesterol, lipid and fatty acid metabolism, in addition to those associated with BA homeostasis, implicating FXR in roles much more diverse than previously appreciated (Kemper, 2011). Significantly, several of these pathways modulated by FXR (summarized in Table 6. 1), are important for the pathophysiological processes that lead to liver steatosis and eventually NASH.

6.1.1.1 FXR and Bile Acid Homeostasis

As the cognate ligands of FXR are BAs, it is inevitable that its primary role is in maintaining enterohepatic BA pool, according to metabolic needs, by exerting transcriptional control over genes responsible for BA synthesis, transport and the refilling of the gallbladder. A key target of FXR is small heterodimer partner (SHP), which acts as an atypical nuclear receptor devoid of its own DBD. Instead, SHP, similarly to coregulators (discussed in Chapter 5), dimerizes with other nuclear receptors via their LXXLL amino acid recognition sites, inhibiting their transcriptional activity. In this way, FXR-activated SHP ultimately inhibits the expression of the BA biosynthetic enzyme CYP7A1, by binding to its activators, Liver Receptor Homolog-1 (LRH-1) and Hepatocyte Nuclear Factor 4 alpha (HNF4 α), and causing subsequent histone deacetylation (Hoeke et al., 2014, Boulias et al., 2005). Equally, another key target of FXR activation is the Fibroblast Growth

Factor 19 (FGF19) and its mouse ortholog (FGF15), which are expressed in response to intestinal FXR activation. Hormone-like FGF19 links the gut and the liver axis, to also downregulate *cyp7a1* via c-Jun Kinase-dependant phosphorylation pathways (Holt *et al.*, 2003; Inagaki *et al.*, 2005). Mice lacking these two independent FXR activated genes, SHP and FGF19, both show an upregulation in CYP7A1 mRNA expression, and an altered BA pool (Boulias *et al.*, 2005; Inagaki *et al.*, 2005). In addition to their actions in BA synthesis, both SHP and FGF19 modulate other important effects. SHP plays a major regulatory role in the negative regulation of several diverse pathways, whereas FGF19(15) has been shown to improve insulin sensitivity and reduce body weight in animal models (Strack and Myers, 2004).

In addition to its role in the negative regulation of *de novo* BA synthesis, FXR, maintains BA levels by also regulating transport into and out of the liver. By directly binding to FXREs on the DNA encoding BSEP, and by SHP-mediated repression of NTCP, FXR can regulate these genes, resulting in both the efflux of BAs into the bile duct, and prevention of BA uptake into the liver, respectively (Ananthanarayanan, *et al.*, 2001). Furthermore, FXR also has been shown to bind to the promoters and upregulate the expression of *ostα* and *ostβ*, which encode transporters responsible for the facilitative diffusion of BA across the basolateral membrane of hepatocytes and cholangiocytes in the liver, as well as epithelial cells in the ileum (Ballatori *et al.*, 2009). By detecting and responding to elevated levels of BAs, FXR in both the intestines and liver can directly interact with target genes, or can mediate the response of other nuclear receptors, to orchestrate a complex network of signalling pathways to protect the liver from the toxic effects arising from BA accumulation.

6.1.1.2 FXR and Lipid Metabolism

Over the last decade, the mechanisms underlying the intimate crosstalk between BA and lipid metabolism have been uncovered. *De novo* BA synthesis is driven by the catabolism of cholesterol. Accordingly, FXR-mediated regulation of BA production has subsidiary effects on cholesterol metabolism. CYP7A1 activation leads to a decreased cholesterol content in hepatic microsomes, which in turn causes the upregulation of LDL receptor (LDLR) and the subsequent LDL harvesting and reductions in plasma-LDL content. In this respect, FXR antagonism has been proposed as a potential treatment for hypercholesterolaemic patients, however, studies using the supposed FXR antagonist, natural extract, guggulsterone, produced conflicting results (Lefebvre *et al.*, 2009). Conversely, treatment of hepatic cells *in vitro*, have suggested that FXR activation increases VLDLR mRNA and expression, and may indeed be beneficial for lowering circulating VLDLs (Nakahara *et al.*, 2002). In addition, FXR has the ability to independently impact the composition of HDLs, which are responsible for the transport of cholesterol from the

periphery to the liver where it can be excreted. FXR is thought to repress Apolipoprotein A1 (*apoA1*) gene expression and negatively regulate HDL cholesterol levels, but can alter their composition by direct induction of the phospholipid transfer protein (PLTP) (Urizar *et al.*, 2000).

Similarly, FXR has been implicated in triglyceride metabolism by transcriptionally regulating various genes. FXREs have been identified on the promoter of Apolipoprotein C2 (*apoc2*), and mRNA levels were upregulated when hepatoma cells, HepG2, were incubated with CDCA (Kast *et al.*, 2001). Activation of APOC2 results in the hydrolysis and clearance of triglycerides in chylomicrons and VLDLs, and previous studies reported a decrease in plasma and hepatic triglyceride content upon treatment with FXR agonists (Kast *et al.*, 2001). Several other mechanisms of triglyceride lowering have been attributed to FXR signalling, including the induction of fatty acid oxidation by PPAR α and pyruvate dehydrogenase kinase 4 (PDK4), and the repression of VLDL-producing microsomal triglyceride transfer protein (MTP) (Xie *et al.*, 2016).

In addition to promoting triglyceride clearance, FXR has been shown to ameliorate triglyceride levels, by also potentiating their production. FXR is thought to participate in *de novo* lipogenesis pathways, affecting free fatty acid content, and their ability to form VLDLs. Acting via SHP, FXR activation leads to the downregulation of sterol regulatory element binding transcription factor (SREBF1) and its lipogenic target genes, leading to concomitant decreases in triglyceride levels (Watanabe *et al.*, 2004).

Taken together, these studies support the role of FXR in lipid metabolism and further suggest that appropriate and selective FXR agonism, could have a beneficial effect on pathophysiologies involving dyslipidaemia, and in particular, those constituting the first 'hit' of NASH.

6.1.1.3 FXR and Glucose Metabolism

Several studies have shown that FXR signalling can also affect hepatic glucose production, intestinal glucose absorption, as well as insulin sensitivity in the peripheral organs. Whilst the cross talk between BA metabolism and glucose metabolism is yet to be fully delineated, some of the advantageous effects of BAs may be in part due to the FXR-independent activation of the other BA receptor, TGR5, which is also known to be a key player in normal glucose homeostasis (Watanabe *et al.*, 2011). However, although the exact mechanism is still controversial, FXR has been shown to induce glycogen synthase kinase 3 β (GSK3) derived glycogen storage in *db/db* diabetic mice (Zhang *et al.*, 2006). Furthermore, FXR appears to be partly responsible for the regulation of gluconeogenic genes, phosphoenolpyruvate carboxykinase (PEPCK), fructose 1,6-biphosphatase 1 (FBP-1) and glucose-6-phosphatase (G6Pase) (Cariou and Staels, 2007). It is

thought that FXR induced SHP can negatively regulate these genes by preventing the recruitment of coactivator proteins PGC1 α and CBP1 to HNF4 α , which in turn lies upstream of both *pepck* and *fbp1* and is responsible for their expression. Similarly, G6Pase also appears to be repressed by the SHP-mediated inhibition of a transcription factor in its promoter region (Yamagata *et al.*, 2004).

The effect of FXR and its agonists on genes involved in gluconeogenesis appear to be dependent on the source of the cells used, the species being tested, or even the nutritional or diabetic status of the subjects, highlighting that FXR is only one part of a complex network of interacting metabolic and hormone-sensing receptors. Nevertheless, the consensus is that there is an intimate link between FXR and insulin sensitivity. FXR deficiency has been associated with insulin resistance, impaired insulin signalling and glucose tolerance in affected tissues such as skeletal muscle; the FXR agonist, GW4064, has been shown to improve insulin sensitivity in both obese and diabetic mice (Cariou *et al.*, 2006). Although the exact underlying mechanism, linking FXR to insulin sensitivity, is yet to be determined it is thought that FXR's role in regulating free fatty acids and triglycerides in circulation and, ultimately, in insulin sensitive tissues, plays an important part. More recently, FXR has been described to play a role in the differentiation and lipid storage in adipocytes, and also has been shown to directly induce the insulin-sensitive glucose transporter GLUT4 (Cariou *et al.*, 2006; Shen *et al.*, 2008). Insulin desensitization, particularly in adipocytes, is a significant contributory factor in the release and circulation of free fatty acids and accumulation of triglycerides in the liver, which leads to the hepatic lipotoxic phenomenon of NASH.

6.1.1.4 FXR and NASH

The FXR-induced alterations in lipid and glucose metabolism have highlighted FXR as a prospective therapeutic target in the treatment of NASH, which arises due to physiological abnormalities in several of these metabolic pathways. FXR null mice, are one of the only mouse models to give rise to the entire spectrum of NAFLD, including hepatocellular carcinoma, further confirming the association between FXR and fatty liver disease. Recent evidence again supports the underlying role of FXR in the progression and potential treatment of the disease.

Inflammation is a parallel process contributing toward the progression of NASH and leads to hepatitis and the initiation of fibrosis. FXR is thought to increase the transcriptional activity of intercellular adhesion molecule 1 (ICAM-1) promoter, whilst also being implicated in a sophisticated pathway antagonizing nuclear factor kappa B (NF κ B) activity, by suppressing its proinflammatory effects (Wang *et al.*, 2008). Moreover, FXR null mice displayed more severe necrosis and inflammation following liposaccharide treatment compared to wild type mice;

with significant increases in Tumour necrosis factor alpha (TNF α), Interleukin 1a (IL1a), Interleukin 6 (IL6) and Cyclo-oxygenase 2 (COX2) expression (Wang *et al.*, 2008). Furthermore, recent studies have identified a FXRE in the promoter of monocyte chemoattractant protein (also known as chemokine (c-c motif) ligand 2 (CCL2)). CCL2 promotes the infiltration of monocytes and memory T-cells to the site of inflammation, and it is thought to be essential in the development of NASH (Armstrong and Guo, 2017). Likewise, FXR-specific decreases in CCL2 expression and inflammatory cell infiltration were seen in methionine/choline deficient, dietary induced 'NASH' mice, treated with an FXR agonist (Zhang *et al.*, 2009).

In addition, the discovery of FXR expression in hepatic stellate cells, suggested a role for the nuclear receptor in the promotion and regulation of fibrosis. HSCs undergo differentiation from their resting, fat storage-type phenotype, to an activated fibroblast-promoting phenotype, whereby the deposition and accumulation of extra-cellular matrix is initiated. Treatment of hepatic cells *in vitro* with OCA augmented HSC transdifferentiation, and supports the idea of using FXR agonists for the reduction or prevention of fibrosis in the latter stages of NASH (Fiorucci *et al.*, 2005). Moreover, FXR/SHP cascades driven by OCA treatment, was able to prevent carbon tetrachloride (CCl₄)-induced upregulation of fibrotic markers, Tissue Inhibitor of Metalloproteinase 1 (TIMP1), Type 1 Collagen (COL1A1) and α -Actin Smooth Muscle (ACTA2), in rats (Fiorucci *et al.*, 2005).

The association between FXR, lipotoxicity and NASH also was observed in double knockout mutant mice fed a high fat diet, lacking both LDLR and FXR (Kong *et al.*, 2009). FXR deficiency was associated with macrosteatosis, hepatocyte ballooning and the additional infiltration of inflammatory cells. Mice livers also displayed the initiation of fibrosis and displayed increased expression of COL1A1, TIMP1 and ACTA2 when FXR was knocked out. Furthermore, transforming growth factor β 1 (TGF β 1), which is considered to be the most potent fibrogenic cytokine produced in several cell types in the liver, and regulates several phosphorylation pathways, leading to the activation and migration of hepatic stellate cells, also was increased with the ablation of FXR, consistent with the conclusion that FXR deficiency may contribute towards the pathologic manifestation of NASH (Kong *et al.*, 2009).

In addition to its effects on metabolic pathways, FXR is thought to play a role in enteroprotection and in maintaining gut microbiota composition. Furthermore, dysbiosis has been observed in NAFLD, implicating the microbiome in the development of NAFLD in mice and humans (Gonzalez *et al.*, 2016). Studies also paradoxically demonstrated that intestinal FXR antagonists led to changes in ceramide metabolism and a decrease in fatty acid synthesis in the

liver, arising from modulated gut bacterial compositions, with an additive effect of altered bile acid pools, ultimately preventing NAFLD (Gonzalez *et al.*, 2016). Likewise, the use of bile acid sequestrants and the subsequent inhibition of FXR signalling in the terminal ileum, resulted in the improvement of NAFLD mice (Mcgettigan *et al.*, 2016). Accordingly, the evidence presented, suggests that FXR antagonism in the small intestine, may prove beneficial for the treatment of NAFLD.

As such, whilst the evidence suggests a clear link between FXR signalling and the progression of NASH, exact pathways are yet to be realised, and responses seen may be dependent on the tissue type in which FXR is expressed. While theoretically, it may be beneficial to globally target FXR in order to treat all aspects of metabolic syndromes, the possible concomitant activation of undesirable side effects reiterates the fact that selective modulation of FXR, in independent tissues, may be the most appropriate therapeutic strategy in the treatment of NASH.

Table 6. 1 Summary of FXR regulatory effects either by direct transcriptional regulation or downstream effects.

Gene Function	Upregulated	Effect	Downregulated	Effect
BA synthesis, transport and metabolism	SHP	Inhibition of BA synthesis and other pathways	CYP7A1	Decreased BA synthesis
	FGF15/19	Inhibition of BA synthesis and increased insulin sensitivity	CYP8B1	Decreased BA synthesis
	BSEP	Increased BA efflux out the cell	LRH-1	Decreased BA synthesis, decreased gluconeogenesis, decreased de novo lipogenesis
	OSTα/β	Increased BA efflux out the cell	NTCP	Decreased uptake of BAs into the liver
	IBABP	Increased BA transport across the cell		
Lipid metabolism	APOC1	Increased cholesterol transport	APOA1	Decreased HDL synthesis
	APOC2	Increased lipoprotein lipase activity and increased triglyceride clearance	APOC3	Increased lipoprotein lipase activity and increased triglyceride clearance
	PLTP	Transfer of lipoproteins from LDLs to HDLs	SREBF1	Reduced fatty acid/ triglyceride synthesis
	PDK4	Increased β-oxidation	HNF4A	Decreased plasma cholesterol and gluconeogenesis
	PPARA	Increased β-oxidation	MTP	Decreased VLDL assembly
	VLDLR	Increased VLDL uptake and triglyceride clearance		
	GSK3	Increased glycogen storage	FBP-1	Decreased gluconeogenesis

Glucose metabolism	GLUT4	Increased insulin sensitivity	G6Pase	
			PEPCK	
Inflammation	ICAM	Increased recruitment of macrophages to endothelium	NFkB	Decreased inflammation pathways
			iNOS	Decreased vascular smooth muscle cell inflammation
			COX-2	Decreased vascular smooth muscle cell inflammation
			IL1A/B	Decreased inflammation pathways
			IL16	
			TNF α	
			CCL2	
Fibrosis			CCR2	Decreased fibrogenesis
			COL1A1	
			ACTA2	
			TIMP1	
			TGF β 1	

6.1.2 Experimental Models of NASH

The rapidly changing NASH drug discovery field has targeted several mechanisms of the disease alone or in combination. Novel drugs either target the metabolic pathways leading to the development of NASH, or the fibrotic processes exacerbating NASH and associated cardiovascular co-morbidities, but FXR presents an attractive target with the potential to affect both strategies. A major obstacle to the pipeline in the discovery of potential NASH therapeutics, is the paucity of appropriate and relevant disease models, on both a cellular and whole organism level, which can replicate the metabolic and histological features of the disease.

Currently *in vitro* models are very simplistic, and do not mimic the complexity and multifaceted systems occurring in whole organs with multiple different cell types; their use in studying the progression of NASH is in its infancy. Nonetheless, these models are often used to primarily identify molecular mechanisms involved in the disease. Human hepatocytes are considered the most appropriate representation of clinical models. However, hepatocytes have a limited role in the fibrotic process, and due to ethical considerations and the ease of reproducibly extracting and preparing primary cultures, hepatic immortalised cell lines are chosen as an alternative (Chavez-Tapia *et al.*, 2011). Cell lines, due to their steady growth, stable phenotype, in an almost unlimited life span, and the ease of reproducibly culturing, make their use more advantageous. Hepatic carcinoma-derived cell lines, Huh7 and HepG2, are well established and the most extensively used cell lines for the study of a wide variety of liver-related pathologies. Despite being shown to resemble foetal hepatocytes, both cell lines are currently being assessed for their changes in cell morphology, function and metabolism under different cell culture conditions designed to mimic NASH (Green *et al.*, 2015). Moreover, due to their expression of FXR-related genes, both cell lines present an accessible medium in which to study the direct effects of FXR agonism.

As with hepatic cell lines, the use of *in vivo* animal models to elucidate the mechanisms involved in the progression of steatosis and fibrosis, as well as to study the potential therapeutic treatment of the disease, is very complex. An ideal model should encompass the histological and pathological hallmarks of human-related lipid metabolism and fibrosis, but currently, there is no single animal model that reflects all aspects of this multifactorial disease.

Only a small percentage of people develop NAFLD due to genetic reasons and, as such, animal models based on hepatic triglyceride accumulation due to dietary factors, are considered a more clinically relevant approach. To induce steatosis, a variety of high energy, high fat (45-60% total energy as fat), high sucrose (65% of body weight), and high fructose diets, with and without

trans fat, have been used (as reviewed in Green *et al.*, 2015). Whilst most regimens were able to induce some form of altered triglyceride metabolism and fat accumulation, the species, strain and sex of the animals, can have pronounced differences on the extent of the effects seen. In addition, despite altered lipid metabolism, most mice do not develop fibrosis, even after long term feeding. Moreover, the commonly used method of feeding mice a methionine and choline -deficient diet, which induces NASH-like symptoms including hepatocyte ballooning, steatosis and inflammation, did not induce weight gain or insulin resistance, and clinical relevance was dependent on the metabolic background of the strain, suggesting that it may only be applicable with use in a diabetic mouse model (Rinella *et al.*, 2008). Furthermore, since most human diets are not deficient in methyl groups, its translatability to human disease has been debated.

A new model for the simple, rapid and reproducible production of NASH in mice has been recently described (Hoshida *et al.*, 2018). The model involves the combinatorial use of a western diet (high fat (21% by weight), high fructose (41% by weight) and high cholesterol (1.25% by weight)), with the chronic administration of liver injury-inducing carbon tetrachloride (CCl₄), over a 12-week period. Mice under this schedule displayed increased body and liver weights, increased serum levels of liver injury markers, histologically defined fibrosis, steatosis, lobular inflammation and ballooning, as well as the prototypical transcriptomic changes in fatty acid metabolism, bile acid metabolism and oxidative phosphorylation, all concordant with changes seen in human NASH (Hoshida *et al.*, 2018). As such, the simplicity and translatability of this model proves it a suitable choice for the study of different compounds in ameliorating NASH pathology.

6.2 Aims and Objectives

FXR has been shown to have pleiotropic roles in the regulation of several diverse signalling cascades and pathways. Accordingly, significant attention has been directed at ways to modulate this receptor and its target genes, in an attempt to mediate the pathways contributing to NASH. Two novel BA-derived compounds, Jed441 and Jed561, have been shown to activate FXR by promoting the recruitment of its coactivators (Chapter 5). Whilst the results of these assays are promising, the artificiality of the cell-free environment means that these compounds require further validation. Assessment of these compounds in a cellular setting, interacting with the full length FXR protein, with the additional regulation of various partner receptors, transcription factors and post translational modifications, is required to determine whether these compounds can promote the physiological response expected.

Thus, the aim of the research presented in this chapter, is to identify whether lead compounds Jed441 and Jed561, can functionally activate FXR, observed by the regulation of its target genes. In addition, a secondary aim is to determine whether the observed responses are specific to FXR activation and, whether the specific fluorination at the 4 β position on Jed561 confers this compound with any favourable effects over the non-fluorinated derivative. Furthermore, work here aims to evaluate the effects of these compounds *in vivo*; firstly, in mice under normal physiology, to determine whether results seen *in vitro* are translatable to a whole organism setting, and secondly in a mouse model of NASH. The results obtained from this research are anticipated to help delineate the mechanisms of action of these compounds, as well as confirming their potential use as pharmacological therapeutics for the treatment of NASH or other metabolic or hepatic-related diseases.

6.3 Methods

6.3.1 Cell Culture and Treatment with Compounds

Huh7 cells were purchased from the Japanese Collection of Research Bioresources (JCRB) and HepG2 cells were a kind gift from Dr Francesca Greco, School of Pharmacy, University of Reading. Huh7 cells were cultured in low glucose containing Dulbecco's Modified Eagle Medium (DMEM) supplemented with 10% Foetal Bovine Serum (FBS) and 2 mM L-glutamine (all Sigma Aldrich). HepG2 cells were grown in Eagle's Essential Minimal Medium (EMEM) supplemented with 10% FBS, 2 mM L-glutamine and 1% non-essential amino acids (all Sigma Aldrich). Cells were grown and maintained at a confluence of 70-80% in an incubator (Autoflow IR direct heat incubator, Nuaire, UK) set to 37°C and a carbon dioxide concentration of 5%. When an appropriate estimated number of cells had been obtained (as determined by visual inspection), the cell culture medium was removed from the flasks and cells washed with Dulbecco's phosphate buffered saline (PBS, Sigma Aldrich). For a 75 cm² flask, 2 mL trypsin (Sigma Aldrich) was added to the flask, and flasks incubated for 2-3 minutes at 37°C to facilitate detachment. An 8 mL aliquot of complete medium was added to the flask to stop the trypsinization reaction, and the entire contents transferred to a 15 mL falcon tube. A 100 µL aliquot of cell suspension was withdrawn and diluted 1:1 with trypan blue solution and analysed in under a light microscope (Nikon) in a haemocytometer chamber, to determine the cell density. The remainder of the cell suspension was spun down by centrifugation at 100 – 125 × g for 5 minutes. The supernatant was removed, and the cells were resuspended in a volume of fresh medium to give a concentration of 1×10⁶ cells/ml. Aliquots of 1 mL were seeded in 6 well plates so that each well contained approximately 1×10⁶ cells, and cells were left to adhere overnight. The following day, the medium was removed from each well and replaced with medium containing either DMSO (1% final concentration), OCA, Jed441 or Jed561 at their respective EC₅₀ and EC₉₀ concentrations. Plates were duplicated, and one plate was incubated for 6 hours, and the other plate for 24 hours, after compound addition. A schedule of treatment can be found in Table 6. 2.

Table 6. 2 Treatment of Huh7 and HepG2 cells with different compounds for gene expression analysis

Incubation time (hours)	Compound	Compound Concentration (nM)
6	Vehicle (control)	-
	OCA	(EC ₅₀) 300
		(EC ₉₀) 1000
	Jed441	(EC ₅₀) 27
		(EC ₉₀) 180
	Jed561	(EC ₅₀) 92
		(EC ₉₀) 400
24	Vehicle (control)	-
	OCA	(EC ₅₀) 300
		(EC ₉₀) 1000
	Jed441	(EC ₅₀) 27
		(EC ₉₀) 180
	Jed561	(EC ₅₀) 92
		(EC ₉₀) 400

* EC₅₀ and EC₉₀ values used for cell culture experiments were determined by cell-based FXR reporter gene assays conducted by Irene Boz, School of Pharmacy, University of Reading.

In guggulsterone inhibition experiments, cells were incubated with either 10 or 100 µM final concentrations of guggulsterone, in the presence or absence of Jed441 and Jed561 at their respective EC₅₀ concentrations. In experiments comparing the effect of the fluorination of the compound, Jed692 was cultured at identical final concentrations as Jed561, 92 nM.

After incubation for the respective amount of time, cells were washed with ice-cold PBS prior to RNA extraction.

6.3.2 Maintenance and Treatment of Mice

All procedures involving the animal husbandry, housing and treatment of mice were carried out by Saretius Ltd but are described here for context. All procedures were performed under license from the Home Office and in accordance with The Animals (Scientific Procedures) Act 1986. C57BL/6J mice (hereafter referred to as wild type (WT) mice) were obtained from Charles River Laboratories (UK) and housed under standard conditions. Age and weight matched groups of 10-12-week-old male mice were used for all experiments.

6.3.2.1 Single Dose Administration of Compounds in Wild Type Mice

Weight matched male WT mice were given a solution of OCA or Jed compounds diluted in DMSO equating to 2mg compound per kg body weight (2mg/kg) by oral gavage (n=3 per treatment group). Control mice were given a vehicle solution of 5% DMSO in 0.5% methylcellulose in water. Mice were housed in standard cages and given free access to food and water. At terminal sampling points 0.5, 2 and 8 hours after administration, mice were sacrificed by a CO₂ euthanasia method (UK Home office guidelines). Mice administered with vehicle were sacrificed 0.5 hours post treatment. In experiments comparing Jed compounds with competitor compounds, mice were treated equally with a final compound concentration of 2mg/kg and were sacrificed 6 hours post administration. Livers were saline-perfused, harvested, and a 0.5g slice, stored at room temperature in RNAlater (Invitrogen) was provided for the RNA extraction (Chapter 6, section 6.3.4) and qPCR work as described (Chapter 6, section 6.4.2.2).

6.3.2.2 Daily Administration of Compounds in Wild Type Mice

Sixty WT mice (n=10 per treatment group) were treated for 5 days with either vehicle (5% v/v DMSO in 0.5% w/v methyl cellulose in water), OCA (30 mg/kg body weight), Jed561 (2 mg/kg, 10 mg/kg or 30 mg/kg body weight) or Jed441 (10 mg/kg body weight). Compounds were administered by oral gavage, once daily, and animals were given free access to food and water. On the 5th day, food was removed 30 minutes before the final treatment, and mice were starved for a further 6 hours before being sacrificed. Strips of specifically the left lobe of the liver, in RNAlater were provided for the RNA extractions and qPCR work.

6.3.2.3 Production of NASH Mouse Model and Treatment with Compounds

Sixty WT mice were fed either a low-fat diet (LFD, 10% fat, n=10) or a high fat diet (HFD, 60% fat, n=50). After 30 days, in order to induce liver fibrosis, mice were subjected to additional intraperitoneal injections of carbon tetrachloride (CCl₄) at doses equivalent to 25% at 0.5 ml/kg, for mice on the HFD, or olive oil sham for mice on the LFD. Treatment with CCl₄ or olive oil was administered every 4 days. After 21 days of CCl₄ treatment, HFD mice were allocated to groups of 10 according to body weight. For the next 29 days, mice were treated orally with either Jed561 (at 2, 10 and 30 mg/kg), OCA 30 mg/kg, or vehicle (5% DMSO in 0.5% w/v methyl cellulose in water). LFD mice were dosed with vehicle. At the end of the study, mice were euthanised by CO₂ (as per home office guidelines) and strips of the left lobe of the liver, stored in RNAlater were provided for the RNA extractions and qPCR work.

6.3.3 RNA Extraction from Adherent Cells

RNAse-free plasticware and RNAse-free water were used in the preparation of total RNA from sources. All surfaces were decontaminated using RNAseZapTM (Invitrogen). For total RNA

extraction a solid phase extraction method was used whereby silica's nucleic acid binding properties are used to facilitate the purification of RNA, as first described by Boom et al., (1990). The RNAqueous™ Total RNA Isolation kit (Ambion) was used, according to manufacturer's instructions. For cultured cells, the medium was removed, and cells were washed with 1×PBS to remove cellular debris and residual medium. Total RNA was extracted from fresh cells. For 1×10^6 cells, 350 μ L lysis buffer was added directly to the well and cells were harvested by scraping with a pipette tip. Lysed cells were combined with an equal volume of 64% ethanol and mixed thoroughly by pipetting. The ethanol-lysate mix was transferred to a column and spun at 12,000 $\times g$ for 1 minute and flow through discarded. The membrane was washed to remove residual DNA, proteins and salt, by adding wash buffer 1 to the column and centrifuging at 12,000 $\times g$ for 1 minute, before discarding the flow through and repeating this step twice with wash buffer 2. An additional spin with the empty cartridge was included to completely dry the membrane of ethanol. Finally, the total RNA was eluted in 2 sequential aliquots of 50 μ L preheated elution buffer (nuclease-free water containing trace amounts of EDTA). Total RNA samples were checked for integrity and quality, and immediately used for cDNA synthesis, before being stored at -80°C.

6.3.4 RNA Extraction from Mouse Livers

A ~150 mg slice of liver was removed from RNA/*later*™ and blotted on tissue paper to remove excess buffer. The slice was added to 1800 μ L of guanidium-based lysis buffer in a glass, round-bottom homogenizer, and lysed on ice by manually grinding the tissue. Once sufficiently liquefied, the homogenate was transferred to an Eppendorf and centrifuged at top speed for 5 minutes to clarify the sample and remove debris. Once clarified, 300 μ L of lysate, equivalent to approximately 30 mg of liver tissue, was combined with an equal volume of 50-64% ethanol and mixed gently by inversion or pipetting (50% ethanol was used in cases where the sample was particularly viscous, to prevent the filter from clogging in subsequent steps). The lysate-ethanol mix was transferred to a column from the RNAqueous™ kit, and the rest of the extraction protocol was followed, according the manufacturer's instructions as detailed previously (Chapter 6, section 6.3.3).

6.3.5 Analysis of RNA Quantity, Purity and Integrity

RNA concentration was quantified by measuring the absorbance at 260 nm, using a Nanodrop Lite spectrophotometer (Thermo Scientific). The Nanodrop also measures the absorbance of light at 280 nm and the purity of RNA was estimated by analysing the $A_{260:280}$ ratio, where a value of between 1.8 and 2.2 was deemed to be free from residual protein contamination and acceptable for downstream applications. The integrity of the RNA was determined by running a

sample on a denaturing formaldehyde agarose gel. A 1% agarose, 1×MOPS, 6.6% formaldehyde gel was made with the addition of 1× SYBR® Safe DNA stain to visualize the nucleic acids. Prior to loading, equal volumes of formaldehyde loading dye (Ambion) was added to 1 µg RNA or 3µL RiboRuler High Range RNA ladder (ThermoScientific) and samples were heated at 70°C for 10 minutes before being immediately snap cooled on ice for 2-3 minutes to denature any intrinsic secondary structure of the RNA. The gel was run at 90V for 1 hour 30 minutes and visualised under UV light, using the NuGenius gel doc system (Syngene). Pure RNA was demonstrated by the presence of 2 sharp, intense bands at approximately 5 kb and 1.9 kb, indicating the ribosomal RNA subunits, 28S and 18S, respectively. The 28S upper band was expected to be twice the intensity of the 18S lower band for intact RNA. Smearing below the 18S rRNA band was taken to indicate degraded RNA, whilst smearing and/or bands above the 28S rRNA was indicative of DNA contamination.

6.3.6 DNase Treatment of RNA

Before reverse transcription, RNA samples were treated with amplification grade DNase I (Invitrogen) to remove any residual contaminating genomic DNA which could interfere with downstream cDNA synthesis and RT-qPCR applications. To 1 µg RNA, 1 µL 10× DNase I reaction buffer and 1 µL DNase I (1 U/µL) was added. DPEC-treated water was added to a final volume of 10µL. Reactions were incubated at room temperature for 15 minutes, followed by inactivation of DNase I by the addition of 1 µL of 25mM EDTA and incubation at 65°C for 10 minutes. The whole of the DNase I reaction was used for cDNA synthesis.

6.3.7 Complementary DNA (cDNA) synthesis

Reverse transcription from RNA to cDNA was carried out using the iScript Advanced cDNA Synthesis Kit for RT-qPCR (Biorad) according to manufacturer's instructions. The iScript reaction mix contains buffer, MgCl₂, enhancers and stabilisers and both random and oligo(dT) primers to allow for the efficient synthesis of long, full length cDNAs from eukaryotic mRNA. The reverse transcriptase used, as part of this kit, is a recombinant Moloney murine leukaemia virus enzyme and RNase inhibitors are included in the reaction mixture. A typical reaction would contain 4 µL 5× iScript advanced reaction buffer, 1 µL iScript advanced reverse transcriptase, 10 µL DNase I treated RNA and nuclease-free water to a final volume 20 µL. The reaction was incubated at 46°C for 20 minutes, before inactivation at 95°C for 1 minute. Newly synthesised cDNA was diluted in TE buffer (10 mM Tris pH 8, 1 mM EDTA), aliquoted and stored at -20°C until use in qPCR experiments.

6.3.8 Quantitative Real-time Polymerase Chain Reaction (qPCR) (Principle)

Gene expression analysis of FXR and its downstream targets was conducted using quantitative real-time PCR, a method which is used to precisely and efficiently monitor real-time increases in amplicon concentration. For all reactions, a StepOnePlus™ Real-time PCR system (Applied Biosystems) was used, which, in brief, comprises a 96-well thermal cycler for the amplification of DNA, a light source for the excitation of fluorescent probes and a charge coupled device camera to detect and record changes in fluorescence. The method used here utilized the fluorescent intercalating dye SYBR® Green, which, in its free, unbound form has undetectable fluorescence. SYBR® Green, however, can bind to double stranded DNA in a sequence independent manner, by intercalating between DNA bases and subsequently emits a fluorescent signal. As such, as the reaction progresses, and the number of cycles increase, there is an increase in fluorescence emission, which is a direct consequence of the increase in target DNA being produced. This increase in fluorescence intensity is measured by the camera and the change in fluorescence intensity (ΔR_n) is automatically plotted versus the cycle number in amplification plots by computer software. The reaction can be considered to have 3 separate stages; an initial baseline stage, where there is little change in fluorescence detected; the exponential phase, where the number of amplicons double with each cycle; and the plateau stage, where reagents become depleted and DNA replication starts to slow. The baseline is determined by the ΔR_n values from the initial cycles of the PCR reaction (cycles 3-15). An arbitrary threshold is then set based on the variability of the baseline, usually a value at 10 times the standard deviation of the baseline. The baseline and threshold are set automatically by the software, but when necessary were adjusted or set manually between experiments. The threshold cycle (C_t) values are then calculated by determining the point, or cycle number, at which the fluorescence, due to amplification, exceeds the threshold limit. This means that samples with higher starting concentrations of template DNA will cross the threshold at earlier cycles than samples with lower starting DNA, therefor resulting in lower C_t values (Giulietti *et al.*, 2001).

6.3.9 Primer Selection

6.3.9.1 Selection of Reference Genes

Candidate reference genes were selected based on data from existing literature (Bruce *et al.*, 2012; Hashemi *et al.*, 2012; Kouadjo *et al.*, 2007; Li *et al.*, 2017; Suzuki *et al.*, 2000) and tested for stability across different experimental conditions.

6.3.9.2 Optimization of Primers for Reference and Target Genes

KiCqStart™ SYBR® green predesigned primers for target genes were purchased from Sigma Life Science. Where possible, primers chosen were those that span an exon splice site to minimise the possibility of co-amplification of gDNA. All primers amplified an 80-150 bp product. Nuclease-free water was added to the lyophilized primers for a stock concentration of 100 µM. Primers were diluted with nuclease-free water for a working concentration of 10 µM. A list of all primers used can be found in Table 6. 3. To test the efficiency, reproducibility and dynamic range of the assay, a ten-fold serial dilution was made; consisting of 5 concentrations of cDNA generated (as outlined above) from either human reference RNA (Agilent) or pooled cDNA from Huh7 cells, HepG2 cells or mouse liver samples. Following qPCR of these samples, using both reference gene primers and target gene primers, a standard curve was constructed using the threshold cycle (C_t) value (y-axis) versus log cDNA concentration (x-axis). The primer amplification efficiency ϵ of one cycle in the exponential phase was determined by the equation $E = 10^{(-1/\text{slope})} - 1$ (Ginzinger, 2002). The accuracy of these qPCR reactions was determined by the correlation coefficient, R² value of the standard curve, with values higher than 0.98 being suitable. The specificity of each primer was determined by melt curve analysis, which was performed at the end of each run, where the production of one peak at one melting temperature indicated the amplification of a single product and, therefore, primers that were highly specific. Amplified products were confirmed by agarose gel electrophoresis (2% Agarose, 1×TAE, run at 100V for 30 minutes) to check amplicon sizes were as expected, and that only one product was seen.

Table 6. 3 KicQStart® SYBR® Green Primers purchased from Sigma Aldrich tested for use in qPCR reactions

Gene	Gene Name	Species	Catalogue number
GAPDH	glyceraldehyde-3-phosphate dehydrogenase	Human	H_GAPDH_1
B-Actin	B-Actin	Human	H_ACTB_1
FXR	Nuclear receptor subfamily 1, group H, member 4/ FXR	Human	H_NR1H4_1
SHP	Nuclear receptor subfamily 0, group B, member 2/ Short heterodimer partner	Human	H_NR0B2_1
BSEP	ATP-binding cassette, sub-family B (MDR/TAP), member 11/ Bile salt export pump	Human	H_ABCB11_1
FGF19	Fibroblast growth factor 19	Human	H_FGF19_1
CYP7A1	Cytochrome P450, family 7, subfamily a, polypeptide 1	Human	H_CYP7A1_1
TGFβ1	Transforming growth factor beta 1	Human	H_TGFB1_1
SREBF1	Sterol regulatory element binding transcription factor 1	Human	H_SREBF1_1

OSTα	Solute carrier family 51 alpha subunit/ Organic solute transporter alpha subunit	Human	H_OSTALPHA_1
OST β	Solute carrier family 51 beta subunit/ Organic solute transporter beta subunit	Human	H_OSTBETA_1
APOC2	Apolipoprotein C-II	Human	H_APOC2_1
NTCP	Solute carrier family 10 member 1/ Sodium Taurocholate co-transporting polypeptide	Human	H_SLC10A1_1
TIMP1	Tissue inhibitor of metalloproteinase 1	Human	H_TIMP1_1
CCR5	C-C motif chemokine receptor 5	Human	H_CCR5_1
CYP8B1	Cytochrome P450 family 8 subfamily B member 1	Human	H_CYP8B1_1
IL1 β	Interleukin 1 beta	Human	H_IL1B_1
ARF1	ADP-ribosylation factor 1	Mouse	M_ARF1_1
COX7A2L	Cytochrome c oxidase subunit VIIa polypeptide 2-like	Mouse	M_COX7A2L_1
YWHAZ	Tyrosine 3-monoxygenase/tryptophan 5-monoxygenase activation protein	Mouse	M_YWHAZ_1
FXR	Nuclear receptor subfamily 1, group H, member 4/ FXR	Mouse	M_NR1H4_1
SHP	Nuclear receptor subfamily 0, group B, member 2/ Short heterodimer partner	Mouse	M_NR0B2_1
BSEP	ATP-binding cassette, sub-family B (MDR/TAP), member 11/ Bile salt export pump	Mouse	M_ABCB11_1
CYP7A1	Cytochrome P450, family 7, subfamily a, polypeptide 1	Mouse	M_CYP7A1_1
FGF15	Fibroblast growth factor 15	Mouse	M_FGF15_1
TGFβ1	Transforming growth factor beta 1	Mouse	M_TGFB1_1
SREBF1	Sterol regulatory element binding transcription factor 1	Mouse	M_SREBF1_1
CYP27A1	Cytochrome P450, family 27, subfamily a, polypeptide 1	Mouse	M_CYP27A1_1
APOC2	Apolipoprotein C-II	Mouse	M_APOC2_1
PLTP	Phospholipid transfer protein	Mouse	M_PLTP_1
FASN	Fatty acid synthase	Mouse	M_FASN_1
ACACA	Acetyl-Coenzyme A carboxylase alpha	Mouse	M_ACACA_1
CPT2	Carnitine palmitoyltransferase 2	Mouse	M_CPT2_1
TIMP1	Tissue inhibitor of metalloproteinase 1	Mouse	M_TIMP1_1
COL1A1	Collagen, type I, alpha 1	Mouse	M_COL1A1_1
COL3A1	Collagen, type 3, alpha 1	Mouse	M_COL3A1
ACTA2	Actin, alpha 2, smooth muscle, aorta	Mouse	M_ACTA2_1
CCL2	Chemokine (C-C motif) ligand 2	Mouse	M_CCL2_1
CCR2	Chemokine (C-C motif) receptor 2	Mouse	M_CCR2_1
CRP	C-reactive protein, pentraxin-related	Mouse	M_CRP_1
IL6RA	Interleukin 6 receptor, alpha	Mouse	M_IL6RA_1
GYS2	Glycogen synthase 2	Mouse	M_GYS2_1
PCK1	Phosphoenolpyruvate carboxykinase 1, cytosolic	Mouse	M_PCK1_1
CASP8	Caspase 8	Mouse	M_CASP8_1
NRF2	Nuclear factor, erythroid derived 2, like 2	Mouse	M_NFE2L2_1
BCL2	B cell leukemia/lymphoma	Mouse	M_BCL2_1
SOD2	superoxide dismutase 2, mitochondrial	Mouse	M_SOD2_1
GPX1	glutathione peroxidase 1	Mouse	M_GPX1_1

* primers used for final reactions chosen due to efficient amplification of single product -bold

6.3.10 Quantitative PCR

The ready-to-use reaction mastermix, iTaq Universal SYBR Green Supermix (Biorad), containing hot-start iTaq DNA Polymerase, dNTPs, MgCl₂, SYBR Green I dye, enhancers, stabilisers and a blend of ROX and Fluorescein passive reference dyes, was used for all qPCR reactions. A typical reaction for each gene contained 5 µL 2× iTaq Universal SYBR Green Supermix, 500 nM forward and 500 nM reverse primers, approximately 5 ng cDNA and nuclease-free water to a final volume of 10 µL. Each target gene, reference gene and negative controls (without template cDNA) were run in duplicate on an Optical MicroAmp 96 well plate (Applied Biosystems). Plates were sealed with an optical adhesive seal (Applied biosystems), briefly placed on a plate shaker to mix the components and centrifuged. Reactions were run using the Applied Biosystem Step One Plus real-time PCR system, using the following cycling conditions; an initial denaturation step at 95°C for 15 minutes, 40 cycles of amplification, consisting of denaturation step at 94°C for 15 seconds, combined annealing and extension step at 60°C for 1 minute, with a single fluorescent measurement. Melting curve analysis was performed straight after each run by increasing the temperature from 60°C to 95°C in 0.3°C increments and measuring fluorescence dissociation.

6.3.10.1 Relative Quantification

Relative changes in gene expression were determined by the $\Delta\Delta C_t$ method, as described by Livak and Schmittgen, whereby the fluorescent signal of the target transcript of a treatment group is compared to the same target in another untreated group (Livak and Schmittgen, 2001). From duplicate or triplicate reactions, a mean C_t value was calculated for each biological sample. The C_t values of target genes were normalised to the C_t value of the reference gene, for both the samples of cells or mice treated with vehicle and the samples from cells/mice treated with compound calculated as:

$$\Delta C_t = \text{mean } C_{t\text{target}} - \text{mean } C_{t\text{reference}}$$

The ΔC_t values for the treated samples were then normalised against the ΔC_t values of the control samples.

$$\Delta\Delta C_t = \Delta C_{t\text{treated}} - \Delta C_{t\text{control}}$$

Finally, the expression ratio of the target gene in the treated group, relative to the target gene in the control group, was calculated, converting the results to the logarithmic scale using the equation:

$$\text{Fold change} = 2^{-\Delta\Delta C_t}$$

6.3.11 Statistical Analysis

Fold change data are reported as mean \pm Standard Error of Mean (SEM). Standard deviation, standard error and all statistical analyses were performed using the ΔCt values rather than the transformed fold change ratios so that the distribution of the data could be normalised to a linear scale. ΔCt values were compared between groups using a One-Way ANOVA test, with Tukey's Post hoc analyses to make all possible pairwise comparisons. In cases where data had extreme outliers or was not normally distributed as assessed by visual inspection of boxplots and a Shapiro-Wilk's test, a non-parametric Kruskal-Wallis H test, or a robust Welch's ANOVA test was performed with a Games-Howell post hoc test. For all analyses, a P value less than 0.05 was considered to be statistically significant.

6.4 Results

6.4.1 Effect of Lead Compounds *in vitro*

6.4.1.1 Selection of Appropriate Reference and Target Genes

In order to assess the effect and influence of lead compounds, Jed441 and Jed561, on endogenous FXR-mediated gene regulation, the robust and sensitive method of real-time qPCR was used. Two hepatocyte-derived, phenotypically stable cell lines, Huh7 and HepG2, were treated with the compounds and the consequential up- or downregulation of specific genes observed. FXR impacts a considerable number of genes, and the mRNA and protein expression distribution of several of these have been described previously (Fiorucci *et al.*, 2007, Dash *et al.*, 2017). However, the Expression Atlas repository (Papatheodorou *et al.*, 2018) was queried for data and information from microarray and RNAseq studies, carried out in these cell lines, in order to determine which genes would be expressed to detectable levels under baseline standard culture conditions. The results from these queries, shown in Appendix 5, collating raw data from RNAseq experiments carried out by Genentech and the Cancer Cell Line Encyclopedia, identified genes that were expressed in only Huh7 cells, HepG2 or both. Genes that displayed detectable transcripts in either or both cell lines, and that had primers that were readily available to purchase from the Kicqstart® SYBR® green primer catalogue (Sigma- Aldrich), were selected for use in real-time qPCR assays. Where possible, primers that overlapped an exon-exon boundary of a gene were chosen.

Prior to experimental work, these predesigned, commercial primers were verified for their specificity and efficiency. Primers were initially checked using the Primer-BLAST tool (Ye *et al.*, 2012) to test for any self-complementarity and possible non-specific targets (results not shown). Once it was satisfied that these primers were unlikely to form primer dimers, or amplify off-target products, they were tested further by performing a qPCR standard curve. Complementary DNA transcribed from either human reference RNA (Agilent), or RNA isolated from untreated, classically cultured Huh7 or HepG2 cells, was serially diluted 10-fold to form a series of standards, whereby the starting standard concentration was comprised of approximately 10 ng cDNA. These standards were used as template DNA in qPCR reactions with each primer, and assays were run under constant automated PCR conditions, consisting of a 95°C denaturation step, followed by a single combined 60°C annealing and elongation step. The C_t values generated from qPCR experiments were used for analysis of gene expression. C_t values are the point at which the fluorescence, due to amplification, exceeds any background fluorescence and crosses the threshold, correlating to initial transcript levels. The human

reference RNA is prepared by pooling RNA from several different whole human tissues (Agilent) and expression of many of the liver-specific genes was very low, especially at the lowest concentrations. This resulted in very high (>35 cycles) and wide-ranging C_t values, making the generation of accurate standard curves for each primer difficult (data not shown). Instead, cDNA from untreated (control samples) Huh7 or HepG2 cells were used in subsequent tests. At the highest cDNA concentration used, approximately 10 ng cDNA template, C_t values ranged from 15 to 31 cycles. A higher C_t value corresponds to a low abundance of starting cDNA, and so primers for genes that produced C_t values above 25 cycles were excluded from final assays.

A standard curve was constructed from the mean C_t values (y-axis) versus the log cDNA dilutions (x-axis). The slope value was calculated from each curve and used to determine the primer amplification efficiency as per the equation described by Ginzinger (2002):

$$\text{Efficiency} = 10^{(-\frac{1}{\text{slope}})} - 1$$

or alternatively,

$$\% \text{Efficiency} = (10^{(-\frac{1}{\text{slope}})} - 1) \times 100$$

In an ideal reaction whereby amplification is completely efficient, the template doubles with each cycle during the exponential stages of amplification, meaning that the efficiency value is 2 (100%), and the slope of the standard curve, approximately -3.2. The primers tested achieved amplification efficiencies of between 70% and 160% (Appendix 7). Primer efficiencies between 80 and 100% are deemed acceptable to produce robust, reproducible and reliable data. In some cases, where the primer efficiencies did not achieve the 80% cut-off threshold, it is possible that the low primer efficiency is as a result of the standard primer concentrations (500 nM) and annealing temperatures (60°C) applied to all assays. An annealing temperature that is 5°C lower than the primers' melting temperature is usually recommended and, in some cases, the estimated melting temperature (as quoted by the manufacturers) was as low as 55.5°C, as with GAPDH, and may account for lower efficiencies than expected. In reality, the actual primer melting temperature will be slightly different due to SYBR® green master mix buffers, salts and reagents, and due to the automated nature of data collection, a 60°C annealing temperature was retained. In other cases, e.g. FXR, a low efficiency value can indicate that the limit of detection was reached during the assays, where in the lowest standards of cDNA, the target

transcripts were negligible and not amplified. In these cases, the correlation coefficient (R^2 value) of the curve, and the amplification plot was observed and were often indicated by plots which had uniformly spaced, sigmoidal amplification traces in the initial standards, but not in the lower concentrations where fluorescence trails off or is not detected, as seen in Figure 6. 1, where fluorescence signal is not increased above baseline, and FXR was not amplified in the later cycles corresponding to lowest concentrations of starting cDNA.

To verify the specificity of the primers, melt curve analysis was included at the end of each qPCR run. The intercalating SYBR® green dye only fluoresces when bound to double-stranded DNA. By incrementally increasing the temperature and continually measuring the fluorescence, as the double stranded DNA begins to denature, the dye dissociates, reducing the fluorescent signal. When fluorescence is plotted as a function of temperature, it is possible to see the homogeneity of the sample being measured. A single peak indicates the presence of a single product, whereas multiple or shouldering peaks indicates the presence of multiple amplicons or non-specific annealing. Products were further confirmed by DNA agarose gel electrophoresis. Data from melt curves, standard curves and amplification plots of primers tested can be found in Appendix 7. Primers that were determined to be sensitive, efficient and specific for genes modulated by FXR were selected for experimental assays.

For the accurate relative quantification of mRNA in qPCR reactions, a suitable endogenous standard or reference gene is essential in order to normalise the target gene expression and exclude any cell to cell variability in transcription levels. An inherent assumption of this method is that reference genes show stable and unregulated expression in the tissue of interest, and importantly, across tissues that may have undergone different treatments. For this reason, housekeeping genes, which are found constitutively expressed in every nucleated cell to maintain basic cellular function, are usually chosen for this purpose. For the selection of a reference gene, in addition to the analyses described here, candidate genes (GAPDH and β -actin) were measured for their stability across treated and untreated cells. RNA from Huh7 cells treated with vehicle (DMSO), or OCA at its EC₅₀ or EC₉₀ concentrations were transcribed to cDNA and used to create dilution standards. The same cDNA samples were used as templates in assays with either GAPDH or β -actin primers. The C_t values displayed in Table 6. 3, show that whilst both primers were able to amplify products to within 1 C_t value across different samples at the same concentration of starting cDNA, the standard deviation for GAPDH values were overall, slightly smaller, suggesting less variation between differentially treated samples. Furthermore, although β -actin had a higher amplification efficiency (87%) compared to GAPDH, (73%), the biggest determinant in choosing GAPDH as a reference gene was due to the fact that

a single melt curve was produced, whereas β -actin appeared to amplify several, non-specific products (Figure 6. 2).

Table 6. 4 Comparison of GAPDH and β -actin C_t values across different human hepatic cell line samples

cDNA dilution	GAPDH C_t Values				B-actin C_t Values			
	Vehicle-treated	OCA (300nM)-treated	OCA (1 μ M)-treated	St. dev	Vehicle-treated	OCA (300nM)-treated	OCA (1 μ M)-treated	St. dev
10^0	16.38	16.20	15.33	0.57	15.50	14.74	14.32	0.60
10^{-1}	20.46	20.03	19.97	0.27	18.81	18.49	17.90	0.46
10^{-2}	24.97	24.65	24.49	0.24	22.88	22.81	22.74	0.07
10^{-3}	28.78	29.75	29.03	0.51	26.84	27.20	26.45	0.37
10^{-4}	33.36	33.70	33.00	0.35	29.82	30.52	30.90	0.55
mean				0.39				0.41
Amplification efficiency (%)	73	67	67		87	77	74	
Accuracy of standard curve (coefficient R^2)	0.999	0.998	0.9992		0.997	0.998	0.998	

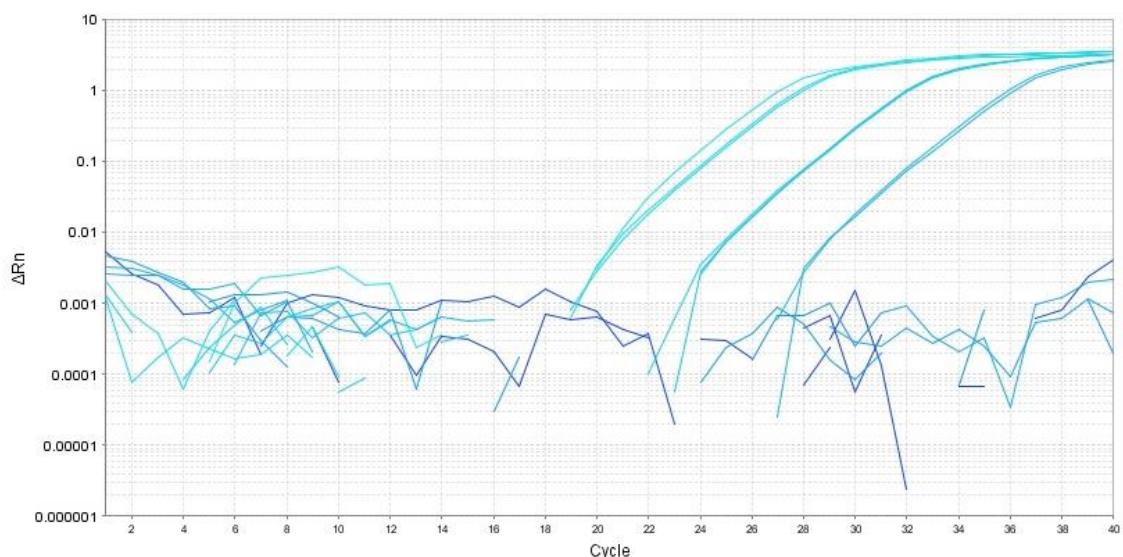


Figure 6. 1 Limit of detection of qPCR assays using primers for FXR. Amplification plot of FXR amplified from 5 cDNA standards. Lowest 2 dilutions did not have detectable levels of FXR gene transcripts and amplification traces were not produced. N=3 per dilution

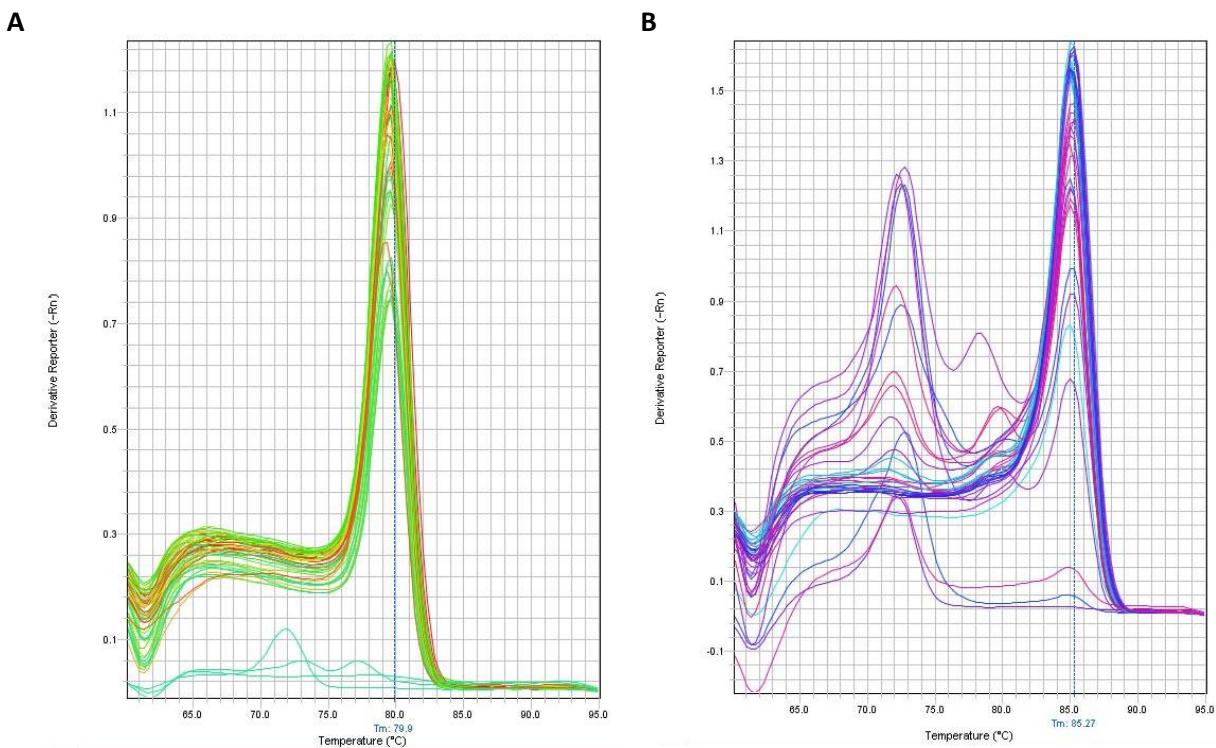


Figure 6.2 Specificity of qPCR assays using reference gene primers. Melt curve plots showing dissociation peaks for (A) GAPDH primers and (B) β -actin primers. N=15

6.4.1.2 Effect of Jed441 in Huh7 cells

In order to observe the effect of novel ligand Jed441 on FXR target genes, Huh7 cells were cultured under the treatment conditions outlined in Table 6. 1. This regime was employed to determine the most effective dose for treatment, as well as to identify any differences in gene expression in response to induction time. Cells were also cultured with OCA at equivalent EC₅₀ and EC₉₀ values, to act as a positive control and representative archetypal steroid FXR agonist, with which to compare efficacy of the novel compounds.

Results displayed in Figure 6. 3, showed that following 6 hours incubation with OCA or Jed441, FXR expression remained relatively stable with no significant up- or downregulation in mRNA levels (Figure 6. 3A)). However, direct FXR target genes, *nr0b2* (SHP), *slc51a* (OST α), and *fgf19* (FGF19), involved in BA synthesis and homeostasis, displayed increases in mRNA levels, as expected, following agonist induced FXR activation. Moreover, Jed441 appeared to be effective at upregulating said target genes, more so than OCA at its corresponding EC₅₀ and EC₉₀ concentrations. Even at its EC₅₀ concentration, Jed441 induced a significant 4-fold rise in SHP mRNA levels, a rise which is similar to the level seen in cells incubated with OCA at its EC₉₀ concentration. This fold change in SHP expression for Jed441 EC₅₀ treatment, doubled when the

EC₉₀ concentration is used (Figure 6. 3B). OST α mRNA expression levels displayed considerable upregulation upon Huh7 incubation with Jed441. Likewise, at its EC₅₀ concentration, Jed441 induced highly significant changes in OST α , increasing mRNA expression by 5-fold, compared to levels detected in untreated control samples. This even outperforms OCA at its EC₉₀ concentration, which induced 4.5-fold rise in OST α expression (Figure 6. 3C). In FGF19 expression, again a dose-response-like effect was observed, where increasing concentrations of both OCA and Jed441 induced increasing changes in mRNA levels. Although this trend was observed, it is not statistically significant, as indicated by the large error bars (Figure 6. 3D).

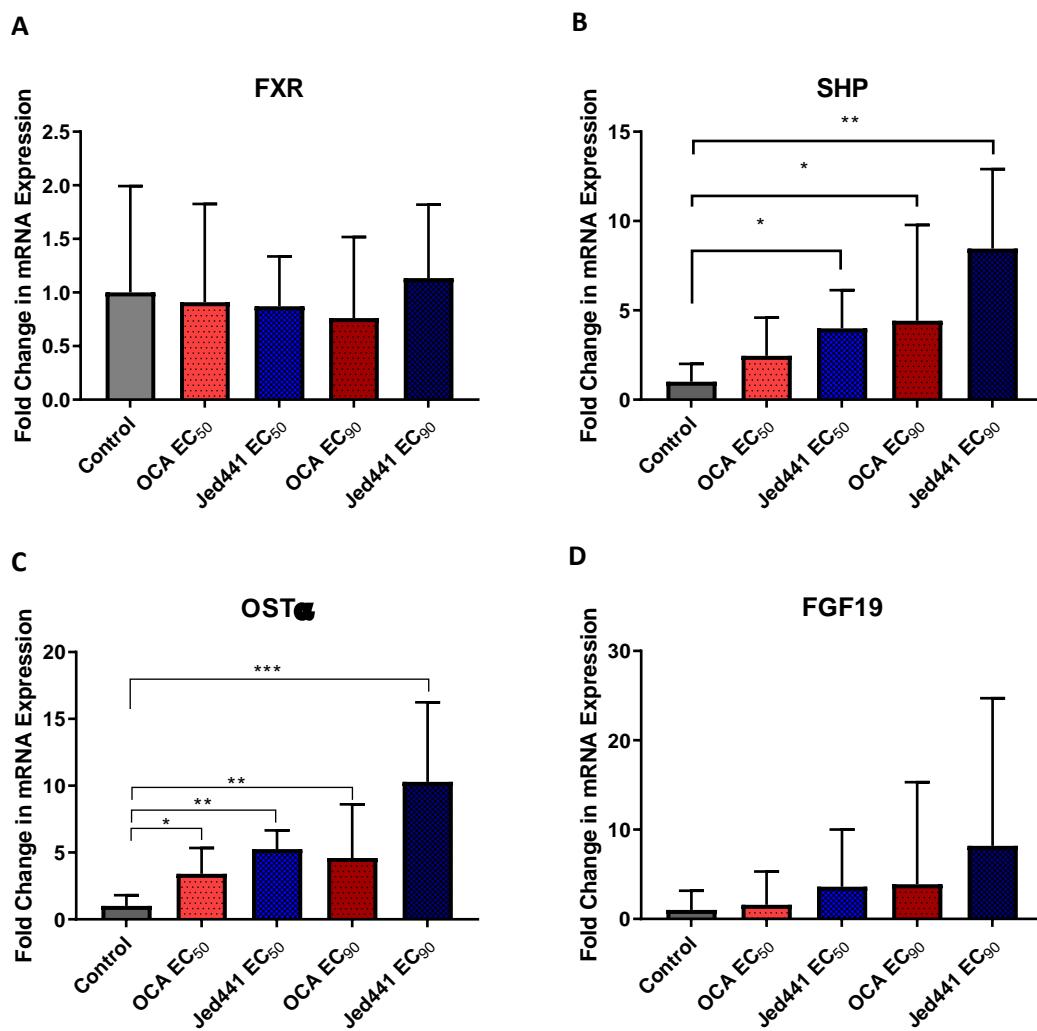


Figure 6. 3 FXR and FXR-target gene expression in Huh7 cells following 6 hours incubation with OCA or Jed441 at their respective EC₅₀ and EC₉₀ concentrations. Huh7 cells were treated with the vehicle DMSO, OCA at its EC₅₀ (300 nM) or EC₉₀ (1 μ M), or Jed441 at its EC₅₀ (27 nM) or EC₉₀ (180 nM) for 6 hours. Relative mRNA expression levels of (A) FXR, (B) SHP, (C) OST α and (D) FGF19 were quantified by real-time qPCR. For each experiment, target gene expression is normalised to GAPDH expression, and the level of target gene mRNA for DMSO-treated cells (control) set to 1, and expression levels in other test conditions shown relative to this. Error bars represent \pm SEM (n=3), statistical significance analysed by one way ANOVA with Tukey's post hoc test. * p<0.05, ** p<0.01, *** p<0.001 versus vehicle control.

Following 24 hours incubation of Huh7 cells with either OCA or Jed441, slight differences in gene expression were observed (Figure 6. 4). FXR mRNA levels in treated cells was not significantly different compared to the untreated control cells (Figure 6. 4A). SHP and FGF19, which are both directly involved in regulating BA synthesis, did not display significant changes in expression after 24 hours. Only Jed441 at its EC₉₀ concentration is able to double the expression levels of both SHP and FGF19, though again, results are not statistically significant (Figure 6. 4B and D). Conversely, FXR agonists were able to induce considerable increases in OST α expression, even after 24 hours of incubation. While OCA was able to upregulate OST α by between 11 and 18-fold, Jed441 was able to increase levels by between 19 and 38-fold, at its EC₅₀ and EC₉₀, respectively (Figure 6. 4C). These results are comparable to GW4064, which, when used at a concentration of 200nM to treat Huh7 cells for 24h, displayed a 30-fold increase in OST α expression (Landrier *et al.*, 2006). In addition to having a very highly statistically significant difference in OST α expression between treated and untreated cells; at their respective EC₉₀ concentrations, Jed441 induced OST α expression is significantly higher than OCA induced expression. Furthermore, expression seen in Jed441 EC₅₀ treated cells, is also slightly higher than that seen in cells treated with OCA EC₉₀.

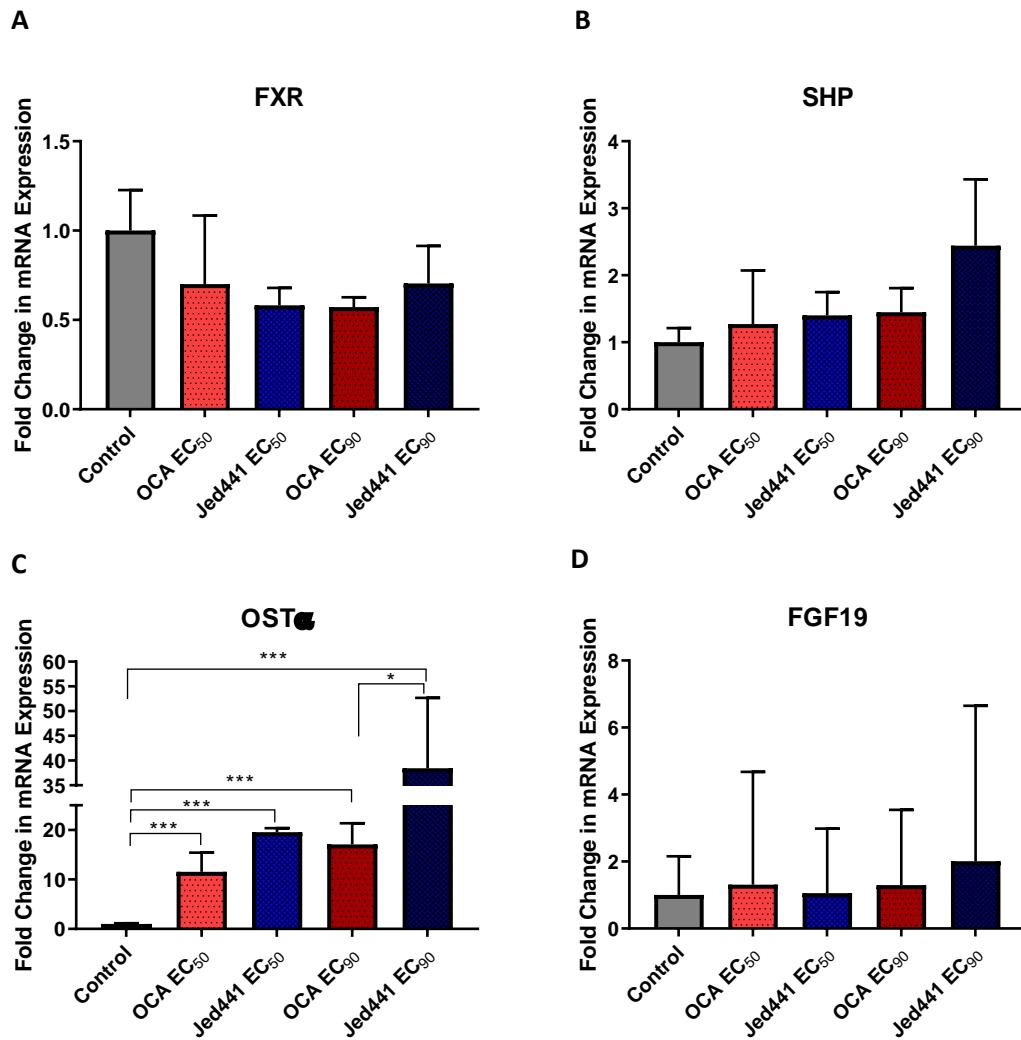


Figure 6. 4 FXR and FXR-target gene expression in Huh7 cells following 24 hours incubation with OCA or Jed441 at their respective EC₅₀ and EC₉₀ concentrations. Huh7 cells were treated with the vehicle DMSO, OCA at its EC₅₀ (300 nM) or EC₉₀ (1 μ M), or Jed441 at its EC₅₀ (27 nM) or EC₉₀ (180 nM) for 24 hours. Relative mRNA expression levels of (A) FXR, (B) SHP, (C) OST α and (D) FGF19 were quantified by real-time qPCR. For each experiment, target gene expression is normalised to GAPDH expression, and the level of target gene mRNA for DMSO-treated cells (control) set to 1, and expression levels in other test conditions shown relative to this. Error bars represent \pm SEM (n=3), statistical significance analysed by one way ANOVA with Tukey's post hoc test. * p<0.05, ** p<0.01, *** p<0.001 versus vehicle control.

Experiments also observed two genes involved in FXR-mediated regulation of lipid metabolism, *Apoc2* (APOC2), which is directly targeted and upregulated by FXR, and *Srebf1* (SREBF1), which is inhibited by SHP. Results in Figure 6.5A and B showed that, after 6 hours incubation with FXR agonists, no significant changes were seen in either APOC2 or SREBF1 mRNA. Although APOC2 is regulated by FXR, it requires the long-range interaction between the APOC2 promoter and upstream FXR-bound hepatic control regions. APOC2 is also under the complex regulation of other receptors such as the thyroid receptor, which may be one reason as to why no significant changes were seen. Likewise, although FXR indirectly regulates SREBF1, by upregulating SHP, it

is thought that SREBF1 is under the additional regulation of non FXR-dependent pathways (Watanabe *et al.*, 2004), which may be a potential reason as to why these agonists are not able to affect change in downstream SREBF1 levels. On the other hand, after 24 hours, Jed441 was able to significantly upregulate APOC2 1.4-fold, although only at its EC₉₀ concentration (Figure 6. 5C). SREBF1 levels appeared to be reduced in cells treated with Jed441 or the highest concentration of OCA, however error bars were fairly high, and these changes were not statistically significant (Figure 6. 5D).

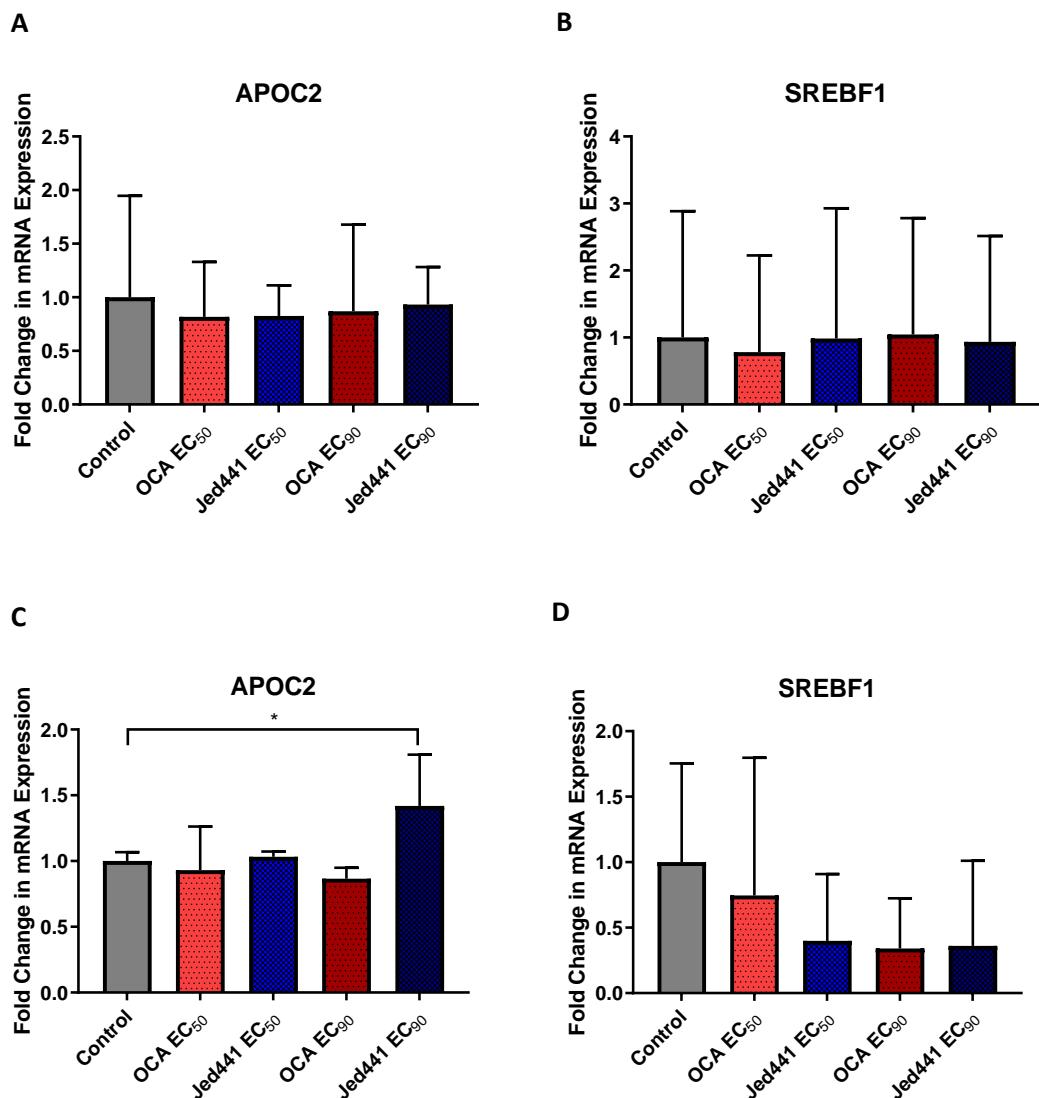


Figure 6. 5 Expression of genes involved in FXR-mediated lipid metabolism pathways in Huh7 cells following 24 hours incubation with OCA or Jed441 at their respective EC₅₀ and EC₉₀ concentrations. Huh7 cells were treated with the vehicle DMSO, OCA at its EC₅₀ (300 nM) or EC₉₀ (1 μ M), or Jed441 at its EC₅₀ (27 nM) or EC₉₀ (180 nM) for (A and B) 6, or (C and D) 24 hours. Relative mRNA expression levels of (A, C) APOC2, (B, D) SREBF1 were quantified by real-time qPCR. For each experiment, target gene expression is normalised to GAPDH expression, and the level of target gene mRNA for DMSO-treated cells (control) set to 1, and expression levels in other test conditions shown relative to this. Error bars represent \pm SEM (n=3),

statistical significance analysed by one way ANOVA with Tukey's post hoc test. * $p<0.05$, ** $p<0.01$, *** $p<0.001$ versus vehicle control.

Although not directly regulated by FXR, TGFB1, is involved in fibrogenesis, and contributes to the development of NASH. Several *in vivo* models of NASH have demonstrated an increase in TGF β 1 mRNA expression and protein levels (Kong *et al.*, 2009; Han, 2018). Moreover, it has also been demonstrated that there is a decrease in TGF β 1 and other inflammatory marker genes in response to activation of FXR by synthetic or semi-synthetic agonists (Goto *et al.*, 2018; Hye Khan *et al.*, 2019). TGF β 1 was tested *in vitro* here, in order to determine whether FXR activation, via Jed compounds may be able affect fibrosis and other pathways involved in the aetiology of NASH. *In vitro* results however, displayed no significant changes in TGF β 1 mRNA expression in response to treatment with FXR agonist for neither 6 hours nor 24 hours (Figure 6. 6).

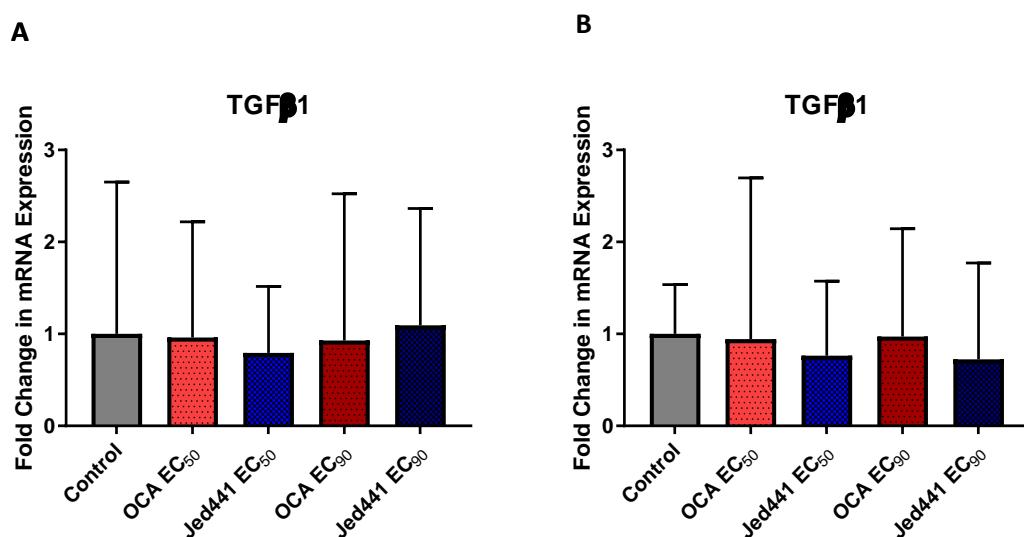


Figure 6. 6 TGF β 1 mRNA expression in Huh7 cells following (A) 6, and (B) 24 hours incubation with OCA or Jed441 at their respective EC₅₀ and EC₉₀ concentrations. Huh7 cells were treated with the vehicle DMSO, OCA at its EC₅₀ (300 nM) or EC₉₀ (1 μ M), or Jed441 at its EC₅₀ (27 nM) or EC₉₀ (180 nM) for (A) 6, or (B) 24 hours. Relative mRNA expression levels of TGF β 1 were quantified by real-time qPCR. For each experiment, target gene expression is normalised to GAPDH expression, and the level of target gene mRNA for DMSO-treated cells (control) set to 1, and expression levels in other test conditions shown relative to this. Error bars represent \pm SEM (n=3), statistical significance analysed by one way ANOVA with Tukey's post hoc test. * $p<0.05$, ** $p<0.01$, *** $p<0.001$ versus vehicle control.

6.4.1.3 Effect of Jed441 in HepG2 cells

The effects of Jed441 in HepG2 cells further support the hypothesis that it is a potent FXR agonist. After 6 hours, incubation of HepG2 with OCA or Jed441 displayed relatively stable expression of FXR mRNA, which did not significantly differ from baseline levels (Figure 6. 7A).

Conversely, FXR-target genes displayed deviations in expression, although somewhat modest changes, when cells were treated with Jed441. Only in the Jed441 EC₉₀ treatment group were statistically significant increases in SHP seen. Although Jed441 EC₉₀ induced 1.8-fold increases in SHP mRNA, this response appears to be dampened compared to the 8-fold increase in response seen in Huh7 cells. Nonetheless Jed441 produces significantly higher levels of SHP than that observed with OCA, both at their respective EC₉₀ concentrations (Figure 6. 7B). The modest increases in SHP mRNA expression, following treatment with FXR agonists, accompanies a downregulation in *cyp7a1* expression. Whilst OCA treatment halves CYP7A1 mRNA levels at both its EC₅₀ and EC₉₀ concentrations, Jed441 reduces expression to 0.3 and 0.2- fold of that seen under control conditions (Figure 6. 7C). As in Huh7 cells, OCA and Jed441 treatment both increased OST α mRNA production in a dose dependent manner. Again, at its EC₅₀ concentration, Jed441 was able to achieve increases in OST α mRNA to levels comparable to those induced by treatment with OCA at its EC₉₀ (Figure 6. 7D).

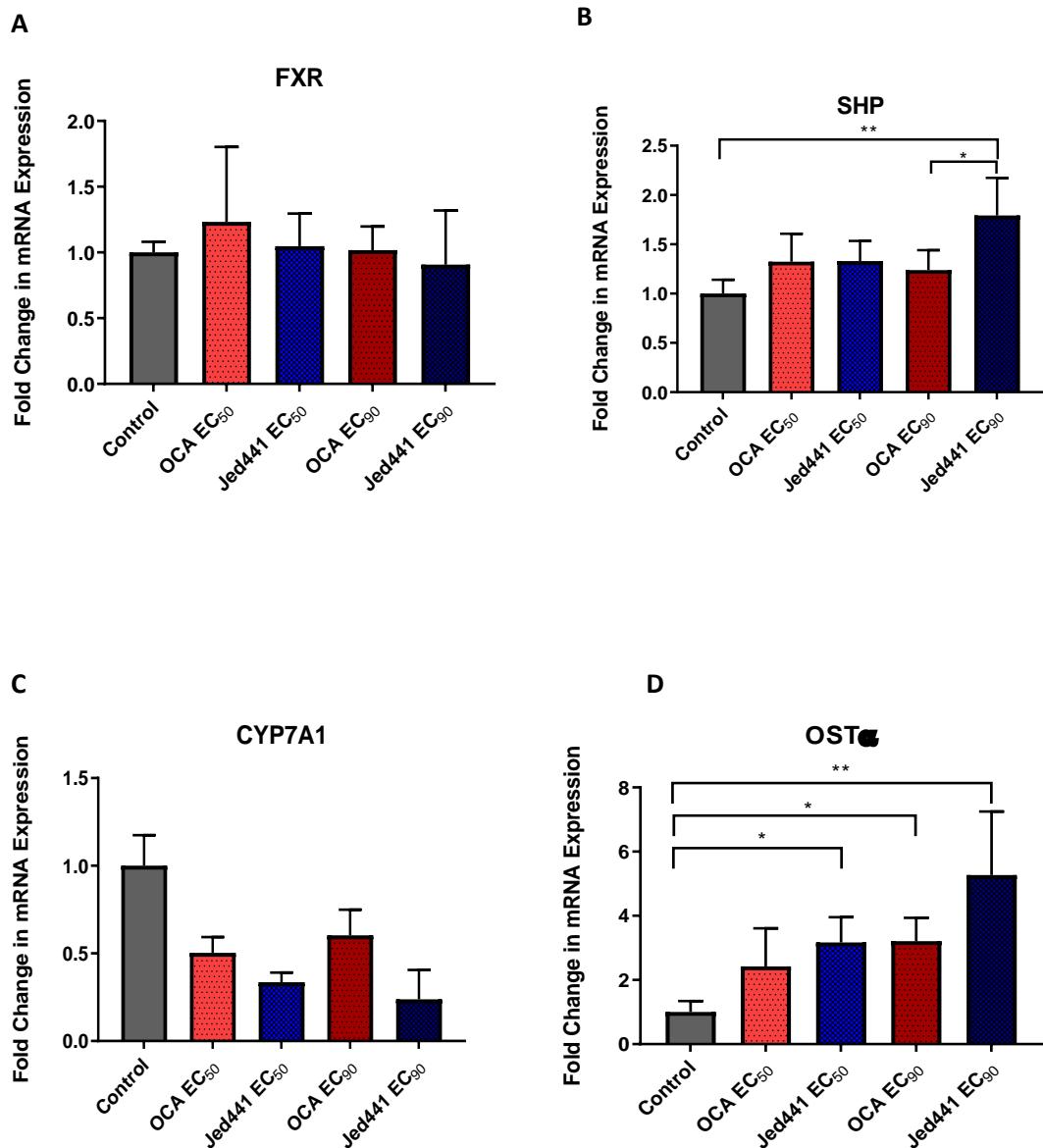


Figure 6. 7 FXR and FXR-target gene expression in HepG2 cells following 6 hours incubation with OCA or Jed441 at their respective EC₅₀ and EC₉₀ concentrations. HepG2 cells were treated with the vehicle DMSO, OCA at its EC₅₀ (300 nM) or EC₉₀ (1 μM), or Jed441 at its EC₅₀ (27 nM) or EC₉₀ (180 nM) for 6 hours. Relative mRNA expression levels of (A) FXR, (B) SHP, (C) CYP7A1 and (D) OST α were quantified by real-time qPCR. For each experiment, target gene expression is normalised to GAPDH expression, and the level of target gene mRNA for DMSO-treated cells (control) set to 1, and expression levels in other test conditions shown relative to this. Error bars represent ±SEM (n=3), statistical significance analysed by one way ANOVA with Tukey's post hoc test. *p<0.05, **p<0.01, ***p<0.001 versus vehicle control.

After 24 hours incubation with the compounds, FXR mRNA expression started to decrease, in cells treated with Jed441 or OCA at its EC₉₀ (Figure 6.8A). These decreases were only statistically significant in the Jed441 EC₅₀ treatment group, which displayed 0.6-fold changes in FXR mRNA levels. Factors contributing to the regulation of FXR expression are not well known. It has been reported that various transcriptional coregulators could potentially affect its expression, as well as the occurrence of posttranslational modifications (Kemper, 2011), but it is also possible that

this downregulation arises due to a negative feedback mechanism, which is compensating for the prolonged activation of FXR. After 24 hours incubation in HepG2, none of the compounds were able to sustain prolonged increases in SHP mRNA expression and levels had returned to basal amounts where there were no differences seen compared to the control group (Figure 6. 8B). This apparent lack of induction of SHP expression may in part be due to a more transient response whereby increases in SHP mRNA are seen shortly after incubation with the FXR agonist, but not sustained after 24 hours. In a previous study with HepG2 cells, SHP mRNA levels were shown to rapidly increase after 6 hours incubation with CDCA, but then gradually decline, almost returning to basal levels by 24 hours (Barbier *et al.*, 2003). Furthermore, SHP protein has been shown to be rapidly degraded, with a half-life of 20-30 minutes in HepG2 cells and its stability is thought to be increased by FGF19 (Miao *et al.*, 2009). RNAseq data from the Expression Atlas database (EMBL_EBI) shows that FGF19 is not intrinsically expressed at very high levels in HepG2 under standard culture conditions, offering another possible suggestion as to why increases in SHP are not seen in this cell line after 24 hours. Nonetheless, although the response in SHP mRNA expression may be short-lived, its effects were observed. The decrease in the downstream SHP target, CYP7A1, persisted at 24 hours in cells treated with Jed441 EC₅₀, where CYP7A1 mRNA is still statistically, significantly lower compared to control cells (Figure 6. 8C).

Both OCA and Jed441 were able to induce prolonged upregulation in OST α expression levels by up to 6- and 9-fold respectively, whereby highly significant increases in OST α were induced in a dose-dependent manner even after 24 hours incubation with compounds (Figure 6. 8D). This, again, is comparable to GW4064 which induced a 5-fold increase in expression in HepG2 cells following 24 hours incubation with a concentration of 200nM (Landrier *et al.*, 2006).

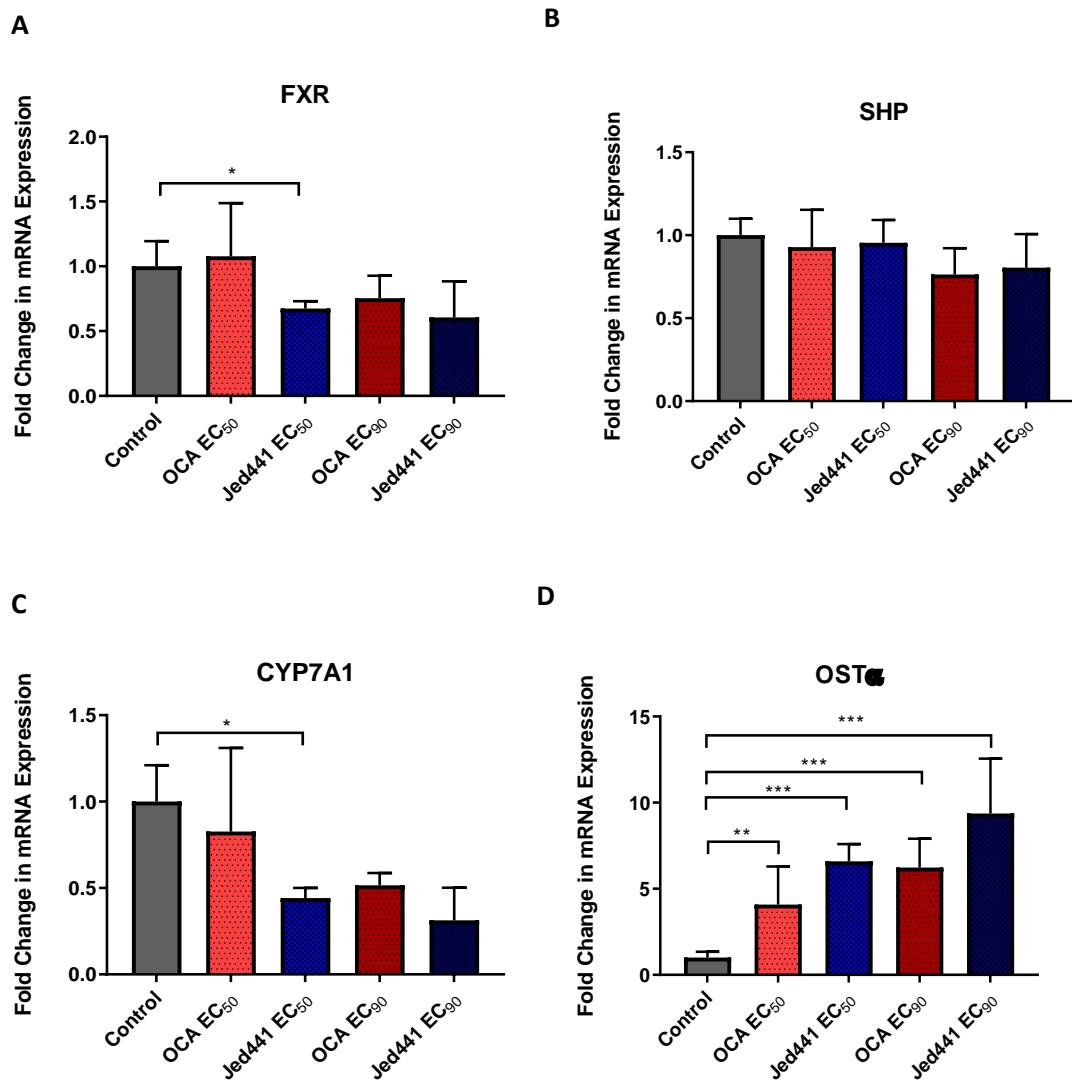


Figure 6. 8 FXR and FXR-target gene expression in HepG2 cells following 24 hours incubation with OCA or Jed441 at their respective EC₅₀ and EC₉₀ concentrations. HepG2 cells were treated with the vehicle DMSO, OCA at its EC₅₀ (300 nM) or EC₉₀ (1 μ M), or Jed441 at its EC₅₀ (27 nM) or EC₉₀ (180 nM) for 24 hours. Relative mRNA expression levels of (A) FXR, (B) SHP, (C) CYP7A1 and (D) OST α were quantified by real-time qPCR. For each experiment, target gene expression is normalised to GAPDH expression, and the level of target gene mRNA for DMSO-treated cells (control) set to 1, and expression levels in other test conditions shown relative to this. Error bars represent \pm SEM (n=3), statistical significance analysed by one way ANOVA with Tukey's post hoc test. *p<0.05, **p<0.01, ***p<0.001 versus vehicle control.

As with Huh7 cells, HepG2 cells did not show any significant changes in TGF β 1 mRNA expression following incubation with FXR agonist compounds. Incubation for neither 6 nor 24 hours was enough to induce any changes to TGF β 1 expression by OCA or Jed441 at their EC₅₀ or EC₉₀ concentrations (Figure 6. 9).

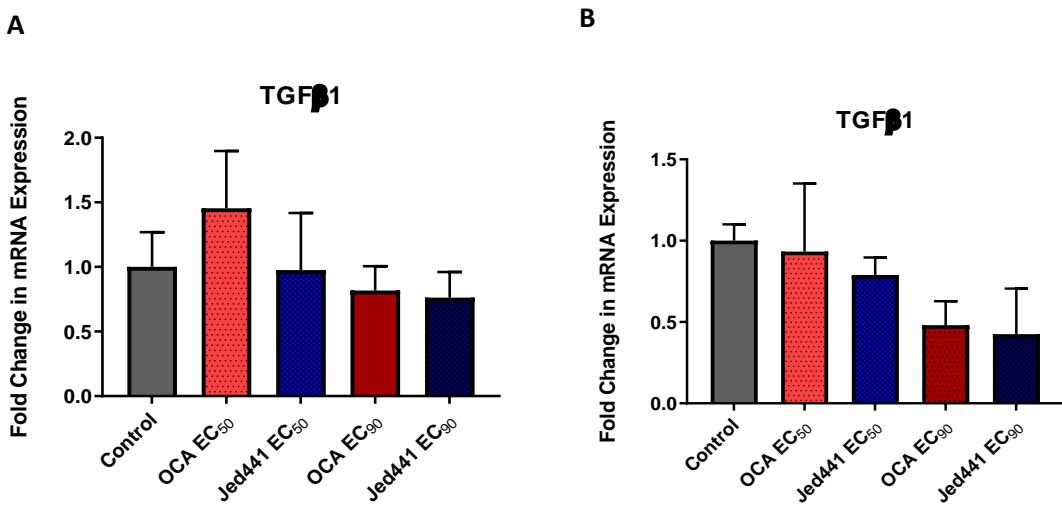


Figure 6.9 *TGF β 1 mRNA expression in HepG2 cells following (A) 6, and (B) 24 hours incubation with OCA or Jed441 at their respective EC₅₀ and EC₉₀ concentrations. HepG2 cells were treated with the vehicle DMSO, OCA at its EC₅₀ (300 nM) or EC₉₀ (1 μ M), or Jed441 at its EC₅₀ (27 nM) or EC₉₀ (180 nM) for (A) 6, or (B) 24 hours. Relative mRNA expression levels of TGF β 1 were quantified by real-time qPCR. For each experiment, target gene expression is normalised to GAPDH expression, and the level of target gene mRNA for DMSO-treated cells (control) set to 1, and expression levels in other test conditions shown relative to this. Error bars represent \pm SEM (n=3), statistical significance analysed by one way ANOVA with Tukey's post hoc test. * p<0.05, ** p<0.01, *** p<0.001 versus vehicle control.*

6.4.1.4 Effect of Jed561 in Huh7 cells

Jed561 also was tested *in vitro* to observe if and how it affects FXR-mediated gene expression. As before, OCA was used in experiments, at corresponding EC₅₀ and EC₉₀ concentrations, to act as a standard of FXR agonism. Once more, human hepatoma cells were incubated for either 6 or 24 hours with the respective compound or DMSO vehicle (control).

Again, after 6 hours incubation in Huh7 cells, FXR expression was stable across all treatment groups, and there were no significant changes in mRNA levels between untreated cells, or those treated with OCA or Jed561 (Figure 6.10A). Figure 6.10B-D, also shows that Jed561 was able to induce statistically significant increases in FXR targets, SHP, OST α and FGF19 in Huh7 cells; and in said genes, Jed561-induced upregulation, was higher than that produced by OCA, at their equivalent concentrations. While Jed561 increased SHP by 3- to 4- fold, OCA-induced increases were only 2 to 3.5-fold; likewise, Jed561- treated cells displayed rises from 4.5-fold to 7-fold in both OST α and FGF19 expression, whereas OCA- treated cells did not achieve these levels of upregulation.

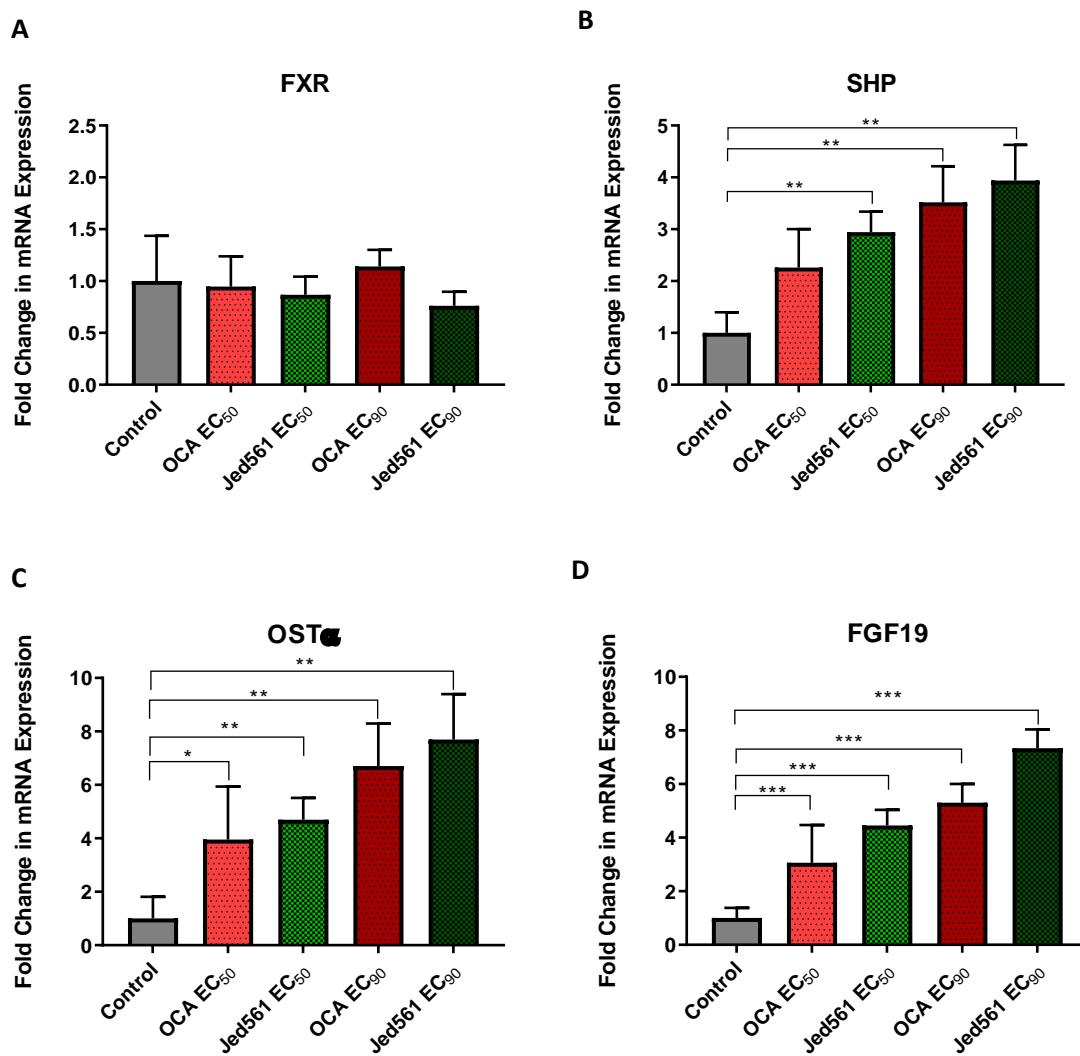


Figure 6. 10 FXR and FXR-target gene expression in Huh7 cells following 6 hours incubation with OCA or Jed561 at their respective EC₅₀ and EC₉₀ concentrations. Huh7 cells were treated with the vehicle DMSO, OCA at its EC₅₀ (300 nM) or EC₉₀ (1 μ M), or Jed561 at its EC₅₀ (92 nM) or EC₉₀ (400 nM) for 6 hours. Relative mRNA expression levels of (A) FXR, (B) SHP, (C) OST α and (D) FGF19 were quantified by real-time qPCR. For each experiment, target gene expression is normalised to GAPDH expression, and the level of target gene mRNA for DMSO-treated cells (control) set to 1, and expression levels in other test conditions shown relative to this. Error bars represent \pm SEM (n=3), statistical significance analysed by one way ANOVA with Tukey's post hoc test. * p<0.05, ** p<0.01, *** p<0.001 versus vehicle control.

Following 24 hours incubation, FXR agonist treated cells displayed FXR mRNA expression levels that weren't statistically different from baseline FXR levels (Figure 6. 11A). SHP continued to be upregulated with FXR agonist treatment, although to a lower extent than seen at 6 hours incubation (Figure 6. 11B, Figure 6. 10B). Again, Jed561 performed slightly better than OCA, inducing increases in SHP expression by 1.7- and 2.6-fold, compared to 1.5 and 2.2-fold changes, at each of their respective EC₅₀ and EC₉₀ concentrations. As seen in Figure 6. 11C, Jed561

treatment caused considerable upregulation in $OST\alpha$, again to levels surpassing those induced by OCA. These increases also are greater than those seen after 6 hours incubation, where, for example, Jed561 EC₉₀ increases $OST\alpha$ by 7-fold after 6 hours treatment (Figure 6. 10C), but 18-fold after 24 hours treatment (Figure 6. 11C). After 24 hours, FGF19 expression in FXR agonist-treated cells weren't statistically different from FGF19 levels in vehicle treated cells (Figure 6. 11D).

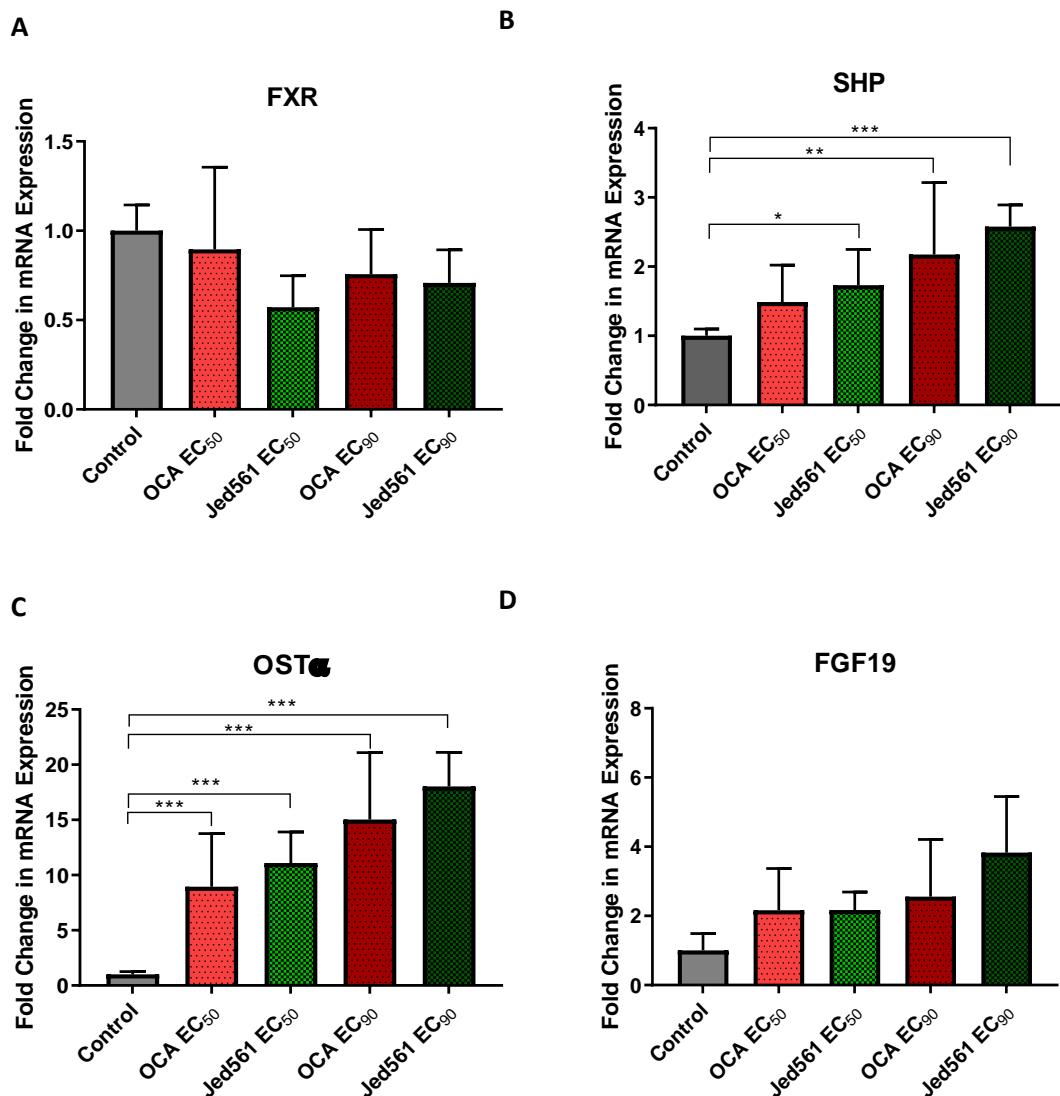


Figure 6. 11 FXR and FXR-target gene expression in Huh7 cells following 24 hours incubation with OCA or Jed561 at their respective EC₅₀ and EC₉₀ concentrations. Huh7 cells were treated with the vehicle DMSO, OCA at its EC₅₀ (300 nM) or EC₉₀ (1 μ M), or Jed561 at its EC₅₀ (92 nM) or EC₉₀ (400 nM) for 24 hours. Relative mRNA expression levels of (A) FXR, (B) SHP, (C) $OST\alpha$ and (D) FGF19 were quantified by real-time qPCR. For each experiment, target gene expression is normalised to GAPDH expression, and the level of target gene mRNA for DMSO-treated cells (control) set to 1, and expression levels in other test conditions shown relative to this. Error bars represent \pm SEM (n=3), statistical significance analysed by one way ANOVA with Tukey's post hoc test. *p<0.05, **p<0.01, ***p<0.001 versus vehicle control.

FXR-mediated genes involved in lipid metabolism remained unaffected by treatment with the FXR agonist compounds, and unlike Jed441, Jed561 did not induce any significant changes to APOC2 or SREBF1, neither at 6 nor 24 hours (Figure 6. 12). Similarly, treatment with neither OCA nor Jed561, for 6 or 24 hours, was able to prompt any changes to TGF β 1 expression in Huh7 cells (Figure 6. 13).

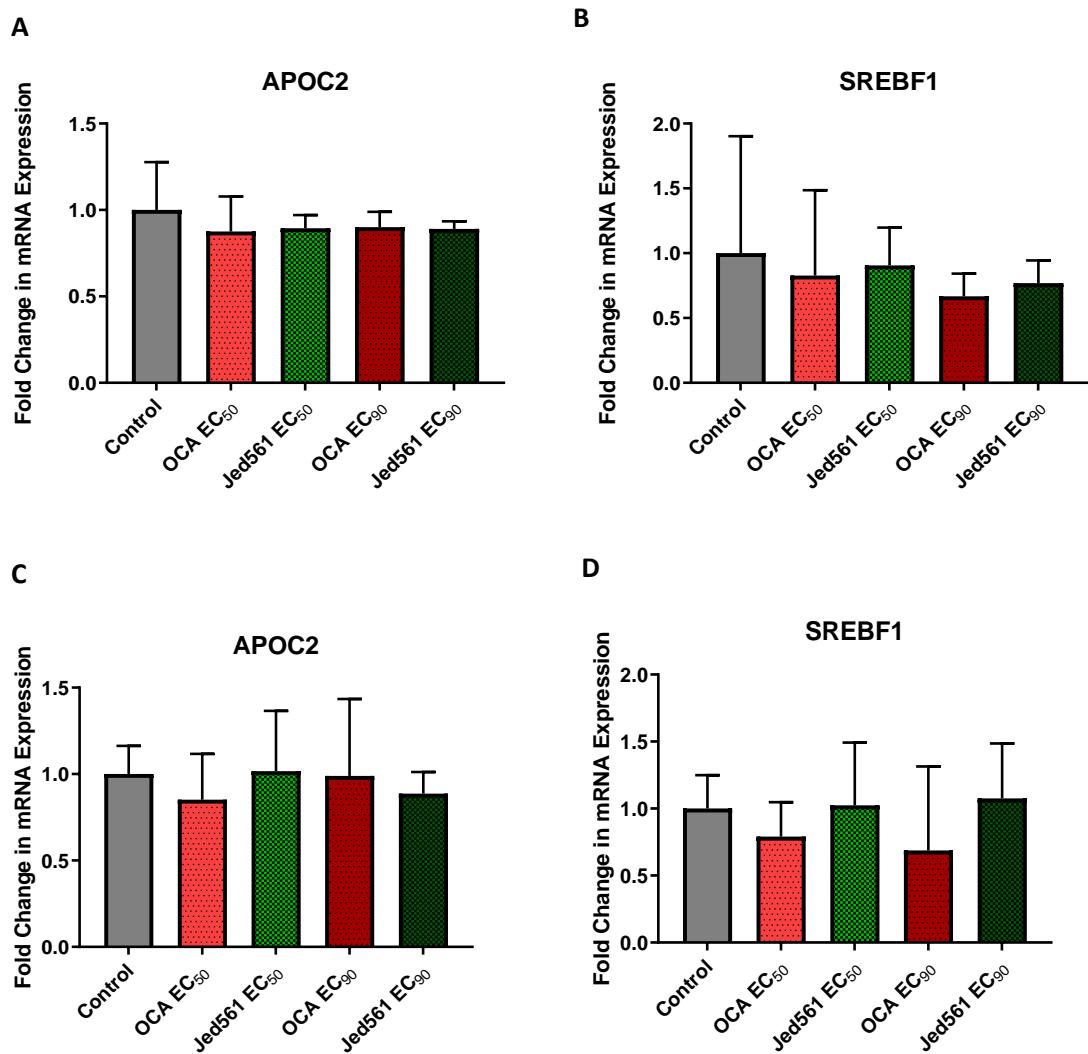


Figure 6. 12 Expression of genes involved in FXR-mediated lipid metabolism pathways in Huh7 cells following 24 hours incubation with OCA or Jed561 at their respective EC₅₀ and EC₉₀ concentrations. Huh7 cells were treated with the vehicle DMSO, OCA at its EC₅₀ (300 nM) or EC₉₀ (1 μ M), or Jed561 at its EC₅₀ (92 nM) or EC₉₀ (400 nM) for (A and B) 6, or (C and D) 24 hours. Relative mRNA expression levels of (A, C) APOC2, (B, D) SREBF1 were quantified by real-time qPCR. For each experiment, target gene expression is normalised to GAPDH expression, and the level of target gene mRNA for DMSO-treated cells (control) set to 1, and expression levels in other test conditions shown relative to this. Error bars represent \pm SEM (n=3), statistical significance analysed by one way ANOVA with Tukey's post hoc test. * p<0.05, ** p<0.01, *** p<0.001 versus vehicle control.

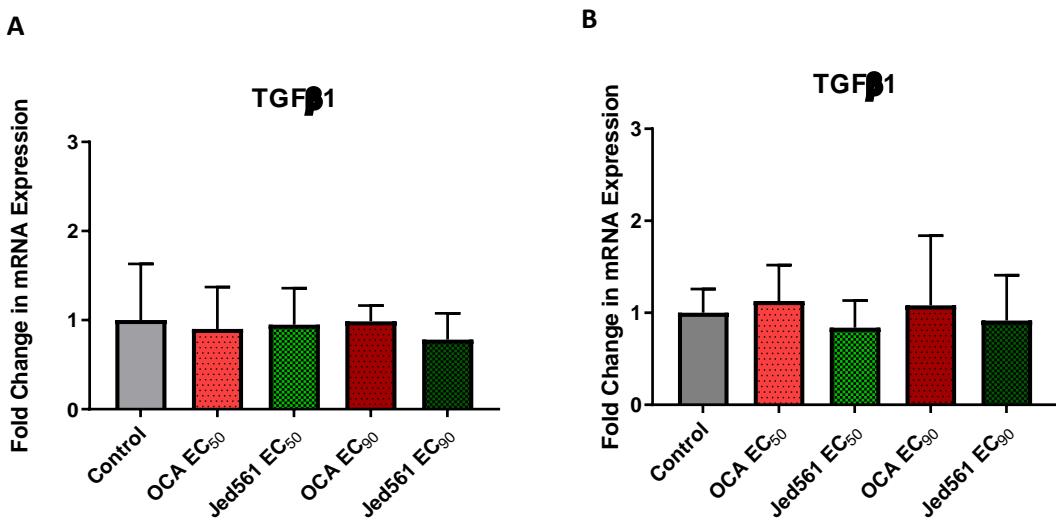


Figure 6.13 *TGF β 1 mRNA expression in Huh7 cells following (A) 6, and (B) 24 hours incubation with OCA or Jed561 at their respective EC₅₀ and EC₉₀ concentrations. Huh7 cells were treated with the vehicle DMSO, OCA at its EC₅₀ (300 nM) or EC₉₀ (1 μ M), or Jed561 at its EC₅₀ (92 nM) or EC₉₀ (400 nM) for (A) 6, or (B) 24 hours. Relative mRNA expression levels of TGF β 1 were quantified by real-time qPCR. For each experiment, target gene expression is normalised to GAPDH expression, and the level of target gene mRNA for DMSO-treated cells (control) set to 1, and expression levels in other test conditions shown relative to this. Error bars represent \pm SEM ($n=3$), statistical significance analysed by one way ANOVA with Tukey's post hoc test. * $p<0.05$, ** $p<0.01$, *** $p<0.001$ versus vehicle control.*

6.4.1.5 Effect of Jed561 in HepG2 cells

In HepG2, the effects of Jed561 on FXR-mediated gene expression were reduced, compared to effects in Huh7. Once again, following 6 hours incubation in HepG2 cells with agonist compounds, FXR gene expression was relatively stable (Figure 6.14A). Surprisingly, in HepG2 cells, neither OCA nor Jed561 were able to induce significant changes in SHP expression above 1.3-fold (Figure 6.14B). Nonetheless, despite this relatively benign response in SHP to these FXR agonists, OCA lowered CYP7A1 expression to 0.5- and 0.6-fold, and Jed561 treatment reduced mRNA to 0.6 and 0.4-fold, respective to its EC₅₀ and EC₉₀ concentrations (Figure 6.14C). As shown in Figure 6.14, OST α is the only direct FXR target gene to show any significant changes in response to culturing HepG2 cells for 6 hours in Jed561. EC₉₀ concentrations of both OCA and Jed561 are sufficient to significantly increase OST α expression 3-fold, with very little difference in fold change achieved.

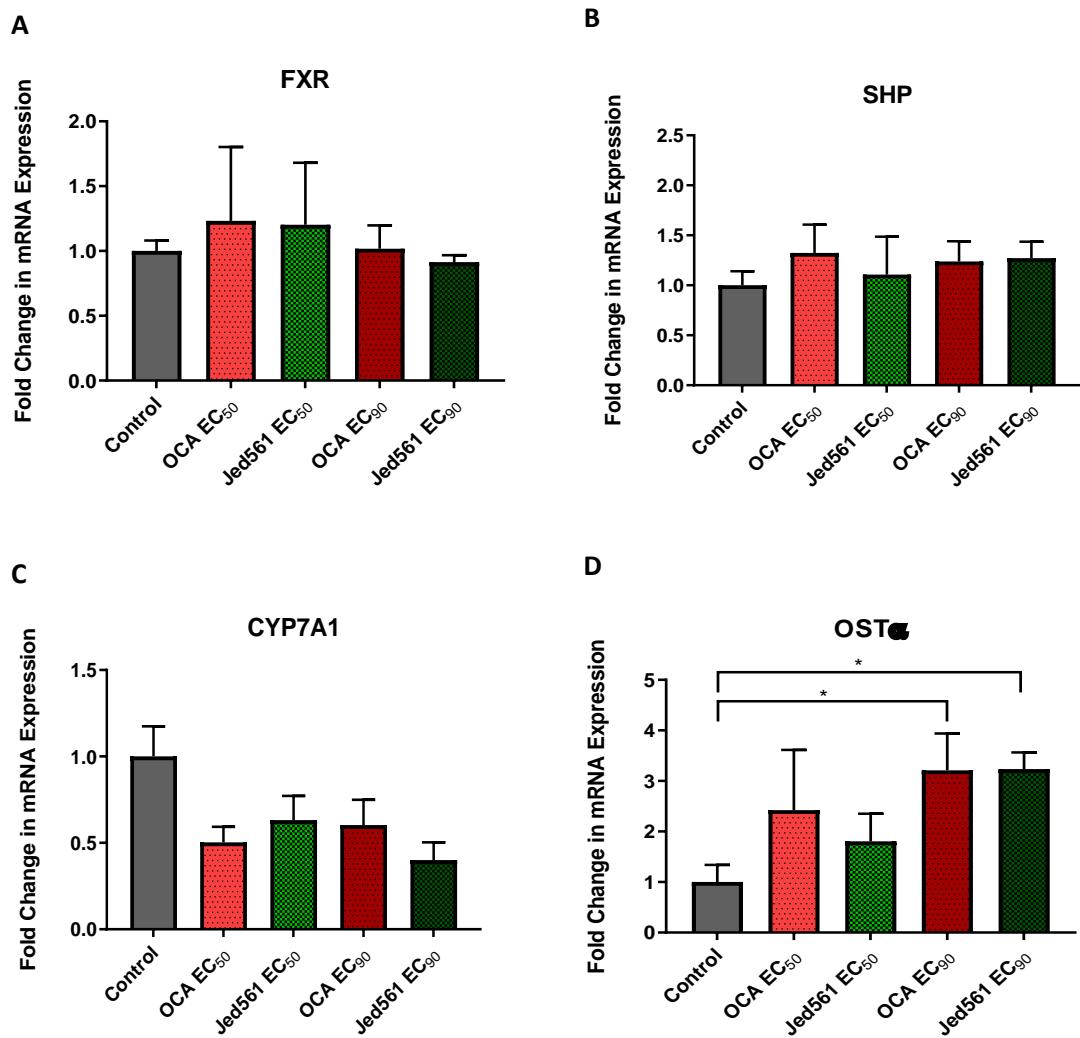


Figure 6. 14 FXR and FXR-target gene expression in HepG2 cells following 6 hours incubation with OCA or Jed561 at their respective EC₅₀ and EC₉₀ concentrations. HepG2 cells were treated with the vehicle DMSO, OCA at its EC₅₀ (300 nM) or EC₉₀ (1 μ M), or Jed561 at its EC₅₀ (92 nM) or EC₉₀ (400 nM) for 6 hours. Relative mRNA expression levels of (A) FXR, (B) SHP, (C) CYP7A1 and (D) OST α were quantified by real-time qPCR. For each experiment, target gene expression is normalised to GAPDH expression, and the level of target gene mRNA for DMSO-treated cells (control) set to 1, and expression levels in other test conditions shown relative to this. Error bars represent \pm SEM (n=3), statistical significance analysed by one way ANOVA with Tukey's post hoc test. *p<0.05, **p<0.01, ***p<0.001 versus vehicle control.

Following 24 hours incubation, HepG2 levels of FXR mRNA remained the same regardless of whether cells were untreated, or treated with OCA/ Jed561 (Figure 6. 15A). Similarly, SHP mRNA expression levels again, appeared to stay close to basal levels in all treatment groups (Figure 6. 15B). CYP7A1 expression, however, appeared to be downregulated with increasing concentrations of either OCA or Jed561, with Jed561 being able to induce a slightly larger response, but as previously, there is no statistical difference between treatment groups (Figure 6. 15C). Once again, OST α is the only gene to significantly increase with treatment by the

compounds. However, this time, OST α appears to be upregulated to a higher extent in OCA-treated cells compared to Jed561, with OCA achieving 4- fold expression and Jed 561 achieving 3.2- fold expression at their EC₅₀ concentrations, and OCA and Jed561 inducing 6- fold and 5.4- fold increases in OST α expression respectively at their EC₉₀ concentrations (Figure 6. 15D).

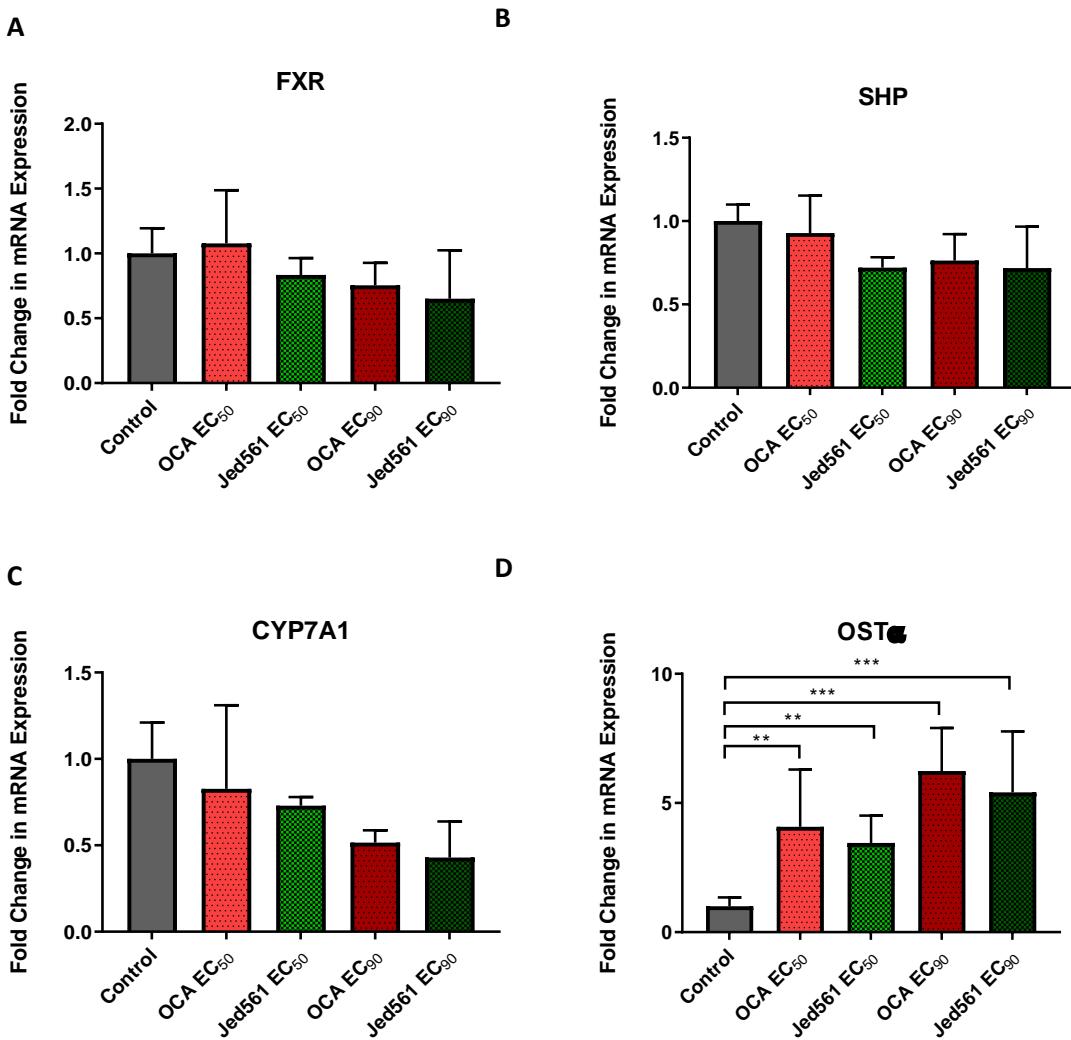


Figure 6. 15 FXR and FXR-target gene expression in HepG2 cells following 24 hours incubation with OCA or Jed561 at their respective EC₅₀ and EC₉₀ concentrations. HepG2 cells were treated with the vehicle DMSO, OCA at its EC₅₀ (300 nM) or EC₉₀ (1 μ M), or Jed561 at its EC₅₀ (92 nM) or EC₉₀ (400 nM) for 24 hours. Relative mRNA expression levels of (A) FXR, (B) SHP, (C) CYP7A1 and (D) OST α were quantified by real-time qPCR. For each experiment, target gene expression is normalised to GAPDH expression, and the level of target gene mRNA for DMSO-treated cells (control) set to 1, and expression levels in other test conditions shown relative to this. Error bars represent \pm SEM (n=3), statistical significance analysed by one way ANOVA with Tukey's post hoc test. * $p<0.05$, ** $p<0.01$, *** $p<0.001$ versus vehicle control.

Interestingly, although Jed561 did not induce any changes in TGF β 1 expression in Huh7 cells, nor in HepG2 cells with only 6 hour incubation time (Figure 6. 16A), after 24 hours treatment of HepG2 cells, there is downregulation of TGF β 1 in a dose-response manner (Figure 6. 16B). With

Jed561 treatment at its EC₅₀ dose, TGF β 1 mRNA levels are significantly reduced to 0.6-fold, and although not statistically significant, treatment with Jed561 at its EC₉₀ reduces expression to just 0.3-fold. Again, Jed561 appears to induce a larger effect than OCA in this response also.

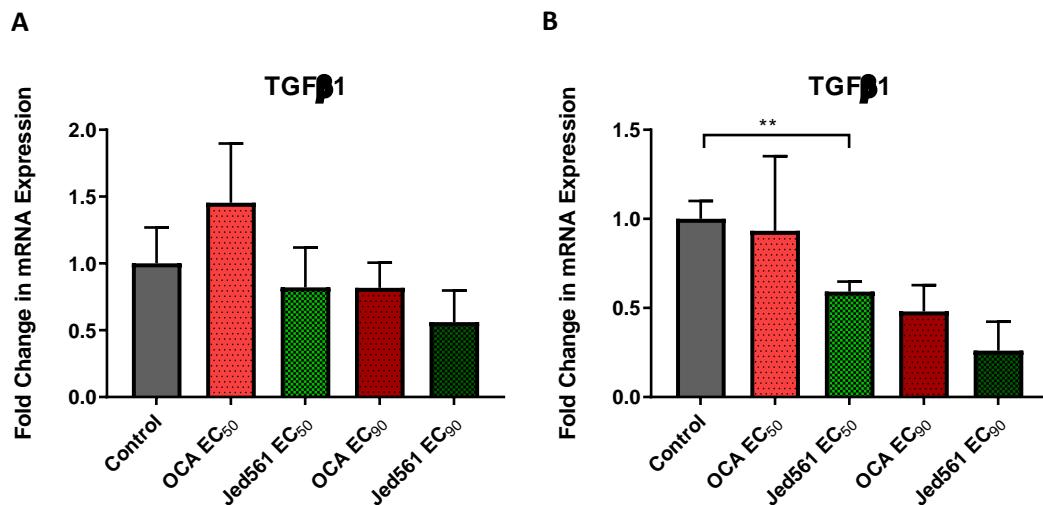


Figure 6.16 TGF β 1 mRNA expression in HepG2 cells following (A) 6, and (B) 24 hours incubation with OCA or Jed561 at their respective EC₅₀ and EC₉₀ concentrations. HepG2 cells were treated with the vehicle DMSO, OCA at its EC₅₀ (300 nM) or EC₉₀ (1 μ M), or Jed561 at its EC₅₀ (92 nM) or EC₉₀ (400 nM) for (A) 6, or (B) 24 hours. Relative mRNA expression levels of TGF β 1 were quantified by real-time qPCR. For each experiment, target gene expression is normalised to GAPDH expression, and the level of target gene mRNA for DMSO-treated cells (control) set to 1, and expression levels in other test conditions shown relative to this. Error bars represent \pm SEM ($n=3$), statistical significance analysed by one way ANOVA with Tukey's post hoc test. * $p<0.05$, ** $p<0.01$, *** $p<0.001$ versus vehicle control.

6.4.1.6 Evaluation of Ligand Specificity in HepG2 Cells

The results from Chapter 6, sections 6.3.1.2 to 6.3.1.5, indicate that novel ligands, Jed441 and Jed561 are indeed able to effect downstream target genes regulated by FXR. However, in order to determine if these responses arise specifically from the binding and activation of FXR, another set of experiments were carried out using α -guggulsterone (herein denoted as guggulsterone). Guggulsterone is a key active compound in the natural exudate of the *Commiphora mukul* tree, and had been classically described as a FXR antagonist, with little activity on other nuclear receptors (Wu *et al.*, 2002). Guggulsterone has been shown to repress CDCA and GW4064-induced FXR activity in a luciferase reporter assay, and to displace an LXXLL-containing coactivator peptide from GW4064-bound FXR (Urizar *et al.*, 2002; Wu *et al.*, 2002). Furthermore, in coactivator recruitment assays shown in Chapter 5, section 5.4.2, guggulsterone was unable to activate FXR. As such, α -guggulsterone was used in HepG2 cells to determine whether it could diminish Jed compound-induced changes in FXR-target gene expression. If agonist induced changes in FXR-mediated gene expression persists in the

presence of guggulsterone, the Jed compounds may be acting via non-FXR specific mechanisms or pathways.

Results in Figure 6. 17, show some intriguing, and varied responses for the effect of guggulsterone in Jed441-induced gene expression. Expression of FXR in HepG2 did not significantly change with treatment of guggulsterone at 10 or 100 μ M, or Jed441, on their own (Figure 6. 17A). Conversely, Jed441 and guggulsterone appear to have a cumulative effect in the downregulation of FXR, and when HepG2 cells were treated with Jed441 and 10 μ M guggulsterone, mRNA levels were reduced to 0.59- fold. Likewise, incubation of cells with Jed441 and 100 μ M guggulsterone decreased FXR expression to 0.62- fold, however, unlike with the 10 μ M concentration, this was not statistically significant.

SHP, on the other hand, was halved in cells treated with only 10 μ M guggulsterone (Figure 6. 17B). Similarly, when treated with 100 μ M guggulsterone alone, SHP levels were also significantly reduced to 0.6-fold. Whilst Jed441 wasn't able to induce any upregulation in SHP, following 6 hours incubation at its EC₅₀ concentration as seen previously (Figure 6. 7B), incubation with both Jed441 and guggulsterone returns SHP more towards basal levels compared to Guggulsterone alone.

When looking at OST α mRNA expression, no significant changes were observed when HepG2 cells were cultured with guggulsterone, at 10 μ M or 100 μ M (Figure 6. 17C). However, as previously, Jed441 was able to induce a highly significant increase in OST α levels at its EC₅₀ concentration. This increase is diminished by culturing HepG2 cells with guggulsterone, and a dose-response inhibition is seen, with 100 μ M guggulsterone reducing OST α mRNA expression more so than 10 μ M.

CYP7A1 expression levels display no statistically significant changes upon culturing HepG2 cells with guggulsterone or Jed441 alone (Figure 6. 17D). Surprisingly, as with FXR expression, treatment of HepG2 cells with both Jed441 and 10 μ M guggulsterone appears to have an added effect, whereby levels are significantly lower compared to baseline mRNA expression. Again, treatment with Jed441 and 100 μ M guggulsterone downregulates CYP7A1 by half, but results are not statistically significant.

Guggulsterone's effect on Jed561-induced changes in gene expression, were akin to those seen for Jed441. No significant changes were seen in FXR expression when HepG2 was cultured with Jed561 or guggulsterone, concurrently or individually (Figure 6. 18A). Similarly to Jed441, cells

treated with Jed561 with 10 μ M guggulsterone had the lowest levels of FXR at 0.4- fold the basal levels, but this downregulation in expression was not statistically significant.

On the other hand, SHP, again showed no significant increases with Jed561 treatment of HepG2 cells for 6 hours (Figure 6. 18B). SHP levels were further decreased by incubation with Jed561 together with 10 μ M guggulsterone, and with 100 μ M, levels were significantly lower than baseline and cells treated with Jed561 on its own.

As with Jed441, Jed561-induced increases in OST α expression were inhibited by co-culturing cells with guggulsterone at 10 and 100 μ M (Figure 6. 18C). Although expression levels did not fully return to baseline levels, a dose-response inhibition was observed.

There were no significant changes to CYP7A1 expression levels following treatment with Jed561 or guggulsterone alone or together (Figure 6. 18D).

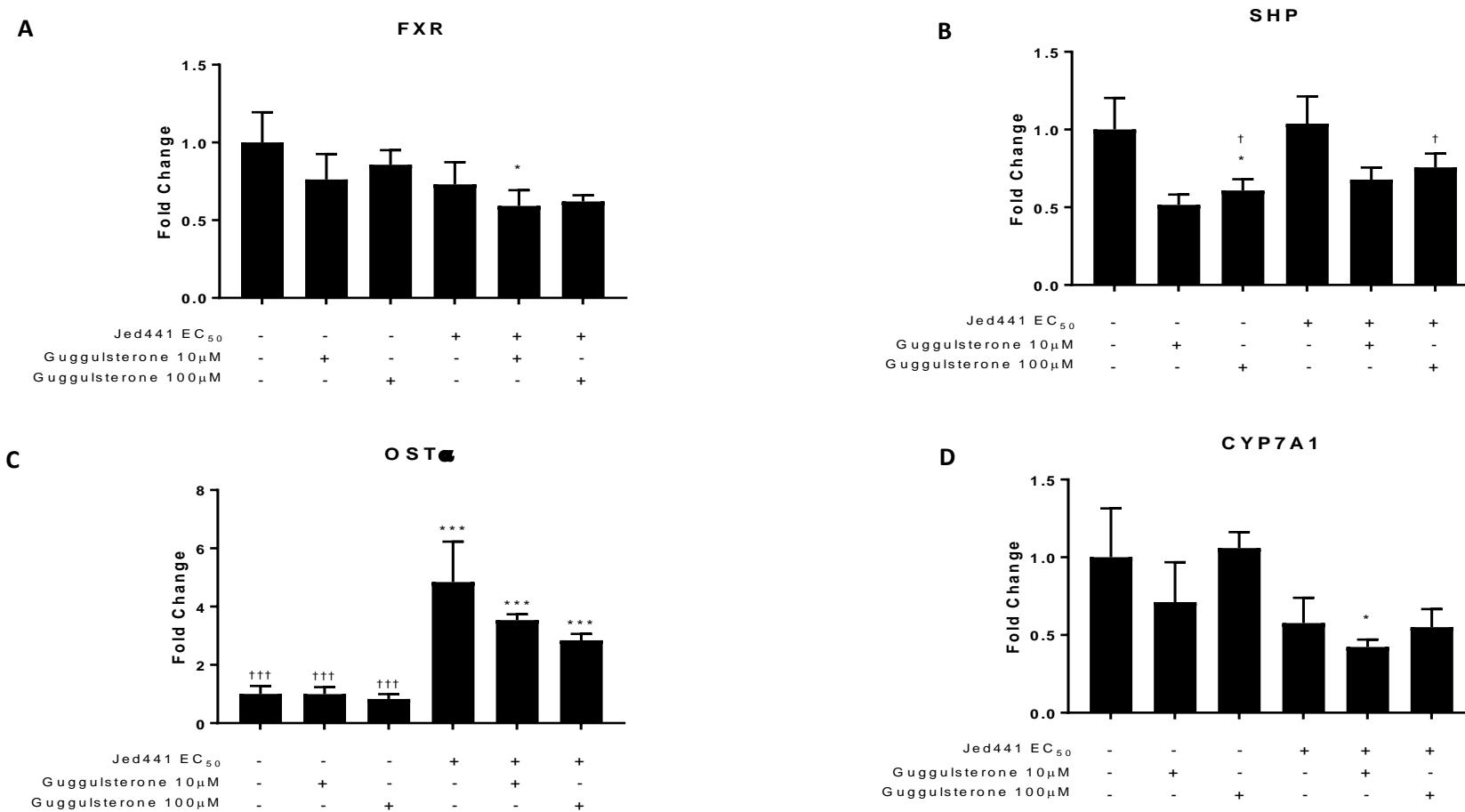


Figure 6. 17 In vitro effect of z-Guggulsterone and Jed441 on FXR target genes. HepG2 cells were treated for 6 hours with either DMSO (control), Jed441 at its EC₅₀ concentration, 10 μM or 100 μM guggulsterone alone, or 10 μM or 100 μM guggulsterone with Jed441 (EC₅₀). Relative mRNA expression levels were assayed by quantitative real-time PCR (qPCR). Fold changes are shown relative to (DMSO-treated) control. Error bars represent ±SEM, n=3, statistical significance analysed by one way ANOVA with Tukey's post hoc test, * p<0.05, ***p<0.001 compared to control, † p<0.05, ††† p<0.001 compared to Jed441 EC₅₀ treatment.

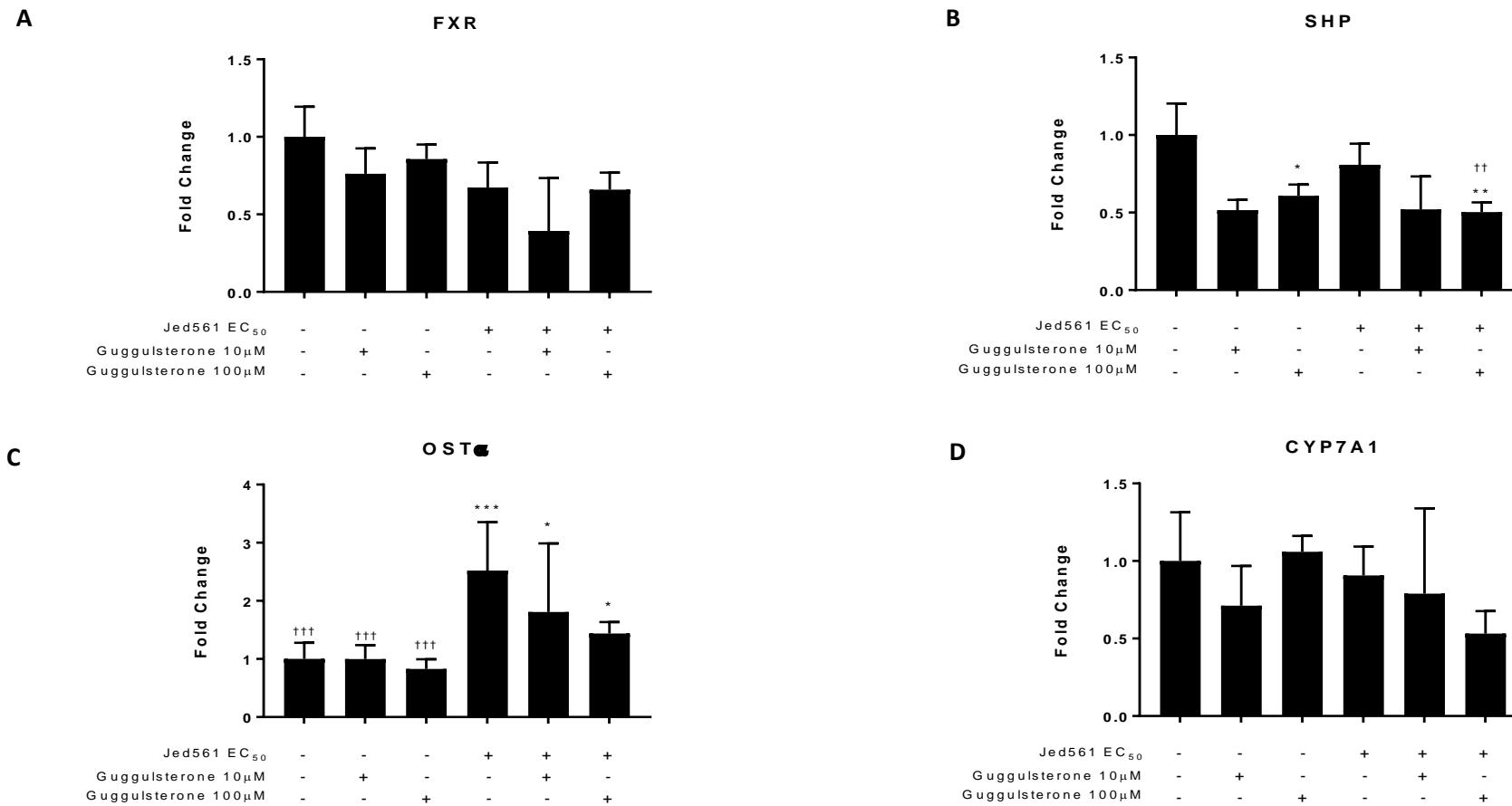


Figure 6. 18 *In vitro effect of z-Guggulsterone and Jed561 on FXR target genes.* HepG2 cells were treated for 6 hours with either DMSO (control), Jed561 at its EC₅₀ concentration, 10 μM or 100 μM guggulsterone alone, or 10 μM or 100 μM guggulsterone with Jed561 (EC₅₀). Relative mRNA expression levels were assayed by quantitative real-time PCR (qPCR). Fold changes are shown relative to (DMSO-treated) control. Error bars represent ±SEM, n=3, statistical significance analysed by one way ANOVA with Tukey's post hoc test. * p<0.05, ***p<0.001 compared to control, † p<0.05, ††† p<0.001 compared to Jed561 EC₅₀ treatment.

6.4.1.7 Effect of Fluorination of Steroid Backbone on FXR-mediated Activity

Lead compound Jed561 contains a fluorine group at the 4β position on the steroid nucleus (Chapter 2, Table 2.9). Compound Jed692, an analogue of Jed561, is identical but does not contain the fluorine group at this position. Whilst coactivator recruitment assays suggest that the EC₅₀ values for both compounds are relatively similar, with Jed561 having a slightly higher affinity for the FXR LBD, both compounds were tested *in vitro* and compared, in order to determine whether the addition of this fluorine is advantageous in mediating the endogenous biological effects of FXR. In order to be completely comparable, both compounds were tested at the EC₅₀ concentration of Jed561, 92 nM.

Results in Huh7 cells showed very similar responses in FXR target gene expression for both compounds, and whilst Jed692 appeared to induce slightly bigger changes in the up or downregulation of genes, there were no statistically significant differences between the responses of Jed692 cells compared to cells treated with Jed561 (Figure 6. 19). Following 6 hours incubation, Huh7 cells displayed significant upregulation in SHP, OST α and FGF19 under treatment with both Jed561 and Jed692 (Figure 6. 19A). FXR, APOC2, SREBF1 and TGF β 1 expression levels remain unchanged following incubation for 6 hours with either compound.

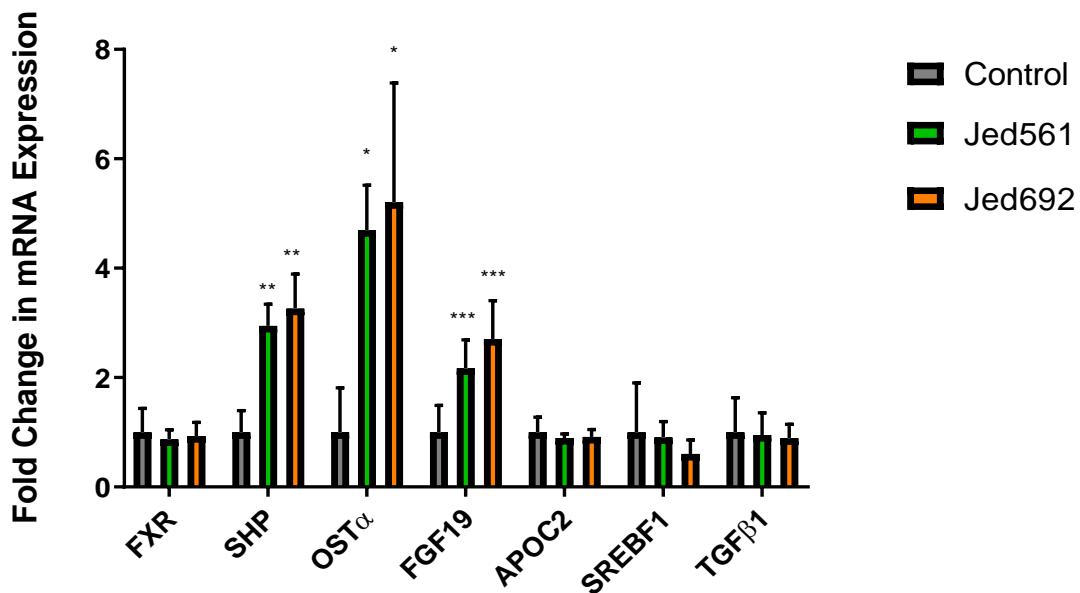
Likewise, following 24 hours treatment with either Jed561 or Jed692, there is an upregulation in FGF19, SHP and OST α , with the latter two displaying statistically significant increases (Figure 6. 19B). Again, Jed692 appears to induce slightly larger increases in upregulation of SHP and FGF19, however there is no significant difference between gene expression in the two treatment groups. After 24 hours, there are no significant changes from basal levels in FXR, APOC2, SREBF1 or TGF β 1 in cells treated with either Jed561 or Jed692.

Meanwhile, in HepG2 cells treated with either Jed561 or Jed692 for 6 hours, SHP mRNA levels remain relatively equal to that of vehicle treated cells (Figure 6. 20A). Regardless of no observed increase in SHP expression, both Jed561 and Jed692 were able to downregulate CYP7A1, but only cells treated with Jed692 display significantly lower levels compared to the control. Expression of OST α and TGF β 1 display no statistically significant changes in cells treated with either compound.

Following 24 hours incubation with these compounds, HepG2 cells yet again show unchanged levels of SHP (Figure 6. 20B). Despite this, both Jed561 and Jed692 significantly downregulate CYP7A1 levels to 0.7-fold and 0.5-fold basal levels, respectively. Furthermore, not only is CYP7A1 expression in Jed692-treated cells highly significantly decreased, compared to control levels, it is also statistically significantly lower than levels found in cells treated with Jed561; the

first indication of any significant difference between the actions of the two analogous compounds. OST α mRNA expression is significantly upregulated by Jed561, and even more so by Jed692, achieving 3.5- and 4.5- fold increases in expression, respectively. TGF β 1 expression is reduced following 24 hours treatment with both Jed561 and Jed692, although this reduction is only statistically significant in Jed561 treatment.

A



B

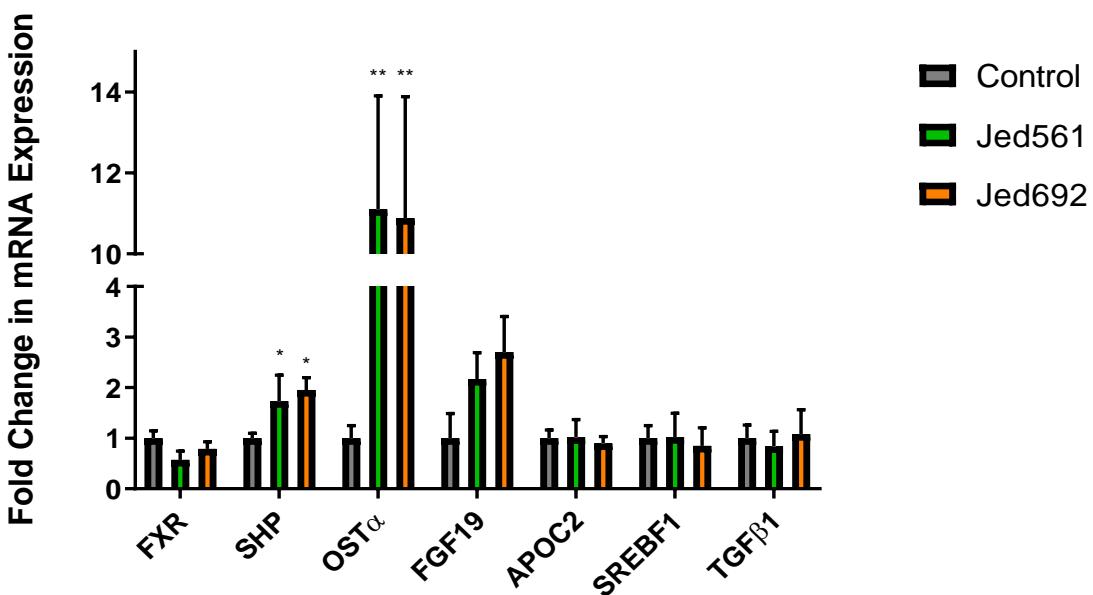


Figure 6. 19 Comparison of *In vitro* effect of Jed561 and Jed692 on FXR target genes. Huh7 cells were treated for (A) 6 or (B) 24 hours with DMSO (control), or 92 nM Jed561 or 92 nM Jed692. Relative mRNA expression levels were assayed by quantitative real-time PCR (qPCR). Fold changes are shown relative to (DMSO-treated) control. Error bars represent \pm SEM, n=3, statistical significance analysed by one way ANOVA with Tukey's post hoc test. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$ compared to control.

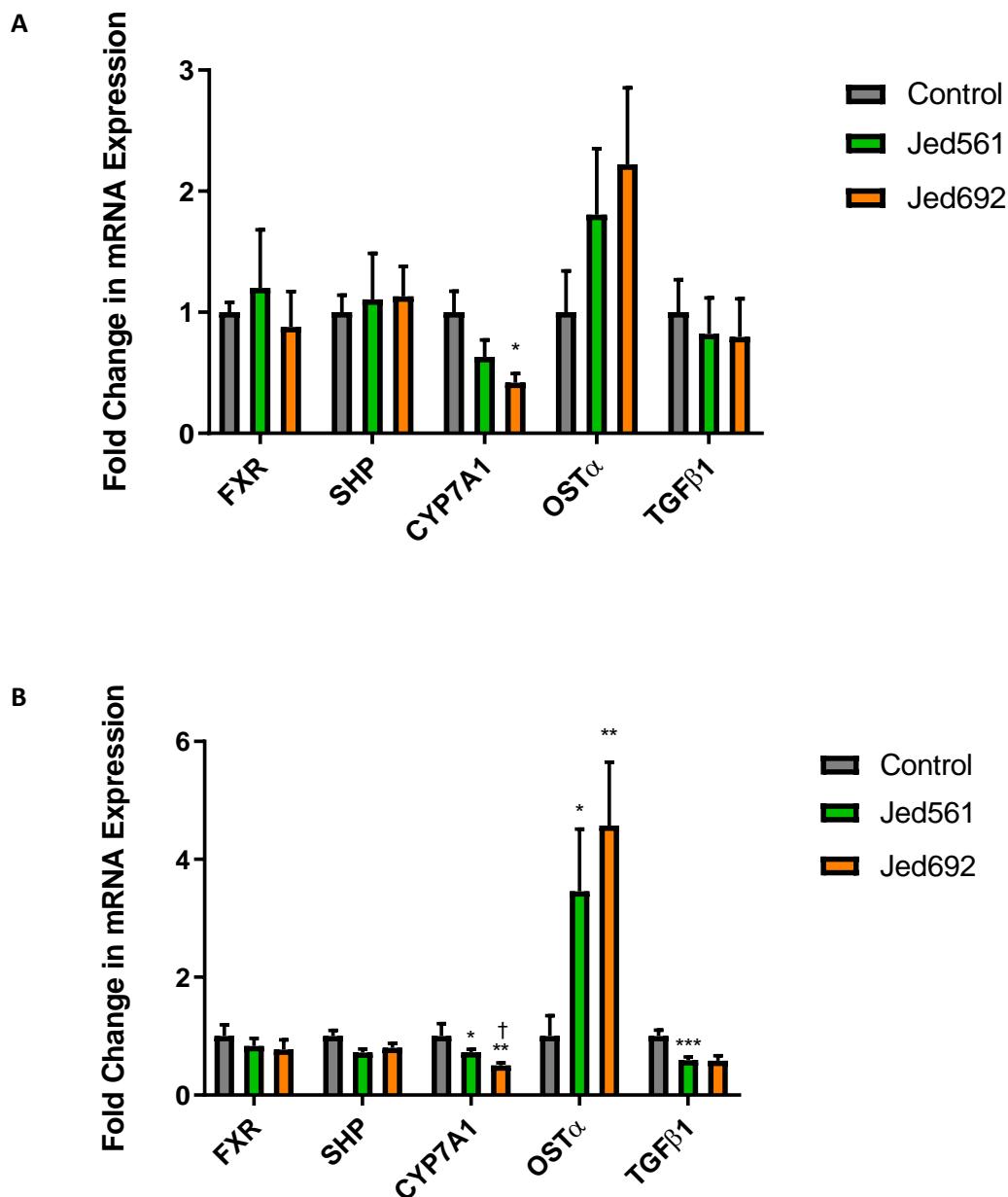


Figure 6. 20 Comparison of *In vitro* effect of Jed561 and Jed692 on FXR target genes. HepG2 cells were treated for 6 (A) or 24 (B) hours with DMSO (control), or 92 nM Jed561 or 92 nM Jed692. Relative mRNA expression levels were assayed by quantitative real-time PCR (qPCR). Fold changes are shown relative to (DMSO-treated) control. Error bars represent \pm SEM, $n=3$, statistical significance analysed by one way ANOVA with Tukey's post hoc test. * $p<0.05$, ** $p<0.01$, *** $p<0.001$ compared to control, † $p<0.05$ compared to Jed561.

6.4.2 Effect of Lead Compounds *in vivo*

Having demonstrated that Jed441 and Jed561 are indeed active and can produce the desired effect of FXR target gene regulation *in vitro*, the two lead compounds were tested *in vivo* to ultimately ensure that they have suitable pharmacological properties; and that they are also able to affect change in genes involved in the metabolic pathways being targeted. Compounds were tested in C57BL/6J mice under normal physiological conditions and under conditions designed to mimic the NASH disease, denoted henceforth as ‘wild type’ (WT) mice or ‘NASH’ mice, respectively.

6.4.2.1 Selection of Appropriate Reference and Target Genes for Mouse Liver RNA

As discussed previously, for qPCR to fulfil its maximum potential as an analytically sensitive and robust method, it is imperative that analysis includes appropriate normalization and validation. In order to do this, an appropriate reference gene, with which to normalize against variances arising from different mRNA extraction methods and reverse transcription efficacies, is needed. Furthermore, these reference genes, usually housekeeping genes, are required to be stably expressed under varying circumstances and treatment conditions, as not to introduce artificial changes or obscure any real differences in expression levels. Unfortunately, no single housekeeping gene manifests stable expression in all tissues or cells and expression levels may vary under different circumstances, emphasizing the need to characterize the expression levels of different transcripts in multiple species or tissues (Kouadjo *et al.*, 2007). To date, several studies have been conducted to extensively characterize transcript levels of numerous genes across different tissues and organs of C57BL/6J mice, to identify ubiquitously expressed genes (Kouadjo *et al.*, 2007; Bruce *et al.*, 2012; B. Li *et al.*, 2017). Results found that traditionally used reference genes, GAPDH, β-actin and β-2 microglobulin, were not suitable for reference purposes in this standard mouse strain (B. Li *et al.*, 2017). Whilst several thousand genes were recognised as being ubiquitously expressed across different sexes, organs and developmental stages (Kouadjo *et al.*, 2007; B. Li *et al.*, 2017), and several genes were identified as being stable across different severities of NASH (Bruce *et al.*, 2012), only 3 were compared for suitability in this study.

ARF1, encoding ADP-ribosylation factor 1; COX7A2L, encoding Cytochrome C Oxidase Subunit 7A2 Like; and YWHAZ, Tyrosine 3-Monooxygenase/Tryptophan 5-Monooxygenase Activation Protein Zeta, were tested for their suitability as an endogenous reference for qPCR experiments. Predesigned, mouse-specific primers for these genes were purchased from Sigma Aldrich. Complementary DNA, derived from the livers of untreated mice, was serially diluted and used to make cDNA template standards. As previously, standards were run under automated qPCR

conditions using the primers for each candidate reference gene. The C_t values generated were used to derive a standard curve and were used for the analysis of the primer's sensitivity and efficiency. Results shown in Figure 6. 21, show that the ARF1 assay had the least sensitivity, with the highest starting C_t values, and concentrations of cDNA lower than 10^{-2} were beyond the limit of detection with these primers (Figure 6. 21A). On the other hand, YWHAZ displayed detectable amplification, and therefore C_t values for the highest 4 concentrations (Figure 6. 21C), and COX7A2L displayed evenly spaced amplification across the entire dilution series tested (Figure 6. 21B).

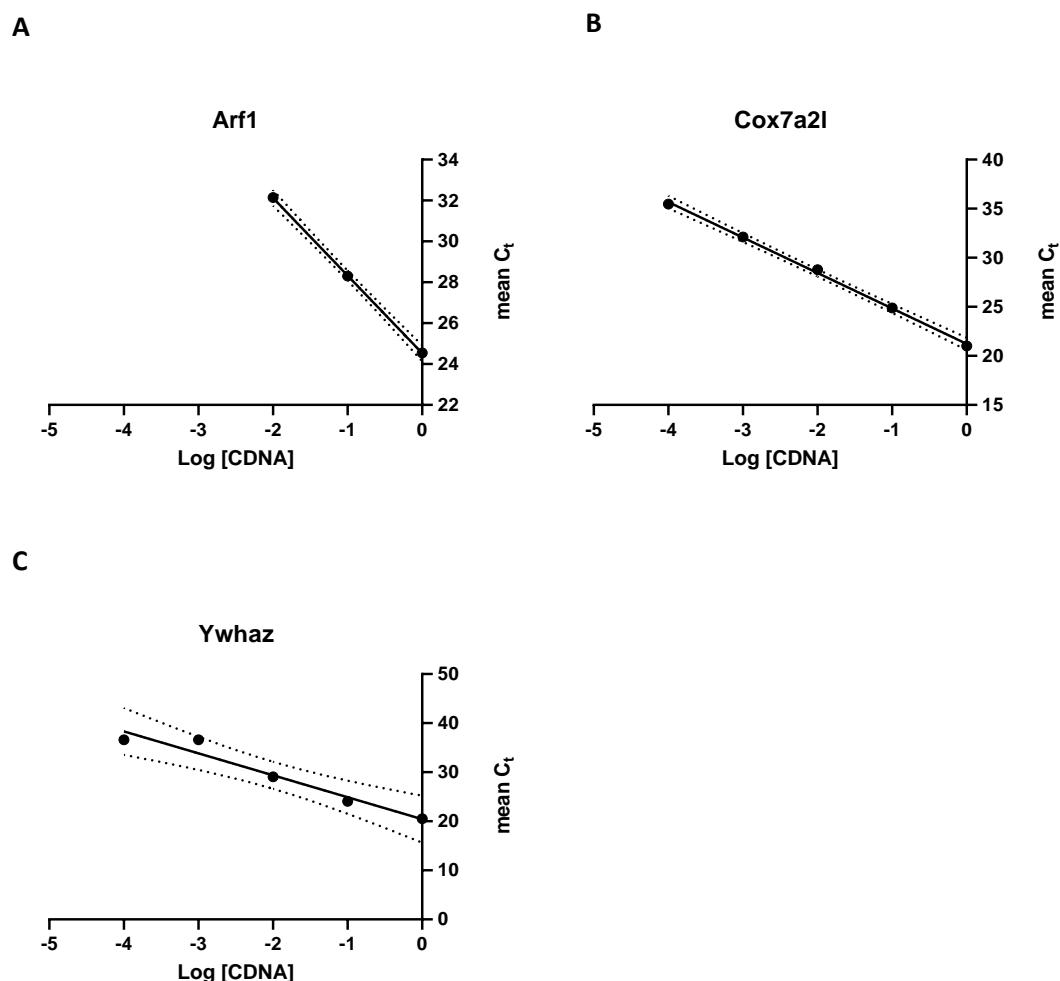


Figure 6. 21 Comparisons of primer sensitivity and efficiencies. 5-fold serial dilutions of target DNA were subject to amplification by (A) ARF1 primers, (B) COX7A2L primers, and (C) YWHAZ primers.

When calculating primer efficiency, the results displayed in Table 6. 5 show that, although not 100%, COX7A2L had the highest primer efficiency for these reagent concentrations and reaction conditions. COX7A2L also produced an R^2 value of 1, meaning that the standard curve is very accurate, and there is high confidence in the correlation between Ct value and target copy number (Bustin and Huggett, 2017). Furthermore, when comparing cDNA from the livers of mice treated with either OCA or Jed compounds, results showed that COX7A2L had the lowest standard error of mean between the three reference gene primers tested, suggesting that it is robust and ubiquitously expressed even under different treatment conditions in C57BL/6J mice. For these reasons COX7A2L was chosen as the endogenous reference gene for relative quantification in C57BL/6J WT mice. In addition, COX7A2L produced only one product as observed by the melt curve plot (not shown) and it was one of the genes described to be stably expressed in different organs (B. Li *et al.*, 2017), meaning that it can potentially be used in future studies looking at different tissue types.

Table 6. 5 Comparison of efficiency, sensitivity and reproducibility of primers for reference genes in C57BL/6J mice samples

Reference Gene	Amplification Efficiency %	Accuracy of standard curve (coefficient R^2)	Standard deviation between Ct values of mice treated with different compounds
ARF1	83	1	0.28
COX7A2L	89	1	0.14
YWHAZ	79	0.995	0.35

Whilst COX7A2L appeared to be the most appropriate reference gene in mice under normal, physiological conditions, it is known that different dietary interventions and treatment can alter expression levels of several housekeeping genes, particularly in models of NAFLD (Bruce *et al.*, 2012). In order to test for stable expression under different conditions, test assays were run using 5 mice from each treatment group of the NASH mice study. COX7A2L and YWHAZ, as the two primers with the superior sensitivity were tested. Messenger RNA levels, characterised by Ct values, were compared across mice on a low fat diet treated with a sham compound (SHAM), NASH model mice treated with vehicle (NASH vehicle), and NASH mice treated with either OCA or Jed561 (NASH OCA 30, NASH Jed561 2, etc.; where the number denotes the concentration in mg/kg of compound used). The results in Figure 6. 22, show that COX7A2L expression was fairly stable and had very little variation within replicates of the same treatment group. A one-way ANOVA was used to test for any differences between the groups, and it was observed that NASH mice treated with OCA (30mg/kg) had significantly higher levels of COX7A2L mRNA compared

to the other groups. This may have arisen due to the larger variation in C_t values in this group compared to the others, and possibly is a consequence of variation in RNA yield from different extractions. C_t values derived from YWHAZ primers, on the other hand, displayed slightly more variation between samples of the same condition; however, when compared statistically across the different groups, showed no significant difference in transcript levels between mice that had undergone different treatments. For this reason, YWHAZ was used as a reference gene in the NASH mice study.

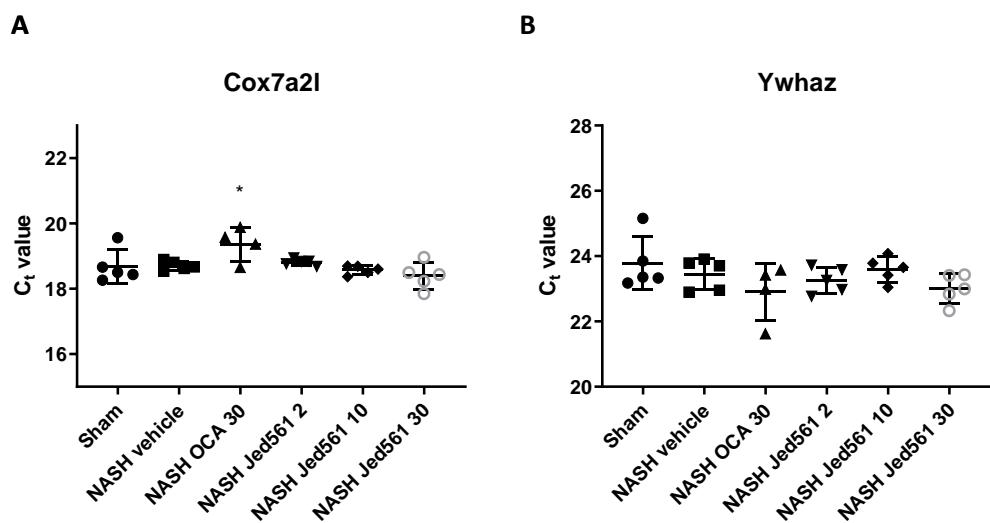


Figure 6. 22 Stability of candidate reference genes across differentially treated mice. mRNA from the left lobe of the liver of C57BL/6J mice fed a low fat diet (sham), fed a high fat diet, induced with CCl_4 and treated with vehicle (NASH Vehicle) or fed a high fat diet, induced with CCl_4 and treated with OCA or Jed561 (NASH OCA 30, NASH Jed561 2, NASH Jed561 10, NASH Jed561 30). (A) COX7A2L and (B) YWHAZ were tested for stable expression across all groups. ($n=5$), statistical significance analysed by one way ANOVA with Tukey's post hoc test. * $p<0.05$ compared to 'sham' control.

Target genes were selected based on information from current literature and, as previously (Chapter 6, section 6.3.1.1), were checked for detectable transcript levels in the Expression Atlas (EMBL-EBI), and based on the availability and suitability of predesigned KICqStart® SYBR® green primers (Sigma Aldrich). The sensitivity, efficiency and specificity of primers were tested as described previously, and data summarized in Appendix 8.

6.4.2.2 Short Term Effect of Jed Compounds on FXR Target Genes in Wild Type Mice

In order to ascertain whether lead compounds would be absorbed in the intestine, and would be transported to the liver (their anticipated site of action), pharmacokinetic analyses were performed on wild type mice by Saretius Ltd., but described here briefly for context. To determine blood/plasma ratios and bioavailability mice were treated by oral gavage with Jed441 or Jed561 at 2 mg/kg body weight and blood samples taken from the tail vein at various time points post administration (0.5, 2, and 8 hours). After euthanasia, terminal blood plasma and organ samples were immediately harvested and prepared for analysis. Compound concentrations were measured by liquid chromatography-mass spectrometry. Results from these analyses (third party data not shown) were taken as indicators for predicting of whole body pharmacokinetics. Liver sections from the mice used by Saretius, were also kindly provided for use in this work presented here. RNA extracted from these livers were used in qPCR assays to examine short term, transient changes arising from treatment with the compound.

Results in Figure 6. 23A, showing changes in expression following administration of Jed441, indicate that after treatment, and presumably activation of FXR, there is a downregulation in its expression. This reduction continues even more dramatically 2 hours after treatment, and by 8 hours, it starts to increase, although only reaching 0.4-fold levels by this time point. SHP, on the other hand, displays an immediate increase in mRNA expression levels, up to 2.5 times more than basal levels (Figure 6. 23B). This upregulation, however, is short lived and by 2 hours after treatment, levels have returned to normal. Moreover, 8 hours post treatment, SHP levels have drastically reduced by more than 10 times the amount seen without treatment. Remarkably, despite being positively regulated by FXR, BSEP expression is downregulated following Jed441 administration. Levels appear to decrease as low as 0.2-fold by 2 hours post treatment (Figure 6. 23C). Notably, the results from these experiments display considerable variation, as seen by the relatively large error bars. In some cases, such as BSEP, the lower limit of expression in control mice overlaps the upper limit seen in treated mice, indicating that the mean expression levels displayed here, may artificially present the data as being downregulated, whereas in reality, there are no real changes in the levels. Accordingly, the results in Figure 6. 23A, B and C, do not show any statistically significant differences before and after treatment with Jed441. CYP7A1 is the only gene to display any significant changes in mRNA levels, whereby following treatment, levels are reduced significantly to 0.03-fold, 2 hours after treatment (Figure 6. 23D). By 8 hours post treatment, CYP7A1 levels start to return to baseline, possibly somewhat due to the parallel reduction in SHP levels.

Jed561 administration induces similar changes in FXR target gene expression as Jed441. Once again, FXR expression shows a decline immediately after administration, however, by 8 hours levels have returned to that of control mice (Figure 6. 24A). Unlike Jed441, Jed561 induces an increase in SHP levels of 2-fold, by 0.5 hours post treatment; and these changes are sustained, with levels continuing to rise, achieving 8-fold increases in expression by 8 hours post treatment (Figure 6. 24B). However, again due to the large error bars and variation between individual mice, the upregulation seen is not statistically significant. Again there are no significant increases in BSEP expression following Jed561 treatment, as would be expected for FXR activation (Figure 6. 24C). Although the mRNA levels would appear to decrease slightly, due to the large error bars between groups, arguably, there are no real changes in BSEP expression. There is, however, significant downregulation in CYP7A1 mRNA expression levels after Jed561 treatment, again with the most significant decreases occurring 2 hours post administration (Figure 6. 24D). Unlike treatment with Jed441, these decreases in CYP7A1 are sustained even at 8 hours after treatment, with expression remaining significantly low, at 0.1-fold lower than basal levels.

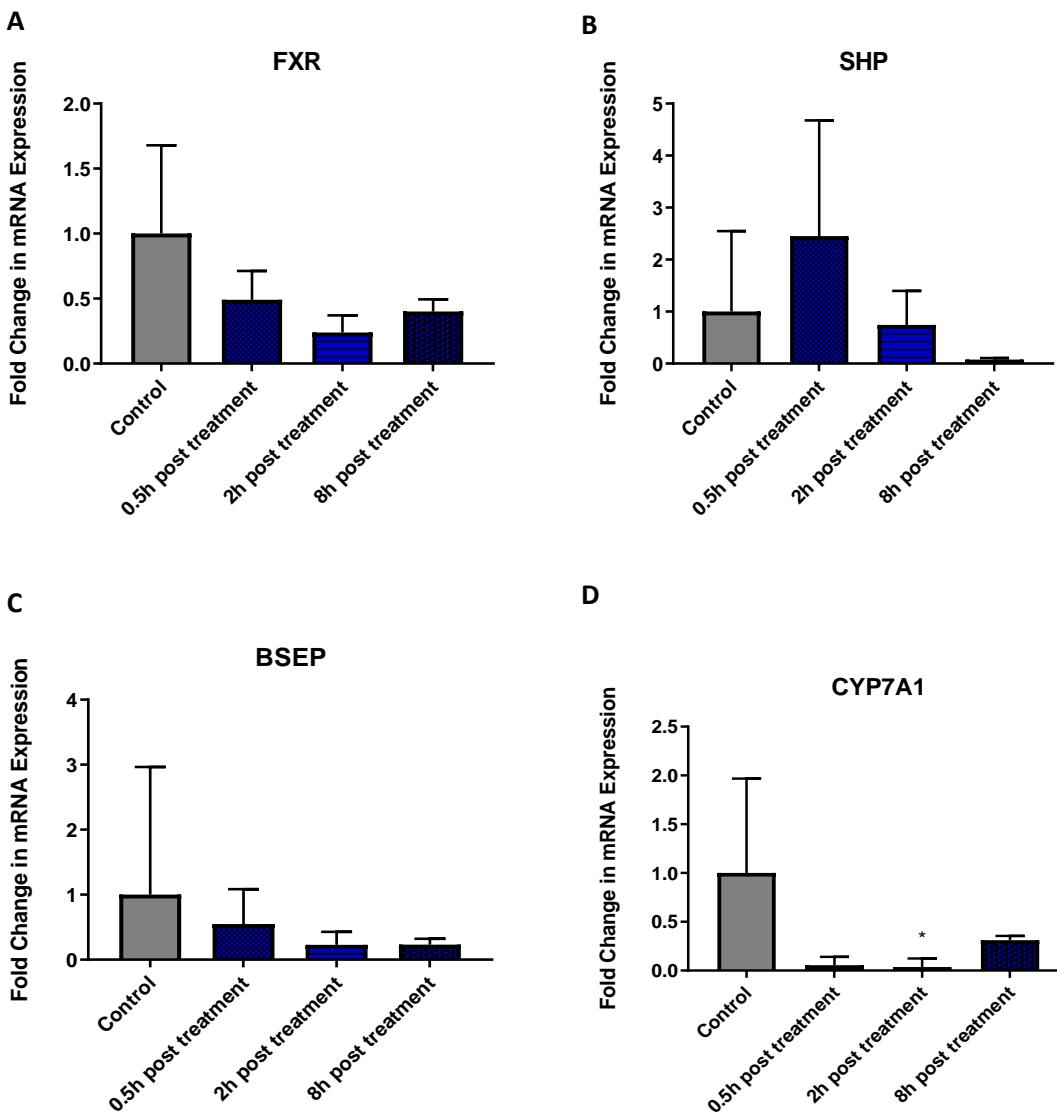


Figure 6. 23 Expression of FXR target genes following single oral dose of Jed441 in C57BL/6J mice. Mice were treated with Jed441 at 2 mg/kg body weight by oral gavage and mice sacrificed 0.5-, 2- and 8- hours post administration. (A) FXR, (B) SHP, (C) BSEP and (D) CYP7A1 mRNA levels assayed by quantitative real-time PCR (qPCR). Fold changes are shown relative to (un-treated mice) control. Error bars represent \pm SEM, n=3, statistical significance analysed by one way ANOVA with Bonferroni's post hoc test. * p<0.05, ** p<0.01, *** p<0.001 compared to control.

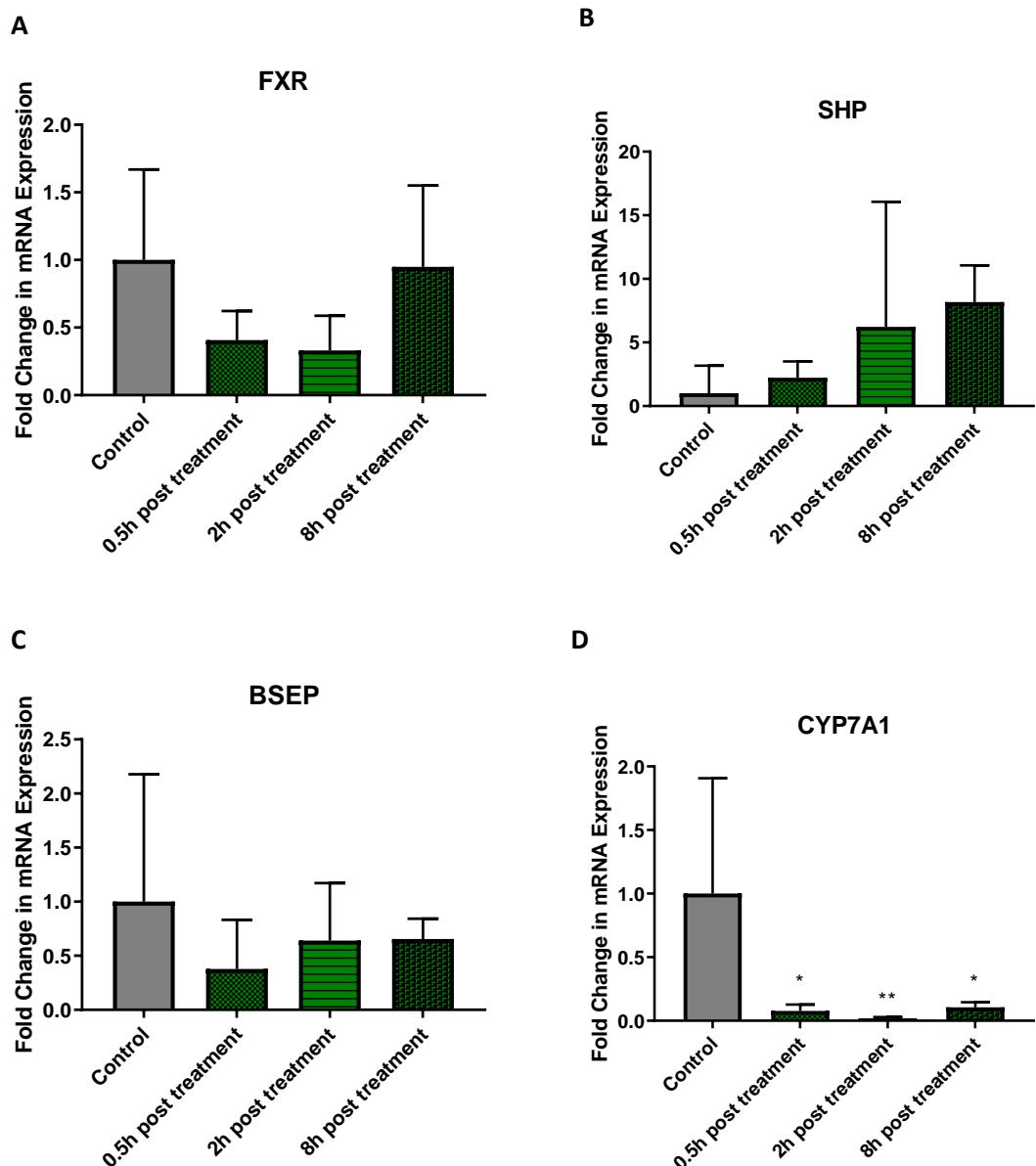


Figure 6. 24 Expression of FXR target genes following single oral dose of Jed561 in C57BL/6J mice. Mice were treated with Jed561 at 2 mg/kg body weight by oral gavage and mice sacrificed 0.5-, 2- and 8- hours post administration. (A) FXR, (B) SHP, (C) BSEP and (D) CYP7A1 mRNA levels assayed by quantitative real-time PCR (qPCR). Fold changes are shown relative to (un-treated mice) control. Error bars represent \pm SEM, n=3, statistical significance analysed by one way ANOVA with Bonferroni's post hoc test. * p<0.05, ** p<0.01, *** p<0.001 compared to control.

In a separate study, mice were treated with lead compound Jed 561, along with the non-fluorinated equivalent, and competitor compounds; steroid FXR agonist, OCA, and non-steroidal agonist, Tropifexor. Again, each compound was tested in triplicate, with a single oral dose of compound, all at 2 mg/kg. Mice were sacrificed 6 hours post administration and livers immediately harvested.

Despite the previous study, after a single treatment with these compounds, results showed that only Tropifexor was able to induce significant changes in FXR target gene expression (Figure 6. 25). Jed561 was unable to sustain increases in SHP 6h after treatment in mice (Figure 6. 25A). Meanwhile Jed692, and similarly, OCA, displayed double the levels of SHP mRNA compared to un-treated mice. However, OCA treated mice in particular, showed a lot of variation between individual replicates, and increases were not statistically significant. Tropifexor at 2mg/kg, however, was able to quadruple SHP mRNA levels. In contrast to the previous study, Jed561 was able to upregulate BSEP 3-fold, 6 hours post administration (Figure 6. 25B). As before, Jed692 appeared to perform better than Jed561, upregulating BSEP expression 4-fold. OCA on the other hand, performed worse than Jed561. Again, the variation between samples negated any statistical significance in these increases, and the only treatment that displayed significant upregulation in BSEP levels, was Tropifexor, which induced 9-fold increases in its mRNA. Remarkably, contrary to expected results, the steroid compounds all appeared to upregulate CYP7A1 levels, whereas only Tropifexor significantly decreased levels. The increases in CYP7A1 seen with Jed561, Jed692 and OCA treatment could be indicative of the activation of off-target, counteracting mechanisms, however, results were not statistically significant, and could even arise due to differences in individual mice, discussed further in Chapter 6, section 6.5.3.

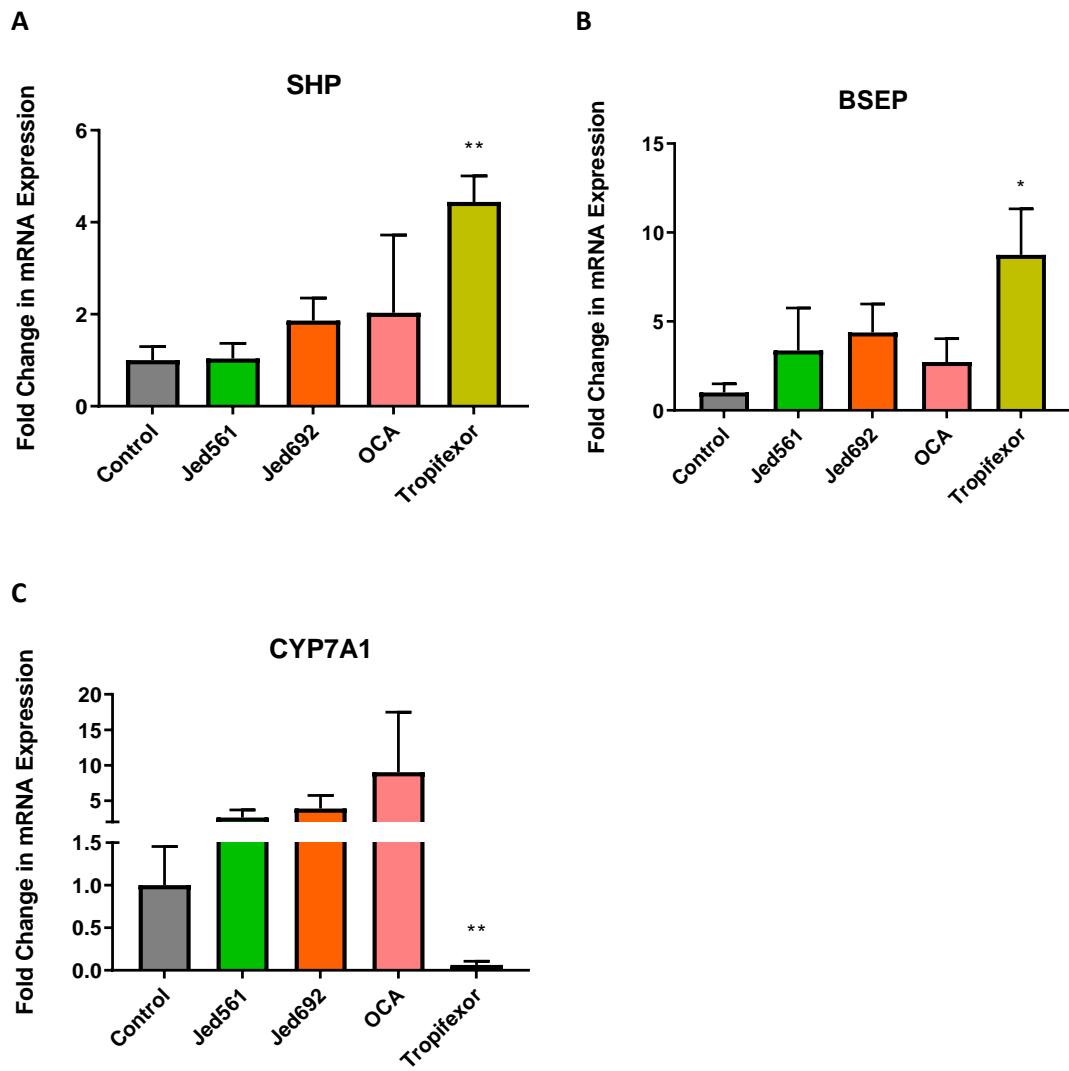


Figure 6. 25 Expression of FXR target genes following single oral dose of Jed and competitor compounds in C57BL/6J mice. Mice were treated with each compound at 2 mg/kg body weight by oral gavage and mice sacrificed 6 hours post administration. (A)SHP, (B) BSEP and (C) CYP7A1 mRNA levels assayed by quantitative real-time PCR (qPCR). Fold changes are shown relative to (un-treated mice) control. Error bars represent \pm SEM, n=3, statistical significance analysed by one way ANOVA with Tukey's post hoc test. * p<0.05, ** p<0.01, *** p<0.001 compared to control.

6.4.2.3 Effect of Continual Treatment with Jed561 on FXR Target Genes in Wild Type Mice

Whilst, lead compounds displayed variable results in mice following a single oral dose, the subsequent study aimed to determine any prolonged changes in genes associated with FXR, following recurrent treatment with steroid agonists. As the appropriate treatment dose of Jed561 was still unresolved, Jed561 was administered at 3 different concentrations, 2, 10 and 30 mg/kg bodyweight of the mice. OCA was again used for comparative purposes, and used at its purported effective dose (Tølbøl *et al.*, 2018). Results from initial pharmacokinetic studies suggested that Jed441 was not properly absorbed in the intestine and displayed relatively short half-life (third party data not shown). Respective of this, Jed441 was only tested at 10mg/kg in order to observe if intestinal FXR activation, or limited hepatic activation was still sufficient to affect change in hepatic target genes. Ten wild type C57BL/6J mice per group, were treated by oral gavage for 5 continuous days with a single daily dose of either vehicle, Jed561, Jed441 or OCA. Mice were fasted for 30 minutes prior to treatment on the 5th day, and then sacrificed 6 hours later. Again, livers were immediately harvested, and specifically the left lobe used for RNA extraction.

Results showed that although both OCA and the Jed compounds are derived from the BA backbone, they displayed differential effects on FXR expression. Mice treated with OCA did not show any differences in FXR expression, whereas Jed compound treated mice had slightly downregulated levels (Figure 6. 26A). Only mice treated with Jed561 at 10mg/kg displayed significantly lower levels of FXR, compared to the untreated mice. However, expression levels of FXR in Jed treated mice were highly significantly decreased compared to levels in OCA treated mice. Interestingly, although Jed561 treated mice appear to have slight reductions in FXR expression levels, direct FXR target genes, SHP and BSEP, are upregulated, as expected from FXR activation.

Both OCA and Jed561 were able to upregulate SHP expression. OCA at 30 mg/kg was able to induce a 7-fold increase in SHP levels, an effect that was highly statistically significant compared to the control mice; whereas doses of Jed561 induced significant increases in SHP mRNA, by as much as 4.5-fold (Figure 6. 26B). Although SHP levels of Jed561 treated mice were significantly lower than those seen in mice treated with OCA, Jed 561 appeared to be effective in upregulating this FXR target gene by concentrations as little as 2 mg/kg. Notably, Jed441 was unable to upregulate SHP expression beyond basal levels.

All compounds were able to upregulate the expression of BSEP, with all treatment groups, except Jed561 (10 mg/kg), showing statistically significant increases compared to vehicle-

treated mice (Figure 6. 26C). This time, Jed561 at 30 mg/kg showed the greatest changes in BSEP expression levels, with a 5.4-fold increase. This was followed closely by OCA treatment, which induced 4.8-fold increases in BSEP. Statistical analyses did not identify significant differences in BSEP expression levels between different concentrations of Jed561 and OCA (30mg/kg), suggesting that administration of just 2 mg/kg may be enough for the upregulation of BSEP to levels comparable to OCA at 30mg/kg. Interestingly, Jed441 produced a highly significant increase in BSEP along with the other compounds.

Results in Figure 6. 26D, showed that 6 hours after administration, mice treated with Jed561 at 30 mg/kg was the only group to display significant decreases in CYP7A1 expression. CYP7A1 levels were decreased by 60%, whereas mice treated with OCA and lower concentrations of Jed561 displayed similar levels of CYP7A1 expression to the control group. Treatment with Jed441 did not significantly change CYP7A1 expression, possibly correlating with the previous results whereby SHP levels were also not upregulated.

When looking at genes involved in the lipid metabolism-regulated mechanisms of FXR, in Figure 6. 27, it was observed that at higher concentrations of Jed561, and with Jed441 at 10 mg/kg, Jed compounds were able to significantly upregulate the direct target gene APOC2, whilst OCA was not. This upregulation was highly significant in comparison to APOC2 levels in control mice and OCA-treated mice, when analysed statistically. Unexpectedly, there was a significant, 7-fold increase in SREBF1 mRNA expression levels following continual treatment with OCA (30 mg/kg). Whilst mice treated with 2 mg/kg doses of Jed561 also showed significant increases in SREBF1 mRNA, Jed561 significantly downregulated this gene and showed no changes at 10 and 30 mg/kg doses, respectively.

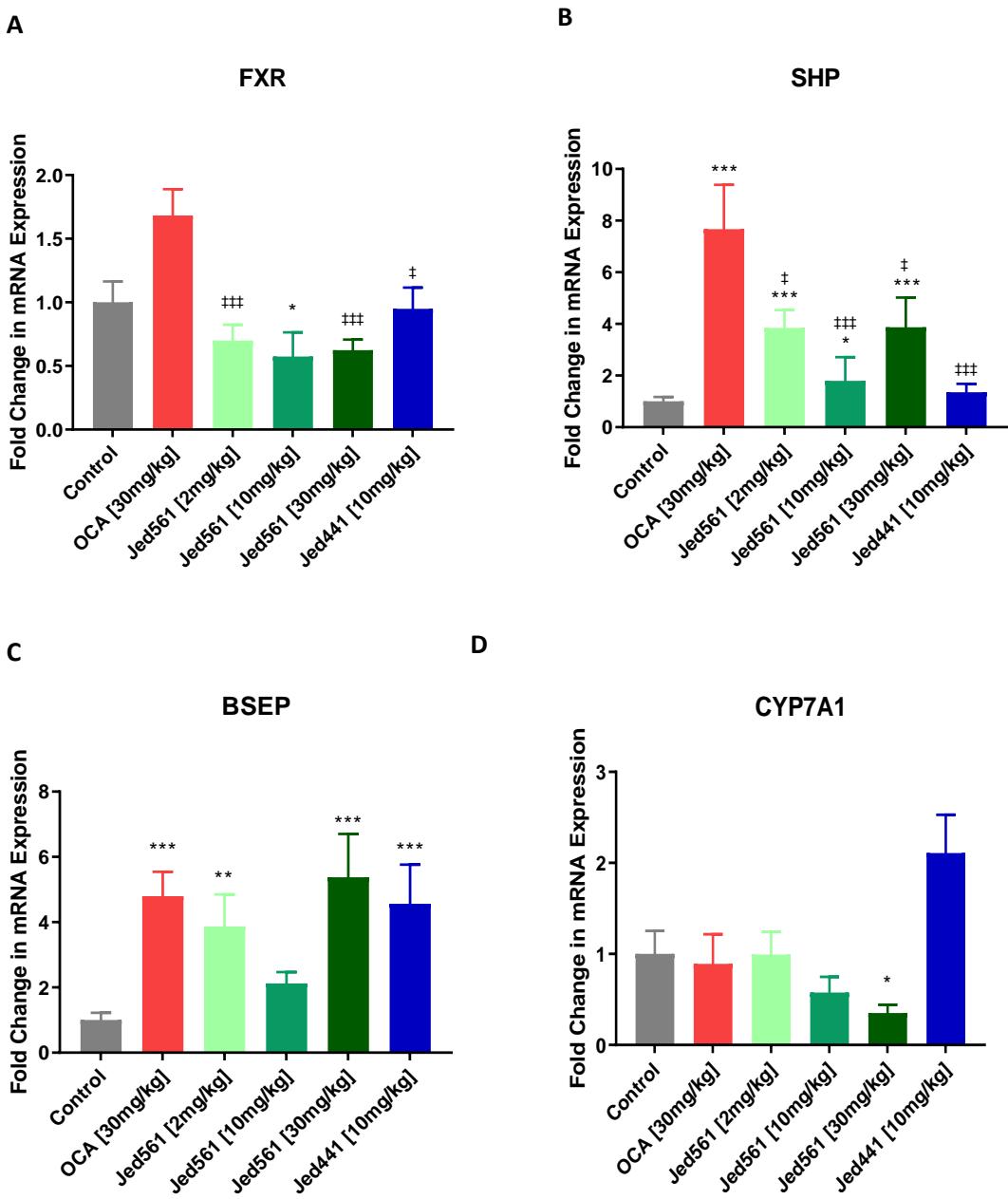


Figure 6. 26 Expression of FXR target genes following continual treatment with Jed compounds in C57BL/6J mice. Mice were treated with OCA at 30 mg/kg, Jed561 at 2, 10 and 30 mg/kg, and Jed441 at 10 mg/kg body weight by oral gavage once daily for 5 continuous days. Mice were sacrificed 6 hours post administration on the final day. (A)FXR, (B), SHP, (C) BSEP and (D) CYP7A1 mRNA levels assayed by quantitative real-time PCR (qPCR). Fold changes are shown relative to (un-treated mice) control. Error bars represent \pm SEM, n=10, statistical significance analysed by one way ANOVA with Tukey's post hoc test. * p<0.05, ** p<0.01, *** p<0.001 compared to control. ‡ p<0.05, ‡‡ p<0.01, ‡‡‡ p<0.001 compared to OCA 30mg/kg treatment.

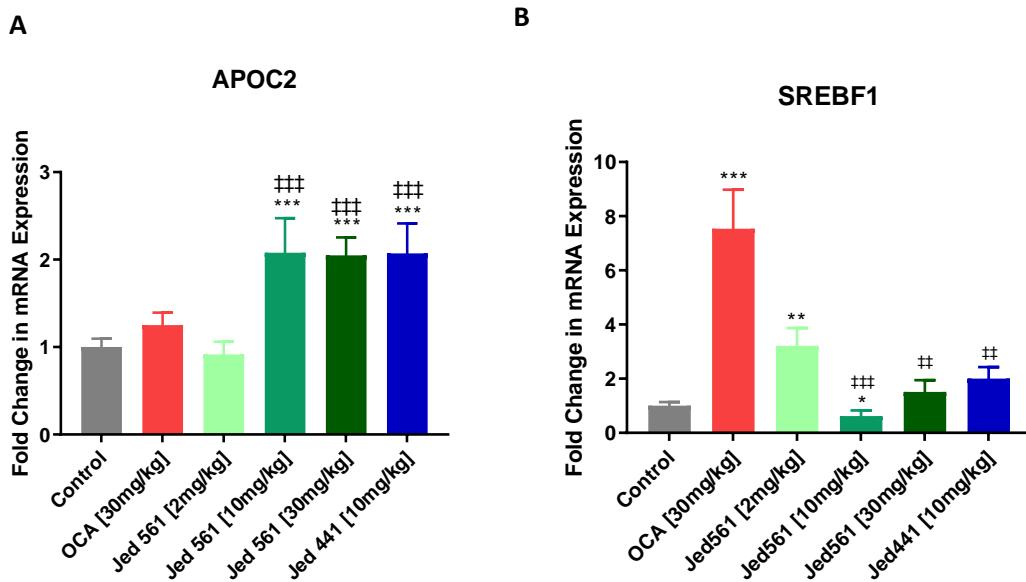


Figure 6. 27 Expression of genes involved in FXR-mediated lipid metabolism following continual treatment with Jed compounds in C57BL/6J mice. Mice were treated with OCA at 30 mg/kg, Jed561 at 2, 10 and 30 mg/kg, and Jed441 at 10 mg/kg body weight by oral gavage once daily for 5 continuous days. Mice were sacrificed 6 hours post administration on the final day. (A) APOC2 and (B) SREBF1 mRNA levels assayed by quantitative real-time PCR (qPCR). Fold changes are shown relative to (un-treated mice) control. Error bars represent ±SEM, n=10, statistical significance analysed by one way ANOVA with Tukey's post hoc test. * p<0.05, ** p<0.01, *** p<0.001 compared to control. # p<0.05, ## p<0.01, ### p<0.001 compared to OCA 30mg/kg treatment.

6.4.2.4 Effect of Jed561 in a mouse model of NASH

The pathogenesis of NASH is very complex and involves several interconnecting mechanisms. Currently, there is no single best NASH murine model which recapitulates and simulates all of the stages contributing to the development of the disease as seen in humans. The model used in this study, was based on a recently described model of NASH and HCC, (Hoshida *et al.*, 2018), and aimed to imitate the progression of the diseases, using a western style dietary aspect in combination with chemically-induced fibrosis. C57BL/6J mice were treated for 30 days with either a low fat (10% fat) diet, or a high fat (60% fat) diet. After 30 days, mice were also dosed intraperitoneally with either olive oil (for mice on the LFD), or carbon tetrachloride (CCl₄) (for mice on the HFD). Carbon tetrachloride treatment, which is a well-established method for inducing liver injury and fibrosis, was executed every 4 days for a total of 21 days. It was forecasted that by this point in the schedule, mice would have developed histological and genetic features of NASH, and for the next 29 days, mice were given a daily oral dose of OCA (30mg/kg), Jed561 (2,10 and 30mg/kg) or vehicle (5% v/v DMSO in 0.5% w/v methyl cellulose in water). Upon completion of the study, blood samples and liver samples were taken for analysis of liver injury markers, triglyceride content quantification, histological determination of NASH (all by third party companies), and the qPCR analyses completed by the author and presented here.

Unfortunately, the treatment prescribed was unable to induce the hallmarks of NASH, as anticipated. Whilst the high fat diet was able to promote changes in body weight and genes associated with lipid metabolism, there were no significant changes in fibrotic and inflammation genes. The failure of the high fat diet and CCl₄ treatment to produce a model of NASH also was reflected in the blind assessed histopathology analysis, which failed to observe any differences between the low fat diet control livers, and those from high fat diet/CCl₄ vehicle-treated mice (Appendix 9).

Interestingly, treatment with OCA 30 mg/kg had markedly different effects on the 'NASH' mice compared to Jed561. When looking at genes involved in fibrogenesis, although the vehicle-treated NASH mice did not show any significant changes in transcript levels compared to the LFD control mice, treatment with OCA at 30 mg/kg, appeared to increase the levels of several genes analysed (Figure 6. 28). Treatment with OCA doubled the expression levels of TGF β 1, a response that was significantly higher than both the LFD control and NASH vehicle control, as tested by statistical analysis. However, TGF β 1 levels with 10 and 30mg/kg Jed561 treatment, although significantly increased compared to LFD controls, were not significantly higher than levels seen in NASH mice treated with vehicle. In fact, TGF β 1 levels were significantly lower than

levels seen in OCA-treated NASH mice (Figure 6. 28A). Likewise, whilst TIMP1 levels remained equal in LFD control mice and NASH-induced mice treated with vehicle, OCA significantly upregulated mRNA expression in NASH mice to 8-fold the levels seen in the LFD control mice. Treatment with Jed561 did not alter TIMP1 expression considerably, but levels in these mice were statistically significantly lower than those seen in mice treated with 30 mg/kg doses of OCA (Figure 6. 28C). Similarly, OCA treatment appeared to upregulate the collagen encoding gene, COL3A1, 2-fold. Although this increase with OCA, was not significantly higher than the controls, levels seen in Jed561 treated mice were actually significantly lower (Figure 6. 28B). ACTA2 expression did not significantly change with the induction of NASH, but levels were upregulated 1.2-fold in NASH vehicle treated mice (Figure 6. 28D). Treatment with OCA appeared to normalise these levels slightly, and treatment with Jed561 appeared to even downregulate ACTA2, with NASH mice treated with Jed561 at 2 mg/kg, displaying significant decreases in mRNA, to half the level seen in LFD control mice.

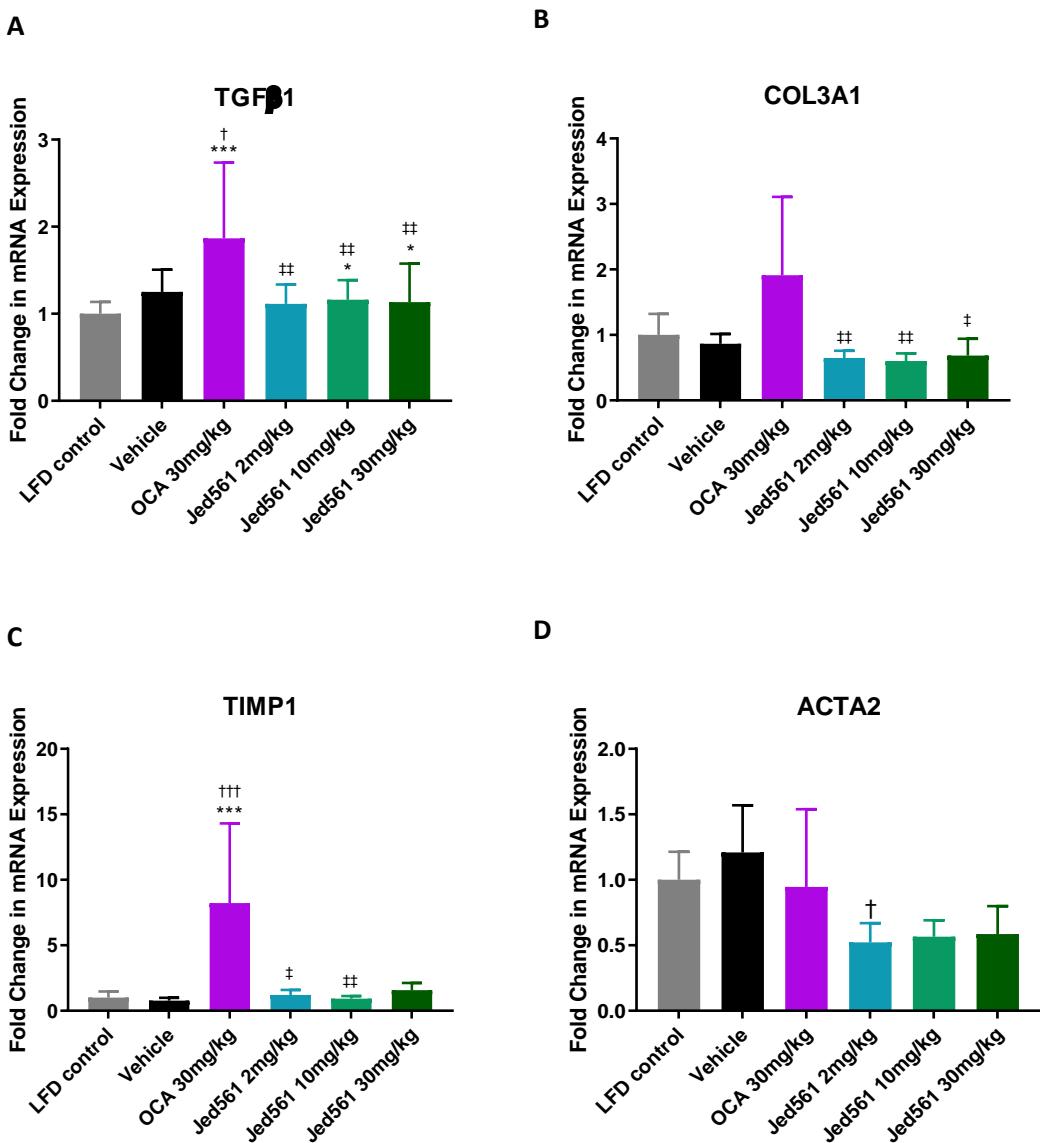


Figure 6.28 Effect of Jed561 on fibrosis genes in 'NASH' mice. C57BL/6J mice were treated with either low fat diet (LFD) or high fat diet for 30 days before CCl₄ induction every 4 days for a further 21 days. Mice were then treated with the respective compound or vehicle for once a day for 29 days. (A) TGF β 1, (B) COL3A1, (C) TIMP1 and (D) ACTA2 mRNA levels assayed by quantitative real-time PCR (qPCR). Fold changes are shown relative to (un-treated LFD mice) control. Error bars represent \pm SEM, n=10, * p<0.05, ** p<0.01, *** p<0.001 compared to low fat diet control. † p<0.05, †† p<0.01, ††† p<0.001 compared to vehicle control. # p<0.05, ## p<0.01, ### p<0.001 compared to OCA 30 mg/kg treatment.

As with the fibrosis genes, OCA also appeared to exacerbate the expression of genes associated with inflammation, in NASH mice (Figure 6.29). Treatment with 30 mg/kg doses of OCA was responsible for a 4.5-fold increase in CCR2 mRNA expression (Figure 6.29A). Meanwhile, there were no significant changes when mice were fed a high fat diet and underwent CCl₄ treatment alone. Similarly, there were no substantial increases in CCR2 expression in NASH mice treated

with Jed561; although Jed561 at 30 mg/kg displayed 1.6-fold changes in mRNA expression, this increase was not significantly different from LFD controls or NASH vehicle controls. CCR2 levels seen in OCA-treated NASH mice, again were significantly higher than those seen in mice treated with Jed561. Interestingly, treating the mice with a high fat diet and CCl₄ with daily administration of vehicle, actually downregulated CCL2 mRNA by 0.5-fold, as seen in Figure 6.29B. Treatment with OCA, however, upregulated CCL2 by 3.7-fold; an increase that was significantly higher than both the LFD control and the NASH vehicle control mice. Treatment with Jed561 at 2 mg/kg and 10 mg/kg also induced increases in CCL2, by 1.2-fold and 2-fold, respectively. Although only somewhat marginally higher than basal levels seen in LFD control mice, these increases were significantly higher than the NASH mice treated with vehicle.

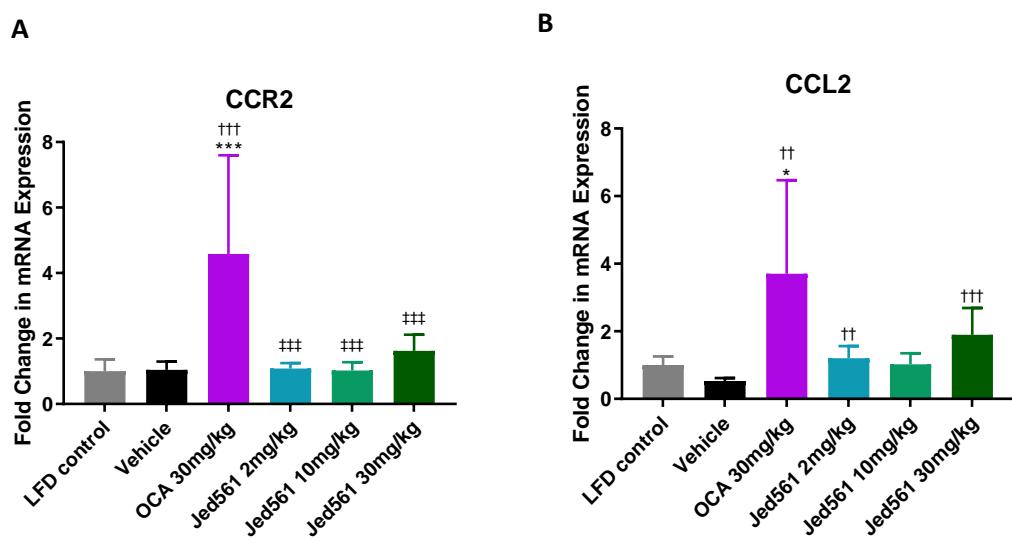


Figure 6. 29 Effect of Jed561 on inflammation genes in 'NASH' mice. C57BL/6J mice were treated with either low fat diet (LFD) or high fat diet for 30 days before CCl₄ induction every 4 days for a further 21 days. Mice were then treated with the respective compound or vehicle for once a day for 29 days. (A) CCR2 and (B) CCL2 mRNA levels assayed by quantitative real-time PCR (qPCR). Fold changes are shown relative to (un-treated LFD mice) control. Error bars represent \pm SEM, n=10, * p<0.05, ** p<0.01, *** p<0.001 compared to low fat diet control. † p<0.05, †† p<0.01, ††† p<0.001 compared to vehicle control. ‡ p<0.05, ‡‡ p<0.01, ‡‡‡ p<0.001 compared to OCA 30 mg/kg treatment.

Observing genes involved in lipid and cholesterol metabolism, as expected, high fat diet mice treated with CCl₄ and vehicle, displayed highly significant increases in the fatty acid synthesis gene, SREBF1 (Figure 6. 30A). SREBF1 expression levels in vehicle treated NASH mice were 2.5-fold higher than expression levels in LFD control mice. These levels are reduced to 2-fold with OCA treatment; and significantly reduced further to 1.6-, 1.2- and 1.3-fold, with Jed561 treatment at 2 mg/kg, 10 mg/kg and 30 mg/kg, respectively. The lipid uptake gene, PLTP, is halved in NASH mice with vehicle treatment (Figure 6. 30B). Although in OCA treated NASH

mice, mRNA expression levels display similar levels to LFD control mice, PLTP mRNA is significantly higher than in NASH vehicle control mice. Equally, NASH mice treated with Jed561 at 30 mg/kg, also display PLTP mRNA levels that are significantly higher than NASH mice treated with vehicle. At lower concentrations, however, Jed561 treatment is without effect on PLTP expression.

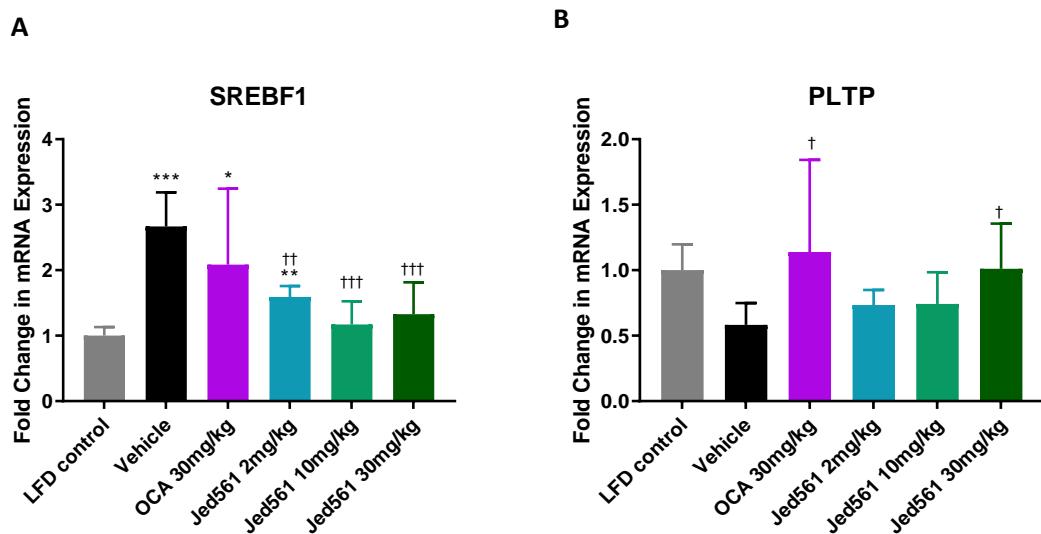


Figure 6. 30 Effect of Jed561 on lipid homeostasis genes in 'NASH' mice. C57BL/6J mice were treated with either low fat diet (LFD) or high fat diet for 30 days before CCl₄ induction every 4 days for a further 21 days. Mice were then treated with the respective compound or vehicle for once a day for 29 days. (A) SREBF1 and (B) PLTP mRNA levels assayed by quantitative real-time PCR (qPCR). Fold changes are shown relative to (un-treated LFD mice) control. Error bars represent \pm SEM, n=10, * p<0.05, ** p<0.01, *** p<0.001 compared to low fat diet control. † p<0.05, †† p<0.01, ††† p<0.001 compared to vehicle control. ‡ p<0.05, ‡‡ p<0.01, ‡‡‡ p<0.001 compared to OCA 30 mg/kg treatment.

An investigation into CASP8 expression, and the cell death pathway showed that, although the NASH model mice treated with vehicle did not show altered expression of CASP8, both OCA and Jed561 treatment at all doses, were able to downregulate mRNA to levels significantly lower than both the LFD control and NASH vehicle control (Figure 6.31).

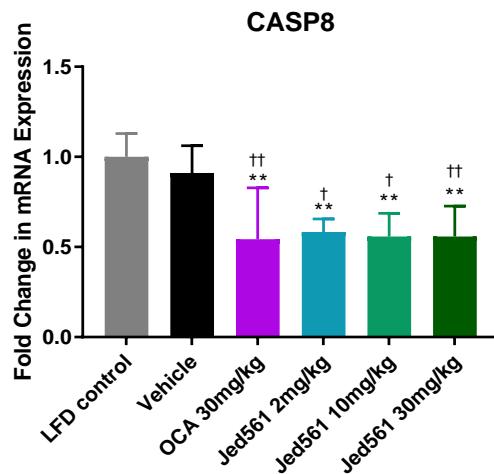


Figure 6. 31 Effect of Jed561 on CASP8 apoptosis gene in 'NASH' mice. C57BL/6J mice were treated with either low fat diet (LFD) or high fat diet for 30 days before CCl_4 induction every 4 days for a further 21 days. Mice were then treated with the respective compound or vehicle for once a day for 29 days. CASP8 mRNA levels assayed by quantitative real-time PCR (qPCR). Fold changes are shown relative to (untreated LFD mice) control. Error bars represent $\pm SEM$, $n=10$, * $p<0.05$, ** $p<0.01$, *** $p<0.001$ compared to low fat diet control. † $p<0.05$, †† $p<0.01$, ††† $p<0.001$ compared to vehicle control. ‡ $p<0.05$, ‡‡ $p<0.01$, ‡‡‡ $p<0.001$ compared to OCA 30 mg/kg treatment.

6.5 Discussion

6.5.1 Effect of Jed441 *in vitro*

Having displayed a high potency for FXR and the potential to selectively recruit coactivators, as determined by the cell free coactivator recruitment assays, Jed441 was examined in hepatic cell lines, for its ability to functionally activate FXR. Assays by qPCR observing *in vitro* transcriptomic changes in FXR-target genes in response to Jed441, showed that this novel compound has the ability to regulate genes involved in bile acid homeostasis, demonstrated by its actions in significantly upregulating SHP, FGF19 and OST α , and downregulating CYP7A1 mRNA expression (Figure 6. 3, 6. 4, 6. 7 and 6. 8). Jed441 was effective at both its EC₅₀ and EC₉₀ concentrations; and the dose-response, as seen in the increasing fold changes in gene expression, as a result of increasing compound concentrations, suggests that Jed441 has full agonist capabilities of FXR-mediated BA metabolism. Moreover, Jed441 appeared to be more efficacious than OCA at its corresponding EC₅₀ and EC₉₀ concentrations; supporting data initially seen in Chapter 5, which suggested that Jed441 had a higher affinity for FXR than OCA.

In addition, prolonged exposure to Jed441, can affect other FXR-mediated genes involved in lipid metabolism and fibrogenesis. In Huh7 cells, after 24 hours incubation, Jed441 at its EC₉₀, induced significant upregulation in APOC2, whilst also appearing to reduce SREBF1 levels. Likewise, following 24 hours incubation in HEPG2 cells, Jed441 was seen to induce decreases in TGF β 1. As discussed previously in Chapter 6, section 6.1.1.2, both APOC2 and SREBF1 contribute to the levels of hepatic and circulating triglycerides, by maintaining their clearance and synthesis, respectively. Furthermore, with evidence of TGF β cytokine signalling being associated with lipid accumulation, in addition to promoting fibrosis in hepatic stellate cells and hepatocytes (Ling Yang *et al.*, 2014); data shown here suggests that Jed441 is a potential mediator of BA-and lipid related pathways.

The upregulation of SHP, FGF19 and the other genes assessed, certainly indicate FXR-mediated activity. However, to ensure that effects were FXR-specific and not due to the activation of TGR5 or hormone receptors, a competition-type assay with guggulsterone was employed. Whilst guggulsterone treatment dampened the Jed441-driven upregulation seen in OST α , as expected, it also had an unanticipated, additive effect on the downregulation of CYP7A1 when used in conjunction with Jed441. The failure of guggulsterone to effectively inhibit the actions of Jed441 on every gene studied, does not necessarily imply that Jed441 is acting via non FXR-specific pathways, but is likely to reflect the inconclusive role of guggulsterone as an antagonist of FXR.

Although guggulsterone was initially described as an antagonist of FXR, in both its failure to recruit certain FXR-associated coactivators, and its failure to induce changes in the expression of certain FXR-target genes (Urizar *et al.*, 2002), guggulsterone is perhaps more accurately described as a selective modulator of FXR. guggulsterone was shown to upregulate FXR-target, BSEP, in both cells and animals, with the provided plausible explanation that it can recruit a different subset of FXR coactivators to the BSEP promoter that are not recruited under classical conditions seen with an FXR agonist (Cui *et al.*, 2003). Furthermore, it is now thought that guggulsterone acts as a promiscuous ligand, binding to other nuclear receptors to exert its response, like that seen by Owsley and Chiang, (2003), who reported a PXR-mediated downregulation of CYP7A1 in HepG2 cells (Owsley and Chiang, 2003).

In order to fully elucidate the FXR-specificity of Jed441, qPCR analysis could be carried out in hepatocellular cell lines whereby FXR has been knocked down or knocked out by RNA interference mechanisms, or by the increasingly popular method of gene editing, CRISPR. Both mechanisms, even with additional microarray analysis, would allow us to fully understand the actions of Jed441, which are either FXR-mediated or arise due to the off-target activation of other receptors. Nonetheless, Jed441 was tested, by external third party companies, and in TGR5 assays it was unable to promote secondary cAMP signalling, at any of the concentrations tested (up to 100 μ M, third party data not shown), indicative of minimal TGR5 activation.

6.5.2 Effect of Jed561 *in vitro*

The coactivator recruitment assays displayed Jed561 as having very low EC₅₀ values in the recruitment of coactivators to FXR. While this data is generally acceptable as an indicator of FXR agonism, Jed561 also needed to be tested for its ability to activate FXR in a genuine cellular environment and for its impact on biological molecular functions. As with Jed441, Jed561 was tested in hepatoma cell lines for a simple, initial indication as to whether it was able to regulate known FXR-target genes. Jed561 was shown to indeed regulate genes involved in BA homeostasis and metabolism. Jed561 significantly upregulated SHP, FGF19 and OST α , and down regulated CYP7A1, in changes that were sustained for up to 24 hours (Figure 6. 10, 6. 11, 6. 14 and 6. 15). Again, Jed561 appeared to be more potent than OCA at regulating these genes, at their respective EC₅₀ and EC₉₀ concentrations, also predicted by the coactivator recruitment data in Chapter 5. Unlike Jed441, however, Jed561 did not appear to be able to regulate genes involved in lipid metabolism; neither APOC2 nor SREBF1 mRNA expression changed significantly from basal levels upon incubation of cells with both concentrations of Jed561 (Figure 6. 12). Meanwhile, Jed561 did induce a downward trend in TGF β 1 expression in HepG2 cells (Figure 6.

16). This suggests that Jed561 may not be as potent as Jed441 in the regulation of all FXR target genes, or that Jed561 is in fact a more selective regulator of FXR.

Again, while guggulsterone was unable to completely ablate Jed561-mediated gene expression in competition assays, instead causing an additive effect in certain genes (Figure 6. 18), Jed561 was tested externally in TGR5 activation assays, where it was shown to be incapable of inducing TGR5 signalling at concentrations up to 100 μ M (third party data not shown), indicating some specificity towards FXR.

Comparisons between Jed561 and its non-fluorinated derivative, Jed692, show that both compounds are able to activate FXR and significantly regulate direct target genes involved in BA homeostasis (Figure 6. 19 and 6. 20). Jed692 appeared to induce an even bigger response in target gene expression than Jed561, which may suggest that the addition of the fluorine atom to the 4 β carbon position on the steroid nucleus, may even be having a detrimental effect on the compound. The addition of the highly electronegative fluorine to organic compounds has been associated with both increases and reductions of hydrogen bond donor capacity of neighbouring alcohol groups (Gillis *et al.*, 2015), and in this respect, may be inhibiting the interaction potential of the C3 hydroxyl group. However, as there is no statistical significance between the responses seen for the two analogues, the data refute the notion that the 4 β fluorine group is solely responsible for improved efficacy of Jed561.

6.5.3 Short Term Effects of Compounds *in vivo*

Despite showing promising results, by effectively regulating FXR-target genes *in vitro*, the simple and isolated nature of cancer cell lines may mean that effects of Jed441 and Jed561 are not directly translatable to a multicellular, complex organ. An inherent limitation of using immortalized cell lines, is that due to their repeated subculture, they may undergo some genetic drift. Despite being used at relatively low passage numbers, both Huh7 and HepG2 cells have been described to resemble foetal hepatocytes (Green *et al.*, 2015), and may not be representative of an adult liver. *In vitro* experiments do not represent the multicellular context of the liver, nor do they take into consideration the effects of hormonal and nutritional status in a whole organism. In addition, the *in vitro* studies do not uncover the bio-kinetic properties of these compounds. The absorption of the compounds in the intestines, the circulation of the compounds in the blood, and the bioavailability and metabolism of the compounds in the liver, is of paramount consideration when targeting a receptor expressed at multiple loci, such as FXR. As discussed previously, the desired effects of FXR-agonists are dependent on whether the

receptor is being activated in the ileum or liver, and this may even determine the clinical applications of these compounds.

Following a single, 2mg/kg bodyweight, oral dose of Jed441 and Jed561 in WT C57BL/6J mice, similar responses were seen to that *in vitro*. Jed441 was able to upregulate SHP immediately after administration (Figure 6. 23). However, with increasing time, SHP levels returned to basal levels and even appeared to be downregulated. This is consistent with previous studies, which have described this transient effect in both SHP mRNA and protein levels (Barbier *et al.*, 2003). A concomitant decrease in CYP7A1 mRNA was also seen, confirming the activity of Jed441 in FXR-mediated BA metabolism. Jed561 also was confirmed to be an agonist of FXR, with the ability to regulate BA metabolism genes; but instead displayed a continually increasing SHP mRNA level post treatment, in addition to an immediate and sustained decrease in CYP7A1. This prolonged response may reflect the bioavailability and half-lives of these compounds. Results from externally conducted caco2-intestinal absorption models and *in vivo* assessment of liver concentrations, suggest that Jed441 isn't absorbed as efficiently as Jed561, and that the half-life of the latter is longer (third party data not shown). Consequently, Jed441 was no longer considered as a lead candidate for the treatment of NASH.

In another study, following the single 2 mg/kg, oral dose of Jed561 in comparison with its non-fluorinated derivative, and competitor compounds, OCA and, non-steroidal, Tropifexor; changes in gene expression in mice livers were assessed 6 hours post administration (Figure 6. 25). Alterations in FXR-target gene expression in response to Jed561 treatment were contradictory to the results seen in the previous study. Whilst SHP expression was not affected by Jed561, BSEP was upregulated, consistent with FXR agonism. Furthermore, the potency of Jed561 compared to OCA differed with each gene. As seen *in vitro*, Jed692 appeared to be slightly more effective at upregulating both SHP and BSEP, but as there was no significant difference between the responses seen in the two treatment groups, again, it is not possible to attribute the enhanced potency of these compounds to the 4 β fluorine. Intriguingly, the steroid FXR agonists displayed increases in CYP7A1 expression, whereas Tropifexor downregulated this enzyme as expected. In mice, CYP7A1 is upregulated in response to high cholesterol diets via a LXR-mediated pathway (Goodwin *et al.*, 2003), however, it's unlikely that a single dose of these BA-like compounds would increase cholesterol levels to levels sufficient enough to induce this pathway. It is possible that these steroid compounds are directly activating LXR, whose cognate ligands are oxysterols. Even so, it has been shown that CYP7A1 regulation in humans, does not work via this same LXR mechanism (Goodwin *et al.*, 2003),

stressing the importance of considering interspecies differences when analysing the effectiveness of novel compounds.

A significant limitation seen in these initial *in vivo* studies, was the considerable variation in gene expression between subjects, especially in control, untreated animals to which expression in treated animals was normalised. This highlighted the limitations of the study design. Firstly, the animals were given free access to food, and whilst mice normally only feed during dark cycles, there were no measures put in place to control different feeding patterns of individual mice, which would stimulate the post-prandial release of endogenous BAs, likely activating FXR. Future studies in mice included a short fasting period prior to euthanasia, in accordance with protocols approved by the Animal Care and Use Committee. Furthermore, the liver sections provided for these initial studies were random. Transcriptomic differences in different lobes of the liver, are well documented in rodents (Corton *et al.*, 2012), and could be a source of variation seen here. The succeeding studies utilized liver sections from the left lobe in order to minimize any inconsistency. Additionally, the use of only three mice per treatment group in these studies, means that any outliers in the data can have a powerful effect on the averages calculated, the distribution of the data, and the variation observed, often masking genuine patterns in gene expression. As such, for later studies, the number of mice per treatment group was increased to ten, to increase the statistical sensitivity of the assessments.

6.5.4 Long Term Effects of Compounds *in vivo*

Despite all compounds being derived from the same BA backbone, continual administration of OCA, Jed561 and Jed441 produced different responses in FXR target genes. This may be in part due to the differential recruitment of different coactivators, as seen in Chapter 5, or in part due to the activation and interference of other pathways.

Comparisons between mice treated with 30 mg/kg doses, of either OCA or Jed561, show that OCA causes a slight upregulation in FXR mRNA expression, whereas repeated administration of Jed561 at the same concentration causes a slight downregulation in FXR mRNA expression. Again, the downregulation of FXR may arise as a result of negative feedback regulation due to continued activation of the receptor, however, if the FXR mRNA levels directly correlate to protein levels, this could have a consequential effect on FXR-target genes.

OCA appears to be more potent at upregulating SHP *in vivo*, as seen previously (Figure 6. 25), however, this increase does not appear to have a subsequent effect on the expression of the downstream SHP target, CYP7A1. Jed561, however, although not as potent as OCA in SHP regulation, downregulates CYP7A1 with 30 mg/kg doses, and upregulates BSEP to higher levels

than OCA. Recent studies by Hoeke *et al.*, (2014), have described a previously unidentified mode of FXR-mediated transcriptional regulation of SHP, which occurs by binding at a LRH-1 site, unlike BSEP, which is regulated through an inverted repeat FXRE. It may be that OCA can more effectively promote this type of regulation to upregulate SHP, compared to Jed561, which may be more efficient at activating the typical FXRE binding consensus. This mechanism, however, does not explain the discrepancy between increased SHP levels in OCA-treated mice, and the failure to downregulate CYP7A1.

Moreover, OCA-related increases in SHP fail to inhibit the expression of SREBF1, and quite significantly, treatment with OCA appears to exacerbate mRNA levels, compared to vehicle treated mice (Figure 6. 27). While at a 2 mg/kg dose, Jed561 treated mice also display significant increases in SREBF1, treatment with 10 and 30 mg/kg Jed561 appears to downregulate or not significantly alter this gene. Like CYP7A1, the stimulation of SREBF1 is due to oxysterol and cholesterol activation of LXR (Watanabe *et al.*, 2004). It's possible that, due to FXR activation and the suppression of CYP7A1, the increase in cholesterol and oxysterol levels induced a counter response by LXR and subsequently upregulated SREBF1, however, OCA treated mice do not display any significant changes in CYP7A1 expression. Again, it's possible that these steroidal compounds are activating LXR directly or activating LXR/SHP-independent pathways to upregulate SREBF1. However, it is unclear why different doses of Jed561 produce conflicting responses. The determination of whether these responses are due to activation of FXR, or due to the specific class of drug, requires further investigation, and long term treatment of the mice with a non-steroidal agonist, such as Tropifexor, is needed. Nevertheless, these results could have serious implications for the utilization of these compounds as treatments for metabolic disorders. Long term treatment with OCA was associated with an increase in total serum cholesterol and LDL, with an additional decrease in HDL in NASH patients (Neuschwander-Tetri *et al.*, 2015), and although the interspecies differences in cholesterol metabolism have been well documented, the effect on sterologenic and lipogenic genes is an important consideration when evaluating the potential therapeutic capabilities of these compounds.

Nevertheless, at 10 and 30 mg/kg, Jed561 is able to significantly upregulate APOC2, unlike OCA, suggesting that this novel compound may be beneficial in the regulation of lipid metabolism at its highest concentrations, and may indeed pose advantages over OCA.

Treatment with Jed561 at 30 mg/kg, appeared to be the most effective at modulating the entirety of FXR effects. Doses as little as 2 mg/kg were adequate to affect changes in BA metabolism associated genes. However, this concentration was not able to effectively regulate

genes involved in lipid metabolism. Nevertheless, fine tuning studies to find an appropriate dose, may mean that Jed561 can be used at concentrations significantly less than those for which OCA is currently being used in clinical trials, further minimising the potential of adverse side effects.

The actions of Jed441 on FXR target genes in mice were conflicting. Whilst upregulating BSEP and APOC2, indicative of FXR activation, Jed441 was not able to significantly upregulate SHP. It also upregulated CYP7A1 and to an extent, SREBF1, which is not expected for an FXR agonist. This differential response in FXR target genes, may be as a result of differential transcriptional regulation by this compound, or arise due to problematic absorption.

6.5.5 Effects of Jed561 in a NASH Mouse Model

In the final study reported here, which aimed to identify any beneficial effects in the amelioration of NASH arising from Jed561 treatment, conclusive results were hindered by the misfortune that the histological and transcriptomic hallmarks of NASH were not achieved. Initially, the mice did not tolerate the CCl₄ treatment well, resulting in the culling of 2 subjects and the revision of the treatment schedule to adopt a 4-day interval between fibrosis treatments instead of 3 as described in (Hoshida *et al.*, 2018). Furthermore, it is possible that the scheduled 3 week CCl₄ treatment was long enough to induce a substantial or sustained fibrotic phenotype; previous studies using this NASH mouse model reported CCl₄ treatment for 4 to 12 weeks (Xiao *et al.*, 2017; Hoshida *et al.*, 2018). Vehicle treated NASH mice in this study displayed increases in total body weight, but not liver weight, and did not have significantly increased levels of triglycerides in the liver. Furthermore, they did not show altered levels of liver injury markers and only displayed mild “NASH” scores and fibrosis scores, as determined by an independent assessor (third party data not shown).

The data from the liver marker assays and histological analyses, were corroborated by the data from the qPCR experiments reported here, which also displayed no significant changes to genes involved in fibrosis and inflammation in NASH mice treated with vehicle. The only changes seen in the untreated NASH mice were in SREBF1 expression levels, which were significantly upregulated, and in PLTP which showed a slight reduction. This suggests that the mice used in this model may have only developed the steatosis component of the disease, and not the advanced inflammation and fibrotic aspects.

Despite the failure of these mice to develop substantial features of NASH, treatment with OCA remarkably appeared to exacerbate NASH, by significantly upregulating genes involved in fibrosis and inflammation (Figure 6. 28 and 6. 29). TGF β 1, COL3A1 and TIMP1 are all markers of

extracellular matrix deposition and were all upregulated by 30 mg/kg doses of OCA. Data reported here contradict previous studies, whereby 3 mg/kg OCA was able to prevent CCl₄-induced fibrosis and led to decreases in the expression of TGF β 1, TIMP1 and ACTA2 in rats (Fiorucci *et al.*, 2005). The treatment of NASH mice with Jed561, however, did not significantly upregulate any fibrosis genes, and instead was associated with a decrease in ACTA2 expression levels compared to untreated NASH mice. Likewise, OCA, but not Jed561, was associated with increases in the expression of CCL2 and its receptor CCR2. Again, this contradicts previous studies, which reported that OCA treatment at 5 mg/kg was able to downregulate the expression of CCL2 in mice subjected to CCl₄ liver injury (Zhang *et al.*, 2017). The upregulation of these fibrogenesis and inflammation genes also translated into the increase in fibrosis and lobular inflammation, and OCA treated mice had the most severe NASH and fibrosis scores determined by histological staining (third party data not shown).

While treatment with Jed561 did not significantly affect fibrosis or inflammation at a genetic or histological level, it did have substantial effects on the expression of genes involved in lipid metabolism. SREBF1, a regulator of *de novo* lipogenesis, was significantly upregulated in untreated NASH mice. This upregulation was improved with treatment by OCA, and more so by treatment with Jed561. In addition, the downregulation in PLTP, caused by the high fat diet and CCl₄ treatment, was normalised by both OCA and Jed561. The regulation of lipid metabolism genes by Jed561 does not appear to be substantially dependent on the concentration used, as responses seen for treatment with 2 mg/kg do not significantly differ from responses seen for treatment with 30 mg/kg.

Hepatocyte cell death is thought to contribute to the progression of fibrosis and inflammation in both murine and human models of NASH. Hatting *et al.*, (2013), recently showed that the cysteine protease, CASP8, is responsible for the apoptosis driven increase in inflammation and fibrosis in NASH mouse models. Moreover, they described the potential role of CASP8 in hepatic fat metabolism. Data reported here shows that both OCA and Jed561 can significantly reduce CASP8 expression compared to NASH mouse controls and LFD mouse controls. This suggests that Jed561 is still a contender for the treatment of NASH by attenuating CASP8 signalling.

While the results shown here do not overtly indicate the beneficial use of Jed561 for the treatment of NASH phenomena such as fibrosis and inflammation, Jed561 could be a potential therapeutic treatment for lipid dysregulation. While the most effective dose of Jed561 is yet to be determined, at 30 mg/kg, it shows distinct differences to the semi-synthetic BA, OCA. Unlike OCA, Jed561 was not associated with significantly increased fibrosis genes or markers, and

Jed561 treatment did not increase liver injury markers, ALT or AST (third party data not shown). These data suggest that Jed561 and OCA are working via independent mechanisms to produce the responses in NASH mice described here.

6.5.6 Conclusions

Over the last two decades, the sensitive and precise technique of real-time qPCR, has proved invaluable to the validation and assessment of transcriptional gene expression changes. As with other techniques, qPCR has its associated limitations, arising from the need to appropriately normalise the data to account for differences in RNA extraction methods, the need to choose abundant and stably expressed reference genes, the requirement of efficient and specific primers, and of course, in recognising the fact that altered levels of mRNA do not always directly correspond with altered levels of functional protein activity. Nonetheless, the use of qPCR for this work here, has helped to identify the transcriptional changes associated with novel compounds, Jed441 and Jed561.

The data presented here confirm the roles of both Jed441 and Jed561 as agonists of FXR. The improved potency of these novel compounds in the regulation of FXR-target genes compared to OCA, at equivalent concentrations *in vitro*, suggest that the addition of chemical groups to the steroid backbone, confers enhanced FXR affinity and activation. Although the results do not conclude that the 4 β F group is responsible for this improved affinity, the chemical substituents on the C24 end of the compound may prove beneficial, and the added fluorine groups may be advantageous in altering the stability or solubility of these compounds.

The different responses in FXR target gene regulation *in vivo*, suggest that OCA and the Jed compounds are working via alternative pathways. In addition, Jed compounds may be capable of regulating specific target genes in a selective manner. The distinct regulation of genes by Jed561 compared to OCA, may prove beneficial to Jed561, which may be able to ameliorate the metabolic components of NASH pathogenesis, without exacerbating fibrosis. The results do not rule out the activation of other nuclear receptors such as LXR, which also play an extensive role in the pathways underlying NASH progression. However, the results suggest that Jed561 is still a contender for pharmacological NASH treatment, if not alone, in conjunction with other therapies.

Chapter 7.

General Discussion and Recommendations for Future Work

7. General Discussion and Recommendations for Future Work

BAs are important molecules that have well-established roles in aiding absorption and digestion in the intestines. More recently, BAs have been identified as potent, hormone-like signalling molecules, with systemic endocrine and cardio-metabolic effects. Through activation of G-protein coupled receptor, TGR5, and nuclear transcription factors, such as FXR, BAs can effectuate diverse signalling networks, many of which contribute to the aetiology of certain diseases. BAs have been shown to regulate their own homeostasis and circulation, as well as controlling energy maintenance, lipid metabolism, and mechanisms mediating inflammation and microbial gut populations. FXR, in particular, has been lauded as an attractive therapeutic target due to its ligand-dependent regulation of several genes and downstream signalling cascades that are pertinent to these functions. Dysregulation of these pathways can lead to the progression of hepatic complications and metabolic disorders, and extensive research has identified a pivotal role of FXR in NASH. The increasing prevalence of NASH and the paucity of effective treatment emphasized the urgency in developing a pharmacological agent. As such, significant efforts have been directed at identifying an FXR agonist with the desired genetic effects and pharmacokinetic properties to treat this disease. Despite the development of BA-derived FXR agonist, the 'first in class', OCA; clinical studies with this drug identified the need for compounds that could activate FXR independently of TGR5 (Neuschwander-Tetri *et al.*, 2015). Furthermore, due to the complexity of FXR signalling and the far-reaching effects of its activation, gene-selective modulation, or methods to limit the systemic effects of FXR are required. The aim of this work was to develop a highly potent FXR agonist, which could activate FXR and mediate its downstream target genes in a manner that would be advantageous in the treatment of NASH. By developing a highly potent agonist, it was thought that the reduced effective dose of the compound, may still be able to have the intended effect on key FXR mediated metabolic pathways, whilst having limited systemic circulation and minimising undesirable off-target effects.

Detailed analysis of FXR-ligand complexes defines important structure-function features that guide design of novel BA compounds

In order to design novel FXR agonists, a preliminary understanding of FXR architecture, ligand binding mechanisms and structural dynamics was required. To date, there have been no published structures of FXR in its entirety, but secondary structure prediction and homology modelling applied in this work, projected an overall FXR structure characteristic of a typical NR; a highly disordered N-terminal region, a zinc-finger fold DBD, and a 12 alpha helix bundle LBD

possessing a hydrophobic binding pocket (Rastinejad *et al.*, 2013). The compilation and evaluation of the numerous, previously published, crystal structures of the FXR LBD highlighted the structural plasticity of the domain, whereby individual loops and helices can move to accommodate diverse ligand scaffolds and to facilitate receptor activation. Comparisons of the FXR LBD structures showed that helix 12 is remarkably flexible, and its assumed conformation is dependent on whether an agonist or antagonist occupied the ligand binding pocket. These results support existing dogma surrounding NR dynamics, whereby the current understanding is that helix 12 conformation determines the availability of coactivator protein interaction surfaces, which ultimately promote the transcription of associated target genes (Rastinejad *et al.*, 2013). In its apo form, the LBD assumes a ‘molten’-like state where helix 12 lacks structural order and is not fixed in a single conformation, but rather exists in a mobile, dynamic state. Binding of an agonist, and ligand filling of the interior pocket, stabilizes the global LBD conformation and helix 12 position against the LBD core surface, forcing corepressor proteins to dissociate, whilst securing the hydrophobic binding groove required for coactivator binding (Rastinejad *et al.*, 2013). On the other hand, the binding of a bulky antagonist compound, acts to prevent the stabilisation of the LBD by an agonist, or to induce a conformation where helix 12 is displaced, suspending coactivator interaction surfaces. This dynamic stabilization mechanism underlies work presented in Chapter 5.

In addition to emphasizing the structural dynamics of receptor activation, evaluation of the ligand binding mechanisms in published FXR co-crystal structures, along with computational analyses, clarified the unique chemical environment of FXR’s ligand binding pocket. The top side of the pocket along helix 3 is predominantly lined by hydrophobic residues, whereas the bottom region of the cavity is well suited to polar interactions, allowing the receptor to easily discriminate the amphiphilic properties of cognate BA ligands. Certain residues, such as H447, Y361 and W469 have been described as the ‘activation trigger’ for FXR, however, interactions with these residues were not crucial for receptor activation, and ultimately, any ligand binding mechanism that stabilizes helix 12, either directly or indirectly, can be considered to have agonist properties.

The most potent, naturally occurring BA, CDCA was exploited for the design of novel ligands and extensions of the steroid backbone were anticipated to engage in previously unoccupied sub-pockets of the LBD. Removal of the carboxylic acid and extensions of the side chain, with the inclusion of different functional groups, was expected to limit the promiscuity of BA ligands and enhance specificity toward FXR. As carboxylic acids are thought to be susceptible to conversion to liver toxic metabolites (Flesch *et al.*, 2017), altering the classic BA scaffold was believed to be

advantageous for numerous reasons. Designs included the addition of various functional groups and moieties, including fluorine atoms and trifluoromethyl groups. Fluorine functional groups, are routinely added to compounds in medicinal chemistry as they have an influence on numerous pharmacodynamic and pharmacokinetic properties; their electron withdrawing capacity can influence interactions in ligand-protein binding resulting in increased affinities, their lipophilic properties afford the compounds a lower pKa and improved bioavailability, and carbon-fluorine covalent bonds are more stable than hydrocarbon bonds, meaning that fluorine containing compounds tend to be more metabolically stable in liver microsomes (Shah and Westwell, 2007). The addition of functional fluorine groups has been observed in FXR agonists, such as the highly potent Tropifexor (Tully *et al.*, 2017), and the effect of their addition to the BA scaffold of novel ligands described in this work was the subject of investigations at a structural, biochemical and molecular level (Chapters 3, 5 and 6).

Identification of key FXR-ligand interactions important for specificity and activation, using in silico docking

In order to elucidate ligand binding details of prospective new compounds and to discern the effects of added functional groups, compounds were grouped according to structural similarity and ‘placed’ into the FXR LBD by *in silico* molecular docking. Docking solutions identified that novel compounds in this library could all potentially make polar interactions with ‘classic’ residues H447, Y361, Y369, S332 and R331, whilst the addition of electronegative atoms or bulky functional groups could contribute to additional halogen bonds or VdW interactions. In addition, the extended structures of many of the novel compounds, occupied the entire ligand binding pocket, and many compounds, mainly those with distal trifluoromethyl side chains, were observed to protrude into a sub-pocket between loop 1/2 and helix 3. Previous studies have posited that this sub-pocket is an alternative binding site for small molecule modulators of FXR (Meyer *et al.*, 2005), and it is possible that many of the novel compounds could display partial, but selective activation of the receptor. Whilst docking into a semi-rigid FXR structure was not able to discriminate the degree of agonism, between full, partial or selective agonists, docking studies did identify that the extended structure of the lead compounds were unlikely to support an antagonist conformation of FXR, and compounds were unable to bind in the proposed binding site for TGR5, implying that these compounds were indeed FXR-specific agonists.

Due to the flexible nature of the receptor and the mechanisms surrounding its activation, software-calculated binding affinities do not necessarily correlate with receptor activation, and so top-ranking compounds, identified by the SYBYL software scoring function, may not have

genuinely been the most active. Whilst providing important molecular detail of ligand binding mechanisms, when used on its own, molecular docking methods utilized here, were unable to accurately determine which interactions were responsible for increased activation, and conclusively, which ligands would be the most potent. Activity data of many of these compounds, obtained from third party sources, helped to postulate potential ligand interactions and binding orientations, which may account for the improved efficacy of these compounds over both CDCA and OCA. By correlating the activity data with the structural features of the compounds, it was observed that in general, ligands with a fluorine atom addition at the C4 β position on the 'A ring' of the BA backbone, outperformed other compounds without this substituent. Docking studies identified that these compounds could make additional halogen bonds with H447 and Y361, in addition to the canonical hydrogen bonds from the C3 hydroxyl, possibly increasing the ligands specificity and affinity for the receptor and enhancing the levering mechanism to constrain helix 12 in its active position. However, this addition was not enough to completely account for the increased potency of these compounds, and the side chains were proposed to have an equally important effect on potency. Side chains of the most potent compounds were positioned towards to the top of the pocket where many, such as Jed441 and Jed561, were observed to make polar interactions with residues along helix 3. Data from docking solutions, collectively with supplementary data from FXR-activation assays, implied that lead compounds identified herein, may have enhanced activity due to their interaction mechanisms with helix 3, which is known to form part of the coactivator binding groove with helix 12.

Attempts to quantify the structure-activity profile proved challenging

In order to rationalise which specific functional groups were responsible for the increased potency of these compounds, a 3D-QSAR was attempted to statistically correlate the structural features of novel ligands with their associated activity. The complex structure of these compounds, in which the addition of different functional groups alters the torsion angles and puckering of the ligand scaffold, and its position within the ligand binding pocket, in addition to the limited range of activity, rendered the QSAR attempts unsuccessful. The QSAR model generated yielded little further insight to what was already established by molecular docking. However, although molecular docking proved to be quick and computationally inexpensive, concerns were raised surrounding its accuracy. The FXR LBD is intrinsically highly flexible, and the limitations of a semi-rigid docking procedure emphasize the need for validation of proposed binding mechanisms and characterization of ligand binding in its native state. Whilst other computational methods, such as molecular dynamics, can simulate ligand binding and the

conformational changes involved in helix 12 stabilization, experimental methods such as X-ray crystallography or NMR were postulated to provide atomic level detail, whereby the overall structure of the LBD helices and side chains could be determined, in addition to the positioning of the ligand and its proximity to proposed binding residues and helices. Furthermore, it was anticipated that the LBD structure obtained with these novel ligands, would be more appropriate for future rounds of *in silico* docking.

A structure determination approach to validate FXR-ligand binding

Although there are currently no published structures of the full length FXR protein or the DBD, several publications have documented the recombinant expression, purification and crystallisation of the FXR LBD in complex with a ligand. Small scale expression screens identified that His₆-tagged FXR LBD could be expressed in a soluble form, using *E.coli* as a host, and with some optimization, expression of the FXR DBD-LBD also may be achieved. With the priority of investigating novel ligand binding mechanisms, work in this thesis focussed on solely the FXR LBD. However, the highly flexible, hydrophobic nature of the FXR LBD presented several challenges in the large-scale overexpression, purification and crystallisation of this protein.

The problem of low target protein yield was fundamentally addressed by the use large scale cultures, in excess of 4L, similarly to what had been reported in literature (Akwabi-Ameyaw *et al.*, 2008). Together with a more efficient and scale-appropriate lysis method of cell disruption, FXR LBD yields were considerably improved. Despite this success, purification attempts were thwarted by the protein's propensity to aggregate and co-elute with similar sized, histidine-rich *E. coli* proteins. Attempts were made at separating the contaminant proteins by various chromatography methods, however, the similar molecular weights of the proteins, and the need to minimize the loss of target protein, proved these measures ineffective. Final protocols included a thrombin cleavage step to remove the His₆-tag, followed by an additional purification by IMAC, to separate resin-bound proteins from the now cleaved FXR LBD. Although this method resulted in a 'cleaner' target protein, the sample was not 100% homogenous and future studies may need to include more stringent wash steps and alternative SEC column resins to address the heterogeneity. It may even be required for the FXR LBD to be co-expressed with obligate partner receptor domain, RXR LBD, to prevent the FXR LBD from dimerizing with non-specific proteins. Time-dependent precipitation has been previously documented by other groups working with this protein (Soisson *et al.*, 2008) but the measures taken to counter this problem, such as site directed mutagenesis of surface exposed cysteines or expression of the coactivator peptide grafted on to helix 12 (unpublished; private communication with Stephen M. Soisson, MERCK), suggest that the challenges are not trivial and not well documented.

Although supplementation of the protein with DTT and addition of the ligand and coactivator peptides shortly after purification, was observed to stabilize the domain, future studies should consider mutagenesis or alkylation of surface cysteines to completely mitigate the formation of disulphide bonds, thus limiting potential aggregation.

The expression construct later used in this work encoded the FXR LBD with 2 mutations of surface glutamate residues. Mutating these residues to alanines, was thought to overcome the energy requirements essential for stabilizing glutamate's highly dynamic side chain in the creation of crystal contacts, increasing the crystallisability of the protein (Cooper *et al.*, 2007). Although diffraction quality crystals were not obtained, the formation of microcrystals and crystalline precipitate clusters, suggests that with comprehensive refinement and optimisation of the buffer reagent conditions, and possibly with the help of microseeding, superior quality crystals may be acquired.

Regardless of not being able to validate the hypothesized ligand-binding mechanisms via X-ray structure determination, the optimised methods established here (Chapter 4) may help future lab members to express and purify the FXR LBD, and knowledge acquired may be transferable to other NR LBDs or other highly flexible, hydrophobic eukaryotic proteins. Furthermore, other structural techniques can be applied to the FXR LBD to provide further insight regarding receptor activation and dynamics. NMR may be able to help discriminate between different mechanisms of receptor activation, and may be able to provide information about whether the novel ligands are promoting a canonical active conformation, or one which resembles partial agonism (Kojetin and Burris, 2013). Likewise, the use of other biochemical techniques such as steady-state fluorescence anisotropy, may allow helix 12 mobility to be monitored upon ligand binding, providing additional evidence of genuinely active compounds, regardless of whether the exact mechanism has been elucidated or not.

Lead compounds selectively recruit FXR coactivators

FRET-based recruitment assays are a biochemical technique that can be employed to monitor coactivator associations with the LBD, by way of measuring receptor activation and dynamics. Recruitment of a coactivator protein essentially determines whether FXR's transcriptional role is executed or not. Coactivators are responsible for the recruitment of other transcription factors, chromatin remodelling processes, post translational modifications, and overall epigenetic changes, that govern the assembly of transcriptional machinery required for the initiation of target gene expression (McKenna and O'Malley, 2002). The mechanism of ligand-induced helix 12 stabilization, and subsequent coactivator recruitment provides the basis for

this assay whereby potential ligands can be experimentally assessed. Titrations of ligands or coactivator peptides can be used to quantitatively determine the affinity of the FXR LBD for that binding partner. Results presented in this thesis demonstrate that both Jed441 and Jed561 are genuine agonists for FXR (Chapter 5). The compounds were able to recruit fluorescein-labelled coactivator peptides to a tagged FXR LBD, resulting in an increase in the measured fluorescence signal. The novel compounds are significantly more efficacious than CDCA, and even had up to 10-fold higher affinities for FXR than OCA. Furthermore, both compounds were able to recruit a number of different coactivators, whilst also causing the slight dissociation with corepressor peptides, further supporting the idea that the conformation adopted by the FXR LBD upon compound binding, is one of an active receptor. Moreover, the narrow range of EC₅₀ values, determined by the recruitment of several different coactivators, upon Jed561 binding, indicate that FXR's affinity for Jed561 remains unchanged regardless of the coactivator used. Contrary to results obtained by cell-based FXR reporter gene assays, conducted by a third party company, initial coactivator recruitment assays here, placed Jed561 as a more potent agonist than Jed441, with EC₅₀ values between 11-12 nM, as opposed to 19-35 nM. These differences may have been due to the cell-free nature of the assay, or due to the lack of integrated signalling from receptor partner RXR, which has been shown to increase the affinity of FXR for ligands, when used as a dimer in coactivator assays (Zheng *et al.*, 2018); and later gene expression studies, showed that Jed441-induced FXR activation resulted in greater responses in target gene regulation than that seen with Jed561 (Chapter 6).

A noteworthy concern in the design of FXR agonists, is the need to not just mimic cognate agonist activity, but to develop selective modulators that can activate target genes independently of one another. In the body, FXR is subject to several different mechanisms, which all serve to limit the receptor's promiscuous actions to a few intended target genes or tissues. One such way in which FXR's actions are restricted, is due to the cooperation between the receptor and different cofactors, to generate specific gene expression activation (Massafra *et al.*, 2018). Structurally distinct ligands have been proposed to differentially modulate the interaction with different coactivators, which in turn, due to their different post-translational modifications, can orchestrate the sub-cellular localization of FXR, its protein-protein and protein-DNA interactions, as well as its protein stability and transcriptional activity (Massafra *et al.*, 2018). Furthermore, the expression and availability of coactivator proteins are often tissue-specific, allowing ligands to initiate distinct effects dependent on the composition and stability of multi-subunit coactivator complexes (Torra *et al.*, 2004; N. Li *et al.*, 2017).

One of the aims of the research presented here, was to identify if top candidate ligands, Jed441 and Jed561, were sufficiently structurally distinct from OCA, that they could specifically recruit different coactivator proteins. By titrating different coactivator peptides, representative of different coactivator family subclasses, recruitment assays were indicative of the FXR receptor affinity for coactivators, when each ligand occupied its binding site. OCA displayed no notable differences in recruitment profiles of the three coactivator peptides tested, SRC2-2, RIP140 L8 and TRAP220/DRIP 1, suggesting that the affinity for all three were similar and that OCA could not selectively recruit (or differentiate between) these coactivators. However, both Jed441 and Jed561 displayed slightly different recruitment profiles. Results suggested that Jed561 had a higher affinity for SRC2-2 than the other coactivators, and Jed441 also displayed dramatically increased associations with SRC2-2, particularly at higher concentrations of the peptide. Although further work is needed to fully support this hypothesis, results shown here provide an early indication that novel compounds, Jed441 and Jed561, may be able to selectively recruit coactivator complexes, thus, achieving regulation of specific genes. Recommendations into gaining further evidence for this, include using co-immunoprecipitation and pull down assays, which may be able to identify differences in OCA bound- and Jed441/Jed561 bound- FXR interactions with different regulatory proteins.

Another mechanism thought to achieve selective gene modulation, is through the occupancy of the auxiliary binding site between loop 1/2 and helix 3 (Pellicciari *et al.*, 2006). This site was identified as the binding site for the small molecule guggulsterone, which was originally thought to be an FXR antagonist, but later redefined as a selective modulator of FXR (Urizar *et al.*, 2002). Although the exact mechanisms linking ligand binding mode, receptor conformation and gene-specific regulation still need to be delineated, it is postulated that extended, carboxylic acid side chains of BA-derived compounds may be able to induce a broad range of functional FXR profiles due to perturbations of this secondary pocket (Pellicciari *et al.*, 2006), paving the way for the development of steroid selective agonists of FXR. Docking results presented in Chapter 3, highlighted that both Jed441 and Jed561 are closely associated with helix 3, with side chain extensions partially occupying the secondary sub-pocket. However, their bona fide structural conformations are yet to be determined, and their putative transcriptomic effects still need to be assessed in order to determine their ability to regulate specific genes and pathways.

Jed compounds activate and modulate target genes in vitro

To identify the transcriptomic changes induced by Jed441 and Jed561, gene expression analysis was conducted by the use of qPCR with focus on FXR regulated genes. Hepatocyte-derived cell lines, Huh7 and HepG2, were used as a simple platform to initially identify if direct FXR target

genes could be up- or down-regulated as a consequence of compound treatment. Cells were treated with either compound, at their EC₅₀ and EC₉₀ concentrations, in parallel to OCA at its respective concentrations, which, as the 'gold standard', was used for comparison. Results showed that both Jed441 and Jed561 are genuine agonists for FXR, in that, they can functionally activate it and modulate its target genes. Both lead compounds were shown to control FXR target genes involved in BA homeostasis, upregulating SHP, FGF19, and OST α , and downregulating CYP7A1, in dose-dependent responses. Furthermore, both compounds were shown to be more effective than OCA at exerting these FXR-mediated changes, and target genes displayed an increased fold changes in expression when cells were treated with Jed compounds compared to OCA, at their corresponding EC₅₀ concentrations. Results initially observed (presented in Chapter 5) suggested that Jed441 and Jed561 both had higher affinities for FXR than OCA. Additionally, evidence presented in Chapter 6 showed that both Jed compounds had a higher efficacy compared to OCA also.

In addition to genes underlying BA metabolism and transport, prolonged exposure to Jed441 was able to affect FXR-mediated genes involved in lipid metabolism and fibrogenesis *in vitro*. At 180nm, Jed441 was associated with the upregulation of APOC2, as well as decreases in SREBF1 and TGF β 1 mRNA. However, this response was not seen with Jed561, which only showed downregulation of TGF β 1 *in vitro*. The results suggest that Jed561 may not be as potent as Jed441, in line with data initially obtained by a third party derived from cell-based FXR reporter assays. Equally, these results may have arisen from the differential activation of target genes by the two lead candidates. Nevertheless, *in vitro* results suggest that the lead candidate compounds, Jed441 and Jed561, could both functionally regulate FXR target genes.

Following an exhaustive analysis of the structure-function relationships of known ligand-FXR complexes (Chapter 3), a number of novel compounds were designed to include extended side chains, which could potentially occupy the entire FXR binding pocket, in an attempt to tailor the compounds towards FXR specificity. To determine if compound-induced transcriptomic changes were due to the specific activation of FXR, competitive inhibition-type assays were performed with guggulsterone. Although guggulsterone was unable to activate FXR in coactivator recruitment assays, its ability to selectively modulate FXR, meant that results regarding the specificity of Jed441 and Jed561, were inconclusive. For a more definitive confirmation that responses to the Jed compounds are FXR-specific, future studies could utilize gene silencing or editing methods to create FXR-null cell lines, and to test gene expression in response to compound treatment of these cells. Nevertheless, assessment of Jed441 and Jed561 in cAMP-response assays by external third parties, confirmed that these compounds were not able to

activate TGR5, as suggested by the docking studies. Accordingly, the primary aim of this work was achieved, in that novel compounds, Jed441 and Jed561, had both been discovered and identified as highly potent, specific agonists of FXR, with higher efficacies than the steroidal competitor, OCA.

Without a 3D structure of the novel ligands in the receptor pocket, the exact mechanisms underlying the molecular recognition and enhanced potency of Jed441 and Jed561 remain unknown. It is postulated that the addition of a fluorine atom to the BA backbone may be positively influencing the compounds affinity for FXR. In addition to fluorine's unique properties in altering the pKa, solubility and hydrogen bond-donating capacity of potential drug candidates, the inclusion of distinct chemical signatures, such as fluorine atoms, in the naturally occurring BA backbone, provides the potential advantage of these candidates being protected under appropriate intellectual property regulations. Initial *in silico* docking solutions of potential compounds with a fluorine atom added to the 'A' ring of the OCA, identified that these additions were not expected to alter the overall binding modes of these compounds in the ligand binding pocket. However, compounds with the addition of fluorine atom at the C4 β position, could benefit from additional halogen interactions with Y369, whereas fluorines added to other positions along the 'A' ring, could not. Although a comprehensive study into the addition of a fluorine atom at different loci along the BA backbone, and the subsequent effect on compound activity, was not carried out in this thesis, coactivator recruitment assays with the 1 β F analogue, Jed678, identified that this compound had a significantly worse EC₅₀ value than OCA. This suggested that the position of the fluorine atom on the steroid nucleus, could have important effects for ligand binding. Nevertheless, comparisons of docking results between Jed561 and its non-fluorinated equivalent, Jed692, identified differences in the sub-locations of the compounds within the binding pocket, where Jed692 was positioned further out of the cavity and Jed561 was seen to occupy the entire pocket, including slight extensions into the allosteric sub-pocket. These differences are supported by evidence from the coactivator recruitment assays, which suggest that Jed561 has a slightly better affinity for the FXR LBD, than Jed692. However, this enhanced potency of Jed561 does not appear to be translated to cellular responses, as the qPCR data suggest that Jed692 has similar or improved gene regulation responses compared to Jed561. The evidence points to the extended side chain tails as the moieties responsible for improved FXR activity, however, the potential advantages of 4 β F additions on the physico-chemical properties of the compound cannot be dismissed.

Jed compounds activate and modulate target genes in vivo

Confirmed as legitimate FXR agonists, Jed441 and Jed561 were tested in mice, for their characterization in a multi-organ, complex organism, and evaluation for their potential use as therapeutic agents for diseases such as NASH. Supplementary data from third party sources, identified that, while Jed561 displayed favourable pharmacokinetic properties, Jed441 was poorly absorbed in the intestines and displayed a relatively short half-life. Accordingly, Jed561 was progressed as the lead compound, whereas Jed441 was used for investigational purposes only. Changes to hepatic mRNA following a single oral dose of Jed561 in C57BL/6J mice supported the results shown *in vitro*, which identified that Jed561 can regulate FXR-specific genes. Jed561 treated mice displayed a sustained upregulation in SHP and concomitant downregulation in CYP7A1 levels. Whilst Jed441 was able to induce changes in SHP and CYP7A1 mRNA, responses were only transient, reflecting the compound's limited bioavailability. Gene expression changes in response to a single oral 2 mg/kg dose of Jed561, compared to a single treatment of OCA at the same concentration, identified potential differences in their gene-regulating mechanisms. Jed561 displays modest changes to SHP mRNA, whereas OCA is more effective at upregulating this gene. On the other hand, Jed561 can increase BSEP levels, but OCA cannot. These disparities between Jed compounds and OCA, were noted in later studies that compared the effects of longer-term compound administration in mice. Specifically, after 5 days of treatment with compounds (administration once daily), SHP mRNA expression was significantly increased in mice treated with 30 mg/kg OCA, more so than mice treated with the same concentration of Jed561. Conversely, 30 mg/kg of Jed561 was able to induce greater increases in BSEP, and greater reductions in CYP7A1, compared to OCA. Both BSEP and CYP7A1 are directly involved in regulating the size of the BA pool. SHP, on the other hand, is an atypical NR, with downstream signalling effects on numerous target proteins. As such, Jed561-mediated FXR activation may be an appropriate selective approach for targeting diseases associated with intrahepatic accumulation of BAs, such as cholestasis, without off-target side effects.

Paradoxically, in the research described here, OCA was associated with effects not previously reported with its use (Dash *et al.*, 2017). Administration of OCA for 5 days at 30 mg/kg resulted in significant upregulation of SREBF1. This lipogenesis-regulating receptor controls the expression of genes involved in fatty acid and triglyceride biosynthesis, and FXR-mediated SHP signalling previously has been associated with its inhibition (Watanabe *et al.*, 2004). However, although OCA displayed significantly increased SHP levels, this response was not accompanied by the downregulation of SREBF1. On the other hand, Jed561 at 30 mg/kg doses, displayed more favourable profiles in lipogenesis genes, with increases in APOC2, responsible for lowering

serum triglyceride levels, and no significant increases in SREBF1. Nevertheless, at 2 mg/kg doses, Jed561 did display increased levels of SREBF1 mRNA, like OCA. The differential responses of Jed561 at different concentrations, highlights the need to extensively investigate the full range of transcriptomic changes in response to this compound, to identify any unanticipated genomic changes associated with its use, and to fine tune an appropriate dose, which can have the desired therapeutic effects whilst minimising adverse reactions. Methods such as RNA-seq, or microarrays allow global transcriptomic changes to be easily profiled. RNA-seq analysis on OCA-treated primary hepatocytes revealed a number of on-target FXR-mediated effects, as well as several novel changes to genes associated with other metabolic pathways (Dash *et al.*, 2017). In addition to identifying transcriptomic changes relevant to liver-specific FXR activation, the analysis also revealed changes to cholesterol synthesis genes, which reinforced the paradigm of abnormal cholesterol levels seen in previous clinical trials with OCA (Neuschwander-Tetri *et al.*, 2015; Dash *et al.*, 2017). By analysing global gene expression or lipidomic profiles in response to Jed561, it may be possible to predict, prior to clinical studies, if this compound, like OCA, will have unfavourable effects on cholesterol synthesis and lipoprotein metabolism.

Jed compounds show efficacy in a NASH model

To further evaluate the therapeutic potential of Jed561, specifically for the amelioration of genes involved in NASH, the compound was tested in a mouse model of the disease. There are over 40 rodent models of NASH that serve as surrogates for the induction of disease-specific physiological changes. Yet, none completely mimic all aspects of the multifactorial disease, and translatability between rodent metabolism and human metabolism are often debated. A model based on feeding mice a high fat ‘western’ diet, followed by chemically inducing fibrosis with CCl₄, first described by Hoshida and colleagues (Hoshida *et al.*, 2018), was expected to deliver the prototypical histological, biochemical and transcriptomic changes associated with human NASH. Unfortunately, in the work described here, the hallmarks of NASH were not achieved, limiting the applicability of the results. However, evidence from the qPCR data identified unique gene expression profiles of OCA, compared to Jed561. QPCR results were corroborated by results from liver marker assays and histological analyses conducted by independent parties that suggested that OCA exacerbated NASH progression. Contrary to previous rodent studies (Fiorucci *et al.*, 2005; Zhang *et al.*, 2017), OCA was associated with the increased expression of fibrogenesis and inflammation genes, TGF β 1, COL3A1, TIMP1, CCL2 and CCR2 in ‘NASH’ model mice. However, this profile was not observed in ‘NASH’ mice treated with Jed561, which did not significantly upregulate any fibrosis gene at any concentration used, and instead, displayed downregulation in the fibrosis marker gene ACTA2. The differential transcriptomic effects

between the two compounds also were translated to a cellular and tissue level, whereby results from independent blinded histological analyses of these mice liver sections, confirmed worsening of the steatosis, inflammatory cell infiltration, vacuolation and fibrosis. The extent of fibrosis is considered to be a major determinant of cardiovascular co-morbidity and mortality in NASH (Angulo *et al.*, 2015), and so the exacerbation of fibrosis markers seen with OCA can have significant impacts on its application as a NASH treatment. Jed561, on the other hand, with its ability to decrease ACTA2 levels, still presents the potential for use as a NASH drug.

Despite the inadequate induction of fibrosis, the NASH mice in this study displayed signs of altered lipid metabolism, with vehicle-treated mice displaying significantly upregulated SREBF1 and decreased PLTP expression. ‘NASH’ mice treated with OCA displayed slightly normalised expression of these genes and profiles were even more improved upon treatment with Jed561.

These results further substantiate the notion that, despite being derived from the same CDCA scaffold, Jed561 and OCA operate by different mechanisms of action. Jed561 is proposed to interact with many of the canonical residues in the FXR LBD, similarly to OCA. However, docking solutions suggest that it can make several novel interactions, and its bulky side chain may partially occupy the secondary binding pocket, which may have specific modulatory effects on FXR activation. Jed561-occupied FXR also was seen to have a higher affinity for SRC2-2 than other coactivator peptides tested. By inducing a novel FXR LBD conformation and exerting specific post-translational effects through its preferential binding of certain coactivators, it is indeed possible that Jed561 may be able to regulate different FXR-target genes distinctively to OCA. Transcriptomic analysis of a perfused organotypic hepatocyte system, or even organ specific FXR knock out mice, will help to build a complete picture of Jed561-mediated actions, allowing profiles to be compared with OCA-driven changes. This will help clarify the possibility of differential gene regulation, due to compartmentalisation and differential bioavailability of these two compounds.

Jed561 as a therapeutic compound

Jed561 holds the capacity for use as a therapeutic agent. An ideal NASH drug should reduce steatosis, hepatic inflammation and liver cell injury, and have antifibrotic effects, whilst improving underlying metabolic dysfunctions (Hansen *et al.*, 2017). Although insufficiently validated for treatment of fibrosis and inflammation in this work, Jed561 shows promising results in correcting lipid metabolism pathways. The NASH field is continuously, rapidly changing, and due to the lack of agents that can target the multiple aspects of NASH injury and disease progression, pharmaceutical considerations now include the use of combinatorial

therapies. FXR agonists, such as Tropifexor, are presently being trialled in combination with apoptosis signalling kinase inhibitors and inflammatory cytokine inhibitors, but FXR agonists are still considered core components of therapy (Carter and Khan, 2019). The use of amphiphilic, non-toxic, BA-based FXR agonists, provides the added advantage of aiding drug delivery and absorption of concomitantly administered drugs. Therefore, while the anti-fibrotic and anti-inflammatory effects of Jed561 still need to be evaluated in another NASH model, in the meantime, its anti-steatotic effects could be assessed in a combinatorial strategy with other prominent drug candidates targeted towards additional NASH pathophysiologies. In addition, results shown in this thesis have identified a potential use for Jed561 in cholestatic diseases such as PBC. OCA was approved by the FDA, in 2016, for use in the treatment of chronic PBC, and it is currently undergoing evaluation for its long term safety and effects in a phase 3 double blind study with PBC patients (Trauner *et al.*, 2019). Jed561, however, may be a more appropriate drug candidate, due to its improved responses in genes regulating the BA pool, and potential limited off-target effects. Furthermore, if selective modulation of FXR can be achieved with Jed561, it poses the capacity for use in treatment targeting metabolic syndromes or cardiometabolic diseases.

Jed441 as a therapeutic compound

Despite being restricted to the intestines, Jed441 may still be able to affect lipid metabolism, insulin sensitivity and weight management, via intestine-specific FXR signalling cascades and actions of the FGF19 hormone. While the therapeutic benefits of intestinal FXR agonism and antagonism are poorly understood and remain controversial (Albert *et al.*, 2014; Fang *et al.*, 2015), further studies into Jed441-mediated transcriptomic changes, may help evaluate the therapeutic potential of this compound, and could lead to its progression in the search for drugs targeting other FXR-related pathologies, such as inflammatory bowel disease (Wildenberg and van den Brink, 2011), or intestinal ischaemia reperfusion injury (Wang *et al.*, 2018).

Conclusion

The field of FXR-mediated biology is extremely complex, and the extent of FXR-mediated signalling is still being uncovered. By acting as a ligand-activated transcription factor, FXR can regulate a plethora of target genes in diverse networks of signalling pathways. Furthermore, the discovery of DNA-independent mechanisms of FXR signalling, emphasizes the vast scope of its regulation. Its role in multiple important physiological pathways, is both a blessing and a curse when considering the therapeutic potential of this receptor, and an ideal strategy targets specific subsets of FXR-responsive genes, without systemic FXR activation. To achieve this,

efforts are now shifting towards allosteric modulation of the FXR receptor and the multiple loci in which its actions can be regulated, for example, interactions with its partner receptor RXR, interactions with different coactivators, and interactions with different DNA response elements (Rastinejad *et al.*, 2013). This is thought to limit NR desensitization (Veras *et al.*, 2019), however, this approach is challenging due to insufficient data regarding FXR DBD and its full length, native structure, and the ability to selectively target specific DNA response elements still poses potential problems.

Although inter-domain modulatory effects are appreciated, efforts herein focused on the interactions between FXR and its ligand, and potential selective recruitment of coactivators. The combination of computational, structural and molecular biology techniques applied in this thesis, allowed a deeper understanding of the FXR activation and dynamics to be attained and made significant contributions toward the development of novel, BA-derived agonists of FXR. Candidates, Jed441 and Jed561, may both selectively regulate FXR, albeit due to different underlying mechanisms, whether limited bioavailability and compartmentalisation of the compound, or due to ligand specific recruitment of coactivators. The selective modulation and increased potency of these novel BA compounds make them equally, if not more, attractive than the 'gold standard' FXR agonist OCA for use as a new pharmaceutical agent.

8. References

Abdelkader, N. F., Safar, M. M. and Salem, H. A. (2016) 'Ursodeoxycholic Acid Ameliorates Apoptotic Cascade in the Rotenone Model of Parkinson's Disease: Modulation of Mitochondrial Perturbations', *Molecular Neurobiology*, 53(2), pp. 810–817. doi: 10.1007/s12035-014-9043-8.

Abràmoff, M. D., Hospitals, I., Magalhães, P. J. and Abràmoff, M. (2007) 'Image Processing with ImageJ', *Biophotonics international*, 11(7), pp. 36–42.

Acharya, C., Coop, A., Polli, J. and MacKerell Jr., A. (2011) 'Recent Advances in Ligand-Based Drug Design: Relevance and Utility of the Conformationally Sampled Pharmacophore Approach', *Current Computer Aided Drug Design*, 7(1), pp. 10–22. doi: 10.1016/j.pestbp.2011.02.012.Investigations.

Adorini, L. (2013) 'The Development of Drugs Interacting with Bile Acid Receptors', *Annual meeting of the European Association for the Study of the Liver (EASL)*, Amsterdam, The Netherlands, 24-28 April 2016

De Aguiar Vallim, T. Q., Tarling, E. J. and Edwards, P. A. (2013) 'Pleiotropic Roles of Bile Acids in Metabolism', *Cell Metabolism*, 17(5), pp. 657–669. doi: 10.1016/j.cmet.2013.03.013.Pleiotropic.

Akwabi-Ameyaw, A., Bass, J., Caldwell, R., Caravella, J., Chen, L., Creech, K., Deaton, D., Jones, S., Kaldor, I., Liu, Y., Madauss, K., Marr, H., McFayden, R., Miler, A., Navas, F., Parks, D., Spearing, P., Todd, D., Williams, S. and Wisely, G. B. (2008) 'Conformationally constrained farnesoid X receptor (FXR) agonists: Naphthoic acid-based analogs of GW 4064', *Bioorganic and Medicinal Chemistry Letters*, 18(15), pp. 4339–4343. doi: 10.1016/j.bmcl.2008.06.073.

Akwabi-Ameyaw, A., Caravella, J., Chen, L., Creech, K., Deaton, D., Madauss, K., Marr, H., Miller, A., Navas, F., Parks, D., Spearing, P., Todd, D., Williams, S. and Wisely, G. B. (2011) 'Conformationally constrained farnesoid X receptor (FXR) agonists: Alternative replacements of the stilbene', *Bioorganic and Medicinal Chemistry Letters*. Elsevier Ltd, 21(20), pp. 6154–6160. doi: 10.1016/j.bmcl.2011.08.034.

Albert, I., Tanaka, N., Li, F., Xie, C., Krausz, K. W., Amin, S. G., Nichols, R. G., Patterson, A. D., Zhang, L., Takahashi, S., Cai, J., Jiang, C., Desai, D., Fang, Z.-Z., Qi, Y. and Gonzalez, F. J. (2014) 'Intestinal farnesoid X receptor signaling promotes nonalcoholic fatty liver disease', *Journal of*

Clinical Investigation, 125(1). doi: 10.1172/jci76738.

Alemi, F., Kwon, E., Poole, D. and Lieu, T. (2013) 'The TGR5 receptor mediates bile acid – induced itch and analgesia', *The Journal of Clinical Investigation*, 123(4), pp. 1513–1530. doi: 10.1172/JCI64551.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W. and Lipman, D. J. (1990) 'Basic local alignment search tool', *Journal of Molecular Biology*, 215(3), pp. 403–410. doi: 10.1016/S0022-2836(05)80360-2.

Ananthanarayanan, M., Balasubramanian, N., Makishima, M., Mangelsdorf, D. J. and Suchy, F. J. (2001) 'Human Bile Salt Export Pump Promoter Is Transactivated by the Farnesoid X Receptor/Bile Acid Receptor', *Journal of Biological Chemistry*, 276(31), pp. 28857–28865. doi: 10.1074/jbc.M011610200.

Andersen, K. R., Leska, N. C. and Schwartz, T. U. (2013) 'Optimized E.coli expression strain LOBSTR eliminates common contaminants from His-tag purification', *Proteins*, 81(11), pp. 1857–1861. doi: 10.1002/prot.24364.

Anderson, A. C. (2003) 'The Process of Structure-Based Drug Design', *Chemistry & Biology*, 10, pp. 787–797. doi: 10.1016/j.

Angulo, P., Kleiner, D. E., Dam-Larsen, S., Adams, L. A., Bjornsson, E. S., Charatcharoenwitthaya, P., Mills, P. R., Keach, J. C., Lafferty, H. D., Stahler, A., Haflidadottir, S. and Bendtsen, F. (2015) 'Liver Fibrosis, but No Other Histologic Features, Is Associated With Long-term Outcomes of Patients With Nonalcoholic Fatty Liver Disease', *Gastroenterology*, 149(2), pp. 389-397.e10. doi: <https://doi.org/10.1053/j.gastro.2015.04.043>.

Aranda, A. and Pascual, A. (2017) 'Nuclear Hormone Receptors and Gene Expression', *Physiological Reviews*, 81(3), pp. 1269–1304. doi: 10.1152/physrev.2001.81.3.1269.

Armstrong, L. E. and Guo, G. L. (2017) 'Role of FXR in Liver Inflammation During Nonalcoholic Steatohepatitis', *Current Pharmacology Reports*, 3(2), pp. 92–100. doi: 10.1007/s40495-017-0085-2.

Bachs, L., Pares, A., Elena, M., Piera, C. and Rodes, J. (1992) 'Effects of long-term rifampicin administration in primary biliary cirrhosis', *Gastroenterology*, 102(6), pp. 2077–2080.

Ballatori, N., Li, N., Fang, F., Boyer, J. L., Christian, W. V and Hammond, C. L. (2009) 'OST alpha-OST beta: a key membrane transporter of bile acids and conjugated steroids.', *Frontiers in bioscience (Landmark edition)*, 14, pp. 2829–44. Available at:

<http://www.ncbi.nlm.nih.gov/pubmed/19273238> <http://www.ncbi.nlm.nih.gov/entrez/fcgi?artid=PMC2694352>.

Barbier, O., Torra, I. P., Sirvent, A., Claudel, T., Blanquart, C., Duran-Sandoval, D., Kuipers, F., Kosykh, V., Fruchart, J. C. and Staels, B. (2003) 'FXR induces the UGT2B4 enzyme in hepatocytes: A potential mechanism of negative feedback control of FXR activity', *Gastroenterology*, 124(7), pp. 1926–1940. doi: 10.1016/S0016-5085(03)00388-3.

Baroni, M., Cruciani, G., Sciabola, S., Perruccio, F. and Mason, J. S. (2007) 'A common reference framework for analyzing/comparing proteins and ligands. Fingerprints for Ligands and Proteins (FLAP): Theory and application', *Journal of Chemical Information and Modeling*, 47(2), pp. 279–294. doi: 10.1021/ci600253e.

Bass, J. Y., Caravella, J. A., Chen, L., Creech, K. L., Deaton, D. N., Madauss, K. P., Marr, H. B., McFadyen, R. B., Miller, A. B., Mills, W. Y., Navas, F., Parks, D. J., Smalley, T. L., Spearing, P. K., Todd, D., Williams, S. P. and Wisely, G. B. (2011) 'Conformationally constrained farnesoid X receptor (FXR) agonists: Heteroaryl replacements of the naphthalene', *Bioorganic and Medicinal Chemistry Letters*. Elsevier Ltd, 21(4), pp. 1206–1213. doi: 10.1016/j.bmcl.2010.12.089.

Begley, M., Gahan, C. G. M. and Hill, C. (2005) 'The interaction between bacteria and bile', *FEMS Microbiology Reviews*, 29(4), pp. 625–651. doi: 10.1016/j.femsre.2004.09.003.

Bellentani, S. (2017) 'The epidemiology of non-alcoholic fatty liver disease', *Liver International*, 37(October 2016), pp. 81–84. doi: 10.1111/liv.13299.

Berg, J., Tymoczko, J. and Stryer, L. (2002) 'Three-Dimensional Protein Structure Can Be Determined by NMR Spectroscopy and X-Ray Crystallography', in *Biochemistry*. 5th edn. New York: W H Freeman.

Bergfors, T. (2003) 'Seeds to crystals', *Journal of Structural Biology*, 142(1), pp. 66–76. doi: 10.1016/S1047-8477(03)00039-X.

Bergfors, T. (2009) *Protein Crystallization*. 2nd edn. Edited by T. Bergfors. La Jolla, CA, USA: International University Line.

Berrow, N., Alderton, D. and Owens, R. J. (2009) 'The Precise Engineering of Expression Vectors Using High-Throughput In-Fusion™ PCR Cloning', in Doyle, S. (ed.) *High Throughput Protein Expression and Purification*. Humana Press.

Beuers, U. and Boyer, J. (1994) 'Bile: A historical review of studies on its form and function', in

Kirsner, J. . (ed.) *Gastroenterology in the 20th Century*. New York: Lea and Ferbiger, pp. 267–288.

Bevan, C. L., Hoare, S., Claessens, F., Heery, D. M. and Parker, M. G. (1999) 'The AF1 and AF2 domains of the androgen receptor interact with distinct regions of SRC1', *Molecular and cellular biology*. American Society for Microbiology, 19(12), pp. 8383–8392. doi: 10.1128/mcb.19.12.8383.

Bhakat, S., Åberg, E. and Söderhjelm, P. (2018) 'Prediction of binding poses to FXR using multi-targeted docking combined with molecular dynamics and enhanced sampling', *Journal of Computer-Aided Molecular Design*, 32(1), pp. 59–73. doi: 10.1007/s10822-017-0074-x.

Billas, I. M., Moulinier, L., Rochel, N. and Moras, D. (2001) 'Crystal Structure of the Ligand-binding Domain of the Ultraspiracle Protein USP, the ortholog of Retinoid X Receptor in Insects', *Journal of Biological Chemistry*, 276, pp. 7465–7474.

Bishop-Bailey, D., Walsh, D. T. and Warner, T. D. (2004) 'Expression and activation of the farnesoid X receptor in the vasculature.', *Proceedings of the National Academy of Sciences of the United States of America*, 101(10), pp. 3668–73. doi: 10.1073/pnas.0400046101.

Bolanos-Garcia, V. M. and Davies, O. R. (2006) 'Structural analysis and classification of native proteins from *E. coli* commonly co-purified by immobilised metal affinity chromatography', *Biochimica et Biophysica Acta - General Subjects*, 1760(9), pp. 1304–1313. doi: 10.1016/j.bbagen.2006.03.027.

Bonthuis, P. J., Huang, W.-C., Stacher Hörndl, C. N., Ferris, E., Cheng, T. and Gregg, C. (2015) 'Noncanonical Genomic Imprinting Effects in Offspring', *Cell reports*, 12(6), p. 979–991. doi: 10.1016/j.celrep.2015.07.017.

Boom, R., Sol, C., SALIMANS, M. M. M., Jansen, C. L., Wertheim-van Dillen, P. and van der Noordaa, J. (1990) 'Rapid and Simple Method for Purification of Nucleic Acids', *Journal of Clinical Microbiology*, 28(3), pp. 495–503. doi: 10.2116/analsci.25.941.

Boulias, K., Katrakili, N., Bamberg, K., Underhill, P., Greenfield, A. and Talianidis, I. (2005) 'Regulation of hepatic metabolic pathways by the orphan nuclear receptor SHP', *EMBO Journal*, 24(14), pp. 2624–2633. doi: 10.1038/sj.emboj.7600728.

Bragg, W. . and Bragg, W. . (1913) 'The Reflexion of X-rays by Crystals', in *Proc. R. Soc. Lond. A*, pp. 428–438.

Bramlett, K. S., Yao, S. and Burris, T. P. (2000) 'Correlation of farnesoid X receptor coactivator

recruitment and cholesterol 7 α -hydroxylase gene repression by bile acids', *Molecular Genetics and Metabolism*, 71(4), pp. 609–615. doi: 10.1006/mgme.2000.3106.

Bruce, K. D., Sihota, K. K., Byrne, C. D. and Cagampang, F. R. (2012) 'The housekeeping gene YWHAZ remains stable in a model of developmentally primed non-alcoholic fatty liver disease', *Liver International*, 32(8), pp. 1315–1321. doi: 10.1111/j.1478-3231.2012.02813.x.

Bustin, S. and Huggett, J. (2017) 'qPCR primer design revisited', *Biomolecular Detection and Quantification*. Elsevier, 14(March), pp. 19–28. doi: 10.1016/j.bdq.2017.11.001.

Buzzetti, E., Pinzani, M. and Tsochatzis, E. A. (2016) 'The multiple-hit pathogenesis of non-alcoholic fatty liver disease (NAFLD)', *Metabolism: Clinical and Experimental*. Elsevier Inc., 65(8), pp. 1038–1048. doi: 10.1016/j.metabol.2015.12.012.

Cai, M., Huang, Y., Sakaguchi, K., Clore, G. M., Gronenborn, A. M. and Craigie, R. (1998) 'An efficient and cost-effective isotope labeling protocol for proteins expressed in Escherichia coli', *Journal of Biomolecular NMR*, 11(1), pp. 97–102. doi: 10.1023/A:1008222131470.

Cariou, B., Charbonnel, B. and Staels, B. (2012) 'Thiazolidinediones and PPAR γ agonists: Time for a reassessment', *Trends in Endocrinology and Metabolism*, 23(5), pp. 205–215. doi: 10.1016/j.tem.2012.03.001.

Cariou, B., Van Harmelen, K., Duran-Sandoval, D., Van Dijk, T. H., Grefhorst, A., Abdelkarim, M., Caron, S., Torpier, G., Fruchart, J. C., Gonzalez, F. J., Kuipers, F. and Staels, B. (2006) 'The farnesoid X receptor modulates adiposity and peripheral insulin sensitivity in mice', *Journal of Biological Chemistry*, 281(16), pp. 11039–11049. doi: 10.1074/jbc.M510258200.

Cariou, B. and Staels, B. (2007) 'FXR : a promising target for the metabolic syndrome ?', *Trends in Pharmacological Sciences*, 28(5), pp. 236–243. doi: 10.1016/j.tips.2007.03.002.

Carter, M. and Khan, S. (2019) 'Novel–novel fixed-dose combination therapies', *Nature Reviews Drug Discovery*. Springer US, 18(6), p. 413. doi: 10.1038/d41573-019-00066-z.

Cave, M. C., Clair, H. B., Hardesty, J. E., Falkner, K. C., Feng, W., Clark, B. J., Sidey, J., Shi, H., Aqel, B. A., McClain, C. J. and Prough, R. A. (2016) 'Biochimica et Biophysica Acta Nuclear receptors and nonalcoholic fatty liver disease 1', *BBA - Gene Regulatory Mechanisms*. Elsevier B.V., 1859(9), pp. 1083–1099. doi: 10.1016/j.bbagr.2016.03.002.

Chandra, V., Huang, P., Hamuro, Y., Raghuram, S., Wang, Y., Burris, T. P. and Rastinejad, F. (2008) 'Structure of the intact PPAR- γ –RXR- α nuclear receptor complex on DNA', *Nature*. Macmillan Publishers Limited. All rights reserved, 456, p. 350. Available at:

[https://doi.org/10.1038/nature07413.](https://doi.org/10.1038/nature07413)

Chandra, V., Huang, P., Potluri, N., Wu, D., Kim, Y. and Rastinejad, F. (2013) 'Multidomain integration in the structure of the HNF-4 α nuclear receptor complex', *Nature*. Nature Publishing Group, a division of Macmillan Publishers Limited. All Rights Reserved., 495, p. 394. Available at: <https://doi.org/10.1038/nature11966>.

Chavez-Tapia, N. C., Rosso, N. and Tiribelli, C. (2011) 'In vitro models for the study of non-alcoholic fatty liver disease.', *Current medicinal chemistry*, 18(7), pp. 1079–84. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/21254970>.

Chiang, J. Y. L. (2013) 'Bile Acid Metabolism and Signalling', *Comprehensive Physiology*, 3(3), pp. 1191–1212. doi: 10.1002/cphy.c120023.Bile.

Choi, J. H., Banks, A. S., Estall, J. L., Kajimura, S., Boström, P., Laznik, D., Ruas, J. L., Chalmers, M. J., Kamenecka, T. M., Blüher, M., Griffin, P. R. and Spiegelman, B. M. (2010) 'Anti-diabetic drugs inhibit obesity-linked phosphorylation of PPARgamma by Cdk5', *Nature*, 466(7305), pp. 451–456. doi: 10.1038/nature09291.

Chung, H. T. (2013) 'RIP140, a Janus metabolic switch involved in defense functions', *Cellular and Molecular Immunology*. Nature Publishing Group, 10(1), pp. 7–9. doi: 10.1038/cmi.2012.53.

Cooper, D. R., Boczek, T., Grelewska, K., Pinkowska, M., Sikorska, M., Zawadzki, M. and Derewenda, Z. (2007) 'Protein crystallization by surface entropy reduction: Optimization of the SER strategy', *Acta Crystallographica Section D: Biological Crystallography*, 63(5), pp. 636–645. doi: 10.1107/S0907444907010931.

Copple, B. L. and Li, T. (2016) 'Pharmacology of bile acid receptors: Evolution of bile acids from simple detergents to complex signaling molecules', *Pharmacological Research*. Elsevier Ltd, 104, pp. 9–21. doi: 10.1016/j.phrs.2015.12.007.

Corton, J. C., Bushel, P. R., Fostel, J. and O'Lone, R. B. (2012) 'Sources of variance in baseline gene expression in the rodent liver', *Mutation research*. 2012/01/05, 746(2), pp. 104–112. doi: 10.1016/j.mrgentox.2011.12.017.

Costantino, G., Entrena-Guadix, A., Macchiarulo, A., Gioiello, A. and Pellicciari, R. (2005) 'Molecular dynamics simulation of the ligand binding domain of farnesoid X receptor. Insights into helix-12 stability and coactivator peptide stabilization in response to agonist binding', *Journal of Medicinal Chemistry*, 48(9), pp. 3251–3259. doi: 10.1021/jm049182o.

Cross, S. and Cruciani, G. (2010) 'Molecular fields in drug discovery : getting old or reaching maturity ?', 15(January), pp. 23–32. doi: 10.1016/j.drudis.2008.12.006.

Crouse, J. R. (1987) 'Hypertriglyceridemia: A contraindication to the use of bile acid binding resins', *The American Journal of Medicine*, 83(2), pp. 243–248.

Cruciani, G., Crivori, P., Carrupt, P.-A. and Testa, B. (2000) 'Molecular fields in quantitative structure–permeation relationships: the VolSurf approach', *Journal of Molecular Structure: THEOCHEM*, 503(1), pp. 17–30. doi: [https://doi.org/10.1016/S0166-1280\(99\)00360-7](https://doi.org/10.1016/S0166-1280(99)00360-7).

Cui, J., Huang, L., Zhao, A., Lew, J. L., Yu, J., Sahoo, S., Meinke, P. T., Royo, I., Peláez, F. and Wright, S. D. (2003) 'Guggulsterone is a farnesoid X receptor antagonist in coactivator association assays but acts to enhance transcription of bile salt export pump', *Journal of Biological Chemistry*, 278(12), pp. 10214–10220. doi: 10.1074/jbc.M209323200.

Dai, S. Y., Chalmers, M. J., Bruning, J., Bramlett, K. S., Osborne, H. E., Montrose-Rafizadeh, C., Barr, R. J., Wang, Y., Wang, M., Burris, T. P., Dodge, J. A. and Griffin, P. R. (2008) 'Prediction of the tissue-specificity of selective estrogen receptor modulators by using a single biochemical method', *Proceedings of the National Academy of Sciences of the United States of America*, 105(20), pp. 7171–7176. doi: 10.1073/pnas.0710802105.

Danzinger, R. G., Hofmann, A. F., Schoenfield, L. J. and Thistle, J. L. (1972) 'Dissolution of Cholesterol Gallstones by Chenodeoxycholic Acid', *New England Journal of Medicine*, 286, pp. 1–8.

Darimont, B. D., Wagner, R. L., Apriletti, J. W., Stallcup, M. R., Kushner, P. J., Baxter, J. D., Fletterick, R. J. and Yamamoto, K. R. (1998) 'Structure and specificity of nuclear receptor – coactivator interactions Structure and specificity of nuclear receptor – coactivator interactions', *Genes and Development*, 12, pp. 3343–3356. doi: 10.1101/gad.12.21.3343.

Dash, A., Figler, R. A., Blackman, B. R., Marukian, S., Collado, M. S., Lawson, M. J., Hoang, S. A., Mackey, A. J., Manka, D., Cole, B. K., Feaver, R. E., Sanyal, A. J. and Wamhoff, B. R. (2017) 'Pharmacotoxicology of clinically-relevant concentrations of obeticholic acid in an organotypic human hepatocyte system', *Toxicology in Vitro*. Elsevier B.V., 39, pp. 93–103. doi: 10.1016/j.tiv.2016.11.014.

Dauter, Z. and Wlodawer, A. (2016) 'Progress in protein crystallography', *Protein Peptide Letters*, 23(3), pp. 201–210. doi: 10.1126/science.1249098.Sleep.

Dearden, J. C., Cronin, M. T. D. and Kaiser, K. L. E. (2009a) 'How not to develop a quantitative

structure–activity or structure–property relationship (QSAR/QSPR)’, *SAR and QSAR in Environmental Research*. Taylor & Francis, 20(3–4), pp. 241–266. doi: 10.1080/10629360902949567.

Dessau, M. A. and Modis, Y. (2011) ‘Protein Crystallization for X-ray Crystallography’, *Journal of Visualized Experiments*, 9(47), pp. 1–6. doi: 10.3791/2285.

Downes, M., Verdecia, M. A., Roecker, A. J., Hughes, R., Hogenesch, J. B., Kast-Woelbern, H. R., Bowman, M. E., Ferrer, J.-L., Anisfeld, A. M. and Edwards, P. A. (2003) ‘A Chemical, Genetic, and Structural Analysis of the Nuclear Bile Acid Receptor FXR’, *Molecular Cell*, 11(4), pp. 1079–1092. doi: 10.1016/S1097-2765(03)00104-7.

Durand, B., Saunders, M., Gaudon, C., Roy, B., Losson, R. and Chambon, P. (1994) ‘Activation function 2 (AF-2) of retinoic acid receptor and 9-cis retinoic acid receptor: presence of a conserved autonomous constitutive activating domain and influence of the nature of the response element on AF-2 activity’, *The EMBO journal*, 13(22), pp. 5370–5382. Available at: <https://www.ncbi.nlm.nih.gov/pubmed/7957103>.

Durrant, J. D. and McCammon, J. A. (2010) ‘Computer-aided drug-discovery techniques that account for receptor flexibility’, *Current Opinion in Pharmacology*. Elsevier Ltd, 10(6), pp. 770–774. doi: 10.1016/j.coph.2010.09.001.

Dussault, I., Beard, R., Lin, M., Hollister, K., Chen, J., Xiao, J. H., Chandraratna, R. and Foman, B. M. (2003) ‘Identification of gene-selective modulators of the bile acid receptor FXR’, *Journal of Biological Chemistry*, 278(9), pp. 7027–7033. doi: 10.1074/jbc.M209863200.

Dyson, J. K., Anstee, Q. M. and McPherson, S. (2014) ‘Non-alcoholic fatty liver disease: a practical approach to diagnosis and staging.’, *Frontline gastroenterology*, 5(3), pp. 211–218. doi: 10.1136/flgastro-2013-100403.

Egelman, E. H. (2016) ‘The Current Revolution in Cryo-EM’, *Biophysical Journal*. Biophysical Society, 110(5), pp. 1008–1012. doi: 10.1016/j.bpj.2016.02.001.

Esterhuysen, C., Heßelmann, A. and Clark, T. (2017) ‘Trifluoromethyl: An Amphiphilic Noncovalent Bonding Partner’, *ChemPhysChem*, 18(7), pp. 772–784. doi: 10.1002/cphc.201700027.

Fang, S., Suh, J. M., Reilly, Shannon M., et al. (2015) ‘Intestinal FXR agonism promotes adipose tissue browning and reduces obesity and insulin resistance’, *Nature Medicine*. Nature Publishing Group, 21(2), pp. 159–165. doi: 10.1038/nm.3760.

Feixas, F., Lindert, S., Sinko, W. and Mccammon, J. A. (2014a) 'Biophysical Chemistry Exploring the role of receptor flexibility in structure-based drug discovery', *Biophysical Chemistry*.

Elsevier B.V., 186, pp. 31–45. doi: 10.1016/j.bpc.2013.10.007.

Feng, W., Ribeiro, R. C. J., Wagner, R. L., Nguyen, H., Apriletti, J. W., Fletterick, R. J., Baxter, J. D., Kushner, P. J. and West, B. L. (1998) 'Hormone-Dependent Coactivator Binding to a Hydrophobic Cleft on Nuclear Receptors', *Science*, 280(5370), pp. 1747 LP – 1749. doi: 10.1126/science.280.5370.1747.

Finn, R. D., Coggill, P., Eberhardt, R. Y., Eddy, S. R., Mistry, J., Mitchell, A. L., Potter, S. C., Punta, M., Qureshi, M., Sangrador-Vegas, A., Salazar, G. A., Tate, J. and Bateman, A. (2016) 'The Pfam protein families database: Towards a more sustainable future', *Nucleic Acids Research*, 44(D1), pp. D279–D285. doi: 10.1093/nar/gkv1344.

Fiorucci, S., Clerici, C., Antonelli, E., Orlandi, S., Goodwin, B., Sadeghpour, B. M., Sabatino, G., Russo, G., Castellani, D., Willson, T. M., Pruzanski, M., Pellicciari, R. and Morelli, A. (2005) 'Protective Effects of 6-Ethyl Chenodeoxycholic Acid, a Farnesoid X Receptor Ligand, in Estrogen-Induced Cholestasis', *Journal of Pharmacology and Experimental Therapeutics*, 313(2), pp. 604–612. doi: 10.1124/jpet.104.079665.

Fiorucci, S., Rizzo, G., Antonelli, E., Renga, B., Mencarelli, A., Riccardi, L., Orlandi, S., Pruzanski, M., Morelli, A. and Pellicciari, R. (2005) 'A farnesoid x receptor-small heterodimer partner regulatory cascade modulates tissue metalloproteinase inhibitor-1 and matrix metalloprotease expression in hepatic stellate cells and promotes resolution of liver fibrosis.', *The Journal of pharmacology and experimental therapeutics*, 314(2), pp. 584–595. doi: 10.1124/jpet.105.084905.

Fiorucci, S., Rizzo, G., Donini, A., Distrutti, E. and Santucci, L. (2007) 'Targeting farnesoid X receptor for liver and metabolic disorders', *Trends in Molecular Medicine*, 13(7), pp. 298–309. doi: 10.1016/j.molmed.2007.06.001.

Flesch, D., Cheung, S.-Y., Schmidt, J., Gabler, M., Heitel, P., Kramer, J., Kaiser, A., Hartmann, M., Lindner, M., Lüddens-Dämgen, K., Heering, J., Lamers, C., Lüddens, H., Wurglits, M., Proschak, E., Schubert-Zsilavecz, M. and Merk, D. (2017) 'Nonacidic Farnesoid X Receptor Modulators', *Journal of Medicinal Chemistry*. American Chemical Society, 60(16), pp. 7199–7205. doi: 10.1021/acs.jmedchem.7b00903.

Forman, B. M., Goode, E., Chen, J., Oro, A. E., Bradley, D. J., Perlmann, T., Noonan, D. J., Burka, L. T., McMorris, T., Lamph, W. W., Evans, R. M. and Weinberger, C. (1995) 'Identification of a

nuclear receptor that is activated by farnesol metabolites', *Cell*, 81(5), pp. 687–693. doi: 10.1016/0092-8674(95)90530-8.

Fritz, T. A., Tondi, D., Finer-Moore, J. S., Costi, M. P. and Stroud, R. M. (2001) 'Predicting and harnessing protein flexibility in the design of species-specific inhibitors of thymidylate synthase1, 21Escherichia coli thymidylate synthase numbering is used unless otherwise noted.2PDB coordinates have been deposited with the RCSB with a', *Chemistry & Biology*, 8(10), pp. 981–995. doi: [https://doi.org/10.1016/S1074-5521\(01\)00067-9](https://doi.org/10.1016/S1074-5521(01)00067-9).

Fu, J., Si, P., Zheng, M., Chen, L., Shen, X., Tang, Y. and Li, W. (2012) 'Bioorganic & Medicinal Chemistry Letters Discovery of new non-steroidal FXR ligands via a virtual screening workflow based on Phase shape and induced fit docking', *Bioorganic & Medicinal Chemistry Letters*. Elsevier Ltd, 22(22), pp. 6848–6853. doi: 10.1016/j.bmcl.2012.09.045.

Gaieb, Z., Liu, S., Gathiaka, S., Chiu, M., Yang, H., Shao, C., Feher, V. A., Walters, W. P., Kuhn, B., Rudolph, M. G., Burley, S. K., Gilson, M. K. and Amaro, R. E. (2018) 'D3R Grand Challenge 2: blind prediction of protein–ligand poses, affinity rankings, and relative binding free energies', *Journal of Computer-Aided Molecular Design*. Springer International Publishing, 32(1), pp. 1–20. doi: 10.1007/s10822-017-0088-4.

Gascon-Barré, M., Demers, C., Mirshahi, A., Néron, S., Zalzal, S. and Nanci, A. (2003) 'The normal liver harbors the vitamin D nuclear receptor in nonparenchymal and biliary epithelial cells', *Hepatology*, 37(5), pp. 1034–1042. doi: 10.1053/jhep.2003.50176.

Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M. R., Appel, R. D. and Bairoch, A. (2005) 'Protein Identification and Analysis Tools on the ExPASy Server', in Walker, J. M. (ed.) *The Proteomics Protocols Handbook*. Totowa, NJ: Humana Press, pp. 571–607.

Gautier, T., De Haan, W., Grober, J., Ye, D., Bahr, M. J., Claudel, T., Nijstad, N., Van Berkel, T. J. C., Havekes, L. M., Manns, M. P., Willems, S. M., Hogendoorn, P. C. W., Lagrost, L., Kuipers, F., Van Eck, M., Rensen, P. C. N. and Tietge, U. J. F. (2013) 'Farnesoid X receptor activation increases cholesteryl ester transfer protein expression in humans and transgenic mice', *Journal of Lipid Research*, 54(8), pp. 2195–2205. doi: 10.1194/jlr.M038141.

Gertzen, C. G. W., Spomer, L., Smits, S. H. J., Haüssinger, D., Keitel, V. and Gohlke, H. (2015) 'Mutational mapping of the transmembrane binding site of the G-protein coupled receptor TGR5 and binding mode prediction of TGR5 agonists', *European Journal of Medicinal Chemistry*, 104, pp. 57–72. doi: 10.1016/j.ejmech.2015.09.024.

Gillis, E. P., Eastman, K. J., Hill, M. D., Donnelly, D. J. and Meanwell, N. A. (2015) 'Applications of Fluorine in Medicinal Chemistry', *Journal of Medicinal Chemistry*, 58(21), pp. 8315–8359. doi: 10.1021/acs.jmedchem.5b00258.

Ginzinger, D. G. (2002) 'Gene quantification using real-time quantitative PCR: an emerging technology hits the mainstream.', *Experimental hematology*, 30(6), pp. 503–12. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/12063017>.

Giulietti, A., Overbergh, L., Valckx, D., Decallonne, B., Bouillon, R. and Mathieu, C. (2001) 'An overview of real-time quantitative PCR: Applications to quantify cytokine gene expression', *Methods*, 25(4), pp. 386–401. doi: 10.1006/meth.2001.1261.

Gonzalez, F. J., Jiang, C. and Patterson, A. D. (2016) 'An Intestinal Microbiota–Farnesoid X Receptor Axis Modulates Metabolic Disease', *Gastroenterology*. Elsevier, Inc, 151(5), pp. 845–859. doi: 10.1053/j.gastro.2016.08.057.

Goodford, P. J. (1985) 'A computational procedure for determining energetically favorable binding sites on biologically important macromolecules', *Journal of Medicinal Chemistry*. American Chemical Society, 28(7), pp. 849–857. doi: 10.1021/jm00145a002.

Goodwin, B., Watson, M. A., Kim, H., Miao, J., Kemper, J. K. and Kliewer, S. A. (2003) 'Differential Regulation of Rat and Human CYP7A1 by the Nuclear Oxysterol Receptor Liver X Receptor- α ', *Molecular Endocrinology*, 17(3), pp. 386–394. doi: 10.1210/me.2002-0246.

Goto, T., Itoh, M., Suganami, T., Kanai, S., Shirakawa, I., Sakai, T., Asakawa, M., Yoneyama, T., Kai, T. and Ogawa, Y. (2018) 'Obeticholic acid protects against hepatocyte death and liver fibrosis in a murine model of nonalcoholic steatohepatitis', *Scientific Reports*, 8(1), pp. 1–13. doi: 10.1038/s41598-018-26383-8.

Goulis, J., Leandro, G. and Burroughs, A. K. (1999) 'Randomised controlled trials of ursodeoxycholic-acid therapy for primary biliary cirrhosis: a meta-analysis', *The Lancet*, 354(9184), pp. 1053–1060. doi: [https://doi.org/10.1016/S0140-6736\(98\)11293-X](https://doi.org/10.1016/S0140-6736(98)11293-X).

Gräslund, S. *et al.* (2008) 'Protein production and purification', *Nature methods*, 5(2), pp. 135–146. doi: 10.1038/nmeth.f.202.Protein.

Graton, J., Wang, Z., Brossard, A. M., Gonçalves Monteiro, D., Le Questel, J. Y. and Linclau, B. (2012) 'An unexpected and significantly lower hydrogen-bond-donating capacity of fluorohydrins compared to nonfluorinated alcohols', *Angewandte Chemie - International Edition*, 51(25), pp. 6176–6180. doi: 10.1002/anie.201202059.

Green, C., Pramfalk, C., Morton, K. and Hodson, L. (2015) 'From whole body to cellular models of hepatic triglyceride metabolism: man has got to know his limitations', *American Journal of Physiology- Endocrinology and Metabolism*, 308(1), pp. 1–20.

Greiner, E. F., Kirfel, J., Greschik, H., Huang, D., Becker, P., Kapfhammer, J. P. and Schu, R. (2000) 'Differential ligand-dependent protein – protein interactions between nuclear receptors and a neuronal-specific cofactor', 97(13).

Guenther, M. G., Barak, O. and Lazar, M. A. (2002) 'The SMRT and N-CoR Corepressors Are Activating Cofactors for Histone Deacetylase 3', *Molecular and Cellular Biology*, 21(18), pp. 6091–6101. doi: 10.1128/mcb.21.18.6091-6101.2001.

Gutierrez, A., Ratliff, E. P., Andres, A. M., Huang, X., McKeehan, W. L. and Davis, R. A. (2006) 'Bile acids decrease hepatic paraoxonase 1 expression and plasma high-density lipoprotein levels via FXR-mediated signaling of FGFR4', *Arteriosclerosis, Thrombosis, and Vascular Biology*, 26(2), pp. 301–306. doi: 10.1161/01.ATV.0000195793.73118.b4.

Haas, J. T., Francque, S. and Staels, B. (2016) 'Pathophysiology and Mechanisms of Nonalcoholic Fatty Liver Disease', *Annual Review of Physiology*, 78, pp. 181–205. doi: 10.1146/annurev-physiol-021115-105331.

Hagey, L. R., Vidal, N., Hofmann, A. F. and Krasowski, M. D. (2010) 'Evolutionary diversity of bile salts in reptiles and mammals , including analysis of ancient human and extinct giant ground sloth coprolites', *BMC Evolutionary Biology*, 10(133), pp. 1–23.

Han, C. Y. (2018) 'Update on FXR biology: Promising therapeutic target?', *International Journal of Molecular Sciences*, 19(7). doi: 10.3390/ijms19072069.

Han, S. J., DeMayo, F. J., Xu, J., Tsai, S. Y., Tsai, M.-J. and O'Malley, B. W. (2006) 'Steroid Receptor Coactivator (SRC)-1 and SRC-3 Differentially Modulate Tissue-Specific Activation Functions of the Progesterone Receptor', *Molecular Endocrinology*, 20(1), pp. 45–55. doi: 10.1210/me.2005-0310.

Hansen, H. H., Feigh, M., Veidal, S. S., Rigbolt, K. T., Vrang, N. and Fosgerau, K. (2017) 'Mouse models of nonalcoholic steatohepatitis in preclinical drug development', *Drug Discovery Today*. Elsevier Ltd, 22(11), pp. 1707–1718. doi: 10.1016/j.drudis.2017.06.007.

Hard, T., Kellenbach, E., Boelens, R., Maler, B. A., Dahlman, K., Freedman, L. P., Carlstedt-Duke, J., Yamamoto, K. R., Gustafsson, J. A. and Kaptein, R. (1990) 'Solution structure of the glucocorticoid receptor DNA-binding domain', *Science*, 249(4965), pp. 157 LP – 160. doi:

10.1126/science.2115209.

Hashemi, A., Roohvand, F. and Ghahremani, M. H. (2012) 'Selection of valid reference genes for expression studies of hepatic cell lines under IFN-?? treatment', *Biochemical and Biophysical Research Communications*. Elsevier Inc., 426(4), pp. 649–653. doi: 10.1016/j.bbrc.2012.09.009.

Hassell, A. M., An, G., Bledsoe, R. K., Bynum, J. M., Carter, H. L., Deng, S. J. J., Gampe, R. T., Grisard, T. E., Madauss, K. P., Nolte, R. T., Rocque, W. J., Wang, L., Weaver, K. L., Williams, S. P., Wisely, G. B., Xu, R. and Shewchuk, L. M. (2006) 'Crystallization of protein-ligand complexes', *Acta Crystallographica Section D: Biological Crystallography*, 63(1), pp. 72–79. doi: 10.1107/S0907444906047020.

Hatting, M., Zhao, G., Schumacher, F., Sellge, G., Al Masaoudi, M., Gaßler, N., Boekschoten, M., Müller, M., Liedtke, C., Cubero, F. J. and Trautwein, C. (2013) 'Hepatocyte caspase-8 is an essential modulator of steatohepatitis in rodents', *Hepatology*, 57(6), pp. 2189–2201. doi: 10.1002/hep.26271.

Heery, D. M., Kalkhoven, E., Hoare, S. and Parker, M. G. (1997) 'A signature motif in transcriptional co-activators mediates binding to nuclear receptors', *Nature*, 387(6634), pp. 733–736. doi: 10.1038/42750.

Hefti, M. H., Toorn, C. J. G. V. V. Der, Dixon, R. and Vervoort, J. (2001) 'A Novel Purification Method for Histidine-Tagged Proteins Containing a Thrombin Cleavage Site', 185, pp. 180–185. doi: 10.1006/abio.2001.5214.

Hegade, V. S., Speight, R. A., Etherington, R. E. and Jones, D. E. J. (2016) 'Novel bile acid therapeutics for the treatment of chronic liver diseases', *Therapeutic Advances in Gastroenterology*, 9(3), pp. 376–391. doi: 10.1177/1756283X16630712.

Henao-mejia, J., Elinav, E., Jin, C., Hao, L., Mehal, W. Z., Strowig, T., Thaiss, C. A., Kau, A. L., Eisenbarth, S. C., Michael, J., Camporez, J., Shulman, G. I., Gordon, J. I., Hal, M. and Flavell, R. A. (2012) 'HHS Public Access', *Nature*, 482(7384), pp. 179–185. doi: 10.1038/nature10809.Inflamasome-mediated.

Herzik Jr, M. A., Wu, M. and Lander, G. C. (2017) 'Achieving better-than-3-Å resolution by single-particle cryo-EM at 200 keV', *Nature methods*. 2017/10/09, 14(11), pp. 1075–1078. doi: 10.1038/nmeth.4461.

Hoeke, M. O., Heegsma, J., Hoekstra, M., Moshage, H. and Faber, K. N. (2014) 'Human FXR

regulates SHP expression through direct binding to an LRH-1 binding site, independent of an IR-1 and LRH-1', *PLoS ONE*, 9(2), pp. 1–10. doi: 10.1371/journal.pone.0088011.

Hofmann, A. F. (1999) 'Bile Acids : The Good , the Bad , and the Ugly', *Physiology*, 14, pp. 24–29. doi: 10.1038/ajg.2009.118.

Hofmann, A. F. (2004) 'Detoxification of lithocholic acid, a toxic bile acid: relevance to drug hepatotoxicity', *Drug Metabolism Reviews*, 36, pp. 703–722.

Hofmann, A. F. and Hagey, L. R. (2008) 'Bile acids: Chemistry, pathochemistry, biology, pathobiology, and therapeutics', *Cellular and Molecular Life Sciences*, 65(16), pp. 2461–2483. doi: 10.1007/s00018-008-7568-6.

Hofmann, A. F. and Hagey, L. R. (2014) 'Key discoveries in bile acid chemistry and biology and their clinical applications: history of the last eight decades', *Journal of Lipid Research*, 55(8), pp. 1553–1595. doi: 10.1194/jlr.R049437.

Holt, J. A., Luo, G., Billin, A. N., Bisi, J., McNeill, Y. Y., Kozarsky, K. F., Donahee, M., Wang, D. Y., Mansfield, T. A., Kliewer, S. A., Goodwin, B. and Jones, S. A. (2003) 'Definition of a novel growth factor-dependent signal cascade for the suppression of bile acid biosynthesis', *Genes and Development*, 17(13), pp. 1581–1591. doi: 10.1101/gad.1083503.

Honório, K. M., Garratt, R. C., Polikarpov, I. and Andricopulo, A. D. (2007) '3D QSAR comparative molecular field analysis on nonsteroidal farnesoid X receptor activators', *Journal of Molecular Graphics and Modelling*, 25, pp. 921–927. doi: 10.1016/j.jmgm.2006.09.003.

Hoshida, Y., Goossens, N., Chou, H.-I., Ybanez, M., Friedman, S. L., Lee, Y. A., Tsuchida, T., Fiel, M. I., Allen, B., Fujiwara, N. and Martins, S. (2018) 'A simple diet- and chemical-induced murine NASH model with rapid progression of steatohepatitis, fibrosis and liver cancer', *Journal of Hepatology*. European Association for the Study of the Liver, 69(2), pp. 385–395. doi: 10.1016/j.jhep.2018.03.011.

Höskuldsson, A. (1988) 'PLS regression methods', *Journal of Chemometrics*. John Wiley & Sons, Ltd, 2(3), pp. 211–228. doi: 10.1002/cem.1180020306.

Hsieh, J.-C., Nakajima, S., Galligan, M. A., Jurutka, P. W., Haussler, C. A., Whitfield, G. K. and Haussler, M. R. (1995) 'Receptor mediated genomic action of the 1,25(OH)2D3 hormone: Expression of the human vitamin D receptor in *E. coli*', *The Journal of Steroid Biochemistry and Molecular Biology*, 53(1), pp. 583–594. doi: [https://doi.org/10.1016/0960-0760\(95\)00112-D](https://doi.org/10.1016/0960-0760(95)00112-D).

Huang, C., Wang, J., Hu, W., Wang, C., Lu, X., Tong, L., Wu, F. and Zhang, W. (2016)

'Identification of functional farnesoid X receptors in brain neurons', *FEBS Letters*, 590(18), pp. 3233–3242. doi: 10.1002/1873-3468.12373.

Huang, N., Shoichet, B. K. and Irwin, J. J. (2006) 'Benchmarking sets for molecular docking', *Journal of medicinal chemistry*, 49(23), pp. 6789–6801. doi: 10.1021/jm0608356.

Huang, W., Johnson, R. L., Huang, R., Wichterman, J., Hwang, J. Y. and Guy, R. K. (2010) *qHTS for Inhibitors of the Interaction of Thyroid Hormone Receptor and Steroid Receptor Coregulator 2*.

Huber, R., Murphy, K., Miao, B., Link, J., Cunningham, M., Rupar, M., Gunyuzlu, P., Haws, T., Kassam, A., Powell, F., Hollis, G., Young, P., Mukherjee, R. and Burn, T. (2002) 'Generation of multiple farnesoid-X-receptor isoforms through the use of alternative promoters', *Gene*, 290(1–2), pp. 35–43. Available at: (null).

Hye Khan, M. A., Schmidt, J., Stavniichuk, A., Imig, J. D. and Merk, D. (2019) 'A dual farnesoid X receptor/soluble epoxide hydrolase modulator treats non-alcoholic steatohepatitis in mice', *Biochemical Pharmacology*. Elsevier, 166(February), pp. 212–221. doi: 10.1016/j.bcp.2019.05.023.

Hylemon, P. B., Zhou, H., Pandak, W. M., Ren, S., Gil, G. and Dent, P. (2009) 'Bile acids as regulatory molecules.', *Journal of lipid research*, 50(8), pp. 1509–1520. doi: 10.1194/jlr.R900007-JLR200.

Ilari, A. and Savino, C. (2008) 'Protein Structure Determination by X-Ray Crystallography BT - Bioinformatics: Data, Sequence Analysis and Evolution', in Keith, J. M. (ed.). Totowa, NJ: Humana Press, pp. 63–87. doi: 10.1007/978-1-60327-159-2_3.

Inagaki, T., Choi, M., Moschetta, A., Peng, L., Cummins, C. L., McDonald, J. G., Luo, G., Jones, S. A., Goodwin, B., Richardson, J. A., Gerard, R. D., Repa, J. J., Mangelsdorf, D. J. and Kliewer, S. A. (2005) 'Fibroblast growth factor 15 functions as an enterohepatic signal to regulate bile acid homeostasis', *Cell Metabolism*, 2(4), pp. 217–225. doi: 10.1016/j.cmet.2005.09.001.

Inagaki, T., Moschetta, A., Lee, Y. K., Peng, L., Zhao, G., Downes, M., Yu, R. T., Shelton, J. M., Richardson, J. A., Repa, J. J., Mangelsdorf, D. J. and Kliewer, S. A. (2006) 'Regulation of antibacterial defense in the small intestine by the nuclear bile acid receptor', *Proc Natl Acad Sci U S A*, 103(10), pp. 3920–3925. doi: 10.1073/pnas.0509592103.

Intercept Pharmaceuticals Inc. (September 2019) www.Interceptpharma.com.

Jiang, Y., Iakova, P., Jin, J., Sullivan, E., Sharin, V., Hong, I.-H., Anakk, S., Mayor, A., Darlington,

G., Finegold, M., Moore, D. and Timchenko, N. A. (2013) 'Farnesoid X receptor inhibits gankyrin in mouse livers and prevents development of liver cancer.', *Hepatology (Baltimore, Md.)*, 57(3), pp. 1098–106. doi: 10.1002/hep.26146.

Juřica, J., Dovrtělová, G., Nosková, K. and Zendulka, O. (2016) 'Bile Acids , Nuclear Receptors and Cytochrome P450', *Physiology Research*, 65(S4), pp. 427–440.

Kandel, B. A., Thomas, M., Winter, S., Damm, G., Seehofer, D., Burk, O., Schwab, M. and Zanger, U. M. (2016) 'Genomewide comparison of the inducible transcriptomes of nuclear receptors CAR, PXR and PPAR?? in primary human hepatocytes', *Biochimica et Biophysica Acta - Gene Regulatory Mechanisms*. Elsevier B.V., 1859(9), pp. 1218–1227. doi: 10.1016/j.bbagr.2016.03.007.

Kapetanovic, I. M. (2008) 'Computer aided drug discovery and development: in silico-chemico-biological approach', *Chemico Biological Interactions*, 171(2), pp. 165–176. doi: 10.1016/j.cbi.2006.12.006.COMPUTER-AIDED.

Karsa, V., Lignini, A., Patnick, J., Lambert, R. and Sauvaget, C. (2010) 'The dimensions of the CRC problem', *Best Practice & Research. Clinical Gastroenterology*, 24(4), pp. 381–396.

Kast, H. R., Nguyen, C. M., Sinal, C. J., Jones, S. A., Laffitte, B. A., Reue, K., Gonzalez, F. J., Willson, T. M. and Edwards, P. A. (2001) 'Farnesoid X-activated receptor induces apolipoprotein C-II transcription: a molecular mechanism linking plasma triglyceride levels to bile acids.', *Molecular endocrinology (Baltimore, Md.)*, 15(10), pp. 1720–1728. doi: 10.1210/me.15.10.1720.

Kawamata, Y., Fujii, R., Hosoya, M., Harada, M., Yoshida, H., Miwa, M., Fukusumi, S., Habata, Y., Itoh, T., Shintani, Y., Hinuma, S., Fujisawa, Y. and Fujino, M. (2003) 'A G protein-coupled receptor responsive to bile acids', *Journal of Biological Chemistry*, 278(11), pp. 9435–9440. doi: 10.1074/jbc.M209706200.

Keitel, V., Reinehr, R., Gatsios, P., Rupprecht, C., Görg, B., Selbach, O., Häussinger, D. and Kubitz, R. (2007) 'The G-protein coupled bile salt receptor TGR5 is expressed in liver sinusoidal endothelial cells', *Hepatology*. John Wiley & Sons, Ltd, 45(3), pp. 695–704. doi: 10.1002/hep.21458.

Kelly, L. A., Mezulis, S., Yates, C., Wass, M. and Sternberg, M. E. (2015a) 'The Phyre2 web portal for protein modelling, prediction, and analysis', *Nature Protocols*. Nature Publishing Group, 10(6), pp. 845–858. doi: 10.1038/nprot.2015-053.

Kemper, J. K. (2011a) 'Regulation of FXR transcriptional activity in health and disease : Emerging roles of FXR cofactors and post-translational modifications', *Biochimica et Biophysica Acta*. Elsevier B.V., 1812(8), pp. 842–850. doi: 10.1016/j.bbadi.2010.11.011.

Khoshouei, M., Radjainia, M., Baumeister, W. and Danev, R. (2017) 'Cryo-EM structure of haemoglobin at 3.2 Å determined with the Volta phase plate', *Nature communications*. Nature Publishing Group, 8, p. 16099. doi: 10.1038/ncomms16099.

Kim, K. H., Greco, G. and Novellino, E. (1998) 'A critical review of recent CoMFA applications', *Perspectives in Drug Discovery and Design*, 12(0), pp. 257–315. doi: 10.1023/A:1017010811581.

Kinzel, O., Steeneck, C., Schlüter, T., Schulz, A., Gege, C., Hahn, U., Hambruch, E., Hornberger, M., Spalwisz, A., Frick, K., Perovic, S., Deuschle, U., Burnet, M. and Kremoser, C. (2016) 'Bioorganic & Medicinal Chemistry Letters Novel substituted isoxazole FXR agonists with cyclopropyl, hydroxycyclobutyl and hydroxyazetidinyl linkers : Understanding and improving key determinants of pharmacological properties', *Bioorganic & Medicinal Chemistry Letters*, 26, pp. 3746–3753. doi: 10.1016/j.bmcl.2016.05.070.

Kirkwood, J., Hargreaves, D., O'keefe, S. and Wilson, J. (2015) 'Analysis of crystallization data in the Protein Data Bank', *Acta Crystallographica Section:F Structural Biology Communications*, 71, pp. 1228–1234. doi: 10.1107/S2053230X15014892.

Knegtel, R. M. A., Katahira, M., Schilthuis, J. G., Bonvin, A. M. J. J., Boelens, R., Eib, D., van der Saag, P. T. and Kaptein, R. (1993) 'The solution structure of the human retinoic acid receptor-β DNA-binding domain', *Journal of Biomolecular NMR*, 3(1), pp. 1–17. doi: 10.1007/BF00242472.

Kojetin, D. J. and Burris, T. P. (2013) 'Small molecule modulation of nuclear receptor conformational dynamics: implications for function and drug discovery.', *Molecular pharmacology*, 83(1), pp. 1–8. doi: 10.1124/mol.112.079285.

Kok, T., Hulzebos, C. V., Wolters, H., Havinga, R., Agellon, L. B., Stellaard, F., Shan, B., Schwarz, M. and Kuipers, F. (2003) 'Enterohepatic circulation of bile salts in farnesoid X receptor-deficient mice: Efficient intestinal bile salt absorption in the absence of ileal bile acid-binding protein', *Journal of Biological Chemistry*, 278(43), pp. 41930–41937. doi: 10.1074/jbc.M306309200.

Kong, B., Luyendyk, J. P., Tawfik, O. and Guo, G. L. (2009) 'Farnesoid X Receptor Deficiency Induces Nonalcoholic Steatohepatitis in Low-Density Lipoprotein Receptor-Knockout Mice Fed

a High-Fat Diet', *Journal of Pharmacology and Experimental Therapeutics*, 328(1), pp. 116–122. doi: 10.1124/jpet.108.144600.of.

Koshland, D. E. (1958) 'Application of a Theory of Enzyme Specificity to Protein Synthesis', *Proceedings of the National Academy of Sciences of the United States of America*, 44(2), pp. 98–104. doi: 10.1073/pnas.44.2.98.

Kouadjo, K. E., Nishida, Y., Cadrin-Girard, J. F., Yoshioka, M. and St-Amand, J. (2007) 'Housekeeping and tissue-specific genes in mouse tissues', *BMC Genomics*, 8. doi: 10.1186/1471-2164-8-127.

Landrier, J.-F., Eloranta, J. J., Vavricka, S. R. and Kullak-Ublick, G. A. (2006) 'The nuclear receptor for bile acids, FXR, transactivates human organic solute transporter- α and - β genes', *American Journal of Physiology-Gastrointestinal and Liver Physiology*, 290(3), pp. G476–G485. doi: 10.1152/ajpgi.00430.2005.

Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., Mcgettigan, P. A., McWilliam, H., Valentin, F., Wallace, I. M., Wilm, A., Lopez, R., Thompson, J. D., Gibson, T. J. and Higgins, D. G. (2007) 'Clustal W and Clustal X version 2.0', *Bioinformatics*, 23(21), pp. 2947–2948. doi: 10.1093/bioinformatics/btm404.

Lawson, C. L. et al. (2011) 'EMDataBank.org: unified data resource for CryoEM', *Nucleic acids research*. 2010/10/08. Oxford University Press, 39(Database issue), pp. D456–D464. doi: 10.1093/nar/gkq880.

Lax, S., Schauer, G., Prein, K., Kapitan, M., Silbert, D., Berghold, A., Berger, A. and Trauner, M. (2012) 'Expression of the nuclear bile acid receptor/farnesoid X receptor is reduced in human colon carcinoma compared to nonneoplastic mucosa independent from site and may be associated with adverse prognosis', *International Journal of Cancer*, 130(10), pp. 2232–2239. doi: 10.1002/ijc.26293.

Lazarević, S., Đanić, M., Goločorbin-Kon, S., Al-Salami, H. and Mikov, M. (2019) 'Semisynthetic bile acids: a new therapeutic option for metabolic syndrome', *Pharmacological Research*, 146(May). doi: 10.1016/j.phrs.2019.104333.

Leach, A. R., Shoichet, B. K. and Peishoff, C. E. (2006) 'Prediction of Protein-Ligand Interactions. Docking and Scoring: Successes and Gaps', *Journal of Medicinal Chemistry*, 49(20), pp. 5851–5855.

Lefebvre, P., Cariou, B. and Lien, F. (2009) 'Role of bile acids and bile acid receptors in

metabolic regulation', *Physiological*, 89, pp. 147–191. doi: 10.1152/physrev.00010.2008.

Lehmann, J. M., McKee, D. D., Watson, M. A., Willson, T. M., Moore, J. T. and Kliewer, S. A. (1998) 'The human orphan nuclear receptor PXR is activated by compounds that regulate CYP3A4 gene expression and cause drug interactions', *Journal of Clinical Investigation*, 102(5), pp. 1016–1023. doi: 10.1172/JCI3703.

Leiss, O. and Von Bermann, K. (1982) 'Different effects of chenodeoxycholic acid and ursodeoxycholic acid on serum lipoprotein concentrations in patients with radiolucent gallstones', *Scandinavian Journal of Gastroenterology*, 17(5), pp. 587–592.

Lesk, A. M. (2016) 'Protein structure determination', in *Introduction to protein science*. 3rd edn. Oxford: Oxford University Press, pp. 110–118.

Li, B., Qing, T., Zhu, J., Wen, Z., Yu, Y., Fukumura, R., Zheng, Y., Gondo, Y. and Shi, L. (2017) 'A Comprehensive Mouse Transcriptomic BodyMap across 17 Tissues by RNA-seq', *Scientific Reports*. Springer US, 7(1), pp. 1–10. doi: 10.1038/s41598-017-04520-z.

Li, N., Wang, X., Xu, Y., Lin, Y., Zhu, N., Liu, P., Lu, D. and Si, S. (2017) 'Identification of a Novel Liver X Receptor Agonist that Regulates the Expression of Key Cholesterol Homeostasis Genes with Distinct Pharmacological Characteristics', *Molecular Pharmacology*, 91(4), pp. 264–276. doi: 10.1124/mol.116.105213.

Li, T. and Chiang, J. Y. L. (2014) 'Nuclear receptors in bile acid metabolism', *Drug Metabolism Reviews*, 45(1), pp. 145–155. doi: 10.3109/03602532.2012.740048.Nuclear.

Li, T., Ma, H. and Chiang, J. Y. L. (2008) 'TGF β 1, TNF α , and insulin signaling crosstalk in regulation of the rat cholesterol 7 α -hydroxylase gene expression', *Journal of Lipid Research*, 49(9), pp. 1981–1989. doi: 10.1194/jlr.M800140-JLR200.

Li, Y. T. Y., Swales, K. E., Thomas, G. J., Warner, T. D. and Bishop-Bailey, D. (2007) 'Farnesoid X receptor ligands inhibit vascular smooth muscle cell inflammation and migration', *Arteriosclerosis, Thrombosis, and Vascular Biology*, 27(12), pp. 2606–2611. doi: 10.1161/ATVBAHA.107.152694.

Liang, H. and Ward, W. F. (2006) 'PGC-1 α : a key regulator of energy metabolism', *Advances in Physiology Education*, 30(4), pp. 145–151. doi: 10.1152/advan.00052.2006.

Lien, F. et al. (2014) 'Metformin interferes with bile acid homeostasis through AMPK-FXR crosstalk', *The Journal of clinical investigation*. 2014/02/17. American Society for Clinical Investigation, 124(3), pp. 1037–1051. doi: 10.1172/JCI68815.

Linclau, B., Wang, Z., Compain, G., Paumelle, V., Fontenelle, C. Q., Wells, N. and Weymouth-Wilson, A. (2016) 'Investigating the Influence of (Deoxy)fluorination on the Lipophilicity of Non-UV-Active Fluorinated Alkanols and Carbohydrates by a New log P Determination Method', *Angewandte Chemie - International Edition*, 55(2), pp. 674–678. doi: 10.1002/anie.201509460.

Liu, Y., Binz, J., Numerick, M. J., Dennis, S., Luo, G., Desai, B., MacKenzie, K. I., Mansfield, T. A., Kliewer, S. A., Goodwin, B. and Jones, S. A. (2003) 'Hepatoprotection by the farnesoid X receptor agonist GW4064 in rat models of intra- and extrahepatic cholestasis', *Journal of Clinical Investigation*, 112(11), pp. 1678–1687. doi: 10.1172/JCI200318945.

Liu, Z., Wong, J., Tsai, S. Y., Tsai, M.-J. and O'Malley, B. W. (2001) 'Sequential recruitment of steroid receptor coactivator-1 (SRC-1) and p300 enhances progesterone receptor-dependent initiation and reinitiation of transcription from chromatin', *Proceedings of the National Academy of Sciences*, 98(22), pp. 12426 LP – 12431. doi: 10.1073/pnas.231474798.

Livak, K. J. and Schmittgen, T. D. (2001) 'Analysis of relative gene expression data using real-time quantitative PCR and the 2- $\Delta\Delta$ CT method', *Methods*, 25(4), pp. 402–408. doi: 10.1006/meth.2001.1262.

Lou, X., Toresson, G., Benod, C., Suh, J. H., Philips, K. J., Webb, P. and Gustafsson, J.-A. (2014) 'Structure of the retinoid X receptor α –liver X receptor β (RXR α –LXR β) heterodimer on DNA', *Nature Structural & Molecular Biology*. Nature Publishing Group, a division of Macmillan Publishers Limited. All Rights Reserved., 21, p. 277. Available at: <https://doi.org/10.1038/nsmb.2778>.

Lu, J., Chen, M., Stanley, S. E. and Li, E. (2008) 'Effect of heterodimer partner RXR α on PPAR γ activation function-2 helix in solution', *Biochemical and Biophysical Research Communications*, 365(1), pp. 42–46. doi: <https://doi.org/10.1016/j.bbrc.2007.10.143>.

Lu, J., Cistola, D. P. and Li, E. (2006) 'Analysis of ligand binding and protein dynamics of human retinoid X receptor alpha ligand-binding domain by nuclear magnetic resonance', *Biochemistry*, 45(6), pp. 1629–1639. doi: 10.1021/bi051474j.

Lu, Y., Zheng, W., Lin, S., Guo, F., Zhu, Y., Wei, Y., Liu, X., Jin, S., Jin, L. and Li, Y. (2018) 'Identification of an Oleanane-Type Triterpene Hedragonic Acid as a Novel Farnesoid X Receptor Ligand with Liver Protective Effects and Anti-inflammatory Activity', *Molecular Pharmacology*, 93(2), pp. 63–72. doi: 10.1124/mol.117.109900.

Luisi, B. F., Xu, W. X., Otwinowski, Z., Freedman, L. P., Yamamoto, K. R. and Sigler, P. B. (1991) 'Crystallographic analysis of the interaction of the glucocorticoid receptor with DNA', *Nature*, 352(6335), pp. 497–505. doi: 10.1038/352497a0.

Ma, K., Saha, P. K., Chan, L. and Moore, D. D. (2006) 'Farnesoid X receptor is essential for normal glucose homeostasis', *Journal of Clinical Investigation*, 116(4), p. 1102. doi: 10.1172/JCI25604.1102.

Mackinnon, J. A. G., Gallastegui, N., Osguthorpe, D. J., Hagler, A. T. and Estébanez-Perpiñá, E. (2014a) 'Allosteric mechanisms of nuclear receptors: Insights from computational simulations', *Molecular and Cellular Endocrinology*. Elsevier Ireland Ltd, 393(1–2), pp. 75–82. doi: 10.1016/j.mce.2014.05.017.

Macrae, F. (2015) 'Could an "imaginary meal" pill solve the obesity crisis? Drug tricks the body into feeling full - AND lowers cholesterol and blood sugar', *The Daily Mail Online*, January.

Makishima, M. (1999) 'Identification of a Nuclear Receptor for Bile Acids', *Science*, 284(5418), pp. 1362–1365. doi: 10.1126/science.284.5418.1362.

Makishima, M., Lu, T. T., Xie, W., Whitfield, G. K., Domoto, H., Evans, R. M., Haussler, M. R. and Mangelsdorf, D. J. (2002) 'Vitamin D receptor as an intestinal bile acid sensor.', *Science*, 296(5571), pp. 1313–6. doi: 10.1126/science.1070477.

Maloney, P. R., Parks, D. J., Haffner, C. D., Fivush, A. M., Chandra, G., Plunket, K. D., Creech, K. L., Moore, L. B., Wilson, J. G., Lewis, M. C., Jones, S. A. and Willson, T. M. (2000) 'Identification of a chemical tool for the orphan nuclear receptor FXR', *Journal of Medicinal Chemistry*, 43(16), pp. 2971–2974. doi: 10.1021/jm0002127.

Maran, Rengasamy R. M., Thomas, A., Roth, M., Sheng, Z., Esterly, N., Pinson, D., Gao, X., Zhang, Y., Ganapathy, V., Gonzalez, F. J. and Guo, G. L. (2009) 'Farnesoid X Receptor Deficiency in Mice Leads to Increased', *Cancer Research*, 328(2), pp. 469–477. doi: 10.1124/jpet.108.145409.

Marschall, H., Wagner, M., Zollner, G., Fickert, P., Diczfalussy, U., Gumhold, J., Silbert, D., Fuchsbechler, A., Benthin, L., Grundström, R., Gustafsson, U., Sahlin, S., Einarsson, C. and Trauner, M. (2005) 'Complementary Stimulation of Hepatobiliary Transport and Detoxification Systems by Rifampicin and Ursodeoxycholic Acid in Humans', *Gastroenterology*, 129(2), pp. 476–485.

Maruyama, T., Miyamoto, Y., Nakamura, Takao, Tamai, Y., Okada, H., Sugiyama, E., Nakamura,

Tatsuji, Itadani, H. and Tanaka, K. (2002) 'Identification of membrane-type receptor for bile acids (M-BAR)', *Biochemical and Biophysical Research Communications*, 298(5), pp. 714–719. doi: 10.1016/S0006-291X(02)02550-0.

Massafra, V., Pellicciari, R., Gioiello, A. and van Mil, S. W. C. (2018a) 'Progress and challenges of selective Farnesoid X Receptor modulation', *Pharmacology and Therapeutics*. The Authors, 191, pp. 162–177. doi: 10.1016/j.pharmthera.2018.06.009.

McGarr, S. E., Ridlon, J. M. and Hylemon, P. B. (2005) 'Diet, anaerobic bacterial metabolism, and colon cancer: A review of the literature', *Journal of Clinical Gastroenterology*, 39, pp. 98–109.

Mcgettigan, B. M., McMahan, R. H., Luo, Y., Wang, X. X., Orlicky, D. J., Porsche, C., Levi, M. and Rosen, H. R. (2016) 'Sevelamer Improves Steatohepatitis, Inhibits Liver and Intestinal Farnesoid X Receptor (FXR), and Reverses Innate Immune Dysregulation in a Mouse Model of Non-alcoholic Fatty Liver Disease *', *Journal of Biological Chemistry*, 291(44), pp. 23058–23067. doi: 10.1074/jbc.M116.731042.

McGuffin, L. J., Atkins, J. D., Salehe, B. R., Shuid, A. N. and Roche, D. B. (2015) 'IntFOLD: An integrated server for modelling protein structures and functions from amino acid sequences', *Nucleic Acids Research*, 43(W1), pp. W169–W173. doi: 10.1093/nar/gkv236.

McKenna, N. J. and O'Malley, B. W. (2002) 'Minireview: Nuclear receptor coactivators - An update', *Endocrinology*, 143(7), pp. 2461–2465. doi: 10.1210/endo.143.7.8892.

McPherson, A. (1991) 'Current approaches to macromolecular crystallography', 23, pp. 49–71. doi: 10.1007/978-3-642-76168-3.

McPherson, A. (1990) 'Current approaches to macromolecular crystallography', *European Journal of Biochemistry*, 189, pp. 1–23. doi: 10.1007/978-3-642-76168-3.

McPherson, A. and Cudney, B. (2014) 'Optimization of crystallization conditions for biological macromolecules', *Acta Crystallographica Section F:Structural Biology Communications*. International Union of Crystallography, 70, pp. 1445–1467. doi: 10.1107/S2053230X14019670.

McPherson, A. and Gavira, J. A. (2014) 'Introduction to protein crystallization', *Acta Crystallographica Section F:Structural Biology Communications*. International Union of Crystallography, 70(1), pp. 2–20. doi: 10.1107/S2053230X13033141.

McPherson, A., Malkin, A. and Kuznetsov, Y. (1995) 'The Science of Macromolecular Crystallization', *Structure*, 3(8), pp. 759–768.

Meijer, F. A., Leijten-van de Gevel, I. A., de Vries, R. M. J. M. and Brunsved, L. (2019) 'Allosteric small molecule modulators of nuclear receptors', *Molecular and Cellular Endocrinology*. Elsevier, 485(November 2018), pp. 20–34. doi: 10.1016/j.mce.2019.01.022.

Melo-filho, C. C., Braga, R. C. and Andrade, C. H. (2014) '3D-QSAR Approaches in Drug Design: Perspectives to Generate Reliable CoMFA Models', *Current Computer Aided Drug Design*, 10, pp. 148–159.

Mencarelli, A., Renga, B., Distrutti, E. and Fiorucci, S. (2009) 'Antiatherosclerotic effect of farnesoid X receptor', *American Journal of Physiology. Heart and Circulatory Physiology*, 296(2), pp. H272-81. doi: 10.1152/ajpheart.01075.2008.

Merk, A., Bartesaghi, A., Banerjee, S., Falconieri, V., Rao, P., Davis, Mindy I., Pragani, R., Boxer, M. B., Earl, L. A., Milne, J. L. S. and Subramaniam, S. (2016) 'Breaking Cryo-EM Resolution Barriers to Facilitate Drug Discovery', *Cell*, 165(7), pp. 1698–1707. doi: <https://doi.org/10.1016/j.cell.2016.05.040>.

Merritt, M. E. and Donaldson, J. R. (2009) 'Effect of bile salts on the DNA and membrane integrity of enteric bacteria', *Journal of Medical Microbiology*, 58(12), pp. 1533–1541. doi: 10.1099/jmm.0.014092-0.

Meyer, M. E., Gronemeyer, H., Turcotte, B., Bocquel, M. T., Tasset, D. and Chambon, P. (1989) 'Steroid hormone receptors compete for factors that mediate their enhancer function', *Cell*, 57(3), pp. 433–442. doi: 10.1016/0092-8674(89)90918-5.

Meyer, U., Costantino, G., Macchiarulo, A. and Pellicciari, R. (2005) 'Is antagonism of E/Z-guggulsterone at the farnesoid X receptor mediated by a noncanonical binding site? A molecular modeling study', *Journal of Medicinal Chemistry*, 48(22), pp. 6948–6955. doi: 10.1021/jm0505056.

Mi, L.-Z., Devarakonda, S., Harp, J. M., Han, Q., Pellicciari, R., Willson, T. M., Khorasanizadeh, S. and Rastinejad, F. (2003) 'Short Article Structural Basis for Bile Acid Binding and Activation of the Nuclear Receptor FXR', *Molecular Cell*, 11, pp. 1093–1100.

Miao, J., Xiao, Z., Kanamaluru, D., Min, G., Yau, P. M., Veenstra, T. D., Ellis, E., Strom, S., Suino-powell, K., Xu, H. E. and Kemper, J. K. (2009) 'Bile acid signalling pathways increase stability of Small Heterodimer Partner (SHP) by inhibiting ubiquitin-proteasomal degradation', *Genes & Development*, 23(8), pp. 1–11. doi: 10.1101/gad.1773909.

Michiels, P., Atkins, K., Ludwig, C., Whittaker, S., van Dongen, M. and Günther, U. (2010)

'Assignment of the orphan nuclear receptor Nurr1 by NMR', *Biomolecular NMR Assignments*, 4(1), pp. 101–105. doi: 10.1007/s12104-010-9210-4.

Mobley, D. L. and Dill, K. a (2009) 'Binding of Small-Molecule Ligands to Proteins: "What You See Is Not Always "What You Get"', *Structure*, 17(4), pp. 489–498. doi: 10.1016/j.str.2009.02.010.Binding.

Modica, S., Gadaleta, R. M. and Moschetta, A. (2010) 'Deciphering the nuclear bile acid receptor FXR paradigm.', *Nuclear receptor signaling*, 8, pp. 1–28. doi: 10.1621/nrs.08005.

Modica, S., Murzilli, S., Salvatore, L., Schmidt, D. R. and Moschetta, A. (2008) 'Nuclear bile acid receptor FXR protects against intestinal tumorigenesis', *Cancer Research*, 68(23), pp. 9589–9594. doi: 10.1158/0008-5472.CAN-08-1791.

Mohan, V., Gibbs, A., Cummings, M., Jaeger, E. and DesJarlais, R. (2005) 'Docking: Successes and Challenges', *Current Pharmaceutical Design*, 11(3), pp. 323–333. doi: 10.2174/1381612053382106.

Monte, M. J., Marin, J. J. G., Antelo, A. and Vazquez-Tato, J. (2009) 'Bile acids: Chemistry, physiology, and pathophysiology', *World Journal of Gastroenterology*, 15(7), pp. 804–816. doi: 10.3748/wjg.15.804.

Moraes, L. A., Unsworth, A. J., Vaiyapuri, S., Ali, M. S., Molendi-coste, O., Dombrowicz, D., Staels, B. and Bishop-bailey, D. (2017) 'Europe PMC Funders Group Europe PMC Funders Author Manuscripts FXR and its ligands inhibit the function of platelets', 36(12), pp. 2324–2333. doi: 10.1161/ATVBAHA.116.308093.FXR.

Moschetta, A., Bookout, A. L. and Mangelsdorf, D. J. (2004) 'Prevention of cholesterol gallstone disease by FXR agonists in a mouse model', *Nature Medicine*, 10, pp. 1352–1358.

Mossakowska, D. E. (1998) 'Expression of nuclear hormone receptors in *Escherichia coli*', *Current Opinion in Biotechnology*, 9(5), pp. 502–505. doi: 10.1016/S0958-1669(98)80036-0.

Mouzaki, M., Comelli, E. M., Arendt, B. M., Bonengel, J., Fung, S. K., Fischer, S. E., Mcgilvray, I. D. and Allard, J. P. (2013) 'Intestinal Microbiota in Patients With Nonalcoholic Fatty Liver Disease', *Hepatology*, 58(1), pp. 120–127. doi: 10.1002/hep.26319.

Mudaliar, S., Henry, R. R., Sanyal, A. J., Morrow, L., Marschall, H. U., Kipnes, M., Adorini, L., Sciacca, C. I., Clopton, P., Castelloe, E., Dillon, P., Pruzanski, M. and Shapiro, D. (2013) 'Efficacy and Safety of the Farnesoid X Receptor Agonist Obeticholic Acid in Patients With Type 2 Diabetes and Nonalcoholic Fatty Liver Disease', *Gastroenterology*. Elsevier, Inc, 145(3), pp.

574-582.e1. doi: 10.1053/j.gastro.2013.05.042.

Murata, K. and Wolf, M. (2018) 'Cryo-electron microscopy for structural analysis of dynamic biological macromolecules', *Biochimica et Biophysica Acta - General Subjects*. Elsevier, 1862(2), pp. 324–334. doi: 10.1016/j.bbagen.2017.07.020.

Nahoum, V., Lipski, A., Quillard, F., Guichou, J. F., Boublik, Y., Pérez, E., Germain, P., de Lera, A. R. and Bourguet, W. (2008) 'Nuclear receptor ligand-binding domains: reduction of helix H12 dynamics to favour crystallization', *Acta crystallographica. Section F, Structural biology and crystallization communications*. 2008/06/11. International Union of Crystallography, 64(Pt 7), pp. 614–616. doi: 10.1107/S1744309108015492.

Nakahara, M., Fujii, H., Maloney, P. R., Shimizu, M. and Sato, R. (2002) 'Bile acids enhance low density lipoprotein receptor gene expression via a MAPK cascade-mediated stabilization of mRNA', *Journal of Biological Chemistry*, 277(40), pp. 37229–37234. doi: 10.1074/jbc.M206749200.

Nettles, K. W. and Greene, G. L. (2003) 'Nuclear receptor ligands and cofactor recruitment: Is there a coactivator "On Deck"?', *Molecular Cell*, 11(4), pp. 850–851. doi: 10.1016/S1097-2765(03)00133-3.

Neuschwander-Tetri, B. A., Loomba, R., Sanyal, A. J., Van Natta, M. L., Abdelmalek, M. F., Chalsani, N. and Dasarathy, S. (2015) 'Farnesoid X nuclear receptor ligand obeticholic acid for non- cirrhotic, non-alcoholic steatohepatitis (FLINT): a multicentre, randomised, placebo-controlled trial Prof.', *Lancet*, 385(9972), pp. 956–965. doi: 10.1016/S0140-6736(14)61933-4.Farnesoid.

Newman, J., Egan, D., Walter, T. S., Meged, R., Berry, I., Jelloul, M. Ben, Sussman, J. L., Stuart, D. I. and Perrakis, A. (2005) 'Towards rationalization of crystallization screening for small- To medium-sized academic laboratories: The PACT/JCSG+ strategy', *Acta Crystallographica Section D: Biological Crystallography*. International Union of Crystallography, 61(10), pp. 1426–1431. doi: 10.1107/S0907444905024984.

Nicholes, K., Guillet, S., Tomlinson, E., Hillan, K., Wright, B., Frantz, G. D., Pham, T. A., Dillard-Telm, L., Tsai, S. P., Stephan, J.-P., Stinson, J., Stewart, T. and French, D. M. (2002) 'A Mouse Model of Hepatocellular Carcinoma: Ectopic Expression of Fibroblast Growth Factor 19 in Skeletal Muscle of Transgenic Mice', *The American Journal of Pathology*, 160(6), pp. 2295–2307. doi: [https://doi.org/10.1016/S0002-9440\(10\)61177-7](https://doi.org/10.1016/S0002-9440(10)61177-7).

Ó'Fágáin, C., Cummins, P. M. and O'Connor, B. (2017) 'Gel-filtration chromatography', *Methods in Molecular Biology*, 1485, pp. 15–25. doi: 10.1007/978-1-4939-6412-3_2.

Orlov, I., Rochel, N., Moras, D. and Klaholz, B. P. (2012) 'Structure of the full human RXR/VDR nuclear receptor heterodimer complex with its DR3 target DNA', *EMBO Journal*. Nature Publishing Group, 31(2), pp. 291–300. doi: 10.1038/emboj.2011.445.

Ottow, E. and Weinmann, H. (2008) 'Nuclear Receptors as Drug Targets: A Historical Perspective of Modern Drug Discovery', in. Weinheim, Germany: Wiley-VCH.

Owsley, E. and Chiang, J. Y. L. (2003) 'Guggulsterone antagonizes farnesoid X receptor induction of bile salt export pump but activates pregnane X receptor to inhibit cholesterol 7 α -hydroxylase gene', *Biochemical and Biophysical Research Communications*, 304(1), pp. 191–195. doi: 10.1016/S0006-291X(03)00551-5.

Ozers, M. S., Ervin, K. M., Steffen, C. L., Fronczak, J. A., Lebakken, C. S., Carnahan, K. a, Lowery, R. G. and Burke, T. J. (2005) 'Analysis of ligand-dependent recruitment of coactivator peptides to estrogen receptor using fluorescence polarization.', *Molecular endocrinology (Baltimore, Md.)*, 19(1), pp. 25–34. doi: 10.1210/me.2004-0256.

Pacana, T. and Sanyal, A. J. (2015) 'Recent advances in understanding / management of non-alcoholic steatohepatitis', *F1000 Prime Reports*, 7(28), pp. 1–8. doi: 10.12703/P7-28.

Pandey, P., Udayakumar, T., Lin, X., Sharma, D., Shapiro, P. and Fondell, J. (2005) 'Activation of TRAP/Mediator Subunit TRAP220/Med1 is regulated by Mitogen-activated protein kinase-dependent phosphorylation', *Molecular and Cellular Biology*, 25(24), pp. 10695–10710. doi: 10.1128/MCB.25.24.10695.

Papatheodorou, I. et al. (2018) 'Expression Atlas: Gene and protein expression across multiple studies and organisms', *Nucleic Acids Research*, 46(D1), pp. D246–D251. doi: 10.1093/nar/gkx1158.

Parks, D. J., Blanchard, S. G., Bledsoe, R. K., Chandra, G., Consler, T. G., Kliewer, S. A., Stimmel, J. B., Willson, T. M., Zavacki, A. M., Moore, D. D. and Lehmann, J. M. (1999) 'Bile acids: natural ligands for an orphan nuclear receptor.', *Science*, 284(5418), pp. 1365–1368. doi: 10.1126/science.284.5418.1365.

Parsy, C. B., Chapman, C. J., Barnes, A. C., Robertson, J. F. and Murray, A. (2007) 'Two-step method to isolate target recombinant protein from co-purified bacterial contaminant SlyD after immobilised metal affinity chromatography', *Journal of Chromatography B: Analytical*

Technologies in the Biomedical and Life Sciences, 853(1–2), pp. 314–319. doi: 10.1016/j.jchromb.2007.03.046.

Paumgartner, G. and Beuers, U. (2002) ‘Ursodeoxycholic acid in cholestatic liver disease: Mechanisms of action and therapeutic use revisited’, *Hepatology*, 36(3), pp. 525–531. doi: 10.1053/jhep.2002.36088.

Pellicciari, R., Costantino, G., Camaioni, E., Sadeghpour, B. M., Entrena, A., Willson, T. M., Fiorucci, S., Clerici, C. and Gioiello, A. (2004) ‘Bile Acid Derivatives as Ligands of the Farnesoid X Receptor . Synthesis , Evaluation , and Structure - Activity Relationship of a Series of Body and Side Chain Modified Analogues of Chenodeoxycholic Acid’, *Journal of Medicinal Chemistry*, 47, pp. 4559–4569.

Pellicciari, R., Fiorucci, S., Camaioni, E., Clerici, C., Costantino, G., Maloney, P. R., Morelli, A., Parks, D. J. and Willson, T. M. (2002) ‘6a-Ethyl-Chenodeoxycholic Acid (6-ECDCA), a Potent and Selective FXR Agonist Endowed with Anticholestatic’, *Journal of medicinal chemistry*, 45(17), pp. 3569–3572.

Pellicciari, R., Gioiello, A., Costantino, G., Sadeghpour, B. M., Rizzo, G., Meyer, U., Parks, D. J., Entrena-guadix, A., Fiorucci, S. and Uni, V. (2006) ‘Back Door Modulation of the Farnesoid X Receptor : Design , Synthesis , and Biological Evaluation of a Series of Side Chain Modified Chenodeoxycholic Acid Derivatives’, *Journal of Medicinal Chemistry*, 49, pp. 4208–4215.

Pellicciari, R., Passeri, D., De Franco, F., Mostarda, S., Filipponi, P., Colliva, C., Gadaleta, R. M., Franco, P., Carotti, A., Macchiarulo, A., Roda, A., Moschetta, A. and Gioiello, A. (2016) ‘Discovery of 3 α ,7 α ,11 β -Trihydroxy-6 α -ethyl-5 β -cholan-24-oic Acid (TC-100), a Novel Bile Acid as Potent and Highly Selective FXR Agonist for Enterohepatic Disorders’, *Journal of Medicinal Chemistry*, 59(19), pp. 9201–9214. doi: 10.1021/acs.jmedchem.6b01126.

Perez, M. J. and Britz, O. (2009) ‘Bile-acid-induced cell injury and protection’, *World Journal of Gastroenterology*, 15(14), pp. 1677–1689. doi: 10.3748/wjg.15.1677.

Peterlik, M. (2008) ‘Role of bile acid secretion in human colorectal cancer.’, *Wiener medizinische Wochenschrift* (1946), 158(19–20), pp. 539–41. doi: 10.1007/s10354-008-0601-4.

Phillips, M. A., Stewart, M. A., Woodling, D. L. and Xie, Z.-R. (2018) ‘Has Molecular Docking Ever Brought us a Medicine?’, *Molecular Docking*. doi: 10.5772/intechopen.72898.

Pols, T. W. H., Nomura, M., Harach, T., Lo Sasso, G., Oosterveer, M. H., Thomas, C., Rizzo, G., Gioiello, A., Adorini, L., Pellicciari, R., Auwerx, J. and Schoonjans, K. (2011) ‘TGR5 activation

inhibits atherosclerosis by reducing macrophage inflammation and lipid loading', *Cell Metabolism*, 14(6), pp. 747–757. doi: 10.1016/j.cmet.2011.11.006.

Porez, G., Prawitt, J., Gross, B. and Staels, B. (2012) 'Bile acid receptors as targets for the treatment of dyslipidemia and cardiovascular disease', *Journal of Lipid Research*, 53(9), pp. 1723–1737. doi: 10.1194/jlr.R024794.

Pullinger, C. R., Eng, C., Salen, G., Shefer, S., Batta, A. K., Erickson, S. K., Verhagen, A., Rivera, C. R., Mulvihill, S. J., Malloy, M. J. and Kane, J. P. (2002) 'Human cholesterol 7a-hydroxylase (CYP7A1) deficiency has a hypercholesterolemic phenotype', *Journal of Clinical Investigation*, 110(1), pp. 109–117. doi: 10.1172/JCI200215387.

Rada, H. (Oxford protein production facility) (2017) 'OPPF standard protocols: Cloning and expression screening', pp. 1–26.

Ramírez, D. and Caballero, J. (2018) 'Is It Reliable to Take the Molecular Docking Top Scoring Position as the Best Solution without Considering Available Structural Data?', *Molecules*, 23(5), pp. 1–17. doi: 10.3390/molecules23051038.

Rastinejad, F., Huang, P., Chandra, V. and Khorasanizadeh, S. (2013) 'Understanding nuclear receptor form and function using structural biology', *Journal of Molecular Endocrinology*, 51(3). doi: 10.1530/JME-13-0173.

Rastinejad, F., Ollendorff, V. and Polikarpov, I. (2015) 'Nuclear receptor full-length architectures: Confronting myth and illusion with high resolution', *Trends in Biochemical Sciences*. Elsevier Ltd, 40(1), pp. 16–24. doi: 10.1016/j.tibs.2014.10.011.

Renaud, J. P. and Moras, D. (2000) 'Structural studies on nuclear receptors.', *Cellular and Molecular Life Sciences*, 57(12), pp. 1748–69. doi: 10.1007/PL00000656.

Rennie, P. S., Bruchovsky, N., Leco, K. J., Sheppard, P. C., McQueen, S. A., Cheng, H., Snoek, R., Hamel, A., Bock, M. E., MacDonald, B. S., Nickel, B. E., Chang, C., Liao, S., Cattini, P. A. and Matusik, R. J. (1993) 'Characterization of two cis-acting DNA elements involved in the androgen regulation of the probasin gene.', *Molecular Endocrinology*, 7(1), pp. 23–36. doi: 10.1210/mend.7.1.8446105.

Reshetnyak, V. I. (2013) 'Physiological and molecular biochemical mechanisms of bile formation.', *World journal of gastroenterology : WJG*, 19(42), pp. 7341–60. doi: 10.3748/wjg.v19.i42.7341.

Ridlon, J. M., Kang, D.-J. and Hylemon, P. B. (2006) 'Bile salt biotransformations by human

intestinal bacteria.', *Journal of lipid research*, 47(2), pp. 241–259. doi: 10.1194/jlr.R500013-JLR200.

Rinella, M. E., Elias, M. S., Smolak, R. R., Fu, T., Borensztajn, J. and Green, R. M. (2008) 'Mechanisms of hepatic steatosis in mice fed a lipogenic methionine choline-deficient diet', *Journal of lipid research*. American Society for Biochemistry and Molecular Biology, 49(5), pp. 1068–1076. doi: 10.1194/jlr.M800042-JLR200.

Rosano, G. L. and Ceccarelli, E. A. (2014) 'Recombinant protein expression in Escherichia coli: advances and challenges', *Frontiers in microbiology*. Frontiers Media S.A., 5, p. 172. doi: 10.3389/fmicb.2014.00172.

Rosenfeld, M. G., Lunyak, V. V. and Glass, C. K. (2006) 'Sensors and signals : a coactivator / corepressor / epigenetic code for integrating signal-dependent programs of transcriptional response', *Genes & development*, 20(11), pp. 1405–1428. doi: 10.1101/gad.1424806.

Russell, D. W. and Setchell, K. D. (1992) 'Bile acid biosynthesis.', *Biochemistry*, 31(20), pp. 4737–49. doi: 10.1021/bi00135a001.

Sanchez-Weatherby, J., Sandy, J., Mikolajek, H., Loble, C. M. C., Mazzorana, M., Kelly, J., Preece, G., Littlewood, R. and Sørensen, T. L.-M. (2019) 'VMXi: a fully automated, fully remote, high-flux {\it in situ} macromolecular crystallography beamline', *Journal of Synchrotron Radiation*, 26(1), pp. 291–301. doi: 10.1107/S1600577518015114.

Sanyal, A. J., Chalasani, N., Kowdley, K. V., McCullough, A., Diehl, A. M., Bass, N. M., Neuschwander-Tetri, B. A., Lavine, J. E., Tonascia, J., Unalp, A., Van Natta, M., Clark, J., Brunt, E. M., Kleiner, D. E., Hoofnagle, J. H. and Robuck, P. R. (2010) 'Pioglitazone, vitamin E, or placebo for nonalcoholic steatohepatitis.', *The New England journal of medicine*, 362(18), pp. 1675–85. doi: 10.1056/NEJMoa0907929.

Schaap, F. G., Trauner, M. and Jansen, P. L. M. (2013) 'Bile acid receptors as targets for drug development', *Nature Reviews Gastroenterology & Hepatology*. Nature Publishing Group, 11(1), pp. 55–67. doi: 10.1038/nrgastro.2013.151.

Schwabe, J. W. R., Chapman, L., Finch, J. T. and Rhodes, D. (1993) 'The crystal structure of the estrogen receptor DNA-binding domain bound to DNA: How receptors discriminate between their response elements', *Cell*, 75(3), pp. 567–578. doi: [https://doi.org/10.1016/0092-8674\(93\)90390-C](https://doi.org/10.1016/0092-8674(93)90390-C).

Schwabe, J. W. R., Neuhaus, D. and Rhodes, D. (1990) 'Solution structure of the DNA-binding

domain of the oestrogen receptor', *Nature*, 348(6300), pp. 458–461. doi: 10.1038/348458a0.

Seol, W., Choi, H. S. and Moore, D. D. (1995) 'Isolation of proteins that interact specifically with the retinoid X receptor: two novel orphan receptors', *Molecular Endocrinology*, 9(1), pp. 72–85.

Shah, P. and Westwell, A. D. (2007) 'The role of fluorine in medicinal chemistry', *Journal of Enzyme Inhibition and Medicinal Chemistry*, 22(5), pp. 527–540. doi: 10.1080/14756360701425014.

Shaik, F. B., Prasad, D. V. R. and Narala, V. R. (2014) 'Role of farnesoid X receptor in inflammation and resolution', *Inflammation Research*, 64(1), pp. 9–20. doi: 10.1007/s00011-014-0780-y.

Shankar, R., Frapaise, X. and Brown, B. (2006) 'Lean Drug Development in R&D', *Drug Discovery and Development*, pp. 57–60.

Shen, H., Zhang, Y., Ding, H., Wang, X., Chen, L., Jiang, H. and Shen, X. (2008) 'Farnesoid X receptor induces GLUT4 expression through FXR response element in the GLUT4 promoter', *Cellular Physiology and Biochemistry*, 22(1–4), pp. 1–14. doi: 10.1159/000149779.

Shen, P. S. (2018) 'The 2017 Nobel Prize in Chemistry: cryo-EM comes of age', *Analytical and bioanalytical chemistry*. Analytical and Bioanalytical Chemistry, 410, pp. 2053–2057. Available at: https://www.nobelprize.org/nobel_prizes/chemistry/laureates/2017/press.html.

Sheppard, H. M., Harries, J. C., Hussain, S., Bevan, C. and Heery, D. M. (2001) 'Analysis of the Steroid Receptor Interaction Interface and Its Importance for the Function of SRC1 Analysis of the Steroid Receptor Coactivator 1 (SRC1) -CREB Binding Protein Interaction Interface and Its Importance for the Function of SRC1', *Molecular and Cellular Biology*, 1(1), pp. 39–50. doi: 10.1128/MCB.21.1.39.

Shi, L. M., Fang, H., Tong, W., Wu, J., Perkins, R., Blair, R. M., Branham, W. S., Dial, S. L., Moland, C. L. and Sheehan, D. M. (2001) 'QSAR Models Using a Large Diverse Set of Estrogens', *Journal of Chemical Information and Computer Sciences*, 41(1), pp. 186–195. doi: 10.1021/ci000066d.

Shoichet, B. K., McGovern, S. L., Wei, B. and Irwin, J. J. (2002) 'Lead Discovery Using Molecular Docking', *Current Opinion in Chemical Biology*, 6, pp. 439–446.

Sinal, C. J., Tohkin, M., Miyata, M., Ward, J. M., Lambert, G. and Gonzalez, F. J. (2000) 'Targeted Disruption of the Nuclear Receptor FXR/BAR Impairs Bile Acid and Lipid

Homeostasis', *Cell*, 102(6), pp. 731–744. doi: 10.1016/S0092-8674(00)00062-3.

Slattery, M. (2000) 'Diet, lifestyle, and colon cancer', *Seminars in Gastrointestinal Disease*, 11(3), pp. 142–146.

Sliwoski, G., Kothiwale, S., Meiler, J. and Lowe Jr, E. W. (2013) 'Computational methods in drug discovery', *Pharmacological reviews*. The American Society for Pharmacology and Experimental Therapeutics, 66(1), pp. 334–395. doi: 10.1124/pr.112.007336.

Smialowski, P., Schmidt, T., Cox, J., Kirschner, A. and Frishman, D. (2006) 'Will my protein crystallize? A sequence-based predictor', *Proteins: Structure, Function and Genetics*, 62(2), pp. 343–355. doi: 10.1002/prot.20789.

Soisson, S. M., Parthasarathy, G., Adams, A. D., Sahoo, S., Sitlani, A., Sparrow, C., Cui, J. and Becker, J. W. (2008) 'Identification of a potent synthetic FXR agonist with an unexpected mode of binding and activation.', *Proceedings of the National Academy of Sciences of the United States of America*, 105(14), pp. 5337–5342. doi: 10.1073/pnas.0710981105.

Stedman, C., Liddle, C., Coulter, S., Sonoda, J., Alvarez, J. G., Evans, R. M. and Downes, M. (2006) 'Benefit of farnesoid X receptor inhibition in obstructive cholestasis.', *Proceedings of the National Academy of Sciences of the United States of America*, 103(30), pp. 11323–11328. doi: 10.1073/pnas.0604772103.

Stojančević, M., Pavlović, N., Goločorbin-Kon, S. and Mikov, M. (2013) 'Application of bile acids in drug formulation and delivery', *Frontiers in Life Science*. Taylor & Francis, 7(3–4), pp. 112–122. doi: 10.1080/21553769.2013.879925.

Strack, A. M. and Myers, R. W. (2004) 'Modulation of Metabolic Syndrome by Fibroblast Growth Factor 19 (FGF19)?', *Endocrinology*, 145(6), pp. 2591–2593. doi: 10.1210/en.2004-0367.

Strecker, A. (1848) 'Bemerkungen zu der Arbeit des Herrn Dr. Völcker.', *Annals of chemistry and pharmacy*, 64(3), p. 30.

Sumida, Y. and Yoneda, M. (2018) 'Current and future pharmacological therapies for NAFLD/NASH', *Journal of Gastroenterology*. Springer Japan, 53(3), pp. 362–376. doi: 10.1007/s00535-017-1415-1.

Sutherland, J. J., Nandigam, R. K., Erickson, J. A. and Vieth, M. (2007) 'Lessons in Molecular Recognition. 2. Assessing and Improving Cross-Docking Accuracy', *Journal of Chemical Information and Modeling*. American Chemical Society, 47(6), pp. 2293–2302. doi:

10.1021/ci700253h.

Suzuki, T., J. P. and R. D. (2000) 'Review Control Selection for RNA Quantitation', *Biotechniques*, 337(August), pp. 332–337.

Takeyama, K., Masuhiro, Y., Fuse, H., Endoh, H., Murayama, A., Kitanaka, S., Suzawa, M., Yanagisawa, J. and Kato, S. (1999) 'Selective interaction of vitamin D receptor with transcriptional coactivators by a vitamin D analog', *Molecular and cellular biology*. American Society for Microbiology, 19(2), pp. 1049–1055. doi: 10.1128/mcb.19.2.1049.

Thomas, A. M., Hart, S. N., Kong, B., Fang, J., Zhong, X. B. and Guo, G. L. (2010) 'Genome-wide tissue-specific farnesoid X receptor binding in mouse liver and intestine', *Hepatology*, 51(4), pp. 1410–1419. doi: 10.1002/hep.23450 ET - 2010/01/22.

Thomas, C., Gioiello, A., Noriega, L., Strehle, A., Oury, J., Rizzo, G., Macchiarulo, A., Yamamoto, H., Mataki, C., Pellicciari, R., Auwerx, J. and Schoonjans, K. (2009) 'TGR5-mediated bile acid sensing controls glucose homeostasis', *Cell Metabolism*, 10(3), pp. 167–177. doi: 10.1016/j.cmet.2009.08.001.TGR5-mediated.

Thomas, C., Pellicciari, R., Pruzanski, M., Auwerx, J. and Schoonjans, K. (2008) 'Targeting bile-acid signalling for metabolic diseases.', *Nature reviews. Drug discovery*, 7(8), pp. 678–693. doi: 10.1038/nrd2619.

Thomas, G. (2003) 'Fundamentals of Medicinal Chemistry', in *Fundamentals of Medicinal Chemistry*. West sussex: Wiley, pp. 95–110.

Tølbøl, K. S., Kristiansen, M. N. B., Hansen, H. H., Veidal, S. S., Rigbolt, K. T. G., Gillum, M. P., Jelsing, J., Vrang, N. and Feigh, M. (2018) 'Metabolic and hepatic effects of liraglutide, obeticholic acid and elafibranor in diet-induced obese mouse models of biopsy-confirmed nonalcoholic steatohepatitis', *World Journal of Gastroenterology*, 24(2), pp. 179–194. doi: 10.3748/wjg.v24.i2.179.

Torra, I. P., Freedman, L. P. and Garabedian, M. J. (2004) 'Identification of DRIP205 as a coactivator for the farnesoid X receptor', *Journal of Biological Chemistry*, 279(35), pp. 36184–36191. doi: 10.1074/jbc.M405126200.

Trauner, M., Nevens, F., Schiffman, M. L., Drenth, J. P. H., Bowlus, C. L., Vargas, V., Andreone, P., Hirschfield, G. M., Pencek, R., Malecha, E. S., MacConell, L. and Shapiro, D. (2019) 'Long-term efficacy and safety of obeticholic acid for patients with primary biliary cholangitis: 3-year results of an international open-label extension study', *The Lancet Gastroenterology &*

Hepatology, 4(6), pp. 445–453. doi: [https://doi.org/10.1016/S2468-1253\(19\)30094-9](https://doi.org/10.1016/S2468-1253(19)30094-9).

Tremaroli, V. and Bäckhed, F. (2012) 'Functional interactions between the gut microbiota and host metabolism', *Nature*. Nature Publishing Group, a division of Macmillan Publishers Limited. All Rights Reserved., 489, p. 242. Available at: <https://doi.org/10.1038/nature11552>.

Tremblay, A., Tremblay, G. B., Labrie, F. and Giguere, V. (1999) 'Ligand-independent recruitment of SRC-1 to estrogen receptor beta through phosphorylation of activation function AF-1.', *Molecular cell*. United States, 3(4), pp. 513–519.

Tully, D. C. *et al.* (2017) 'Discovery of Tropifexor (LJN452), a Highly Potent Non-bile Acid FXR Agonist for the Treatment of Cholestatic Liver Diseases and Nonalcoholic Steatohepatitis (NASH)', *Journal of Medicinal Chemistry*, 60(24), pp. 9960–9973. doi: [10.1021/acs.jmedchem.7b00907](https://doi.org/10.1021/acs.jmedchem.7b00907).

Urizar, N. L., Dowhan, D. H. and Moore, D. D. (2000) 'The farnesoid X-activated receptor mediates bile acid activation of phospholipid transfer protein gene expression', *Journal of Biological Chemistry*, 275(50), pp. 39313–39317. doi: [10.1074/jbc.M007998200](https://doi.org/10.1074/jbc.M007998200).

Urizar, N. L., Liverman, A. B., Dodds, D. N. T., Silva, F. V., Ordentlich, P., Yan, Y., Gonzalez, F. J., Heyman, R. A., Mangelsdorf, D. J. and Moore, D. D. (2002) 'A Natural Product That Lowers Cholesterol As an Antagonist Ligand for FXR -- Ur ... Página 1 de 4 A Natural Product That Lowers Cholesterol As an Antagonist Ligand for FXR A Natural Product That Lowers Cholesterol As an Antagonist Ligand for FXR -- Ur ...', 296(2010), pp. 2010–2013. doi: [10.1126/stke.2002.135.tw201](https://doi.org/10.1126/stke.2002.135.tw201).

Vagenende, V., Yap, M. G. S. and Trout, B. L. (2009) 'Mechanisms of Protein Stabilization and Prevention of Protein Aggregation by Glycerol', *Biochemistry*. American Chemical Society, 48(46), pp. 11084–11096. doi: [10.1021/bi900649t](https://doi.org/10.1021/bi900649t).

Vanwagner, L. B., Bhave, M., Te, H. S., Feinglass, J., Alvarez, L. and Rinella, M. E. (2012) 'Patients transplanted for nonalcoholic steatohepatitis are at increased risk for postoperative cardiovascular events', *Hepatology*, 56(5), pp. 1741–1750. doi: [10.1002/hep.25855](https://doi.org/10.1002/hep.25855).

Vaquero, J., Monte, M. J., Dominguez, M., Muntané, J. and Marin, J. J. G. (2013) 'Differential activation of the human farnesoid X receptor depends on the pattern of expressed isoforms and the bile acid pool composition', *Biochemical Pharmacology*, 86(7), pp. 926–939. doi: [10.1016/j.bcp.2013.07.022](https://doi.org/10.1016/j.bcp.2013.07.022).

Vavassori, P., Mencarelli, A., Renga, B., Distrutti, E. and Fiorucci, S. (2009) 'The bile acid

receptor FXR is a modulator of intestinal innate immunity', *Journal of Immunology*, 183(10), pp. 6251–61. doi: 10.4049/jimmunol.0803978.

de Vera, I. M. S., Zheng, J., Novick, S., Shang, J., Hughes, T. S., Brust, R., Munoz-Tello, P., Gardner, W. J., Marciano, D. P., Kong, X., Griffin, P. R. and Kojetin, D. J. (2017) 'Synergistic Regulation of Coregulator/Nuclear Receptor Interaction by Ligand and DNA', *Structure*. Elsevier Ltd., 25(10), pp. 1506–1518.e4. doi: 10.1016/j.str.2017.07.019.

Veras, H., Filho, R., Luisa, I., Mariano, M., Dias, G., Bernardi, N., Bruder, M., Amorim, A., Carolina, A. and Figueira, M. (2019) 'Molecular and Cellular Endocrinology Modulation of nuclear receptor function : Targeting the protein-DNA interface', *Molecular and Cellular Endocrinology*. Elsevier, 484(January), pp. 1–14. doi: 10.1016/j.mce.2019.01.023.

Veselovsky, A. V. and Ivanov, A. S. (2003) 'Strategy of Computer-Aided Drug Design Strategy of Computer-Aided Drug Design', *Current Drug Targets - Infectious Disorders*, 3, pp. 33–40. doi: 10.2174/1568005033342145.

Wagner, M., Fickert, P., Zollner, G., Fuchsbichler, A., Silbert, D., Tsybrovskyy, O., Zatloukal, K., Guo, G. L., Schuetz, J. D., Gonzalez, F. J., Marschall, H.-U., Denk, H. and Trauner, M. (2003) 'Role of farnesoid X receptor in determining hepatic ABC transporter expression and liver injury in bile duct-ligated mice', *Gastroenterology*, 125(3), pp. 825–838. doi: [https://doi.org/10.1016/S0016-5085\(03\)01068-0](https://doi.org/10.1016/S0016-5085(03)01068-0).

Wallace, A. C., Laskowski, R. A. and Thornton, J. . (1995) 'LIGPLOT: a program to generate schematic diagrams of protein-ligand interactions', *Protein Engineering*, 8(2), pp. 127–34.

Wang, H., Chen, J., Hollister, K., Sowers, L. C. and Forman, B. M. (1999) 'Endogenous bile acids are ligands for the nuclear receptor FXR/BAR', *Molecular Cell*, 3(5), pp. 543–553. doi: 10.1016/S1097-2765(00)80348-2.

Wang, H. W. and Wang, J. W. (2017) 'How cryo-electron microscopy and X-ray crystallography complement each other', *Protein Science*, 26(1), pp. 32–39. doi: 10.1002/pro.3022.

Wang, H., Zhao, Z., Zhou, J., Guo, Y., Wang, G., Hao, H. and Xu, X. (2017) 'A novel intestinal-restricted FXR agonist', *Bioorganic and Medicinal Chemistry Letters*. Elsevier Ltd, 27(15), pp. 3386–3390. doi: 10.1016/j.bmcl.2017.06.003.

Wang, N., Zou, Q., Xu, J., Zhang, J. and Liu, J. (2018) 'Ligand binding and heterodimerization with retinoid X receptor α (RXR α) induce farnesoid X receptor (FXR) conformational changes affecting coactivator binding', *Journal of Biological Chemistry*, 293(47), pp. 18180–18191. doi:

10.1074/jbc.RA118.004652.

Wang, X., Li, S., Chen, M., Liu, J., Dong, R., Wang, H. and Zhu, S. (2018) 'Activation of the Nuclear Receptor Fxr Improves Intestinal Cell Tolerance to Ischemia–Reperfusion Injury', *Shock*, 50(3).

Wang, Y., Chen, W., Wang, M., Yu, D. and Forman, B. M. (2008) 'Farnesoid X receptor antagonizes NF- κ B in hepatic inflammatory response', *Hepatology*, 48(5), pp. 1632–1643. doi: 10.1002/hep.22519. Farnesoid.

Wang, Y. D., Chen, W. D., Yu, D., Forman, B. M. and Huang, W. (2011) 'The G-Protein-coupled bile acid receptor, Gpbar1 (TGR5), negatively regulates hepatic inflammatory response through antagonizing nuclear factor kappa light-chain enhancer of activated B cells (NF- κ B) in mice', *Hepatology*, 54(4), pp. 1421–1432. doi: 10.1002/hep.24525.

Watanabe, M., Horai, Y., Houten, S. M., Morimoto, K., Sugizaki, T., Arita, E., Mataki, C., Sato, H., Tanigawara, Y., Schoonjans, K., Itoh, H. and Auwerx, J. (2011) 'Lowering bile acid pool size with a synthetic farnesoid X receptor (FXR) agonist induces obesity and diabetes through reduced energy expenditure', *Journal of Biological Chemistry*, 286(30), pp. 26913–26920. doi: 10.1074/jbc.M111.248203.

Watanabe, M., Houten, S. M., Wang, L., Moschetta, A., Mangelsdorf, D. J., Heyman, R. A., Moore, D. D. and Auwerx, J. (2004) 'Bile acids lower triglyceride levels via a pathway involving FXR, SHP, and SREBP-1c', *Journal of Clinical Investigation*, 113(10), pp. 1408–1418. doi: 10.1172/JCI200421025.

Waterhouse, A. M., Procter, J. B., Martin, D. M. A., Clamp, M. and Barton, G. J. (2009) 'Jalview Version 2-A multiple sequence alignment editor and analysis workbench', *Bioinformatics*, 25(9), pp. 1189–1191. doi: 10.1093/bioinformatics/btp033.

Waugh, D. S. (2011) 'An overview of enzymatic reagents for the removal of affinity tags', *Protein Expression and Purification*, 80(2), pp. 283–293. doi: 10.1016/j.pep.2011.08.005.

Wildenberg, M. E. and van den Brink, G. R. (2011) 'FXR activation inhibits inflammation and preserves the intestinal barrier in IBD', *Gut*, 60(4), pp. 432 LP – 433. doi: 10.1136/gut.2010.233304.

Wlodawer, A., Minor, W., Dauter, Z. and Jaskolski, M. (2013) 'Protein crystallography for aspiring crystallographers or how to avoid pitfalls and traps in macromolecular structure determination', *FEBS Journal*, 280(22), pp. 5705–5736. doi: 10.1111/febs.12495.

Woestenenk, E. A., Hammarström, M., van den Berg, S., Härd, T. and Berglund, H. (2004) 'His tag effect on solubility of human proteins produced in *Escherichia coli*: a comparison between four expression vectors', *Journal of Structural and Functional Genomics*, 5(3), pp. 217–229. doi: 10.1023/B:jsfg.0000031965.37625.0e.

Wolf, I. M., Heitzer, M. D., Grubisha, M. and DeFranco, D. B. (2008) 'Coactivators and nuclear receptor transactivation', *Journal of Cellular Biochemistry*, 104(5), pp. 1580–1586. doi: 10.1002/jcb.21755.

Wong, C. W., Komm, B. and Cheskis, B. J. (2001) 'Structure - Function evaluation of ER α and β interplay with SRC family coactivators. ER selective ligands', *Biochemistry*, 40(23), pp. 6756–6765. doi: 10.1021/bi010379h.

Wu, J., Xia, C., Meier, J., Li, S., Hu, X. and Lala, D. S. (2002) 'The Hypolipidemic Natural Product Guggulsterone Acts as an Antagonist of the Bile Acid Receptor', *Molecular Endocrinology*, 16(7), pp. 1590–1597. doi: 10.1210/mend.16.7.0894.

Xiao, H. et al. (2017) 'Synthesis and Biological Evaluation of a Series of Bile Acid Derivatives as FXR Agonists for Treatment of NASH', *ACS Medicinal Chemistry Letters*, 8(12), pp. 1246–1251. doi: 10.1021/acsmedchemlett.7b00318.

Xie, Y., Wang, H., Cheng, X., Wu, Y., Cao, L., Wu, M., Xie, W., Wang, G. and Hao, H. (2016) 'Farnesoid X receptor activation promotes cell proliferation via PDK4-controlled metabolic reprogramming', *Scientific Reports*. The Author(s), 6, p. 18751. Available at: <https://doi.org/10.1038/srep18751>.

Yamagata, K., Daitoku, H., Shimamoto, Y., Matsuzaki, H., Hirota, K., Ishida, J. and Fukamizu, A. (2004) 'Bile acids regulate gluconeogenic gene expression via small heterodimer partner-mediated repression of hepatocyte nuclear factor 4 and Foxo1', *Journal of Biological Chemistry*, 279(22), pp. 23158–23165. doi: 10.1074/jbc.M314322200.

Yang, F., Huang, X., Yi, T., Yen, Y., Moore, D. D. and Huang, W. (2007) 'Spontaneous development of liver tumors in the absence of the bile acid receptor farnesoid X receptor', *Cancer Research*, 67(3), pp. 863–867. doi: 10.1158/0008-5472.CAN-06-1078.

Yang, L., Broderick, D., Campbell, Y., Gombart, A. F., Stevens, J. F., Jiang, Y., Hsu, V. L., Bisson, W. H. and Maier, C. S. (2016) 'Conformational Modulation of the Farnesoid X Receptor by Prenylflavonoids: Insights from Hydrogen Deuterium Exchange Mass Spectrometry (HDX-MS), Fluorescence Titration and Molecular Docking Studies', *Biochimica et Biophysica Acta (BBA) -*

Proteins and Proteomics. Elsevier B.V., 1864(12), pp. 1667–1677. doi: 10.1016/j.bbapap.2016.08.019.

Yang, Liping, Broderick, D., Jiang, Y., Hsu, V. and Maier, C. S. (2014) 'Conformational dynamics of human FXR-LBD ligand interactions studied by hydrogen/deuterium exchange mass spectrometry: Insights into the antagonism of the hypolipidemic agent Z-guggulsterone', *Biochimica et Biophysica Acta - Proteins and Proteomics*. Elsevier B.V., 1844(9), pp. 1684–1693. doi: 10.1016/j.bbapap.2014.06.007.

Yang, Ling, Roh, Y. S., Song, J., Zhang, B., Liu, C., Loomba, R. and Seki, E. (2014) 'Transforming growth factor beta signaling in hepatocytes participates in steatohepatitis through regulation of cell death and lipid metabolism in mice', *Hepatology (Baltimore, Md.)*. 2013/12/18, 59(2), pp. 483–495. doi: 10.1002/hep.26698.

Ye, J., Coulouris, G., Zaretskaya, I., Cutcutache, I., Rozen, S. and Madden, T. L. (2012) 'Primer-BLAST: A tool to design target-specific primers for polymerase chain reaction', *BMC Bioinformatics*, 13(1), p. 134. doi: 10.1186/1471-2105-13-134.

Yeh, S. and Chang, C. (1996) 'Cloning and characterization of a specific coactivator, ARA70, for the androgen receptor in human prostate cells', *Proceedings of the National Academy of Sciences of the United States of America*, 93(11), pp. 5517–5521. doi: 10.1073/pnas.93.11.5517.

Younossi, Z., Anstee, Q. M., Marietti, M., Hardy, T., Henry, L., Eslam, M., George, J. and Bugianesi, E. (2018) 'Global burden of NAFLD and NASH: Trends, predictions, risk factors and prevention', *Nature Reviews Gastroenterology and Hepatology*. Nature Publishing Group, 15(1), pp. 11–20. doi: 10.1038/nrgastro.2017.109.

Yu, H. (1999) 'Extending the size limit of protein nuclear magnetic resonance', *Proceedings of the National Academy of Sciences*, 96(2), pp. 332 LP – 334. doi: 10.1073/pnas.96.2.332.

Zhang, D. G., Zhang, C., Wang, J. X., Wang, B. W., Wang, H., Zhang, Z. H., Chen, Y. H., Lu, Y., Tao, L., Wang, J. Q., Chen, X. and Xu, D. X. (2017a) 'Obeticholic acid protects against carbon tetrachloride-induced acute liver injury and inflammation', *Toxicology and Applied Pharmacology*. Elsevier Inc., 314, pp. 39–47. doi: 10.1016/j.taap.2016.11.006.

Zhang, J.-H., Chung, T. D. Y. and Oldenburg, K. R. (1999) 'A Simple Statistical Parameter for Use in Evaluation and Validation of High Throughput Screening Assays', *Journal of Biomolecular Screening*, 4(2), pp. 1–8. Available at: papers2://publication/uuid/D3051551-18C0-4225-98D3-

Zhang, J., Chalmers, M. J., Stayrook, K. R., Burris, L. L., Wang, Y., Busby, S. A., Pascal, B. D., Garcia-Ordonez, R. D., Bruning, J. B., Istrate, M. A., Kojetin, D. J., Dodge, J. A., Burris, T. P. and Griffin, P. R. (2011) 'DNA binding alters coactivator interaction surfaces of the intact VDR–RXR complex', *Nature Structural & Molecular Biology*. Nature Publishing Group, a division of Macmillan Publishers Limited. All Rights Reserved., 18, p. 556. Available at: <https://doi.org/10.1038/nsmb.2046>.

Zhang, J., Huang, W., Qatanani, M., Evans, R. M. and Moore, D. D. (2004) 'The constitutive androstane receptor and pregnane X receptor function coordinately to prevent bile acid-induced hepatotoxicity', *Journal of Biological Chemistry*, 279(47), pp. 49517–49522. doi: 10.1074/jbc.M409041200.

Zhang, S., Wang, J., Liu, Q. and Harnish, D. C. (2009) 'Farnesoid X receptor agonist WAY-362450 attenuates liver inflammation and fibrosis in murine model of non-alcoholic steatohepatitis', *Journal of Hepatology*. European Association for the Study of the Liver, 51(2), pp. 380–388. doi: 10.1016/j.jhep.2009.03.025.

Zhang, T., Zhou, J., Shi, L., Zhu, R. and Chen, M. (2007) '3D-QSAR studies with the aid of molecular docking for a series of non-steroidal FXR agonists', *Bioorganic & Medicinal Chemistry Letters*, 17, pp. 2156–2160. doi: 10.1016/j.bmcl.2007.01.079.

Zhang, Y., Kast-woelbern, H. R. and Edwards, P. A. (2003) 'Natural Structural Variants of the Nuclear Receptor Farnesoid X Receptor Affect Transcriptional Activation', *Journal of Biological Chemistry*, 278(1), pp. 104–110. doi: 10.1074/jbc.M209505200.

Zhang, Y., Lee, F. Y., Barrera, G., Lee, H., Vales, C., Gonzalez, F. J., Willson, T. M. and Edwards, P. A. (2006) 'Activation of the nuclear receptor FXR improves hyperglycemia and hyperlipidemia in diabetic mice.', *Proceedings of the National Academy of Sciences of the United States of America*, 103(4), pp. 1006–1011. doi: 10.1073/pnas.0506982103.

Zhang, Y., Yin, L., Anderson, J., Ma, H., Gonzalez, F. J., Willson, T. M. and Edwards, P. A. (2010) 'Identification of novel pathways that control farnesoid X receptor-mediated hypocholesterolemia', *Journal of Biological Chemistry*, 285(5), pp. 3035–3043. doi: 10.1074/jbc.M109.083899.

Zheng, W., Lu, Y., Tian, S., Ma, F., Wei, Y., Xu, S. and Li, Y. (2018) 'Structural insights into the heterodimeric complex of the nuclear receptors FXR and RXR', *Journal of Biological Chemistry*,

293(32), pp. 12535–12541. doi: 10.1074/jbc.RA118.004188.

Zhou, H. and Hylemon, P. B. (2014) 'Bile acids are Nutrient Signalling Hormones', *Steroids*, 0, pp. 62–68. doi: 10.1016/j.immuni.2010.12.017.Two-stage.

Appendix 1. Primers used for In Fusion cloning FXR domain constructs

Construct	Forward primer sequence (5' – 3')	Reverse primer sequence (5' – 3')
A1	AGGAGATATACCATGGTATGCAATTCAAGG TCTG	GTGATGGTATGTTCTGAACATCCCAAATC TCGC
B1	AAGTTCTGTTCAGGGCCCGGTATGCAATT AAGGTCTG	ATGGTCTAGAAAGCTTACTGAACATCCAA ATCTCGC
C1	AAGTTCTGTTCAGGGCCCGGTATGCAATT AAGGTCTG	ATGGTCTAGAAAGCTTACTGAACATCCAA ATCTCGC
D1	AAGTTCTGTTCAGGGCCCGGTATGCAATT AAGGTCTG	ATGGTCTAGAAAGCTTACTGAACATCCAA ATCTCGC
E1	AAGTTCTGTTCAGGGCCCGGTATGCAATT AAGGTCTG	ATGGTCTAGAAAGCTTACTGAACATCCAA ATCTCGC
F1	AGGAGATATACCATGGACGAAGTGTGCGTGG TTG	GTGATGGTATGTTCCAAATCTGCACAGC AGC
G1	AAGTTCTGTTCAGGGCCCGGACGAAGTGTGC GTGGTTTG	ATGGTCTAGAAAGCTTACCAAATCTGCAC AGCAGC
H1	AAGTTCTGTTCAGGGCCCGGACGAAGTGTGC GTGGTTTG	ATGGTCTAGAAAGCTTACCAAATCTGCAC AGCAGC
A2	AAGTTCTGTTCAGGGCCCGGACGAAGTGTGC GTGGTTTG	ATGGTCTAGAAAGCTTACCAAATCTGCAC AGCAGC
B2	AAGTTCTGTTCAGGGCCCGGACGAAGTGTGC GTGGTTTG	ATGGTCTAGAAAGCTTACCAAATCTGCAC AGCAGC
C2	AGGAGATATACCATGGTCGTATCTCGTCA AG	GTGATGGTATGTTCCAAATCTGCACAGC AGC
D2	AAGTTCTGTTCAGGGCCCGGTCGTATCTGC GTCAAG	ATGGTCTAGAAAGCTTACCAAATCTGCAC AGCAGC
E2	AAGTTCTGTTCAGGGCCCGGTCGTATCTGC GTCAAG	ATGGTCTAGAAAGCTTACCAAATCTGCAC AGCAGC
F2	AAGTTCTGTTCAGGGCCCGGTCGTATCTGC GTCAAG	ATGGTCTAGAAAGCTTACCAAATCTGCAC AGCAGC
G2	AGGAGATATACCATGGTATGCAATTCAAGG TCTG	GTGATGGTATGTTCAGTCGTACCCCTTA ATACG
H2	AAGTTCTGTTCAGGGCCCGGTATGCAATT AAGGTCTG	ATGGTCTAGAAAGCTTACAGTCGTACCC TTAATACG
A3	AAGTTCTGTTCAGGGCCCGGTATGCAATT AAGGTCTG	ATGGTCTAGAAAGCTTACAGTCGTACCC TTAATACG
B3	AAGTTCTGTTCAGGGCCCGGTATGCAATT AAGGTCTG	ATGGTCTAGAAAGCTTACAGTCGTACCC TTAATACG
C3	AGGAGATATACCATGGTATGCAATTCAAGG TCTG	GTGATGGTATGTTTCGCTGTCCTCGTTC ACG
D3	AAGTTCTGTTCAGGGCCCGGTATGCAATT AAGGTCTG	ATGGTCTAGAAAGCTTATTCGCTGTCCTCG TTCACG
E3	AAGTTCTGTTCAGGGCCCGGTATGCAATT AAGGTCTG	ATGGTCTAGAAAGCTTATTCGCTGTCCTCG TTCACG
F3	AAGTTCTGTTCAGGGCCCGGTATGCAATT AAGGTCTG	ATGGTCTAGAAAGCTTATTCGCTGTCCTCG TTCACG
G3	AAGTTCTGTTCAGGGCCCGGTATGCAATT AAGGTCTG	ATGGTCTAGAAAGCTTATTCGCTGTCCTCG TTCACG

Appendix 2. Amino Acid Sequence of FXR Isoform 1(+)



N.B. LBD binding residues in Chapter 3 uses alternative aa numbering sequence. Alternative residue numbers are shown here after the '/'

Appendix 3. Vector Maps of Constructs used in this Thesis



Figure A3. 1 Plasmid map of pOPINE-FXR as used for constructs A1,F1,C2,G2,C3. Created in Snap Gene



Figure A3. 2 Plasmid map of pOPINF-FXR as used for constructs B1,G1,D2,H2,D3. Created in Snap Gene.

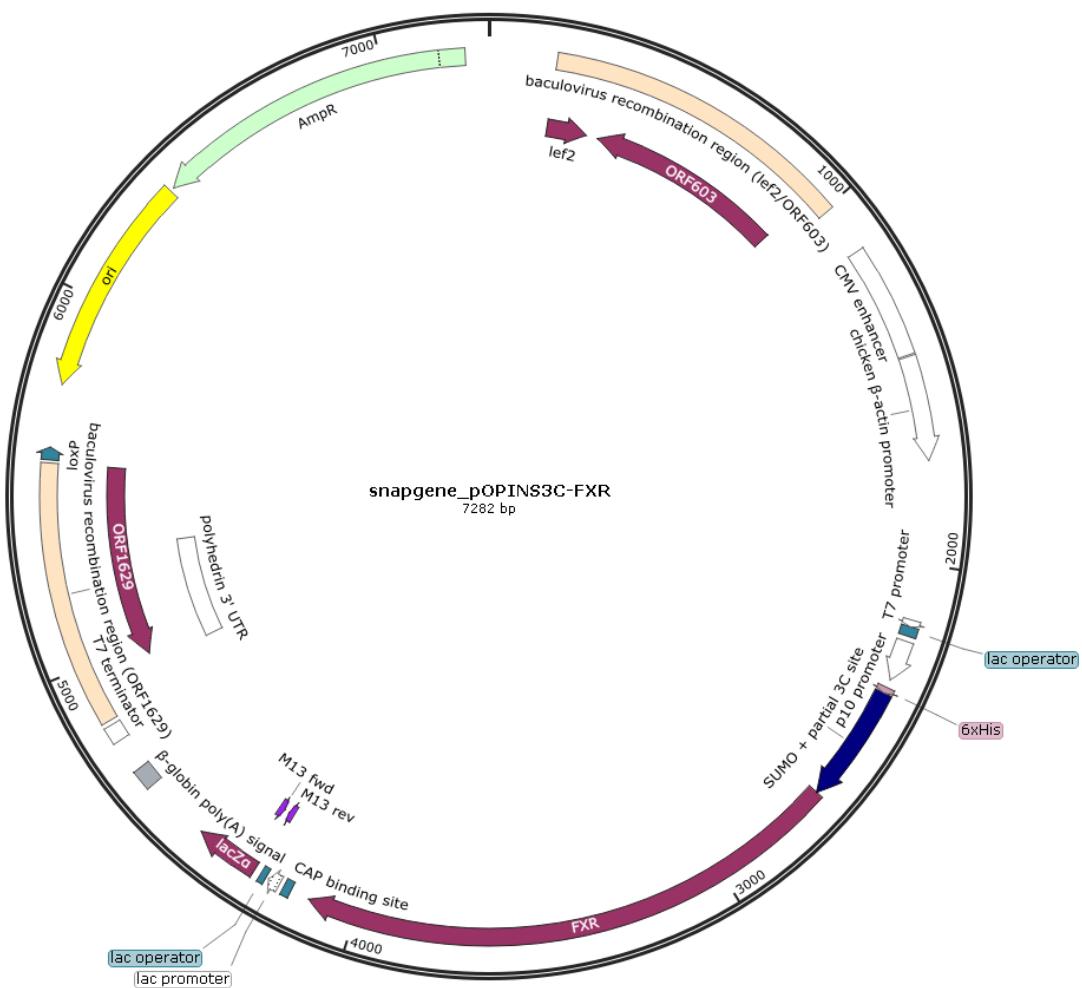


Figure A3. 3 Plasmid map of pOPINS3C-FXR as used for constructs C1,H1,E2,A3,E3. Created in Snap Gene.

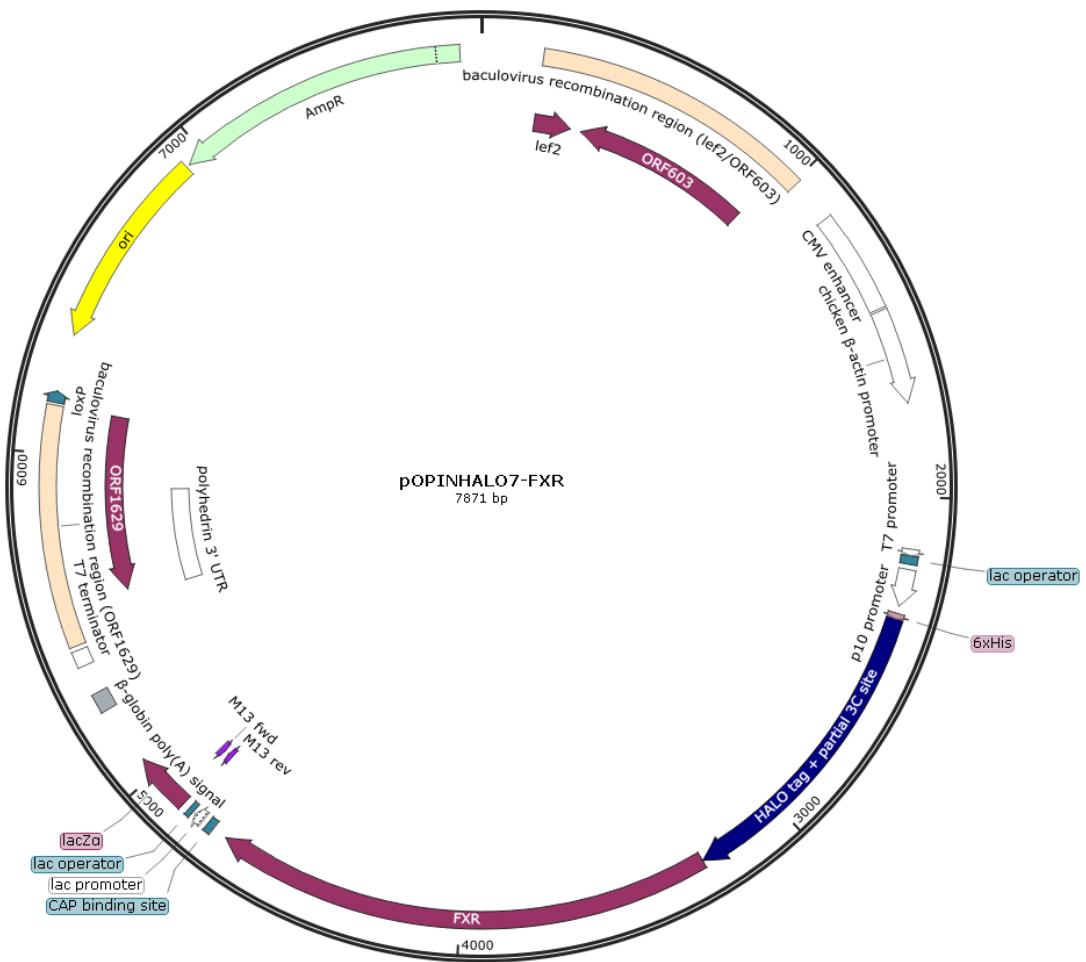


Figure A3. 4 Plasmid map of pOPINHALO7-FXR as used for constructs D1,A2,F2,D3,F3. Created in Snap Gene.

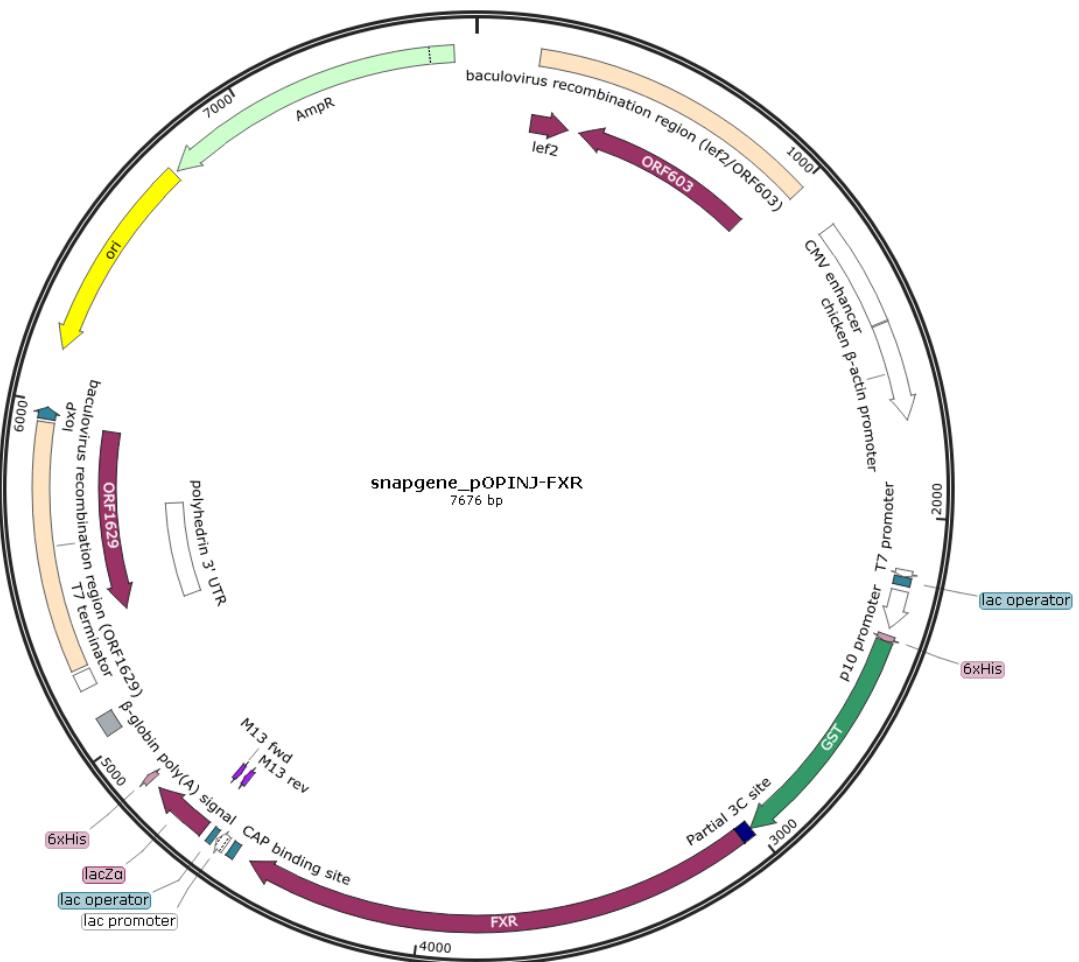


Figure A3. 5 Plasmid map of *pOPINJ-FXR* as used for constructs *E1, B2, F2, B3, G3*. Created in Snap Gene.

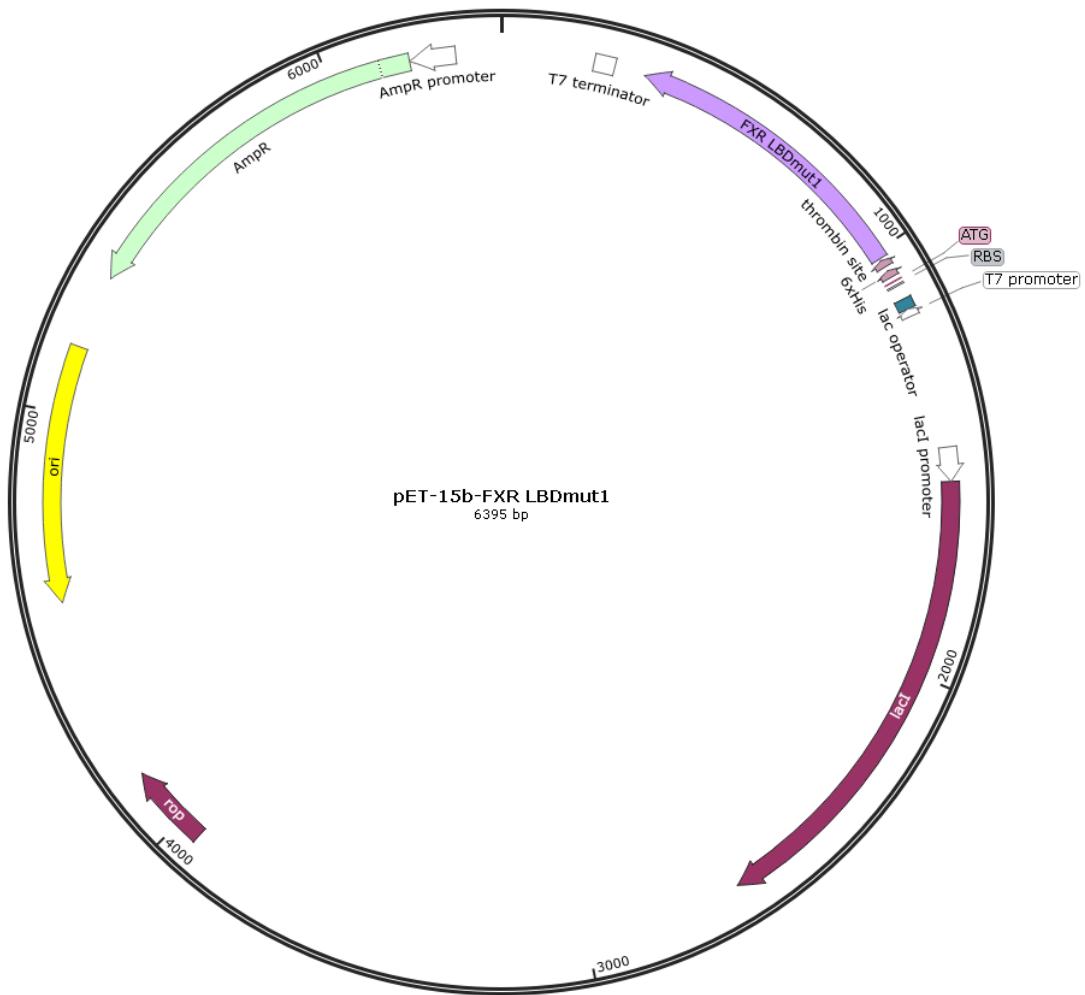


Figure A3. 6 Plasmid map of pET15b-FXR LBDmut1. Created in Snap Gene.

Appendix 4. RNAseq data of coactivator mRNA expression from Expression Atlas data base, as provided by Genotype Tissue Expression (Gtex) Project

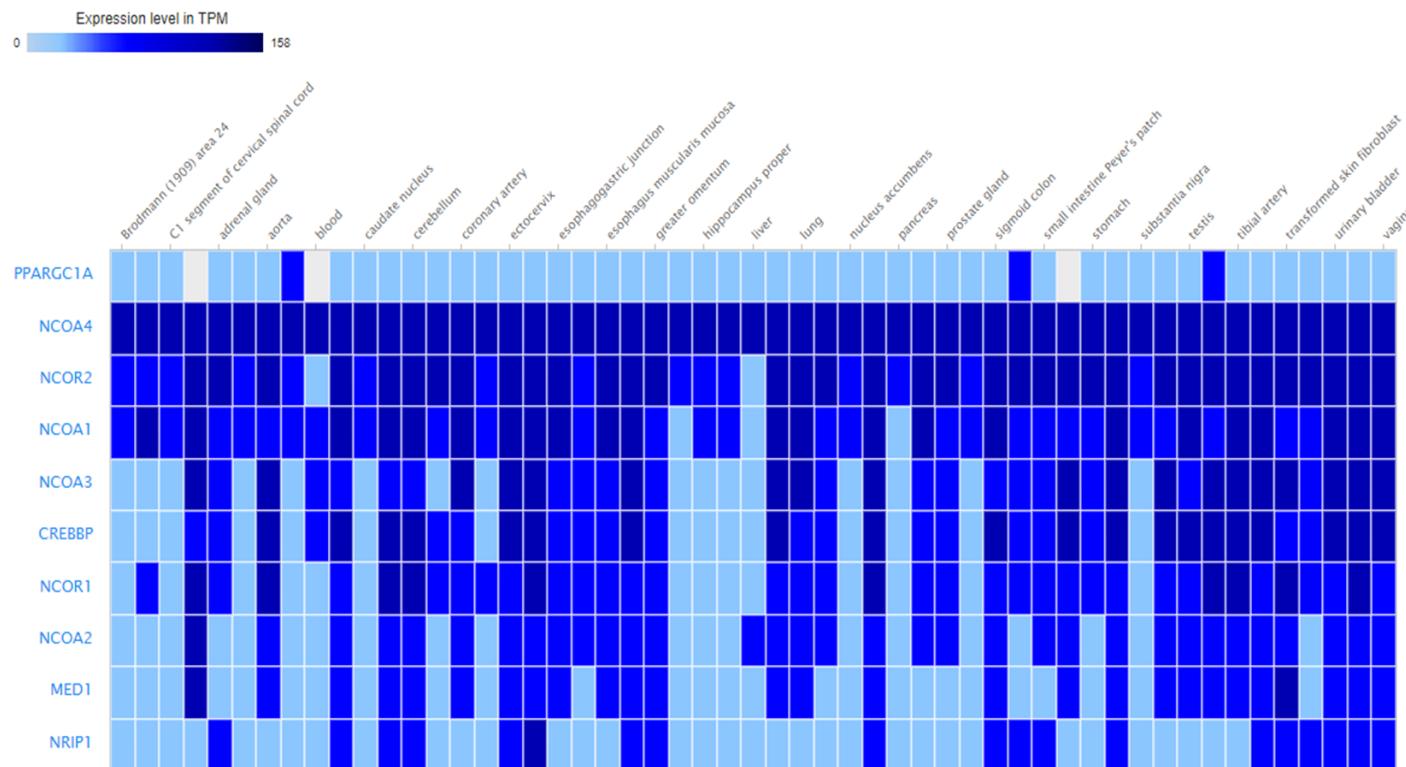


Figure A4. 1 Expression levels of coregulator proteins in different tissues. Data from Genotype Tissue Expression (Gtex) project, accessed on Gene expression atlas website (www.ebi.ac.uk/gxa). Data displayed as transcripts per kilobase million (TPM). Where PPARGC1A is PGC1a; NCOA4 is ARA70; NCOR2 is SMRT; NCOA1 is SRC1; NCOA3 is SRC3; CREBBP is CBP1; NCOA2 is SRC2; MED1 is TRAP220; and NRIP1 is RIP140.

Appendix 5. RNAseq data from HepG2 and Huh7 cells under basal conditions

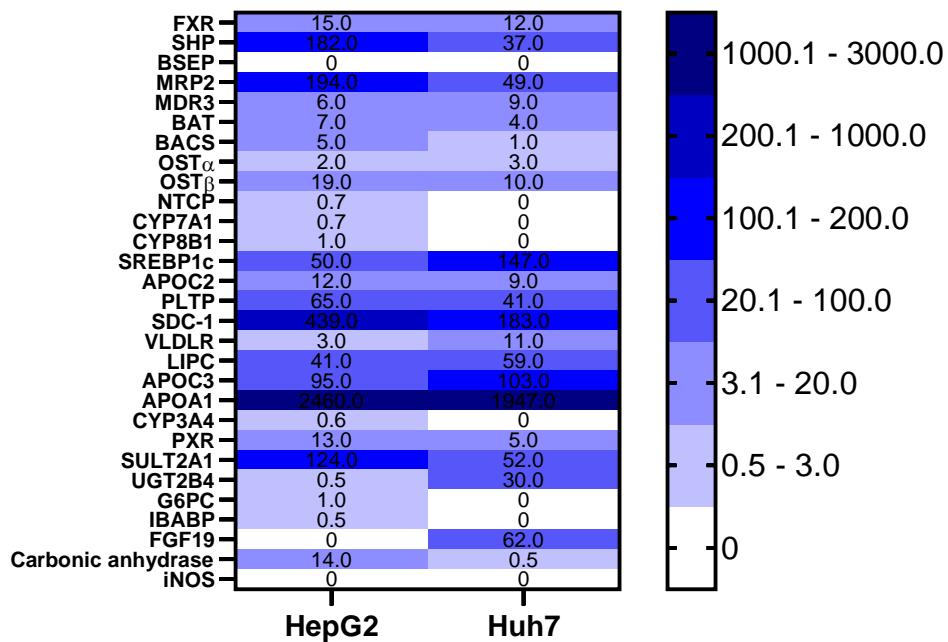


Figure A5. 1 Transcript levels of FXR target genes found in HepG2 and Huh7 cell lines under basal conditions. Data collated from Expression Atlas database provided by Genentech and Cancer Cell Line Encyclopaedia projects. Data shown as transcripts per million.

Appendix 6. RNAseq data from C57BL/6J adult Mice Livers under normal physiological conditions

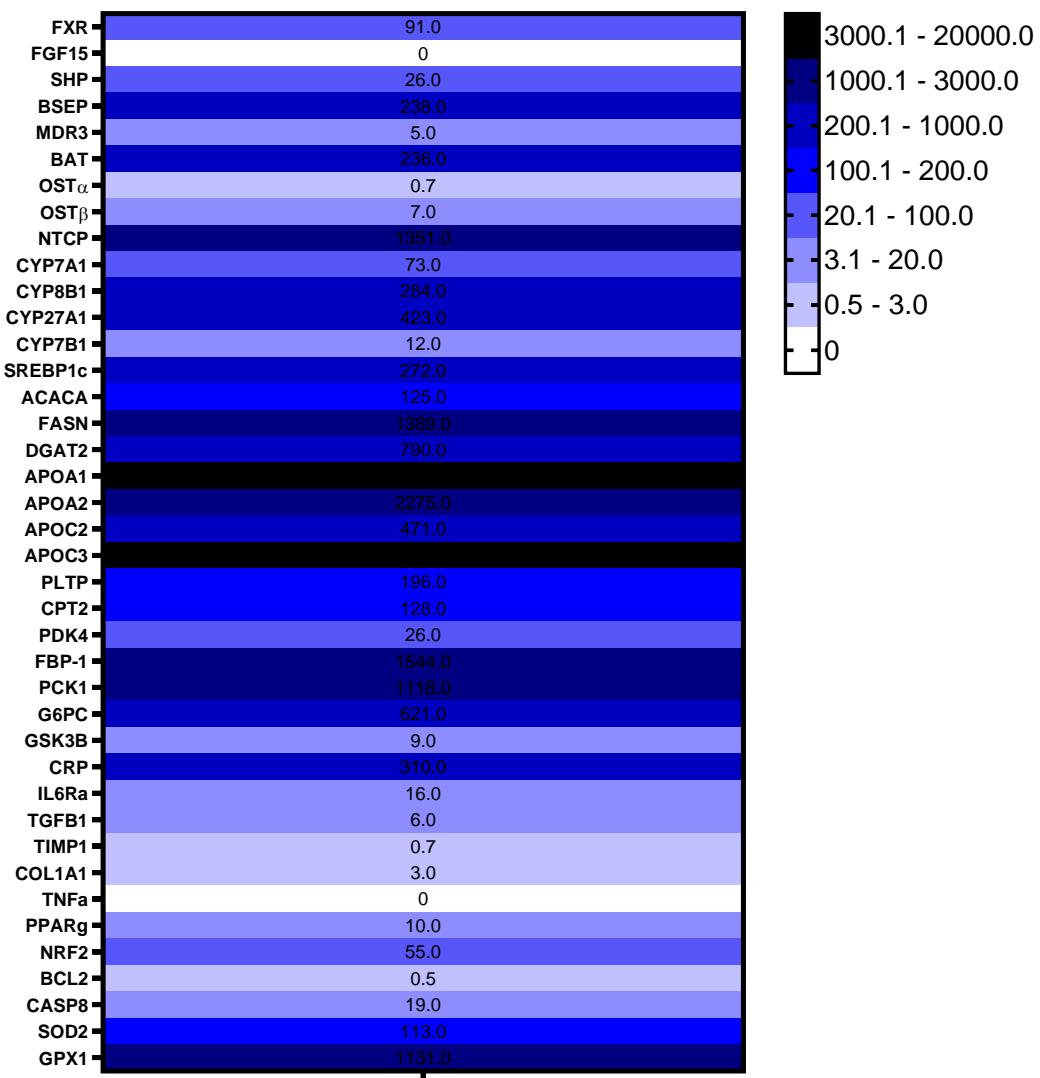


Figure A6. 1 Transcript levels of FXR target genes found in C57BL/6J adult mice under normal physiological conditions. Data collated from Expression Atlas database provided by (Bonthuis et al., 2015). Data shown as transcripts per million.

Appendix 7. Validation of primers for use on cDNA from Huh7 and HepG2 cells

Table A7. 1 Primers validated for use on cDNA from Human cancer cell lines

Primer	Pathway	cDNA from Huh7			cDNA from HepG2		
		R2 Value	Amplification Efficiency %	Single Melt Curve Peak?	R2 Value	Amplification Efficiency %	Single Melt Curve Peak?
FXR_H_NR1H4_1	FXR Signalling	0.995	72	Y	0.988	77	Y
SHP_H_NROB2_1	FXR Signalling, BA secretion	0.999	83	Y	0.998	66	Y
BSEP_H_ABCB11_1	BA secretion	0.959	163	N	0.962	270	N
CYP7A1_H_CYP7A1_1	BA synthesis	0.722	436	N	0.98	84	N
CYP8B1_H_CYP8B1_1	BA synthesis	0.965	125	N	0.71	272	N
FGF19_H_FGF19_1	FXR Signalling, BA synthesis	0.996	97	Y	0.008	-100	N
IL1B_H_IL1B_1	Inflammation	0.645	715	N	0.055	-99	N
CCR5_H_CCR5_1	Inflammation	0.959	158	N	0.187	3396	N
TGFB1_H_TGFB1_1	Fibrosis, tissue remodelling, Monocyte signalling	0.985	97	Y	0.98	80	Y
TIMP1_H_TIMP1_1	Fibrosis, Monocyte activation	0.982	109	N	0.996	86	N
APOC2_H_APOC2_1	Lipoprotein Metabolism	0.999	80	Y	0.998	64	Y
SREBF1_H_SREBF1_1	Lipid metabolism, cholesterol synthesis	0.999	89	Y	0.594	1182	N
NTCP_H_SLC10A1_1	BA transport	0.834	165	N	0.810	1877	N
OSTA_H_OSTALPHA_1	BA transport	0.999	92	Y	0.995	87	Y
OSTB_H_OSTBETA_1	BA transport	0.940	82	Y	0.870	130	N
GAPDH_H_GAPDH_1	Housekeeping,	1.000	73	Y	0.999	68	Y
B-actin_H_ACTB_1	Housekeeping	0.998	87	N	0.994	76	N

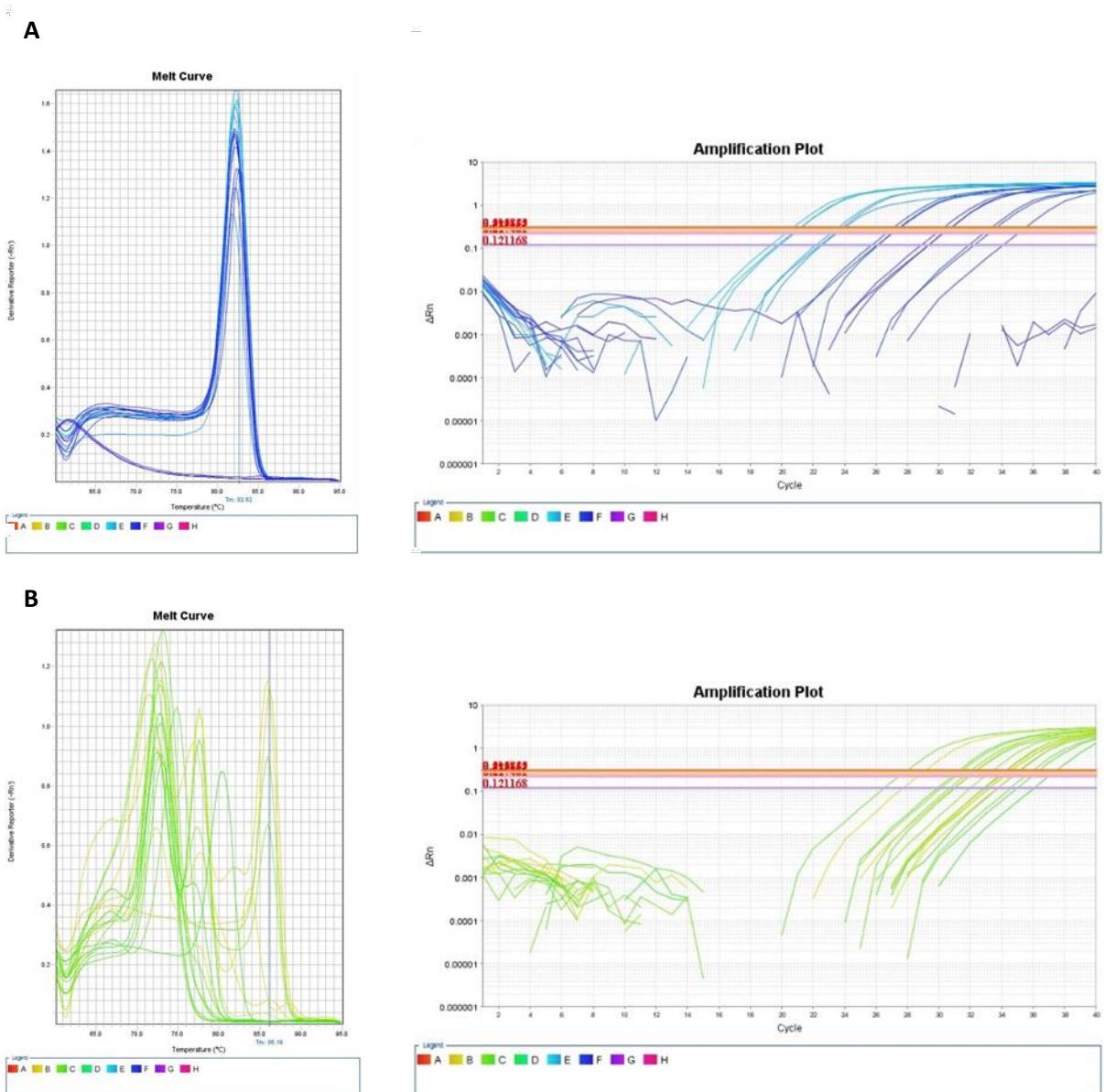


Figure A7. 1 Example of melt curve and amplification plots generated from primer validation tests using qPCR. cDNA was serially diluted and used in reactions with each primer. A fluorescence dissociation melt curve was performed at the end of each run. Primers were expected to produce a single melt curve as seen in (A) for the SHP primer, and amplification plots that were evenly spaced indicating each dilution of the series. Some primers did not achieve this, as seen in (B) for FGF19.

Appendix 8. Validation of Primers for Use on Mouse Liver cDNA

Table A8. 1 Table A7. 1 Primers validated for use on cDNA from C57BL/6J mouse livers

Primer	Pathway	R2 Value	Primer Efficiency (%)	Single Melt curve Peak Y/N
FXR	FXR Signalling	0.998	84	(Y)
FGF15	FXR Signalling	-	-91	N
SHP	FXR Signalling	1	83	(Y)
BSEP	FXR Signalling	0.998	89	(Y)
CYP7A1	BA synthesis	0.992	91	(Y)
CYP27A1	BA synthesis	1	83	(Y)
PLTP	Lipid/ Cholesterol Metabolism	0.996	102	(Y)
APOC2	Lipoprotein/Cholesterol metabolism	0.999	99	(Y)
SREBF1	Lipid/Cholesterol Metabolism	1	77	(Y)
ACACA	Triglyceride synthesis	1	89	(Y)
FASN	Triglyceride synthesis	1	81	(Y)
CPT2	Fatty Acid B oxidation	0.992	101	N
TGFB1	Fibrosis, tissue remodelling, Monocyte signalling	-	-	Y
TIMP1	Fibrosis, tissue remodelling, Monocyte signalling	-	-	Y
ACTA2	Fibrosis, tissue remodelling, Monocyte signalling	-	-	Y
COL1A1	Fibrosis, tissue remodelling,	-	-	N
COL3A1	Fibrosis, tissue remodelling,	-	-	Y
BCL2	Apoptosis	0.743	523	N
CASP8	Apoptosis	0.935	101	(Y)
SOD2	Oxidative Stress	1	75	Y
GPX1	Oxidative Stress	1	94	(Y)
NFE2L2 (Nrf2)	Oxidative Stress	0.990	100	(Y)
GYS2	Glycogen Synthesis	1	88	Y
PCK1	Gluconeogenesis	1	78	(Y)

IL6Ra	Inflammation	0.985	79	(Y)
CRP	Inflammation	1	86	Y
CCL2	Inflammation			Y
CCR2	Inflammation			Y
ARF1	Housekeeping,	1	83	(Y)
COX7A2L	Housekeeping	1	89	Y
YWHAZ	Housekeeping	0.995	79	Y

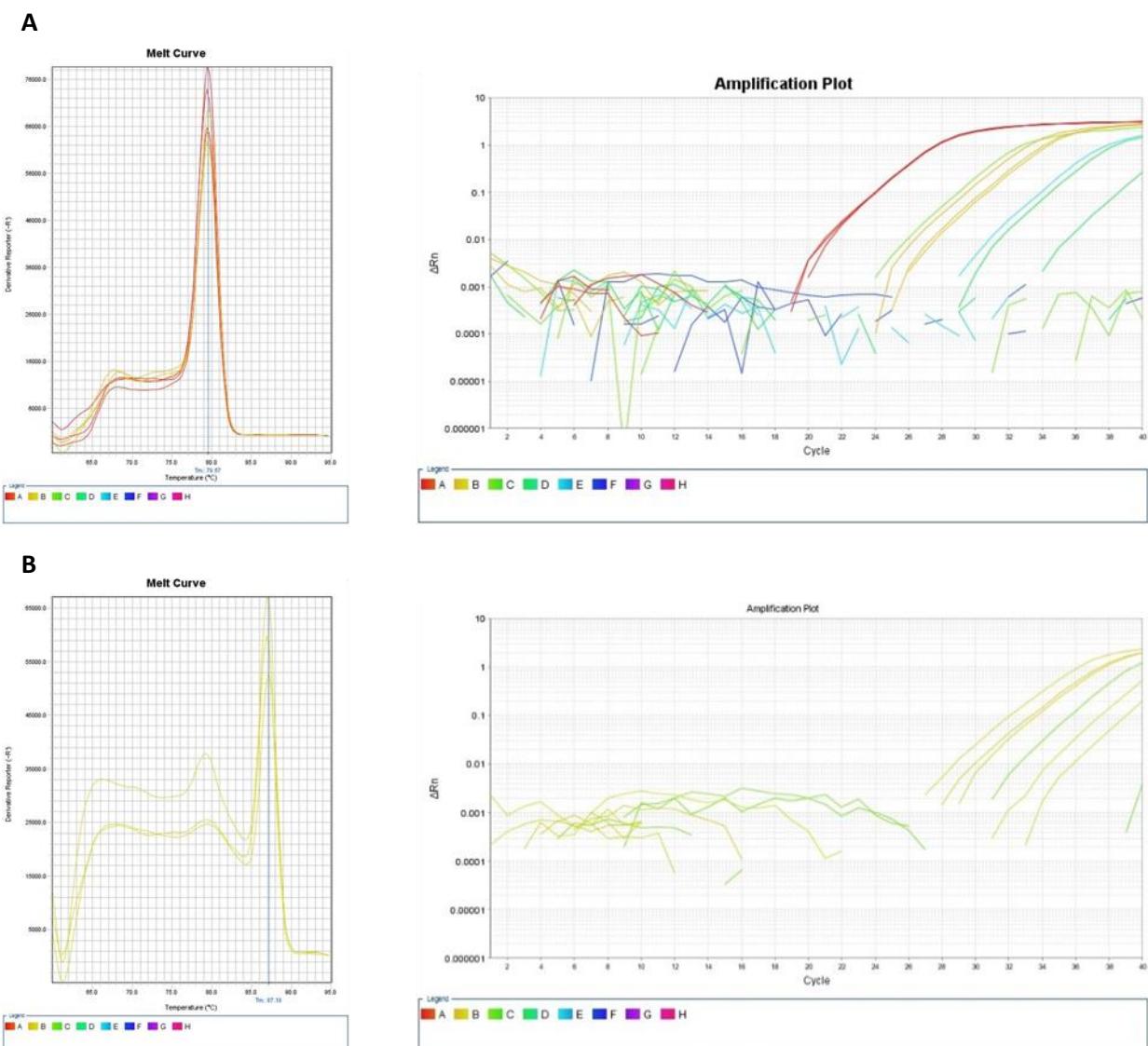


Figure A8. 1 Example of melt curve and amplification plots generated from primer validation tests using qPCR. cDNA was serially diluted and used in reactions with each primer. A fluorescence dissociation melt curve was performed at the end of each run. Primers were expected to produce a single melt curve as seen in (A) for the COL3A1 primer, and amplification plots that were evenly spaced indicating each dilution of the series. Some primers did not achieve this, as seen in (B) for COL1A1.

Appendix 9. Analysis of 'NASH' mice: body weights and histology scores – data provided by Sarettus Ltd.

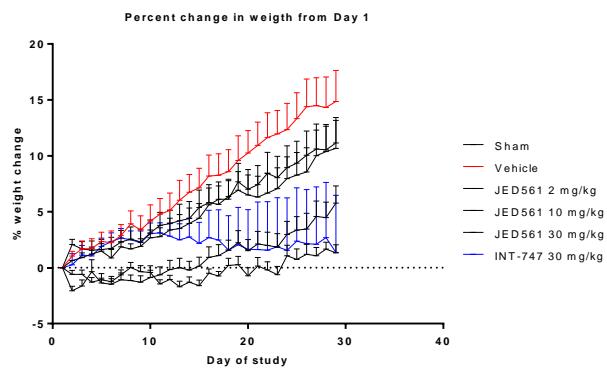


Figure A9. 2 The effects of JED561 (2, 10 and 30 mg/kg) and INT-747 (OCA, 30 mg/kg) on high fat diet and CCl4-induced steatohepatitis in male C57BL/6J mice- Daily weights. Data are mean \pm sem, n=9-10. ** p<0.01 and *** p<0.001 significantly different from HFD/vehicle treatment by Dunnett's post hoc test following significant one-way ANOVA

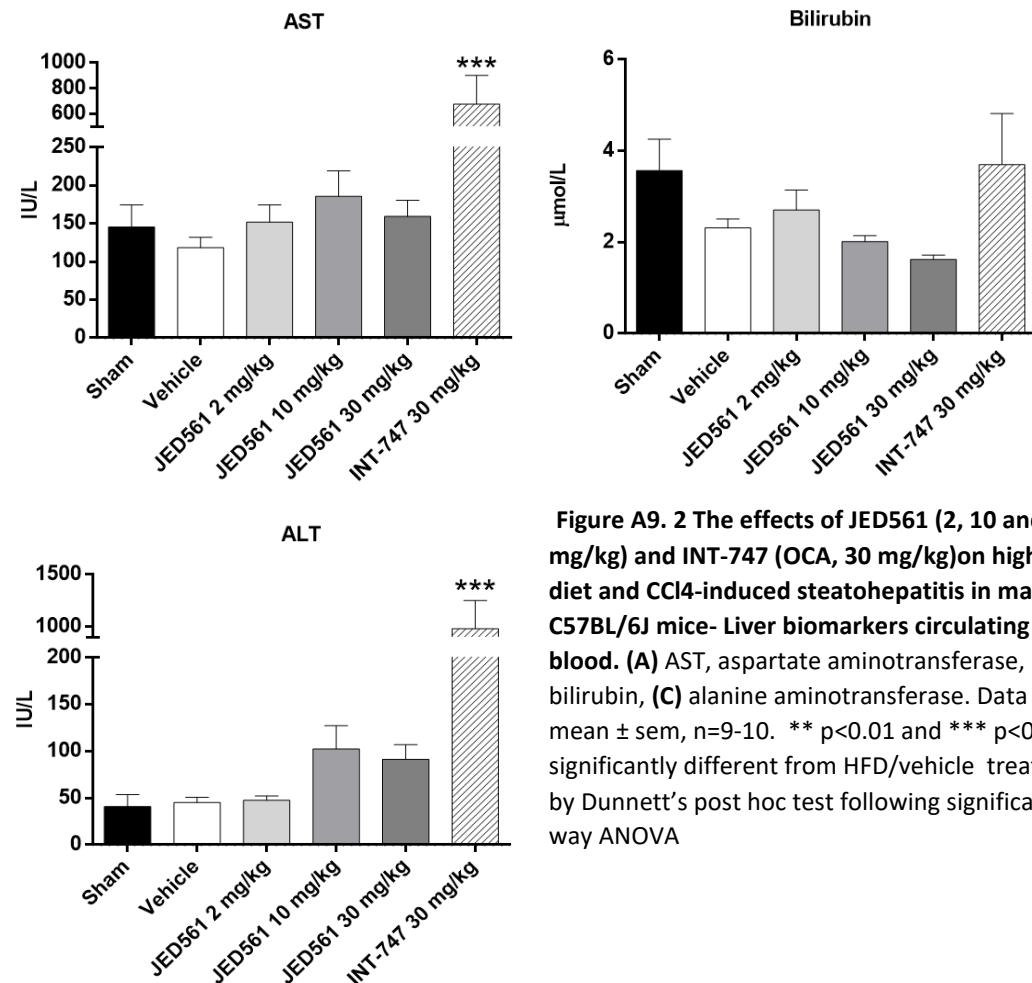


Figure A9. 2 The effects of JED561 (2, 10 and 30 mg/kg) and INT-747 (OCA, 30 mg/kg) on high fat diet and CCl4-induced steatohepatitis in male C57BL/6J mice- Liver biomarkers circulating in blood. (A) AST, aspartate aminotransferase, (B) bilirubin, (C) alanine aminotransferase. Data are mean \pm sem, n=9-10. ** p<0.01 and *** p<0.001 significantly different from HFD/vehicle treatment by Dunnett's post hoc test following significant one-way ANOVA

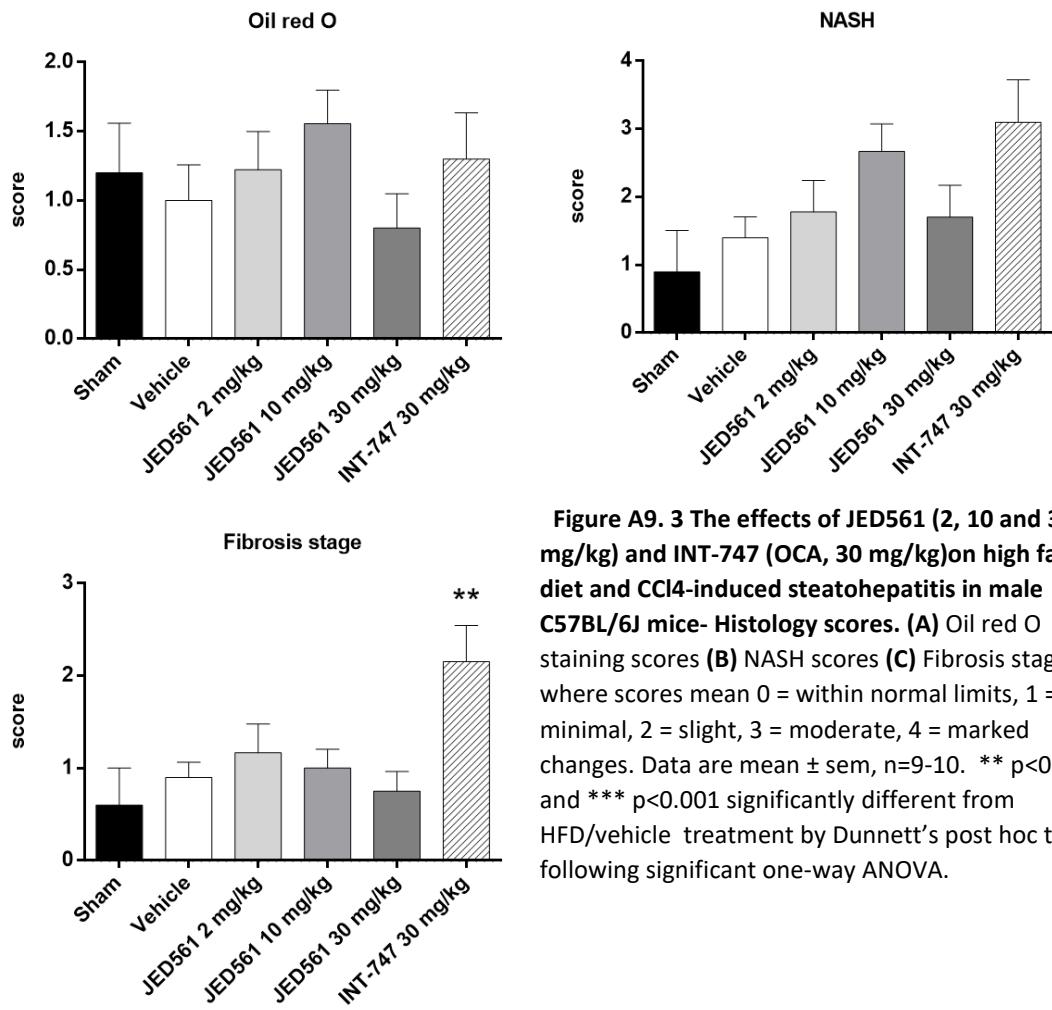


Figure A9. 3 The effects of JED561 (2, 10 and 30 mg/kg) and INT-747 (OCA, 30 mg/kg) on high fat diet and CCl4-induced steatohepatitis in male C57BL/6J mice- Histology scores. (A) Oil red O staining scores (B) NASH scores (C) Fibrosis stage, where scores mean 0 = within normal limits, 1 = minimal, 2 = slight, 3 = moderate, 4 = marked changes. Data are mean \pm sem, n=9-10. ** p<0.01 and * p<0.001 significantly different from HFD/vehicle treatment by Dunnett's post hoc test following significant one-way ANOVA.**