Purification of Liver X Receptors

“A slippery slope through half-charted territory”

Øyvind B. Langgård

Master’s thesis – Department of Chemistry

UNIVERSITY OF OSLO

05.06.2014

Purification of Liver X Receptors

“A slippery slope through half-charted territory”

”I have not failed. I've just found 10,000 ways that won't work.”

Thomas A. Edison

2014

Preparing the Liver X Nuclear Receptors for Structural Studies

http://www.duo.uio.no

Trykk: Reprosentralen, Universitetet i Oslo

Acknowledgements

This thesis concludes my Master degree at the Department of Chemistry, University of Oslo from September 2012 to July 2014. Though many challenges and frustrations were encountered, it has been a joyful two years. I would never had imagined that I could learn and grow so much in such a short time. First of all I would like to thank my supervisors Professor Ute Krengel, Dr. Pål Rongved and Dr. Daniel Burschowsky.

To Ute: I am grateful for you taking me on as a master's student. You are an excellent supervisor and group leader, always with good ideas for solving unexpected problems encountered in the lab, whether the good idea is to try out something different or just taking a cup of coffee and clear the mind. I am glad I have been a part of your group, and am confident enough to say that it is the best group at the Department of Chemistry. "Coffee tiiiiiime!!"

Secondly I would like to thank all former and present members of the Protein Dungeon; Tyskerkroken with Steffi, Joel, André (and Dani). Carl-Henrik, though you're not a listed member, you still brighten up the dungeon with your many languages, quirky facts, excellent birthdaysong-voice and journal club presentation style. Svenskakontoret with Paula, Hedda, Gabriele, Dipankar and Betty - You guys are awesome, and I wouldn't have had as many fond memories and been as wise as I am today without you. Hedda, thank you for always turning my frustrations and tantrums into productive work. Paula, thank you for being so swedish and adorable. Gabriele, pastapastapizzapizza. Dipankar, keep on practicing that Norwegian accent. And finally to my mates at l'office awesome. Mr beerbottle cap HattihattihattiHoe! You came in to l'office awesome and was immediately a part of the gang. You brighten our day with your lovely dialect and catchphrases. I will always treasure our nitrogen-research, and hope that our patent will one day be reality (this also applies to Heggis). Your invaluable stories and partyskills shall never be forgotten!
My science sister Heggis/Muselund/Heggelund/Fyllelund/Frydenlund, you have made the past two years great, though I was the weird kid at first, you still let me get into the cool club. You are always prepared to answer my weird/dumb/intelligent questions and have a debate. You have helped me throughout the project and taught all the secret tricks of conferences. You have actually been as a science bigsister.

Dani, though you left us for a quieter corner of the dungeon, I still see you as a member of l'office awesome. I am forever grateful for all your help and discussions on the project. During my two years here you have become a true friend. Your infamous whiskey-tastings are awesome, and I feel honoured to have been invited to them. You and Nadja are excellent hosts and friends.

And to y'all: On the 26th, the beer "Langgård *et al.* HümleBümle" is on me!!

Love
– Øyvind

Abstract

More and more people suffer from lifestyle diseases, such as cardiovascular diseases, diabetes and obesity. All of these diseases are in some way related to metabolism. Thousands of proteins and factors are involved in metabolism, making the pathways and interactions extremely complex. However, underlying it all are the genes that are encoding the proteins involved. Liver X Receptors (LXRs) are transcription factors in the nuclear receptor family. The LXRs are sensors in the cell and are activated by binding metabolites from cholesterol, glucose and fatty acid metabolism, before binding a huge variety of genes encoding proteins involved in lipid homeostasis, thus regulating pathways and interactions. Two isoforms of LXR have been identified, the LXRα and LXRβ. Animal and cell studies have shown that manipulation of LXR expression has an effect on the expected diseases such as atherosclerosis, diabetes, inflammation and obesity, but also Alzheimer's disease and several types of cancer. Most of the ligands identified for LXRs are oxysterols, making LXRs ideal pharmaceutical targets for drugs against metabolic maladies. For structure-guided drug design it is important to obtain the structures of the LXR ligand-binding domains (LBD) with a ligand bound. The LBDs of the LXRs have been purified and crystallized by several groups, but no structures have been solved with both of the isoforms bound to the same endogenous ligand. The aim of this Master's project was to express, purify and crystallize the LXRs ligand binding domains alone and in complex with the same endogenous and synthetic ligands, to be ready for structural studies by X-ray crystallography. The project was expected to be straightforward, however this was not the case as several problems arose during the expression, purification and crystallization. The original construct yielded little or no soluble protein after harvest. We tried out different varying the conditions during expression, had a ligand present in all solutions and transformed the constructs into cold-adapted cells, but it yielded little to no soluble protein. The proteins were expressed in huge amounts, but in the form of inclusion bodies. We tried to purify the aggregated protein and refold the proteis *in vitro*, but the proteins aggregated when concentrated to a medium concentration. The small soluble fraction of proteins was purified and used for cleavage tests with enterokinase to cleave off the purification tag. Enterokinase turned out to be highly unspecific making the cleavage of the tag difficult. Crystallization trials were done with the tag still on, but did not yield any crystals. So we decided to clone the genes into a new vector, encoding periplasmic expression. Although the ultimate goal of the project never was reached, the road towards it has been highly educational.

Sammendrag

Stadig flere lider av livsstilssykdommer som forkalkning av blodårene, diabetes og fedme. Disse lidelsene fører ofte med seg de to hovedårsakene for dødsfall blant menn og kvinner, nemlig hjerte- og karsykdommer, og kreft. Alle disse lidelsene og de dødelige sykdommene som følger med dem, kan før eller siden relateres til metabolisme, hvor tusenvis av proteiner og faktorer er involvert. Syklusene og interaksjonene innenfor metabolismen kan være svært komplekse, men bakom er genene som styrer alt. Lever X Reseptorer (LXR) er transkripsjonsfaktorer innenfor kjernereseptorfamilien. To isoformer av LXR har blitt identifisert, LXRα og LXRβ, med 78% likhet i proteinsekvensene i deres to ligand-bindingsdomener. Dyrestudier og *in vivo* studier har vist at manipulering av LXR ekspresjon har en effekt på de forventede sykdommene som aterosklerose, diabetes, inflammasjon, fedme, og overraskende nok også Alzheimers sykdom og visse typer kreft. De fleste ligander identifisert for LXR er endogene oksysteroler, noe som gjør LXR til ideelle farmasøytiske mål for medisiner mot metabolisme relaterte sykdommer. For strukturbasert "drug design" er det viktig å oppnå strukturer av de ligand-bindene domenene (LBD) i LXR med ligand. LBDene har blitt produsert, renset og krystallisert av flere grupper, men ingen strukturer av LXR LBD er løst der isoformene er bundet til samme endogene ligand. Målet med dette masterprosjektet var å rekombinant produsere, rense og krystallisere LXR LBD i kompleks med samme endogene og syntetiske ligander. Det var forventet at prosjektet skulle være ukomplisert, men det viste seg å ikke være tilfellet. Under produksjon av proteinene oppsto det mange problemer. De originale konstruktene gav lite eller ikke noe løselig protein under produksjon, men store mengder av protein i uløselige inklusjonslegemer. Vi forsøkte å variere kondisjonene og ha en ligand tilstedeværende i alle løsninger under ekspresjon, og transformere konstruktene inn i kuldetilpassede celler, men alle forsøk gav de samme resultatene med uløselig protein. Vi forsøkte dermed å rense protein fra inklusjonlegemene og folde proteinene *in vitro*. Dette resulterte i ustabile proteiner som ble utfelt ved konsentrering. Det proteinet renset fra den lille løselige fraksjonen ble brukt til krystalliseringsforsøk. Dette gav ingen krystaller, antageligvis på grunn av rensetag'en som forsatt satt på proteinet. Rensetag'en viste seg å være vanskelig å få av ettersom enterokinase er svært uspesifikk. Genene for LXR LBD ble klonet inn i en annen vektor som koder for periplasmisk ekspresjon.

Table of Contents

List of Abbreviations

AE	ArcticExpress
Amp	ampicillin
Bis-Tris	(2-[Bis(2-hydroxyethyl)amino]-2-(hydroxymethyl)-1,3-propanediol
bp	base pair
BSA	bovine serum albumin
CV	column volume
CIP	calf-intestinal alkaline phosphatase
DNA	deoxyribonucleic acid
dNTP	deoxyribonucleoside triphosphate
dsDNA/ds	double-stranded DNA/double stranded
DTT	dithiothreitol
EDTA	ethylenediaminotetraacetic acid
EKMax™	EnterokinaseMax™ (Invitrogen)
ESRF	European Synchrotron Radiation Facility
FD	FastDigest
FPLC	fast performance liquid chromatography
HEPES	*N*-2-hydroxyethylpiperazine-*N*'2-ethanesulfonic acid
IPTG	isopropyl β-D-1-thiogalactopyranoside
kbp	kilo base pairs
LB	lysogeny broth
LBD	ligand binding domain
LMP	low melting point
LXR	Liver X nuclear Receptor
AU	absorption units
MES	2-(*N*-morpholino)ethanesulfonic acid
MQ-H_2O	Milli-Q filtered and ion-exchanged water
MW	molecular weight
MWCO	molecular weight cut off
NEB	New England Biolabs

Ni-NTA	nickel-nitrilotriacetic acid
NR	nuclear receptor
OD	optical density
PAGE	polyacrylamide gel electrophoresis
PBS	phosphate buffered saline
PCR	polymerase chain reaction
PDB	Protein Data Bank
pI	isoelectric point
PMSF	phenylmethanesulfonyl fluoride
rpm	revolutions per minute
SDS	sodium dodecyl sulfate
SOC	super optimal broth with glucose
ssDNA	single-stranded DNA
TAE	tris acetate EDTA
TCEP	tris-(2-carboxyethyl)phosphine
TEV	Tobacco Etch Virus
Tris	tris(hydroxymethyl)aminomethane
UV	ultraviolet
v/v	volume per volume
w/v	weight per volume
$(His)_6$-LXRα/β LBD	ligand binding domain of Liver X Receptor as fusion-protein with the N-terminal hexahistidine-tagged fusion peptide from pRSET B.

1 Introduction

Over weight and obesity is a well-known problem in the western world (*1*), and as rest of the world is getting "westernized", more and more young adults in Asia are overweight or obese (*2*). A consequence of this increase in obesity and overweight is an increase in obesity-related diseases like diabetes, atherosclerosis and other cardiovascular diseases (*3*). A known effect of obesity is the increase in insulin resistance, which heightens the risk of developing the previously mentioned diseases. According to the National Heart, Lung and Blood institute in the US, atherosclerosis in the coronary arteries is the main cause of death of both men and women in the United States of America (*4*). Studies have also shown that an unhealthy diet, which often is the cause of obesity, may lead to cancer (*5*). A correlation has been shown between obesity and colon cancer (*6*). Cancer is a leading cause of death worldwide and is a complex disease dependent on many factors. Naturally there is a high demand for new and improved drugs against these diseases, but unfortunately, approval of new drugs is a long and costly process. From discovery and modification of potential new drugs to an actual drug on the market, it takes usually 12-15 years, with an average cost of 4-11 billion dollars per approved drug. Even though the pharmaceutical industry is a billion dollar business, competition is high and the discovery of novel drugs is difficult. To find novel drugs, it is important to identify the drug targets. With the drug target identified it is possible to design drugs that will bind the target and hopefully not much else. Most drugs target proteins as agonists or antagonists and manipulate pathways, in which the protein is involved. Drug design has traditionally been done in the laboratory by screening for lead compounds, then modifying the hits and again screening for the best compounds, but as computers are getting powerful it is increasingly common to design and screen for new drugs virtually. Virtual screening has significantly decreased the time spent on discovering and designing new drugs, as the computer can sort out promising compounds from hundreds of thousands of compounds, leaving only modification and testing of the promising compounds as the time consuming steps. However, virtual screening is completely dependent on having the structure of the target at hand. Structure determination is achieved by various techniques, with X-ray crystallography being the most important. The Protein Data Bank (PDB) is an extremely important tool for 21st century drug designers (*7*). As of 2014, over 100000 structures are deposited in the databank, of which almost 90% were solved by X-ray crystallography, and the number of deposited crystal structures is growing exponentially; from ~4400 structures in 2004 to ~89000 structures in 2014 (*8*). A little more than 23000 of the structures are of human origin, so there is still a long way to go before structures of all the human proteins and variants of them, are available for drug design and modeling.

1.1 Transcription Factors

A group of very important human proteins are the transcription factors (*9*). Transcription factors are proteins that modulate the transcription of a gene by either binding to or letting go of the DNA sequence adjacent to the regulated gene. Hence the transcription factors are vital for many important cellular processes such as development, intercellular signaling and cell cycle (*10*). As the transcription factor family is so large, individual transcription factors work in many different ways. A gene can be directly regulated by the transcription factor itself, working as a repressor or activator, or by regulation of the synthesis of the named transcription factor. Though many proteins play crucial roles in gene transcription, the classification of transcription factors is the presence of one or more DNA-binding domains (DBDs). Over 2600 proteins in humans have been identified to contain such DBDs (*9*). A DBD recognizes specific sequences of DNA referred to as response elements and binds them by hydrogen bonds and van der Waals forces. The interaction between the transcription factor and the sequence is specific, but not all of the bases in the sequence are always involved in the binding, thus making the transcription factor able to bind to several sequences that are closely related, with different strengths of interaction. However, there are several other factors such as accessibility of response element, and obligatory binding partners, making the transcription factor specific, even though response elements could randomly occur due to the length of the genome. DBDs are classed in major families e.g. zinc fingers, helix-turn-helix motif and basic helix-loop-helix motif to mention a few. As the transcription factors are involved in a wide variety of cellular processes, it is not surprising that the alteration in these factors can lead to human diseases. Mutations of the DNA or in the transcription factor can lead to increased or decreased activity. Development and hormone response disorders, and cancer are related to mutations in DNA. Cancer arises due to abnormal cell growth. Cell growth is controlled by a variety of factors and proteins, some of which inhibit growth while others stimulate it. This is of course a shallow view of cancer, but it is likely that transcription factors are in some way involved in all forms of human cancer. Hence transcription factors are interesting as pharmaceutical drug targets. However, it is difficult to specifically target some transcription factors with small molecules, as not all transcription factors have a ligand-binding domain (LBD). Approximately 13% of all prescription drugs already target the class of transcription factors with a ligand-binding domain – the nuclear receptors (*11*).

1.2 Nuclear Receptors

The nuclear receptors are signal-sensing proteins specific to animals. When a signal molecule is bound to the nuclear receptor, it is activated and works as a transcription factor in the nucleus. The two domains of the nuclear receptors, DNA-binding domain and ligand-binding domain, are highly and moderately conserved, respectively, throughout the animal kingdom. Nuclear receptors are believed to have appeared with the first animals, approximately 635 million years ago according to fossil records (*12*). Now more than 900 nuclear receptor genes have been identified in the animal kingdom, but the number of nuclear receptor genes varies between species. There are approximately 48 nuclear receptor genes identified in human, mouse and rat (*13*). Nuclear receptors are classified according to the ligands associated with them and are called orphan receptors if they have no ligand associated with them. When the ligand associated with the receptor is identified, the receptor is "adopted". The ligands associated with the nuclear receptors are hydrophobic molecules that are metabolites from several metabolic pathways, as well as synthetic drugs, antibiotics, xenobiotics and lipophilic hormones. Depending on the mechanism of action of the nuclear receptor, they are divided into four broad classes. Type I nuclear receptors reside in the cytosol as complexes with heat shock proteins (HSPs). When a ligand is bound to the type I receptor, the heat shock protein dissociates and the nuclear receptors homo-dimerize. The homodimer then translocates from the cytoplasm into the nucleus and binds to specific inverted repeat sequence on the DNA known as response element (RE). Other proteins recruited by the receptor-DNA complex transcribe DNA downstream from the binding site, thus leading to cellular change (*14*). Type III nuclear receptors have the same mechanism of action as Type I, but instead of binding the inverted repeat RE, they bind the direct repeat RE. Type II nuclear receptors reside in the nucleus as complexes with co-repressor proteins. Ligand binding causes the co-repressor to dissociate from the nuclear receptor, which then hetero-dimerizes with another nuclear receptor, usually Retinoid X Receptor (RXR), and binds to the RE. Co-factors are recruited and transcription starts. Type IV nuclear receptors also reside in the nucleus and bind DNA either as monomers or dimers, but only one DNA-binding domain binds to a single half site RE. The DNA-binding domain, on the N-terminal side of the protein, is the most conserved domain of the nuclear receptors and contains two zinc fingers in tandem that consist of approximately 80 amino acid residues in total. The zinc fingers each coordinate a zinc ion through four cysteines and are responsible for recognizing the correlating RE. The DBD is connected to the ligand-binding domain by a flexible hinge region that allows dimerization and DNA binding simultaneously. The ligand-binding domain is also conserved, but with variations through the subclasses of nuclear receptors. The first crystal structure of the ligand-binding domain of a nuclear receptor is a heterodimer of the Retinoic acid receptor α (RAR)

from human and the Retinoid X receptor α from mouse (*15*). It showed that the LBD consists of 12 α-helices making a hydrophobic cavity in which hydrophobic ligands can bind. Once a ligand enters the cavity, helix 12 closes the cavity by hydrogen bonding to the polar end of the steroid. Although 48 nuclear receptor genes have been identified in humans, only approximately half of them have been classified according to their ligands. The remaining are orphan receptors without known ligands and target genes. However, the importance of the nuclear receptors is widely known, as almost every field of medicine is affected by disorders related to inappropriate nuclear receptor signaling (*16*). Nuclear receptor disorders include diabetes, atherosclerosis, inflammation, obesity and cancer. The hydrophobicity of the ligand-binding domain in the nuclear receptors makes them perfect pharmaceutical drug targets, and 13% of modern day pharmacopeia target nuclear receptors. This makes nuclear receptors the second biggest drug target, beaten only by the Rhodopsin-like G-protein Coupled Receptors (*11*). The hydrophobicity of the ligands makes it easy to discover and improve new compounds (*17*). Therefore it is important to "adopt" the orphan receptors and obtain structures of the already known nuclear receptors. Some of the adopted orphan receptors include fatty acid sensors - the peroxisome proliferator-activated receptors (PPAR), the bile acid sensor – Farnesoid X receptor (FXR) and the sterol sensors – Liver X Receptor (LXR).

1.3 Liver X Nuclear Receptors

The Liver X Receptors were discovered in 1994 and 1995 by isolation from a liver cDNA library, and consist of two members, LXRα (NR1H3) and LXRβ (NR1H2). They were given their name from the original discovery in the liver and were classified as orphan receptors, but were later adopted when organic tissue extracts and natural compounds were screened. The most potent activators identified were a group of oxysterols derived from cholesterol metabolism in the brain, adrenal, liver, macrophages and gonads. The oxysterols include 22(S/R)-hydroxycholesterol, 24(S),25-epoxycholesterol, 27-hydroxycholesterol and some bile acids (*16*). The identification of LXR ligands led to a clearer understanding of their roles in metabolism, and structure-activity relationship studies led to the discovery of two powerfull synthetic LXR agonists named T0901317 and GW3965 (*18, 19*). Although identified in the liver, the LXRs are expressed in several tissues, with LXRα expressed to a high extent in the liver and lower levels in adrenal gland, adipose tissue, intestine, kidney and spleen. LXRβ is ubiquitously expressed (*20*), hence LXRβ is sometimes referred to as ubiquitous receptor. The human LXR isoforms are encoded by separate genes, and have been identified to have molecular weights of approximately 50 kDa with a sequence similarity of 78% in their two domains (*21*). Reschly *et al.* showed that LXRs are similar throughout the animal kingdom, but that their LDBs have different specificity for vertebrates and invertebrates (*22*). The LXRs are type II nuclear receptors and, when activated, heterodimerize with retinoid X receptors (NR2B1, 2B2 and 2B3). The heterodimer binds to a response element characterized by direct repeats separated by four nucleotides. The LXR response elements regulate lipid homeostasis. LXRs act as cholesterol sensors in the cell and respond to elevated sterol concentrations by induction of genes encoding several sterol transporters of the ATP binding cassette transporters A and G (ABCA/ABCG), cholesterol ester transfer protein (CETP), cholesterol 7 alpha-hydroxylase (CYP7A) apolipoprotein-E (ApoE), fatty acid synthase and acetyl-CoA carboxylase (ACC) (*23*). This enormous variety of target genes makes LXR involved in numerous pathways. Figure 1.3.1 (*24*) shows an illustration of all the metabolic pathways and cell/tissue types modified in response to LXR signaling. Several studies with knockout mice have shown the different roles of LXRα and LXRβ (*25-27*). LXR activation with the agonists T0901317 and GW3965 has been shown to be anti-diabetic, induce reverse cholesterol transport, suppression of human cancer cell lines of colon, skin and prostate cancer, and a suppressing effect on the production of β-amyloid (*28-31*). However, LXRα has been shown to be involved in apoptosis of germ cells, while LXRβ is involved in the proliferation of germ cells (*32*). The activation of LXRα has been shown to raise triglyceride levels in liver and plasma. These joint and

separate involvements of LXRs make them ideal and challenging pharmaceutical targets. To be able to design drugs that successfully target one of the LXRs both not the other it is important to obtain several structures of the LXR LBDs bound to ligands. And several structures of the LXR LBDs are reported with different synthetic ligands bound (*33-37*).

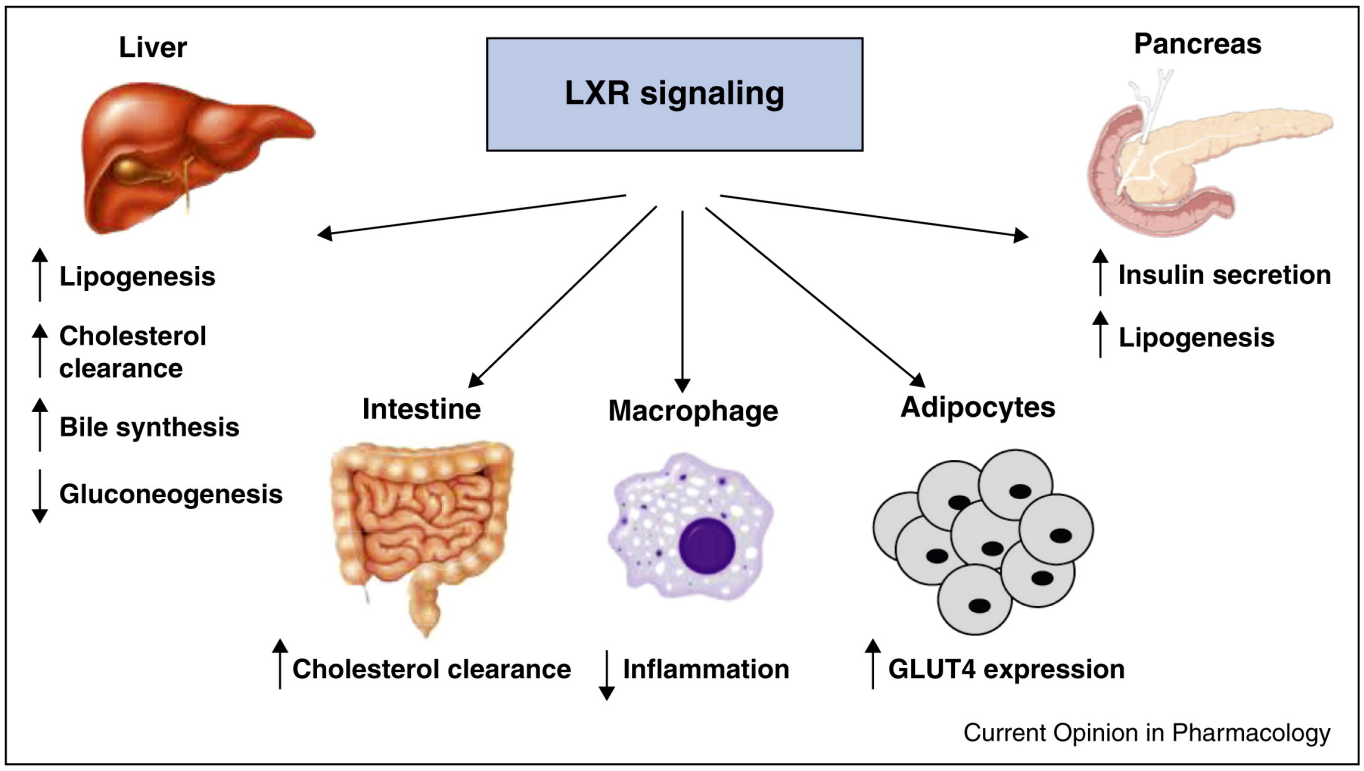


Figure 1.3.1. Metabolic pathways and cell/tissue types modified in response to LXR signaling *(24)*.

1.4 X-ray crystallography and protein purification

1.4.1 X-ray crystallography

X-ray crystallography is an extremely powerful tool for obtaining 3D structures of macromolecules, and is the most common method used for obtaining structures of proteins. The method makes the study of molecules on an atomic level with distinction between separate atoms possible. This allows researchers to study interaction between molecules in great detail, which is invaluable for our understanding of biological molecules and their functions, mechanisms of activation and role in pathways in the body.

The method is based on the fact that in microscopy the wavelength defines the fineness of detail that can be observed. Atoms and the bonds between them are in the range of 0.08-0.25 nm, so an appropriate wavelength of 0.1 nm has to be used to be able to distinguish details. This wavelength is in the range of γ- and X-rays. There is of course no lens that can focus waves in this order of magnitude; the best optical microscope has a lower limit of 500 nm, so to be able to distinguish details on an atomic level, coherent scattering of X-rays from electrons in the atoms has to be detected. The scattering pattern of the X-rays can then be used to calculate the atoms position in space. Scattering from single molecules is not strong enough to be detected, thus the scattering has to be amplified. Amplification happens if the molecule is symmetrically repeated in a lattice – a crystal (*38, 39*). To obtain a crystal the molecule has to be obtained in pure form. When a crystal of the molecule is obtained it is exposed to X-rays and data sets of individual diffraction patterns are collected. These diffraction patterns are used to create electron density maps through a series Fourier-transform calculations. The electron density map is then used to build a 3D model of the molecule. A basic set up of crystallography is shown in figure 1.4.1.

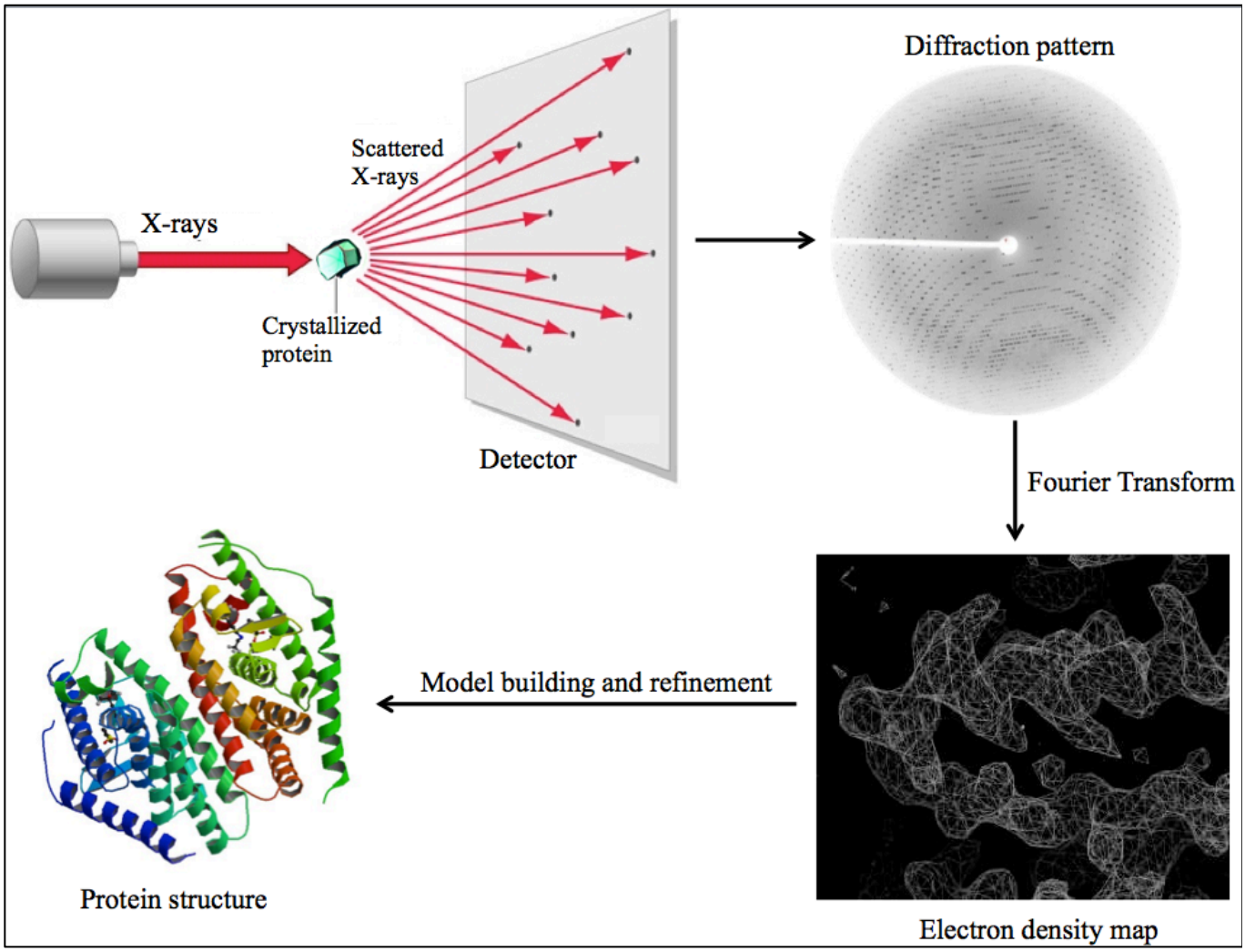


Figure 1.4.1. Basic schematic representation of protein crystallography. The X-ray beam is diffracted by the crystal, creating a diffraction pattern that can be used to create an electron density map. The electron density map is then used to build a 3D model.

1.4.2 **Protein Purification**

There are many ways of purifying proteins from their source by taking advantage of the different properties of the proteins. Chromatography is used to separate proteins by e.g. size, affinity, charge or hydrophobic interactions, to name a few. The rule of thumb being to use as few purification steps as possible to obtain high yield of pure, active protein. If the target protein has an unusual isoelectric point (pI) and/or molecular weight compared to the indigenous proteins in the production host, the purification can be performed with only few steps including e.g. solubility-, pI- or salting out precipitation and size-exclusion chromatography. However as most proteins are in the same range of pI and molecular weight, it is common to fuse an affinity tag onto the target protein by recombinant DNA technology. Affinity chromatography separates proteins by reversible interaction between the tagged protein and immobilized metal ions or a ligand attached to the chromatography matrix. Affinity chromatography is most discriminating and offers high selectivity, intermediate to high capacity and is sometimes efficient enough to yield pure protein. If needed, affinity chromatography can be followed by ion exchange chromatography (IEX) as a second bulk purification step, and gel filtration (GF), also known as size-exclusion chromatography, as a final polishing step. This three-step purification strategy is commonly used in small-scale purification as well as industrial purification, and in theory will yield highly pure active protein. There are many purification tags with corresponding affinity columns available, with histidine tagging being one of the most common. Histidine tags are used in immobilized metal ion affinity chromatography (IMAC). IMAC is based on the interaction between immobilized divalent metal ions and polyhistidine tags in proteins. The affinity for the metal ions is higher with increasing numbers of histidine residues. A typical purification using IMAC begins with equilibrating the column with a low concentration of imidazole. The imidazole molecules will bind to the immobilized metal ions and become the counter ligand. The column is washed with binding buffer to elute the naturally binding proteins in the cell lysate. The protein sample is applied to the column in a binding buffer with low concentration of imidazole. Then the proteins are eluted using a linear gradient of increasing imidazole concentration. Imidazole is the functional group of histidine, hence the imidazole molecules will compete for the nickel ions with the His-tagged proteins and displace them from the column. A typical IMAC purification chromatogram is shown in Figure 1.4.2 (*40*). The first peak from the washing of the column comes from all the naturally binding proteins, the second peak in the elution comes from the His-tagged fusion protein.

IEX separates proteins based on the reversible interaction between a charged protein and an oppositely charged chromatography medium. The proteins have to be solubilized in a buffer with an appropriate pH according to their pI and applied to either an anion- or cation exchanger. The target protein is concentrated during the chromatography and can be eluted from the column either by increasing salt concentration and/or a changing pH gradient. IEX can be used in all steps of purification, but is not as discriminating as IMAC.

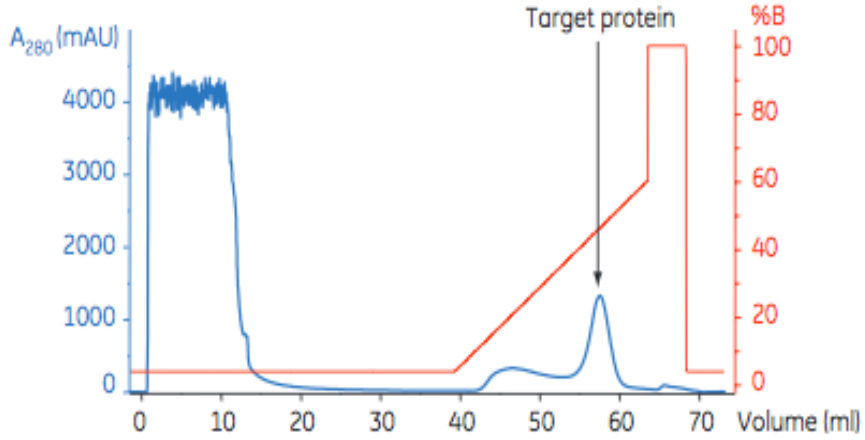


Figure 1.4.2. A typical chromatogram from IMAC purification with gradient elution of His-tagged fusion proteins *(40)*.

GF separates molecules based on their size, under mild conditions. The sample goes through a column packed with material with pores. The molecules are adsorbed in the pores and thus separated by size. Small molecules will travel in and out of pores appropriate to their size, while larger molecules will not fit in the pores and are eluted right away. GF is a simple purification technique that only requires a single buffer and no gradient, giving GF the ability of working within broad ionic strength, pH and temperature range. A typical example of His-tagged GF chromatography is shown in figure 1.4.3 (*40*).

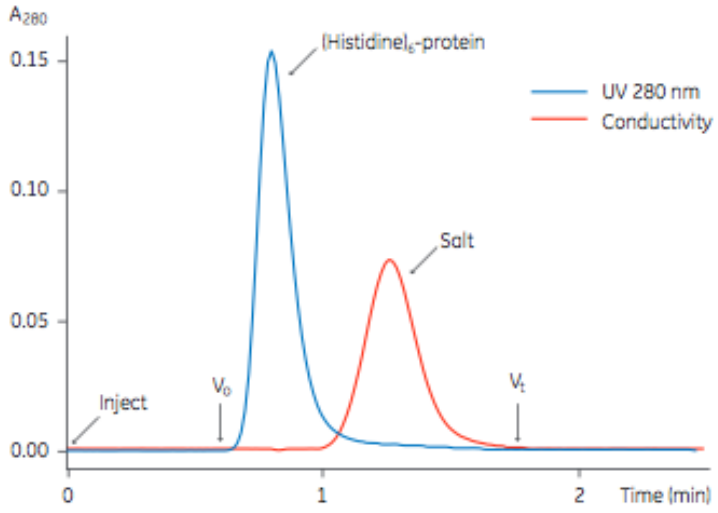


Figure 1.4.3. A typical chromatogram from GF purification of a His-tagged target protein *(40)*

1.5 Aims of Master's Project

This Master's project is a collaboration project with the School of Pharmacy to study the molecular mechanism of activation and inhibition of the Liver X Receptors. At the School of Pharmacy there is established a synthetic program to identify new, innovative chemical entities that are possible LXR ligands. A number of structures of the LXR LBDs with ligands bound are solved, however, most the ligands are synthetic. So the current knowledge about the molecular activation mechanism and most of the interpretations is based on the structures with synthetic ligands bound. To get a better understanding of the LXRs role in lipid homeostasis and be able to do structure-guided drug design, it is important to obtain structures of the LXRs bound to endogenous sterols, synthetic sterol derivatives as well as novel synthetic ligands. Therefore obtaining a high yield of pure active protein is of interest, as the proteins are going to be used in co-crystallization/soaking trials with a diverse library of lead compounds synthesized at the School of Pharmacy.

This Master's project focuses on the first steps, of this big project, with the following aims:

- **Production of the proteins**

  Express the proteins recombinant from *E. coli* cells, purify the proteins by chromatographic methods and optimize a protocol for obtaining high yield of pure active protein.

- **Crystallize the proteins**

  Screen for the optimal crystallization conditions of the proteins before doing co-crystallization/soaking trials with endogenous oxysterols, synthetic oxysterol derivatives, plant phytosterols and synthetic novel ligands synthesized at the School of Pharmacy.

1.6 Strategies

The Master's project was designed based on the previously reported structure-related publications of the LXR LBDs. The articles were compared and all the reported constructs were the basis for the choice of construct. An overview over the articles with their constructs, expression vectors, purification tags, ligands in the crystal structures, and PBD entry codes are shown in Table 1.6. The Master's project is subsectioned into different strategies for obtaining pure active protein in high yield. The first strategy was designed based on the survey compiled in Table 1.6 and was expected to be straightforward. However, it turned out no to be so simple. Hence, several strategies were investigated to obtain pure active protein in high yield. Strategies 1, 1B and 1C are based on the original construct, strategy 2 is based on a different expression construct.

Table 1.6. A compilation of the techniques used for expression and purification of the previously reported structures of LXR LBDs

Group and year	Protein and amino acids	Expression vector	Purification tag	Ligands used in crystallization	PDB code
Färnegård *et al.* 2003	LXRβ: 213-461	pET28a	Thrombin cleavable hexahistidine tag	T0901317 and GW3965	1PQ9, 1PQC and 1PQ6
Fradera *et al.* 2010	LXRα: 182-447	N/A	Non-cleavable hexa-(histidine-glutamine) tag	GW3965, F_3methylAA and benzisoxazole urea	3IPQ, 3IPS and 3IPU
Hoerer *et al.* 2003	LXRβ: 209-461	pET28a(+)	Thrombin cleavable hexahistidine tag	T0901317	1UPV and 1UPW
Svensson *et al.* 2003	LXRα: 207-447 in complex with RXRβ	pET15	Thrombin cleavable hexahistidine tag	LXR agonist T-17 and RXR agonist MPA	1UHL
Williams *et al.* 2003	LXRβ: 214-461	pRSET A	Thrombin cleavable hexahistidine tag	T0901317 and 24(S),25-epoxycholesterol	1P8D

1.6.1 **Strategy 1: Expression of the LXR LBDs from the purchased vector**

Based on the information compiled in table 1.6 the constructs of the LXR LBDs were decided to be 182-447 and 216-460 for LXRα and LXRβ, respectively. The theoretical isoelectric points (pI) and molecular weights of the proteins are approximately 5.9/30.6 kDa and 5.9/28.6 kDa for LXRα LBD and LXRβ LBD, respectively. A quick look at the plot of theoretical molecular weight against pI of the indigenous proteins in *E. coli* in Figure 1.6.1 (*41*), shows that the LXR LBDs are almost exactly in the middle of the plot where most of the native *E. coli* proteins are. Since the LXR LBDs were going to be expressed recombinantly in *E. coli*, purification would require multiple purification steps without any guarantee of yielding highly pure protein. Hence it was desirable to fuse the LXR LBDs to a purification tag, making purification possible with very few steps. The three-step purification strategy of purification was used in all of the articles of LXR LBD crystal structures (*33-37, 42*). Recombinant production and purification of the LXR LBDs seemed straightforward with a cleavable purification tag fused to the proteins. Conveniently LifeTechnologies sold the synthesis and cloning into an expression vector as a package deal, and because a cleavable purification tag was desirable, the genes were purchased in the pRSET B expression vector (Invitrogen™). This is a pUC-derived expression vector designed for high-level protein expression in *E. coli.* The expression of the gene of interest is controlled by the phage T7 promoter, which is recognized by T7 RNA polymerase. For expression to occur, the *lac* operon-controlled T7 RNA polymerase has to be expressed in sufficient amounts. This is inducible by addition of isopropyl β-D-1-thiogalactopyranoside (IPTG) to the growth medium. In addition, the vector contains a sequence that encodes for an N-terminal fusion peptide. This sequence includes a translation initiation codon, a hexahistidine tag, the T7 phage gene 10 leader sequence that stabilizes the transcription of the foreign gene, an epitope for the Xpress™ antibody, and an enterokinase cleavage recognition sequence. A map and the multiple cloning site of the pRSET B vector are shown in Figure 1.6.2 (*43*). The LXR LBDs were cloned into the vector using the *Xho*I and *Kpn*I cloning sites. The expected sizes of the fusion-proteins are 34.6 kDa and 32.6 kDa for $(His)_6$-LXRα LBD and $(His)_6$-LXRβ LBD, respectively. See Appendix; Section A for the full-length sequences.

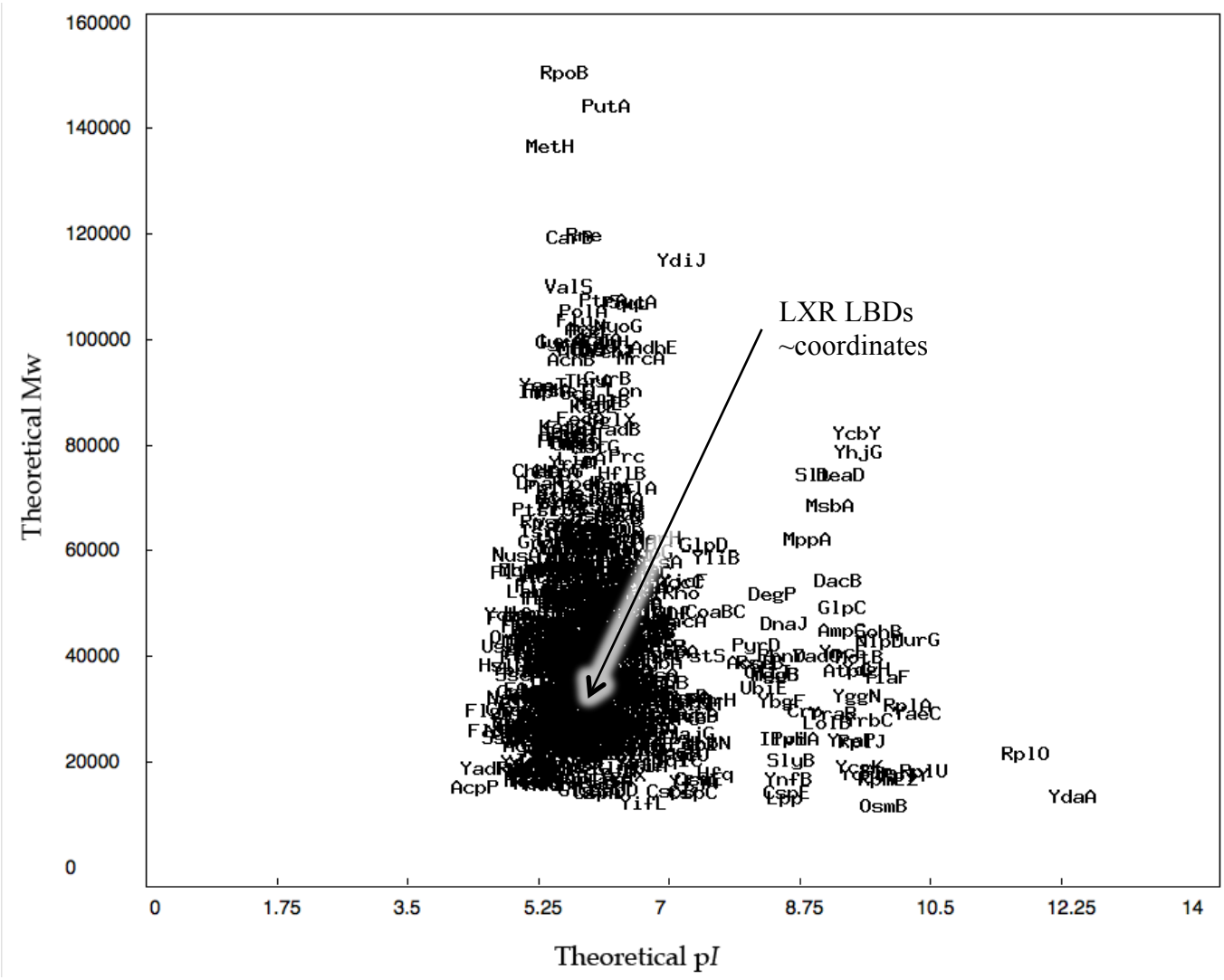


Figure 1.6.1. A plot showing the indigenous proteins of *E. coli* plotted by theoretical molecular weight against theoretical pI *(41)*.

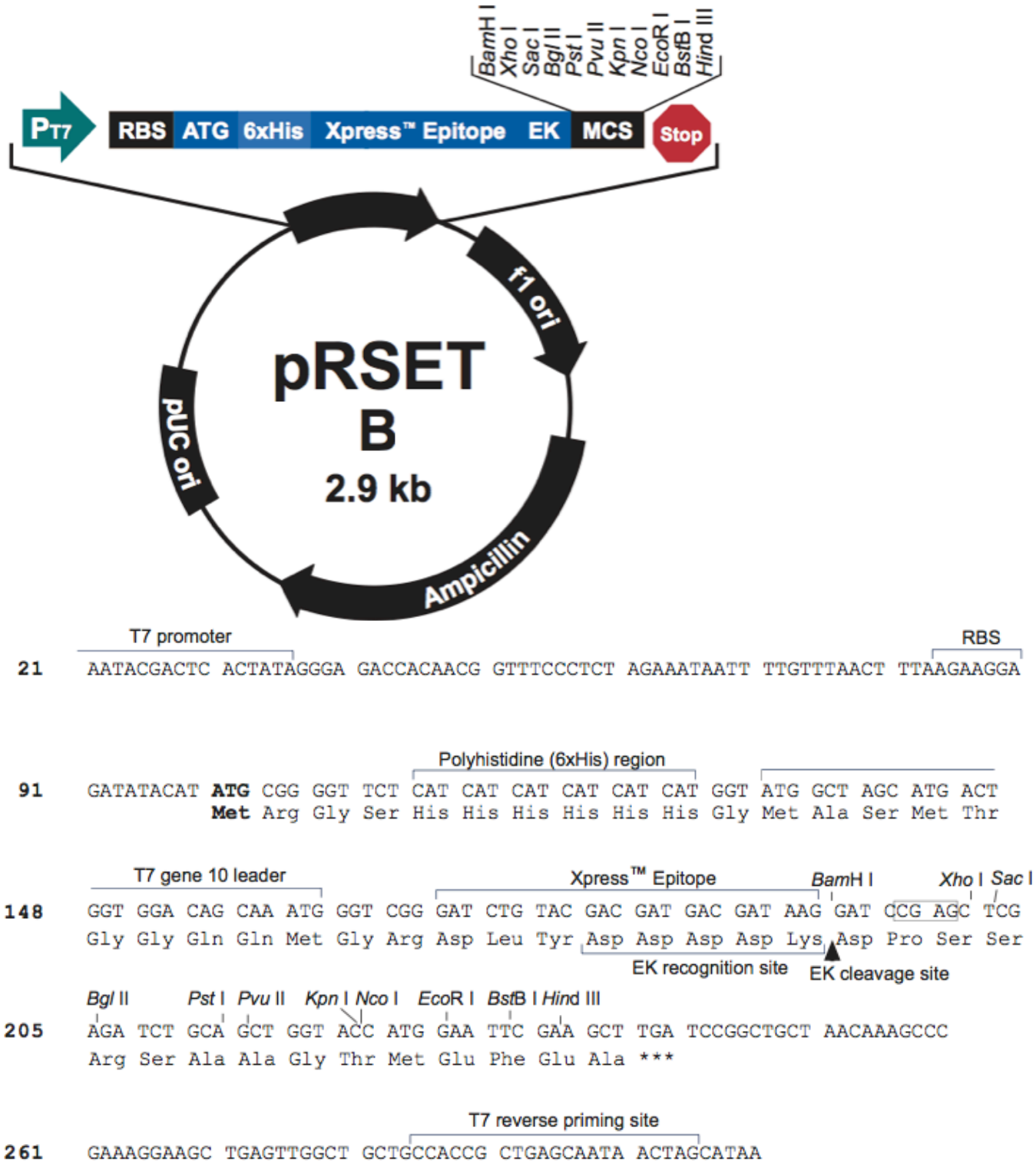


Figure 1.6.2. The map and multiple cloning site of the expression vector pRSET B. *(43)*

1.6.2 Strategy 1B: *in vitro* refolding of proteins purified from inclusion bodies

The experiments with recombinant production of the $(His)_6$-LXR LBDs showed that the cells expressed the proteins in high amounts, but the proteins aggregated and formed insoluble inclusion bodies. It is known that reduced growth rate usually leads to a more soluble expression and reduce inclusion body formation. We tried varying the expression time and temperature during the expression, transformed the constructs into cold-adapted cells engineered to improve the solubility of proteins prone to aggregation, and had a ligand omnipresent during expression, to investigate if this would improve the solubility of the $(His)_6$-LXR LBDs, but all the experiments resulted in a high amount of insoluble protein. Inclusion bodies are often very pure and if refolding of the proteins *in vitro* is possible, inclusion body purification is a direct and effective method of producing highly purified protein (*44*). Hence we decided to try inclusion body purification and refolding of the proteins *in vitro* in refolding buffer with and without cholesterol. REFOLD is an analytical database that contains methods and protocols for refolding of recombinant proteins *in vitro* (*44*). The database has over a thousand refolding records stored. Unfortunately the LXR LBDs are not among the entries.

1.6.3 Strategy 1C: subcloning of Tobacco Etch Virus protease cleavage site into the pRSET B construct

To circumvent problems experienced with enterokinase cleavage of the His-tag in the pRSET B construct, we decided to subclone a novel cleavage site into the construct upstream from the enterokinase cleavage site. The new cleavage site would be recognized by Tobacco Etch Virus (TEV) protease. Tobacco Etch Virus protease (TEVp) is a highly sequence-specific cysteine protease that recognizes the sequence ENLYFQS and cleaves the peptide bond between glutamine and serine. The TEVp cleavage site would be subcloned into the construct between the epitope and the enterokinase cleavage site. TEVp was the protease of choice because of its high sequence specificity (*45*), and an in-house bachelor's thesis on the production and purification of Green Fluorescent Protein (GPF)-labeled, hexahistidine-tagged TEV protease.

1.6.4 **Strategy 2: cloning of the genes in the periplasmic expression vector pFKPEN**

Expression the LXR LBDs in the pRSET B plasmid were shown to be difficult and rate limiting. Strategies 1, 1B and 1C were investigated to optimize and circumvent the problems experienced with the construct from the pRSET B plasmids. All of the strategies led to the same conclusion of little or no protein purified. For every solution more problems were encountered, so we decided to clone the LXR LBD genes into another expression named pFKPEN. Gunnarsen *et al.* engineered this vector to improve the expression of T cell receptors, which were prone to aggregation and proteolysis in bacterial systems (*46*). pFKPEN is optimized for periplasmic bacterial expression and carries the gene encoding the periplasmic chaperone FkpA. FkpA is a peptidyl-propyl *cis,trans*-isomerase that has been shown to reactivate inactive proteins and prevent premature aggregation of early folding intermediates.

2 Experimental Procedures

2.1 Expression of $(His)_6$-LXR LBDs

2.1.1 Expression in BL21DE3

Transformation of *E. coli* BL21DE3 competent cells with the LXR constructs

Approximately 50 ng of each construct were transformed into 50 µl of competent *E. coli* BL21 DE3 cells by heat shocking the cells at 42 °C for 45 seconds. 500 µl of SOC medium, pre-warmed to 37 °C, was added to the cells and the samples were incubated at 37 °C with shaking at 200 rpm, for 1 hour. The cells were centrifuged at 16000 × *g,* and 350 µl of the medium was decanted. The cell pellet was re-suspended in the medium remaining in the tube and plated onto LB agar plates containing 100 mg/l ampicillin. Plating non-transformed cells onto plates with and without ampicillin added were used as controls. The plates were incubated at 37 °C overnight. If the control plates showed positive results (i.e. no growth on the ampicillin plate and massive growth on the agar plate), a colony was picked from each transformation and inoculated in 5 ml of LB medium with added 100 mg/ml of ampicillin (LBamp) and incubated overnight at 37 °C and 120 rpm. 1 ml of each of the overnight cultures were mixed with 0.5 ml 60% glycerol and stored at -80 °C as a glycerol stock.

Expression of $(His)_6$-LXR LBDs

Overnight cultures were made from the LXR glycerol stock by inoculating 50 ml of LBamp with a scrap of cells, and incubating the samples on a rotary shaker at 37 °C/120 rpm. The overnight cultures were used to make main cultures by inoculating 1 l pre-warmed LBamp medium to an OD_{600} of 0.05. The main cultures were incubated on a rotary shaker at 37 °C/110 rpm. OD_{600} was measured hourly during the growth. When the OD_{600} reached 0.8, the cultures were induced with 0.2 mM IPTG. After 4 hours, for normal expression, or 12-24 hours for the overnight expression, the cells were harvested by centrifugation at 6000 × *g* for 15 minutes at 4 °C. The cell pellets were collected and frozen in -80 °C.

Harvest of proteins from the cells

The frozen pellets were thawed and re-suspended in 15-25 ml lysis buffer (50 mM Tris pH 8.0, 50 mM NaCl, 150 µM PMSF with Roche c*O*mplete Protease Inhibitor Cocktail tablet). Lysozyme was added to the sample to a final concentration of 2-4 mg/ml and the samples were incubated at room temperature on a shaker for 2-4 hours. A small scrap of deoxyribonuclease I was added to the sample if the sample was very viscous. The samples were centrifuged at 45000 × *g* for 20 minutes at 4 °C, and the supernatant was decanted from the pellet. To confirm that the induction had worked, samples of the medium prior to the induction, the supernatant and the pellet after harvest were analyzed by SDS-PAGE as described in the SDS PAGE protocol below.

SDS-PAGE protocol

The SDS-PAGE samples taken during the expression and the lysis were dissolved in sample buffer (2% SDS, 5% 0.2 mM β-mercaptoethanol, 10% glycerol, 60 mM Tris pH 6.8, 1.4 mg/ml bromphenolblue and 4M urea) to a total volume of 60 µl, respectively, and heated to 90 °C for 5 minutes. 15 µl of the samples were applied to a NuPAGE 4-12% Bis-Tris gel. 5 µl of SeeBlue®Plus 2 (Invitrogen) were used as marker, and the gel was run in NuPAGE® Tris MES SDS running buffer for 35 minutes at 200 V/400 mA/100 W. The gel was stained with Coomassie Blue or using the SilverQuest™ staining kit (Invitrogen)

2.1.2 Expression in ArcticExpress cells

The LXR LBD constructs were transformed into ArcticExpress competent cells (Agilent Technologies) as described in Section 2.1.1. Two different protocols were tested for the expression of the His-tagged LXR LBDs in the ArcticExpress cells. The protocol described in Section 2.1.2, and Agilent Technologies' ArcticExpress optimized-expression protocol (*47*).

Standard expression protocol

An expression test was set up for both BL21DE3 cells and ArcticExpress cells transformed with the pRSET B LXR LBD constructs. The expression was performed overnight at 30 °C and 12 °C, for BL21DE3 and ArcticExpress cells, respectively. The overnight cultures were harvested and lysed.

Expression protocol optimized for ArcticExpress

Pre-cultures of LBamp medium with gentamycin added to a concentration of 20 µg/l were inoculated with a scrap of ArcticExpress cells from the glycerol stocks. The pre-cultures were incubated overnight on a rotary shaker at 37 °C/200 rpm. The overnight pre-cultures were diluted 1:50 into LB medium with no selection of antibiotics as main cultures and incubated for 3 hours on a rotary shaker at 30 °C/200 rpm. The main cultures were transferred to a rotary shaker with a temperature of 12 °C and incubated for approximately 10 minutes (so the cells had some time to acclimatize) at 200 rpm. IPTG was added to a final concentration of 1 mM and the cultures were incubated for 24 hours. The cells were harvested and lysed as described in Section 2.1.2

2.2 Purification of $(His)_6$-LXR LBDs

The recombinant proteins were fused with an N-terminal hexahistidine tag for purification by immobilized metal ion affinity chromatography (IMAC) as a first step for capturing the proteins. Gel filtration chromatography was used as a second purification step.

2.2.1 Column preparation

HiTrap Chelating HP (GE) 5 ml columns were used for capturing the proteins of interest, with a separate column prepared for each protein. Prior to use the columns were pre-washed with 3 column volumes (CV) of distilled water using a syringe. For charging with Ni^{2+} ions, the columns were loaded with ½ CV of 0.1 M $NiSO_4$. The columns were then washed with 3 CV of distilled water. The columns were placed on the ÄKTApurifier FPLC system (GE Healthcare) and equilibrated by washing with 5-10 CV of binding buffer (50 mM sodium phosphate, pH 7.5, 400 mM NaCl, 100 mM KCl, 0.25 mM TCEP, 150 µM PMSF, 20 mM imidazole).

2.2.2 Protein purification by FPLC

The supernatant from the lysed cells was applied to the columns using a pump, and the columns were washed with 5-10 CV of binding buffer (400 mM NaCl, 150 µM PMSF, 50 mM NaH_2PO_4, 100 mM KCl, 20 mM imidazole). The proteins were eluted and collected in fractions of 1.0 ml, using a linear gradient of increasing concentration of elution buffer (binding buffer with 500 mM imidazole and 1 mM DTT). 10 µl of the fractions from the peaks in the chromatogram were mixed with 10 µl of SDS sample buffer and analyzed by SDS-PAGE as described in Section 2.1.2. Fractions containing the protein of interest were pooled and concentrated in a centrifugal concentrator unit with polyethersulfone membrane and either 5 kDa or 10 kDa molecular weight cut-off filters (MWCO) (Vivaspin 20 from Sartorius Group). The buffer was exchanged to a low-salt, low-imidazole buffer (25 mM Tris pH7.6, 100 mM NaCl, 1 mM DTT). The concentrated protein sample (0.5 ml) was applied to a SuperDex200 10/300 GL column (GE Healthcare) using a 1 ml loop and eluted with elution buffer (25 mM Tris pH 7.6, 100 mM NaCl, 1 mM DTT) in fractions of 1 ml. The peaks in the chromatogram were analyzed by SDS-PAGE. The fractions containing the protein of interest were pooled and concentrated in a centrifugal concentrator.

2.2.1 **Cleavage of the His-tag with enterokinase**

The ligand binding domains of the LXR isoforms were cloned into a vector as a fusion protein with an N-terminal purification peptide, as suggested by the LifeTechnologies. The fusion peptide contains a hexahistidine tag separated from the proteins by an epitope and an enterokinase cleavage site. The cleavage tests were carried out according to the protocol for EnterokinaseMax™ (EKMax™) (*48*). Purified protein concentrated to 0.4 mg/ml was used for a cleavage test with EKMax™. Six reactions were set up at four different temperatures, respectively. The temperatures were 37 °C, room temperature, 16 °C and 4 °C. The room temperature sample was put on the lab bench. The setup for the reactions is shown in Table 2.2.1. The reactions were mixed and placed in their respective temperatures overnight. The digestion samples were analyzed by SDS-PAGE.

Table 2.2.1: The setup for the EKMax™ digestion test of $(His)_6$-LXRα LBD

	A	B	C	D	E	F
EKMax™ (µl)	0.0	1.0 µl of D	1.0 µl of D	1.0 µl of E	1.0	4.0
1X EKMax™-buffer (µl)	0.0	9.0	9.0	9.0	9.0	0.0
10X EKMax™-buffer (µl)	3.0	2.0	2.0	2.0	2.0	3.0
MQ-H_2O (µl)	14.5	5.5	5.5	5.5	5.5	10.5
$(His)_6$-LXRα LBD (µl)	12.5	12.5	12.5	12.5	12.5	12.5
Total (µl)	30.0	30.0	30.0	30.0	30.0	30.0

2.3 Robotic crystal screening

Screening for crystallization conditions was performed using the vapor diffusion technique with a sitting-drop setup using the JCSG+ suite and Morpheus® crystallization screens (Qiagen and Molecular Dimension, respectively). The sitting drops were dispensed in MRC 2 Well Crystallization Plates (Swissci from Hampton Research) by mixing 0.15 µl and 0.3 µl of the $(His)_6$-LXRα LBD protein mix and 0.15 µl and 0.3 µl of the precipitant solution from pre-dispensed reservoirs for JCSG+ suite and Morpheus® screens, respectively. The plates were sealed with crystal clear tape to prevent evaporation. The plates were stored in a dark room at 20 °C.

2.4 Refolding of the proteins *in vitro*, and avoiding aggregation

2.4.1 Purification and refolding of proteins expressed as inclusion bodies

The pellets from the protein expression were re-suspended in CelLytic™ B (Sigma) at a ratio of 10 ml per gram of wet pellet. The suspensions were incubated on a shaker in room temperature for 15 minutes and then centrifuged at 16000 × *g* for 10 minutes to pellet the residual insoluble material. The supernatants were saved for analysis and the pellets re-suspended in CelLytic™ B 10x. The extraction was repeated until it was difficult to dissolve the residual pellets, which took four to five repetitions. Finally the pellets were dissolved by sonication in denaturing buffer (8 M urea, 0.5 M NaCl, 20 mM sodium phosphate buffer, pH 7.4). The samples were centrifuged at 16000 × *g* and the supernatants were decanted. The proteins were refolded by slowly dripping the supernatants into rapidly stirred refolding buffer (0.1 M Tris, 0.4 M L-arginine, 0.150 mM PMSF, 1.0 mM DTT). The refolded protein samples were concentrated in a centrifugal concentrator with 10 kDa MWCO and stored overnight at 10 °C. Precipitation occurred overnight, so the samples were filtered through a sterile syringe filter (0.45 µm polyethersulfone membrane from VWR®) and stored overnight at 10 °C. The samples were further purified by IMAC.

2.4.2 Expression, inclusion body purification and refolding of proteins in presence of cholesterol

Expression

An overnight pre-culture was set up for $(His)_6$-LXRα LBD and $(His)_6$-LXRβ LBD from the glycerol stocks as described in 2.1.2. Main cultures of 1 l LBamp were inoculated and grown to an OD_{600} of 1.0. Ethanol was saturated with cholesterol at a concentration of 0.052 M (*49*). The stock solution of cholesterol was added to the cultures to a final cholesterol concentration of 20 µM. The samples were induced by addition of 0.2 mM IPTG. 1.5 hours later more stock solution was added to the cultures to a final cholesterol concentration of 40 µM. After two hours the temperature in the incubator was lowered from 37 °C to 18 °C and the expression continued overnight. The cells were harvested and lysed as described in Section 2.1.2.

Inclusion body purification

The pellets were re-suspended in 20 ml buffer containing 50 mM HEPES-NaOH pH 7.5, 0.5 M NaCl, 1 mM PMSF, 5 mM DTT and 0.35 mg/ml lysozyme, and incubated at 20 °C for 30 minutes. Triton X-100 was added to a concentration of 1% (v/v), and the samples were sonicated until the solutions were clear. Deoxyribonuclease I was added to a concentration of 20 mg/l and the samples were incubated at 37 °C for 1 hour. The inclusion bodies were pelleted by centrifugation at 30 min/30000 × *g*/4 °C. The pellets were washed twice with PBS containing 1% Triton X-100 followed by centrifugation at 30 min/30000 × *g*/4 °C. The pellets were solubilized in 2 ml of buffer containing 50 mM HEPES-NaOH pH 7.5, 25 mM DTT, 6 M guanidine HCl and incubated at 4 °C for 1 hour. The samples were centrifuged at maximum speed (75000 × *g*) for 10 minutes to remove insoluble material that could act as nuclei to trigger aggregation during folding.

Refolding in presence of cholesterol

The solubilized proteins were rapidly diluted 1:100 in stirred refolding buffer (50 mM HEPES-NaOH pH 7.5, 0.2 M NaCl, 1 mM DTT, 0.5 M L-arginine, 50 µM cholesterol). The mix was stirred for 2 minutes after addition of protein. The samples were kept at 4 °C for 1 hour. Samples from the inclusion body purification and refolding were analyzed by SDS-PAGE.

2.5 Subcloning of Tobacco Etch Virus (TEV) protease cleavage site into the pRSET construct

2.5.1 Linearization of the vector

Enzyme master mixes were made with 4 units of *Xho*I and *Bam*HI. 6 µl of the enzyme master mixes were mixed with 750 ng of the plasmids with $(His)_6$-LXRα LBD and $(His)_6$-LXRβ LBD, respectively, and incubated at 37 °C overnight. Calf-intestinal alkaline phosphatase (CIP) was added to every sample to a concentration of 0.5-1 units per µg DNA, and the samples were incubated at 37 °C for one hour. The reaction was quenched by heating to 80 °C for 20 minutes. The samples were purified by agarose gel electrophoresis as described in the following section. Samples of uncut plasmid were used as a reference.

Agarose gel electrophoresis

The samples were mixed with loading buffer and applied to a 0.8% low melting point agarose gel with 0.5 µg/ml ethidium bromide added. 5 µl of Perfect DNA™ 1 kbp ladder (Novagen) was applied to the gel and the electrophoresis ran for 1 hour at 80 V. The gel was examined using the PhotoDoc-It™ 60 Imaging system with Benchtop UV M-20 Transilluminator (UVP), photographed with Canon PowerShot A480, and the interesting DNA bands were cut out. The DNA was purified using the QIAquick Gel Extraction Kit (Qiagen) following the protocol from the kit.

2.5.2 Hybridization of ssDNA oligonucleotides to form double stranded insert

2 µl of the forward and backward oligonucleotides, respectively, coding for the TEV protease cleavage site were mixed with 2 µl of 10X ligase buffer and 2 µl of 0.5 M NaCl to a total volume of 20 µl with MQ-H_2O. The sample was placed in boiling water for 2 minutes, before cooling to room temperature in the water bath.

2.5.3 **Ligation**

The concentrations of linearized plasmids and hybridized insert were determined by measuring the absorption at 260 nm using NanoPhotometer® from IMPLEN. The insert and plasmid were mixed in ratios of 1:1 and 1:1.5 and added 0.5 µl T4 DNA ligase and 1 µl 10X T4 DNA ligase buffer in a total volume of 10 µl. The ligation reactions were incubated at 16 °C overnight. The reaction was quenched by heating to 80 °C for 20 minutes.

The samples were transformed into Stellar™ cells (Clontech) as described in Section 2.1.1. The plasmids were transformed into the cells in amounts of 1 µl and 2 µl. The cells were plated onto LBamp agar plates and incubated at 37 °C. The plates were checked for growth, and promising colonies were picked and inoculated in LBamp and incubated overnight at 37 °C/120 rpm. The plasmids were purified from the overnight samples with QIAprep Spin Miniprep Kit (Qiagen) and the concentrations of the plasmids were determined by spectrophotometry. The samples were mixed with T7 promoter standard primer and sent to an in-house sequencing service.

2.6 Cloning of the LXR LBD genes into the pFKPEN periplasmic expression vector

To improve the solubility of the LXR LBDs it was decided to clone the LXR LBD genes into a new vector, pFKPEN, coding for periplasmic translocation and co-expression of a protein to help with folding, the periplasmic chaperone FkpA (*46*). The vector was a kind gift from Geir Åge Løset (Department of Biosciences – The Faculty of Mathematics and Natural Sciences, University of Oslo). The cloning was done using the In-Fusion® HD cloning kit (Clontech). Primers for amplifying the LXR LBD genes were designed according to the In-Fusion kit's specifications at the Clontech homepage and ordered from Eurofins MWG Operon. Forward and reverse primers were designed for cloning the LXR LBD genes into the new vector both with a C-terminal hexahistidine tag and without a tag. For the primer sequences see Appendix, Section C.

2.6.1 PCR to amplify the LXR LBD genes and add overlapping ends

The concentrations of the primers were determined by spectrophotometry and stock solutions of 10 µM were made. PCR mixes were prepared for the template DNA of $(His)_6$-LXRα LBD and $(His)_6$-LXRβ LBD from the pRSET B constructs, to generate the inserts without and with His-tag before the In-Fusion cloning procedure. The setup for the mixes is shown in Table 2.6.1. To generate the His-tag, the PCR was performed twice with different primers. For the PCR conditions see Appendix, Section C.

Table 2.6.1. The setup for the PCR of $(His)_6$-LXRα LBD and $(His)_6$-LXRβ LBD to generate inserts without and with His-tag before the In-Fusion reaction

Component	**50 µl reaction (µl)**
10X standard *Taq* reaction buffer	5.0
10 mM dNTP's	1.0
10 µM forward primer	1.0
10 µM reverse primer	1.0
Template DNA	2.0
Taq DNA Polymerase	0.25
MQ-H_2O	39.75

2.6.2 Linearization of pFKPEN expression vector

XL1 Blue $CuCl_2$ cells (Stratagene) transformed with the pFKPEN plasmid were obtained from Geir Åge Løset, via the hands of Master's student Hedda Johannesen. The pFKPEN plasmid was isolated from the cells using the miniprep kit (Qiagen) with a final elution in 30 µl of elution buffer. 1000 ng of the plasmid were used for linearization. The plasmid was linearized by double digestion with the restriction enzymes *Nco*I and *Not*I (FastDigest from Fermentas) according to the FastDigest protocol (*50*). The digestion reactions were incubated for 30 minutes at 37 °C. The PCR products and the linearized vector were purified by agarose gel electrophoresis, as described in Section 2.5.1. Undigested pFKPEN was used as a reference in the electrophoresis. The fragments were purified from the gel pieces using the miniprep kit with a final elution of 30 µl of elution buffer.

2.6.3 Cloning procedure

The PCR fragments for the LXR LBDs, with and without His-tag, and the linearized vector were used for the In-Fusion cloning procedure. The control insert and the pUC19 control vector included in the kit were used as positive control. The linearized pFKPEN vector was used as negative control. The cloning procedure was performed following protocol I:B of the In-Fusion® HD Cloning Kit User Manual (*51*). The samples were incubated for 15 minutes at 50 °C, then placed on ice. The samples from were transformed into Stellar™ cells (Clontech) as described in Section 2.1.1. The transformed cells were plated out on LBamp agar plates and incubated at 37 °C overnight. The plates were checked for colonies and a number of positive colonies from the LXR LBDs plates were inoculated in 5 ml of LBamp. The inoculated samples were incubated at 37 °C overnight. The LXR LBD plasmids were purified from the overnight cultures using the miniprep kit with a final elution of 30 µl of elution buffer.

2.6.4 Verification by restriction digest and validation by sequencing

The purified plasmids were analyzed by diagnostic restriction digest before sequencing. The NEBcutter V2.0 web tool (*52*) was used to analyze all of the original sequences and new sequences for the LXR LBD genes in their respective plasmids, against restriction enzymes available in the lab and their respective number of cleavage sites. Compatible restriction enzymes and their cleavage sites in the plasmids are shown in Table 2.6.1. Restriction enzymes *Acc*I and *Kpn*I (FastDigest from Fermentas) were used for all plasmids of the digestion test. The digestion reactions were incubated for 30 minutes at 37 °C, before the samples were analyzed by agarose gel electrophoresis, as described in Section 2.5.1. Five random cloning samples were mixed with forward and backward sequencing primers, respectively, and sent to an in-house sequencing service. For the primers used see Appendix, Section C.

Table 2.6.1. Compilation of restriction enzyme cleavage sites in the plasmids.

	Enzyme and cleavage site									
Plasmid	*Acc*I	*Bam*HI	*Eco*RI	*Hind*III	*Kpn*I	*Nco*I	*Nde*I	*Not*I	*Xba*I	*Xho*I
Original pFKPEN (4806bp)	431, 699, 1202, 2263	1034, 1344	131, 1334	638	808, 828, 906, 1005	216	-	1026	1097	1
pFKPEN LXRα LBD (4806 bp)	1202, 2263	1034, 1344	131, 1344	-	-	216	853	1026	1097	1
pFKPEN LXRα LBD-$(His)_6$ (4824bp)	1220, 2281	1052, 1362	131, 1352	-	-	216	853	1044	1115	1
pRSET $(His)_6$-LXRα LBD (3685bp)	-	192	1009, 1024	1031	1019	1019	837	-	158	201
pFKPEN LXRβ LBD (4755bp)	1151, 2212	983, 1293	131, 1283	-	-	216	-	975	1046	1
pFKPEN LXRβ LBD-$(His)_6$ (4773)	1169, 2230	1001, 1311	131, 1301	-	-	216	-	1993	1046	1
pRSET $(His)_6$-LXRβ LBD (3628bp)	-	192	967	974	962	962	98	-	58	201

3 Results and Discussion

The aim of this master's project was to produce, purify and crystallize the ligand binding domains (LBDs) of the two isoforms, α and β, of Liver X Nuclear Receptors (LXRs). Obtaining crystals for X-ray experiments with ligands soaked in or co-crystallized was the ultimate goal. LXR LBDs have been successfully purified, crystallized and their structure solved by X-ray crystallography in several groups (*33-37, 42, 53*). The results and discussions will be subsectioned based on three strategies for expressing and purifying pure LXR LBDs.

3.1 Strategy 1: Expression, purification and crystallization of LXR LBDs in pRSET B vector

3.1.1 Expression of $(His)_6$-LXRα LBD and $(His)_6$-LXRβ LBD in BL21DE3 cells

The pRSET B vectors with the LXR LBD genes were transformed into BL21DE3 cells and expressed in LBamp medium for 4 hours, or 12-24 hours. The cells were harvested and lysed as described in Section 2.1.2. The induction and harvesting was analyzed by SDS-PAGE. The gels are shown in Figures 3.1.1.1 and 3.1.1.2. The gels showed that the proteins were inducible, and expressed in high amounts. The soluble fractions of the proteins were used for the purification experiments. Figure 3.1.1.2 shows that there is a lot of protein remaining in the cell pellet after lysis. This indicates that the proteins aggregate and form inclusion bodies in the cells. It is possible to purify the aggregated protein from the inclusion bodies, but refolding the protein *in vitro* gives no guarantee when it comes to structural integrity of the protein. We decided to transform the constructs into a bacterial strain engineered for expressing recombinant protein at low temperatures called ArcticExpress (Agilent Technologies) to investigate if this would help with the aggregation issue.

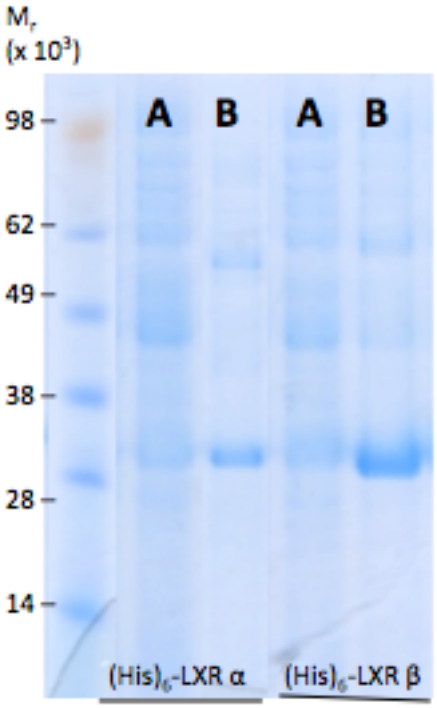
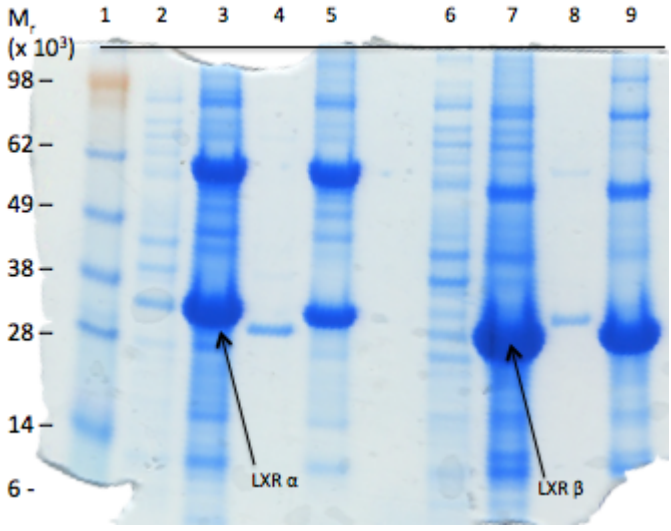


Figure 3.1.1.1 Coomassie stained SDS gel showing the expression of $(His)_6$-LXRα LBD and $(His)_6$-LXRβ LBD in BL21DE3 E. coli cells. A: Samples pre-induction, B: Samples taken after cell harvesting.



Lanes:
1. Molecular weight marker;
2. Supernatant from the first pelleting of LXRα.
3. Pellet from the first pelleting of LXRα.
4. Supernatant from the second pelleting of LXRα.
5. Pellet from the second pelleting of LXRα.
6. Supernatant from the first pelleting of LXRβ.
7. Pellet from the first pelleting of LXRβ.
8. Supernatant from the second pelleting of LXRβ.
9. Pellet from the second pelleting of LXRβ.

Figure 3.1.1.2. Coomassie stained SDS gel showing the lysis stages of BL21DE3 cells expressing $(His)_6$-LXRα LBD and $(His)_6$-LXRβ LBD. The samples were pelleted twice to extract more protein in the supernatant.

3.1.2 Expression of $(His)_6$-LXRα and $(His)_6$-LXRβ in ArcticExpress competent cells.

ArcticExpress (AE) is *E. coli* cells engineered for improved protein processing at low temperatures. These cells co-express the cold-adapted chaperonins Cpn10 and Cpn60 from the psychrophilic bacterium *Oleispira Antarctica*. These chaperonins have 74% and 54% sequence identity to the *E. coli* chaperonins GroEL and GroES, respectively. The sizes of the chaperonins are, as the names suggest, 10 kDa and 60 kDa respectively. An expression test of the ArcticExpress cells was performed according to the standard expression protocol described in Section 2.1. The ArcticExpress cells were incubated at 12 °C during expression. The result of the SDS-PAGE analysis is shown in Figure 3.1.2.1. The gels show that there is no expression of the LXR LBDs in the ArcticExpress cells when following the standard Bl21DE3 expression protocol at low temperature. An ArcticExpress-optimized expression protocol, from Agilent Technologies, was investigated (*47*).

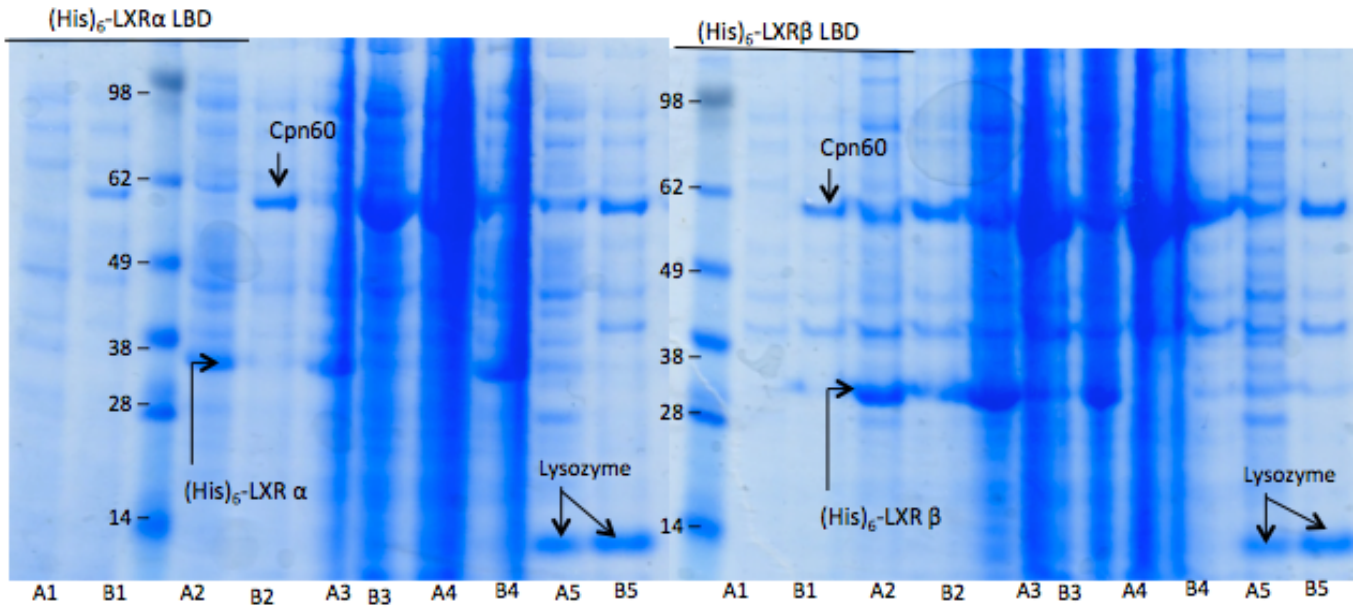
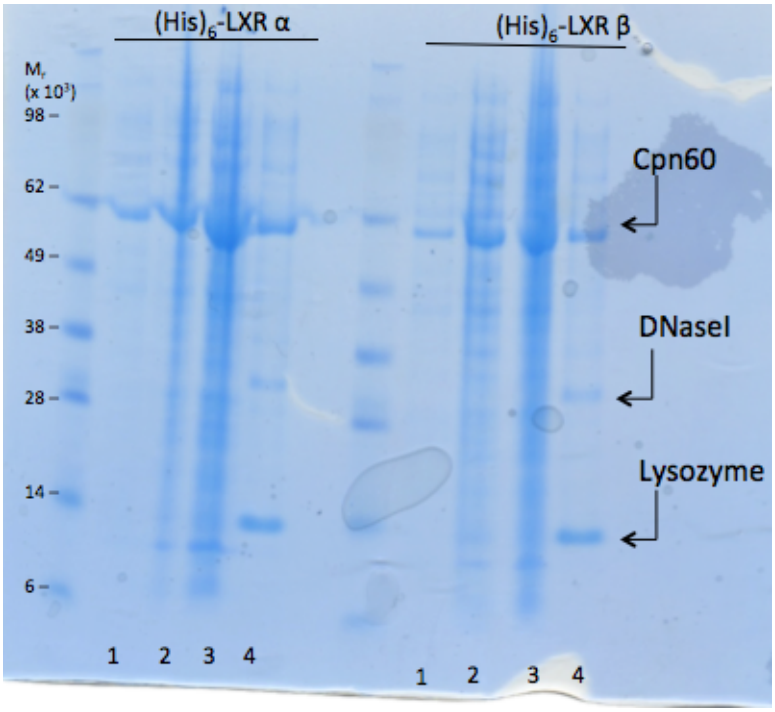


Figure 3.1.2.1. Coomassie stained SDS gels showing the stages of expression and lysis of $(His)_6$-LXRα LBD and $(His)_6$-LXRβ LBD in BL21DE3 (A#) and ArcticExpress (B#) cells. Lanes: 1, Pre-induction; 2, Before harvesting the cells; 3, Pellet before lysis; 4, Pellet after lysis; 5, Supernatant after lysis.

The expression of the LXR LBDs in ArcticExpress was performed at 12 °C for 24 hours, before the cells were harvested and lysed as described in Section 2.1.2. The result of the SDS-PAGE analysis is shown in Figure 3.1.2.2. The gel shows that the ArcticExpress cells do not express the LXR LBDs. The small band visible on the gel around approximately 31 kDa is due to the addition of deoxyribonuclease I during the lysis. We do not know why the ArcticExpress cells did not express the proteins. An explanation would be that the cells might have lost the plasmid. We decided it was not worth the time to pursue the ArcticExpress experiment since the choice of constructs probably were responsible for the aggregation of the proteins, hence ArcticExpress cells give no guarantee of improving the aggregation.

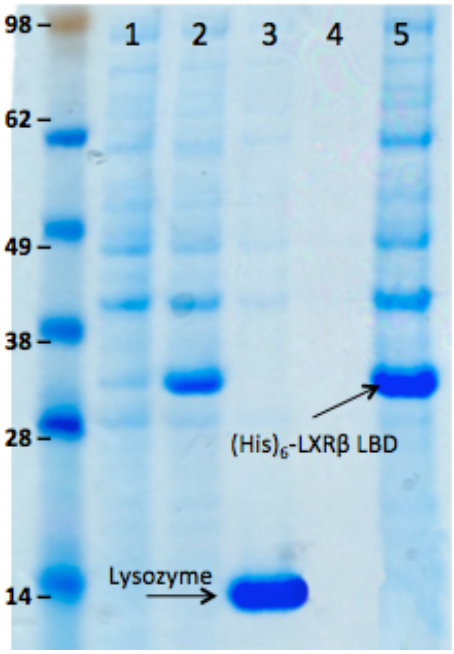


Lanes:
1. Pre-induction
2. The cell medium before harvest
3. Cell pellet after lysis
4. Supernatant after lysis

Figure 3.1.2.2. Coomassie of SDS gel showing the stages of expression and lysis of $(His)_6$-LXRα and $(His)_6$-LXRβ expressed in ArcticExpress cells at 12 °C.

3.1.3 **Expression of $(His)_6$-LXRβ LBD within one day at steady temperatures.**

Experiments with expression of $(His)_6$-LXR LBDs showed high aggregation *in vivo,* and precipitation *in vitro* under varying temperatures (See 3.2 – Inclusion body purification). Fradera *et al.* successfully purified the LXRα LBD, but mentions under the methods and materials section that all of the purification steps were performed within one day at 4 °C. This sentence might indicate that Fradera *et al.* experienced solubility and precipitation issues. We decided to investigate if expression and purification of $(His)_6$-LXRβ LBD within one day, at steady temperature would significantly improve the issue of *in vivo* aggregation and *in vitro* precipitation. The expression of $(His)_6$-LXRβ LBD was performed according to the original protocol (Section 2.1), except that the sample was immediately taken for the next step without any storing in fridge or freezer. Deoxyribonuclease I was added during lysis due to high viscosity, the cells were pelleted twice to extract more protein, and the supernatants were pooled. After harvest, the supernatant was immediately applied on an equilibrated, nickel loaded IMAC column and eluted in 1.0 ml fractions, without knowing if the supernatant contained any of the protein. A zoomed version of the chromatogram from the IMAC purification is shown in Figure 3.1.3.1. The peak in the flow-through shoulder and the early elution peak were analyzed by SDS-PAGE. The results of the SDS-PAGE analyses of expression and harvest samples and the elution peaks are shown in the Figures 3.1.3.2 and 3.1.3.3.



Lanes:
1. Pre-induction
2. The cell medium before harvest
3. First supernatant after lysis
4. Second supernatant after lysis
5. Cell pellet after lysis.

Figure 3.1.3.2. Coomassie stained SDS gel showing the stages of expression and lysis of $(His)_6$-LXRβ LBD all performed within one day. The sample was pelleted twice to extract more protein and the supernatants pooled.

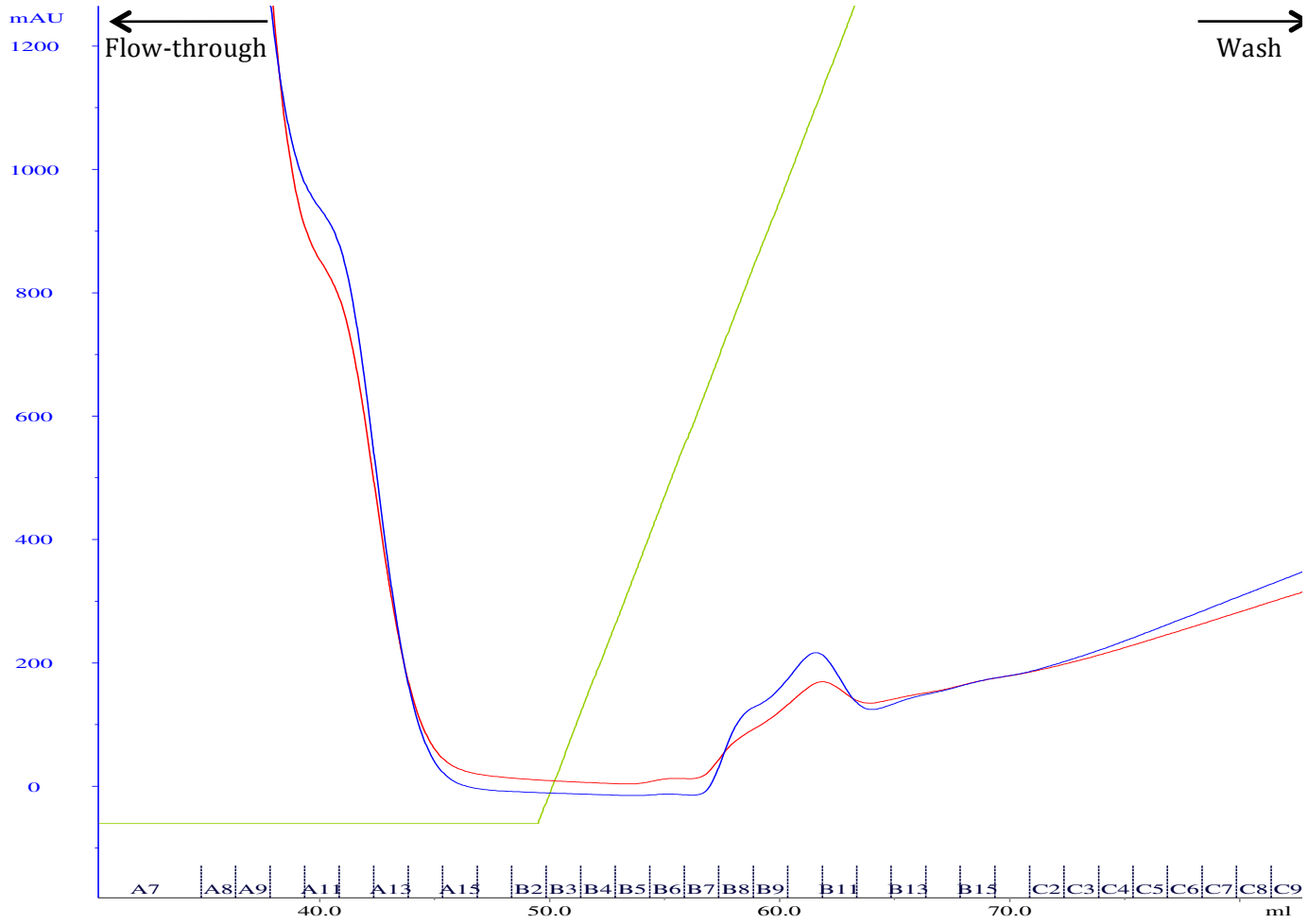


Figure 3.1.3.1. The zoomed chromatogram of the peaks from elution of $(His)_6$-LXRβ LBD from a HiTrap Chelating HP column loaded with nickel. Green graph: concentration of elution buffer. Blue graph: absorption at 280 nm. Red graph: absorption at 254 nm. Vertical-axis: absorbance displayed in milli absorbance units. Horizontal-axis: elution buffer volume measured in milliliters.

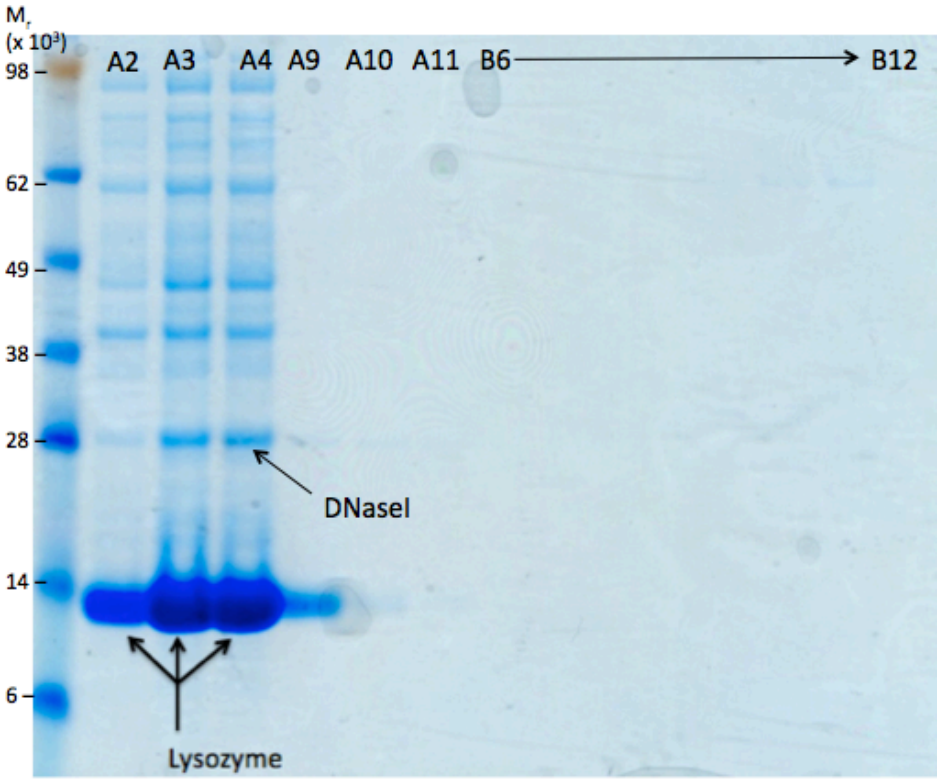
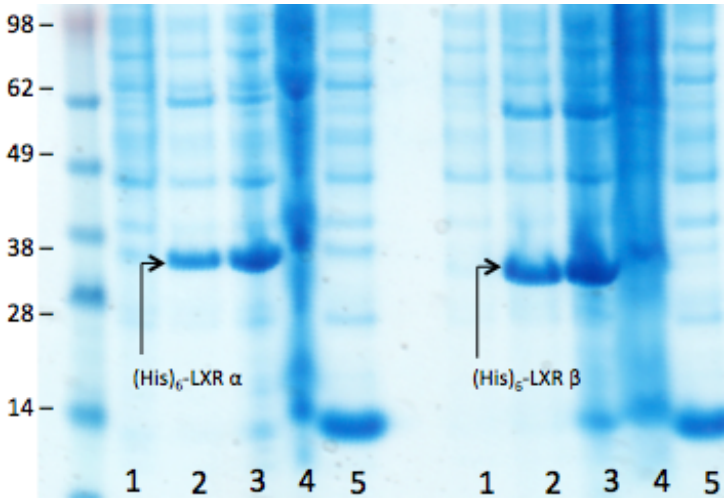


Figure 3.1.3.3. Coomassie stained SDS gel showing the analyzed peaks from the chromatogram in Figure 3.1.3.1. The fractions from the chromatogram are shown in the figure.

The gel in Figure 3.1.3.2 shows that the protein is highly expressed after induction, but that all of the protein remains in the pellet after lysis even though all temperature variations were kept to a minimum during the experiment. The gel in Figure 3.2.3.3 shows that the first peak (flow-through) in the chromatogram is due to the high concentration of lysozyme in the sample and that the second peak in the chromatogram is not even protein. This experiment shows that the construct for $(His)_6$-LXRβ LBD is highly unstable and aggregates *in vivo*.

3.1.4 **Expression of pRSET B constructs in presence of cholesterol**

Hoerer *et al.* found that the original expression test of LXRβ LBD yielded predominantly insoluble protein, and the soluble fraction was unstable and could not be suitably concentrated for protein crystallization (*35*). Addition of the synthetic ligand, T0901317, omnipresent during expression, purification and crystallization stabilized the protein significantly. We decided to do an experiment with ligand omnipresent during expression, and added cholesterol to all solutions and buffers during the expression. Cholesterol was chosen as the ligand because it is cheap and the aim of this master's project is to obtain structures of the LXR LBDs with oxysterols bound. If the result from the cholesterol experiments looked promising, other oxysterols could tried be used during expression. Cholesterol is a lipid molecule that is required to maintain both membrane structural integrity and fluidity in eukaryotic cell, however, it is almost completely absent among prokaryotes, so supplying cholesterol in the growth medium of the *E. coli* cells will lead to a decrease in the membrane fluidity. A publication by Moreno *et al.* investigated the influence of incorporating cholesterol in the growth medium of *E. coli* (*54*). They found that cholesterol in ethanol solution could be supplemented to the cells to a concentration of 40 µM before co-precipitation of cholesterol and cells occurred. LXR LBDs were produced as according to the original protocol (Section 2.1). Ethanol saturated with cholesterol was added to the cultures before induction to a final concentration of 40 µM. The expression was incubated overnight at 18 °C. The result of the SDS-PAGE analysis is shown in Figure 3.1.4.1. The gel shows that the incorporation of cholesterol during the expression had no effect on the aggregation of the proteins.



Lanes:
1. Pre-induction
2. The cell medium before harvest
3. Cell pellet after lysis
4. Error in application
5. Supernatant after lysis

Figure 3.1.4.1. Coomassie of SDS gel showing the stages of expression and lysis of His-tagged LXR LBDs expressed in the presence of cholesterol at 18 °C. The molecular weight marker is given in kDa.

3.1.5 **Purification of the soluble $(His)_6$-LXR LBDs by IMAC and GF**

The purification was performed for both isoforms of the LXR LBDs. The cell extract from the expression was applied to a nickel-loaded affinity column (HiTrap Chelating HP from GE Healthcare). His-tags have a high affinity for Ni^{2+}, this ensures that the fusion proteins bind to the column resin. To elute the proteins, imidazole-containing buffers were used. A linear gradient of increasing imidazole concentration was used for the elution, and fractions were collected. A zoomed version of a chromatogram from one of the $(His)_6$-LXRα LBD IMAC purifications is shown in Figure 3.1.5.1. In the figure a small peak is visible in the middle of the elution. Fractions from the wash, the first peak and the elution peak were analyzed by SDS-PAGE. The result of the analysis is shown in Figure 3.1.5.2. Fractions B9 to C1 were pooled and concentrated in a centrifugal concentrator with 5 kDa MWCO (Vivaspin 20 from Sartorius) to a volume of 200 µl. The sample was loaded on a size-exclusion column (SuperDex200 10/300 GL from GE Healthcare) and eluted in 1 ml fractions. A zoomed version of the chromatogram from the size-exclusion chromatography is shown in Figure 3.1.5.3. Fractions from the first two peaks in the chromatogram were analyzed by SDS-PAGE. After Coomassie Blue staining of the gels, the bands were barely visible, so the gels were stained using the SilverQuest™ staining kit (Invitrogen). The gel from the second peak showed nothing, indicating that this peak is due to salt, which is supported by the conductivity. The gel from the first peak is shown in Figure 3.1.5.4. Fractions A4-A12 were pooled, diluted 1:2 with MQ-H_2O and concentrated to a volume of approximately 125 µl. The concentration of $(His)_6$-LXRα LBD was determined to be 6 mg/ml.

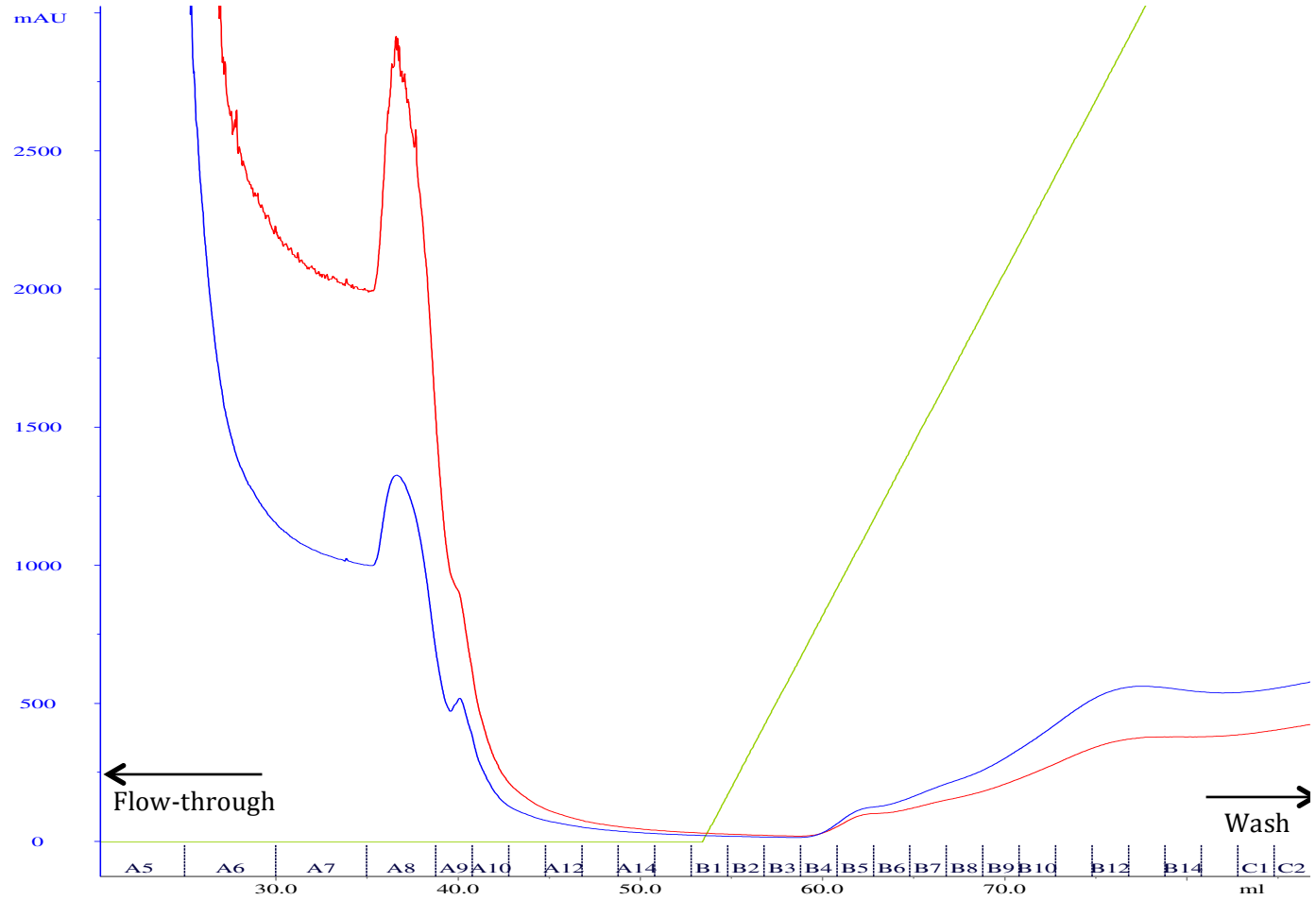


Figure 3.1.5.1. The zoomed chromatogram of the peaks from elution of $(His)_6$-LXRα LBD from a HiTrap Chelating HP column loaded with nickel. Green graph: concentration of elution buffer. Blue graph: absorption at 280 nm. Red graph: absorption at 254 nm. Vertical-axis: absorbance displayed in milli absorbance units. Horizontal-axis: elution buffer volume measured in milliliters.

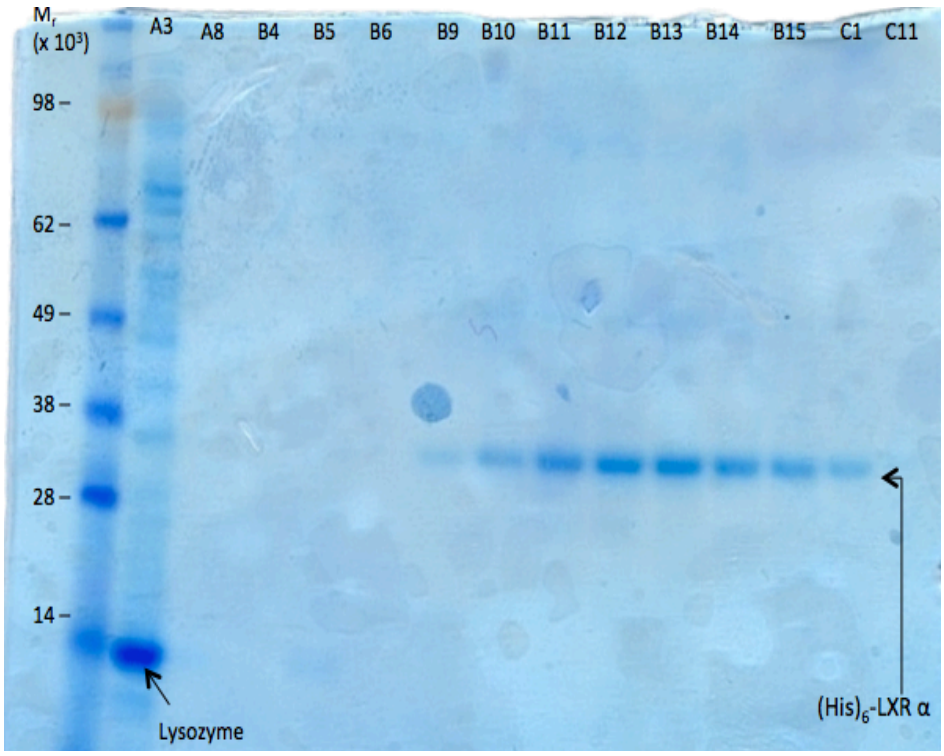


Figure 3.1.5.2. Coomassie stained SDS gel showing the analyzed fractions possibly containing $(His)_6$-LXRα LBD from the chromatogram in Figure 3.1.5.1. The fraction numbers from the chromatogram are shown in the figure.

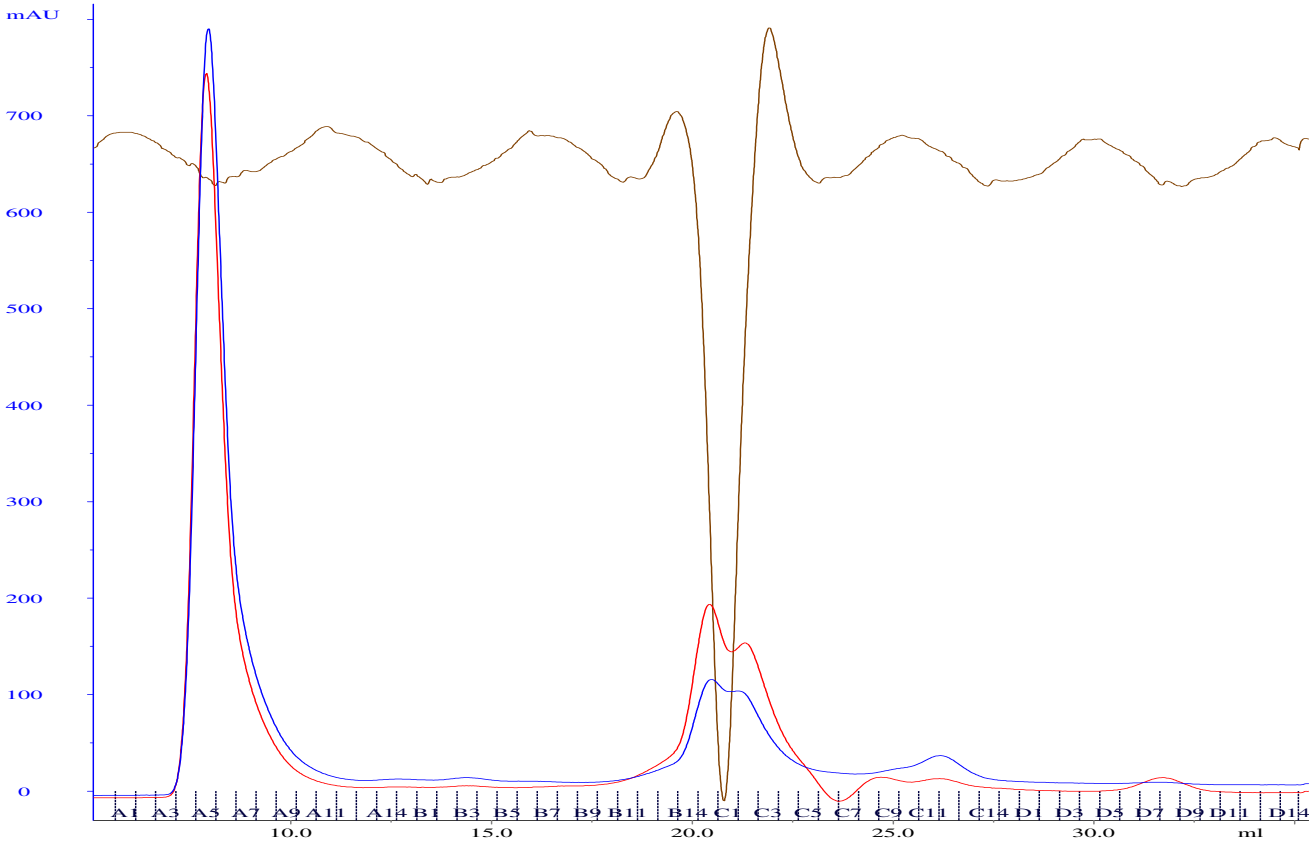
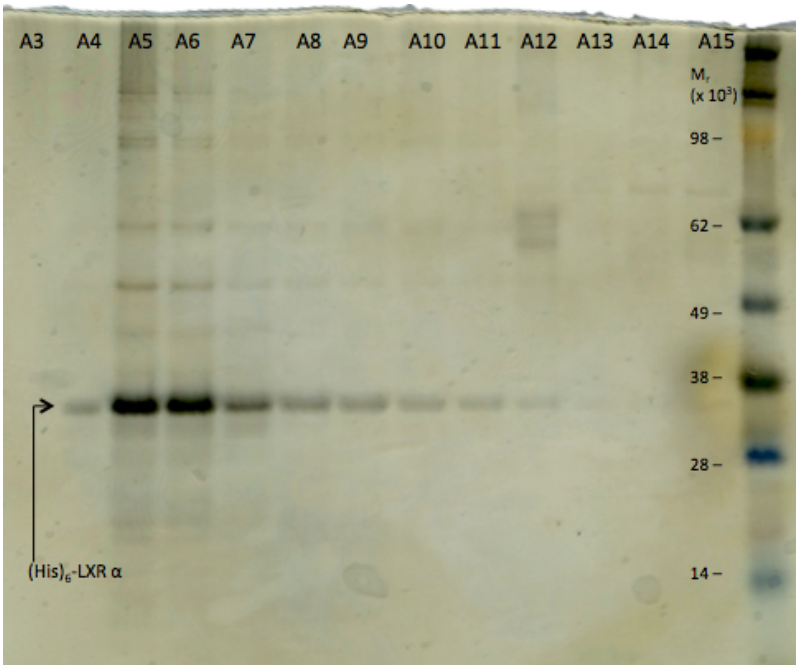


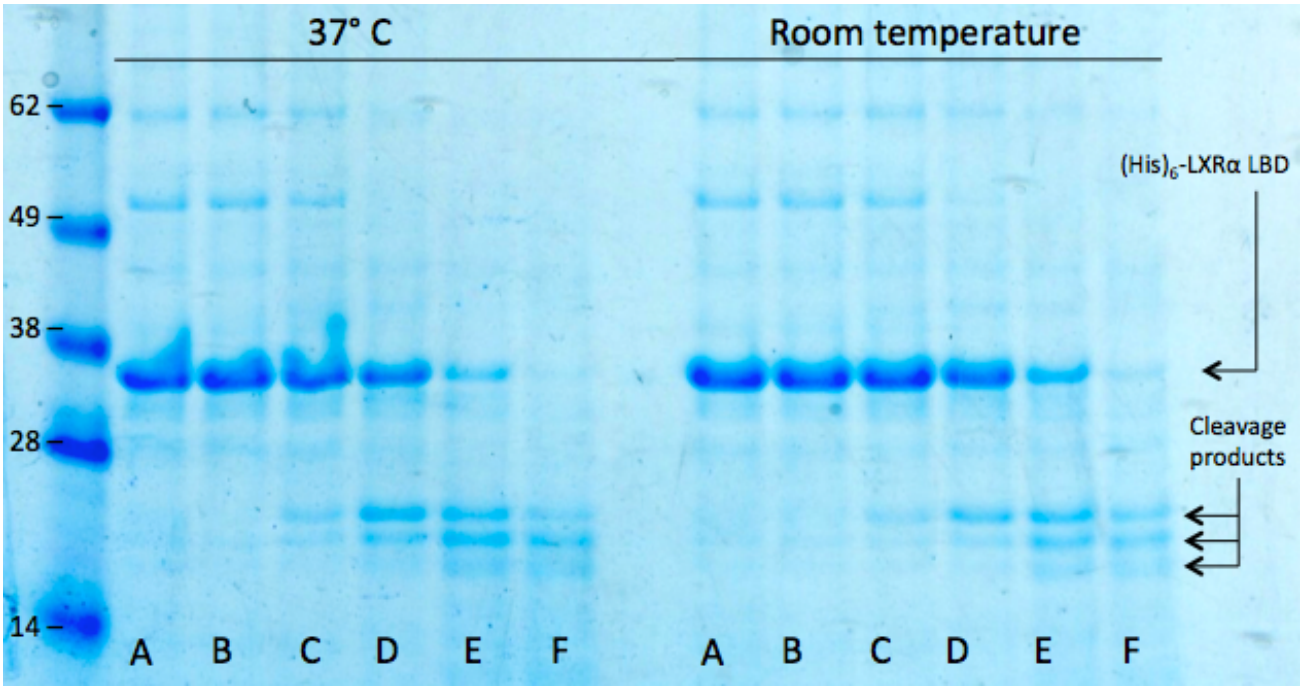
Figure 3.1.5.3. Zoomed chromatogram from the SuperDex200 10/300 GL gel filtration of $(His)_6$-LXRα LBD. Green graph: concentration of elution buffer. Brown graph: conductivity. Blue graph: absorption at 280 nm. Red graph: absorption at 254 nm. Vertical-axis: absorbance displayed in milli absorbance units. Horizontal-axis: elution buffer volume in milliliters.



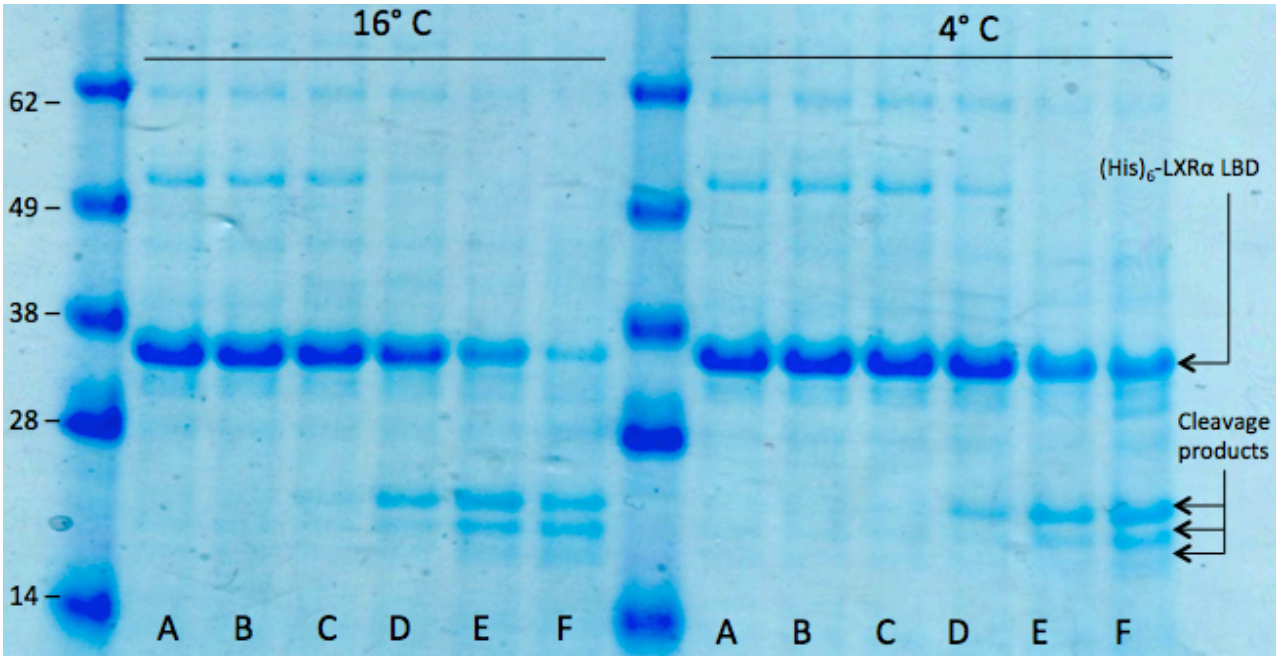
Figures 3.1.5.4. Silver stain of SDS gel showing the analyzed fractions from the purification of $(His)_6$-LXRα LBD on a SuperDex200 10/300 column. The corresponding chromatogram is shown in Figure 3.1.5.3.

3.1.6 **Cleavage of the His-tag with enterokinase**

The LXR LBDs were purchased as fusion proteins with an N-terminal hexahistidine tag separated from the proteins by an epitope and an enterokinase cleavage site. Enterokinase is a serine protease that recognizes the sequence Asp-Asp-Asp-Asp-Lys and cleaves the peptide bond after the lysine residue. EnterokinaseMax™ (Invitrogen) was used for cleaving off the His-tag. EKMax™ is a recombinant preparation of the catalytic subunit of bovine enterokinase produced and purified from the yeast, *Pichia pastoris*, yielding an enzyme with higher specific activity than the native version (*48*). The construct $(His)_6$-LXRα LBD has an expected theoretical size of approximately 34.6 kDa, and after cleavage of the His-tag the expected theoretical size would be 30.6 kDa. Purified soluble $(His)_6$-LXRα LDB was used for the enterokinase cleavage test. An assay was setup to screen for the suitable cleavage conditions varying the amount of protease used and temperature (See Table 2.2.1). The samples were analyzed by SDS-PAGE. The result of the analysis is shown in Figures 3.1.6.1 and 3.1.6.2.



Figures 3.1.6.1. Coomassie stained SDS gel showing the result of a cleavage test performed at 37 °C and room temperature, with enterokinase on purified $(His)_6$-LXRα LBD. A to F shows increasing amounts of protease used for the digestion. Table 2.2.1 shows detailed reaction parameters for the cleavage test. The molecular weight marker is in kDa



Figures 3.1.6.2. Coomassie stained SDS gel showing the result of a cleavage test performed at 16 °C and 4 °C, with enterokinase on purified $(His)_6$-LXRα LBD. A to F shows increasing amounts of protease used for the digestion. Table 2.2.1 shows detailed reaction parameters for the cleavage test. The molecular weight marker is in kDa.

The expected results from the cleavage test would be that the 34.6 kDa fusion protein band fading and another band appearing at 30.6 kDa for increasing amounts of enterokinase. The figures show that the 34.6 kDa fusion protein band is fading, but there is no band appearing right below. There are however several bands appearing at approximately half the size of the fusion protein band. These results indicate that enterokinase is unspecific when cleaving the protein. Although the manufacturer claims enterokinase to be highly specific, it has been shown that this is not the case (*55-57*). Liew *et al.* showed that enterokinase was compromised by non-specific cleavage at three other sites than the expected DDDDK, and that one of the secondary sequences was cleaved at greater efficiency. Shahravan *et al.* found that changing several parameters, pH, temperature, the amount of EK used, did not prevent the unspecific hydrolysis. These authors suspected that aggregation impeded the accessibility of the enterokinase cleavage site and tried to add denaturant to the cleavage buffer. This addition reduced the unspecific hydrolysis, but did not prevent it entirely. Furthermore, the addition of denaturants like urea to the reaction buffer has an inhibitory effect on the activity of enterokinase at higher concentrations. Hosfield and Lu found that by addition of aspartic acid directly downstream of the enterokinase target site, they obtained 84% cleavage. Shahravan *et al.* had an aspartic acid directly downstream of the target site in their sequence, and this did not affect the cleavage rate. The additional cleavage sequences discovered by Liew *et al.* are not in the LXR LBD sequences, but as enterokinase is so highly unspecific, there might be additional unknown enterokinase cleavage sites in the LXR LBD sequences that are similarly or more efficient than the canonical DDDDK.

3.1.7 **Robotic crystal screening**

The purified $(His)_6$-LXRα LBD from the soluble fraction was used for a crystallization trial, to investigate if the N-terminally tagged protein could be directly used for crystallization trials, without removing the tag. The protein mix contained approximately 5.9 mg/ml $(His)_6$-LXRα LBD, 25 mM NaCl, 7 mM Tris pH 8.0 and >0.5 mM DTT. The crystal screens used were the JCSG+ suite (Molecular Dimenstions) and Morpheus® (Molecular Dimensions). The screens were set up in 96 well plates (MRC 2 Well Crystallization Plates from Swissci) using the Oryx 4 crystallization robot. They were stored in a dark room at 20 °C, and regularly inspected. After four months a crystal appeared in the JCSG+suite under the conditions: 0.2 M lithium sulfate, 0.1 M Bis-Tris pH 5.5, 50% PEG400. The droplet is shown in Figure 3.1.7.1. The biggest crystal in the droplet was fished, frozen in a cryostream and tested at European Synchrotron Research Facility in Grenoble France. The initial X-ray data showed that the crystal was salt. Even with big amounts of pure protein, obtaining protein crystals is one of the biggest challenges in protein crystallography. In this case, the concentration of pure protein was relatively low, and in addition the N-terminal purification tag was still on the protein. This flexible tag might prevent the proteins from coming together in an ordered way and forming crystals.

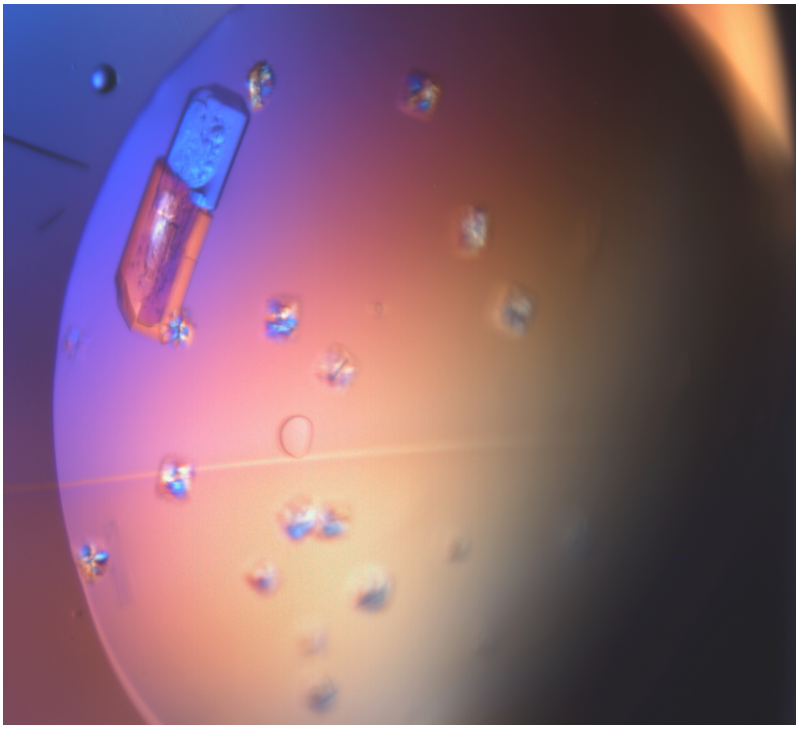


Figure 3.1.7.1 A picture of $(His)_6$-LXRα LDB in the JCSG+ Suite crystallization screen under the conditions: 0.2 M lithium sulfate, 0.1 M Bis-Tris pH 5.5, 50% PEG400

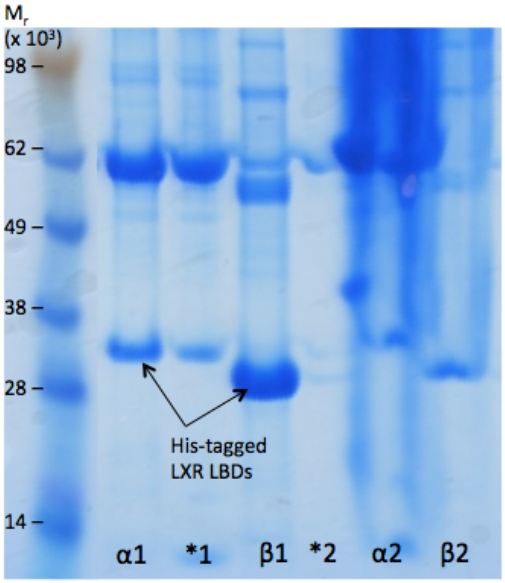
Conclusion of Strategy 1

Strategy 1 was based on the assumption that the expression and purification of the LXR LBDs would be straightforward. The gene lengths were designed based on the articles on crystal structures of LXR LBDs (*33-35, 37, 42*) and out of convenience purchased in the pRSET B expression vector as this vector encodes a N-terminal fusion peptide on the protein with a hexahistidine purification tag separated from the protein by an enterokinase cleavage site. The enterokinase was reported it to be highly specific. The original expression tests with the constructs in 1 l medium yielded little or no soluble protein. For $(His)_6$-LXRα LBD, the final concentration after two-step chromatography purification was determined to be 5.9 mg/ml in a volume of 125 µl, giving a yield of 0.73 mg. For $(His)_6$-LXRβ LBD, no pure protein was obtained from the soluble fraction, as the protein was lost during purification. Scaling the expression up was considered, but cleavage of the purification tag had to be facilitated with enterokinase. The enterokinase digestion test on $(His)_6$-LXRα LBD showed that enterokinase is highly unspecific at a variety of temperatures and concentrations. Enterokinase has also been found to be highly unspecific when varying the pH (*55*). Optimization of enterokinase digestion is possible by adding or deleting residues downstream or upstream of the canonical cleavage site of enterokinase, but appears to be protein-specific, and may not be applied to all fusion proteins. Shahravan *et al.* found that denaturing conditions under enterokinase digestion could improve the unspecific cleavage, which would denature the target protein and require refolding *in vitro*. The experiments with expression showed that the proteins were overexpressed, but most of the proteins were in an insoluble fraction after harvest. We decided to isolate the $(His)_6$-LXR LBDs from the inclusion bodies and do an *in vitro* refolding of the proteins (Strategy 1B). We decided to circumvent the enterokinase issues by introducing a new cleavage site for Tobacco Etch Virus protease (TEVp) into the sequence by subcloning (Strategy 1C). The crystallization test with $(His)_6$-LXRα LBD did not yield protein crystals. Obtaining protein crystals can often be unpredictable, so it cannot be claimed that the absence of crystals is due to crystallization conditions, protein concentration or the flexible N-terminal fusion peptide on the LXRα LBD. Histidine tags may be advantageous in crystallization as the tag allows coordination of the protein to nickel ions, if present, in the mother liquor. In this construct the hexahistidine section is separated from the protein by an epitope and an enterokinase cleavage site. When the sequences of the constructs were run through the Multilayered Fusion-Based Disorder Predictor – Web Server (*58*), the results predicted with a score of 99 that the fusion peptide tail (47 residues) and the N-terminal regions of the proteins are disordered. This high level of disordered residues on the terminal of the proteins is definitely not advantageous for crystal formation and points to that the purification tag has to be removed from the protein before crystallization. We decided to optimize purification and purification-tag cleavage before doing more crystallization trials.

3.2 Strategy 1B: *in vitro* refolding of inclusion body purified proteins

3.2.1 Purification and refolding of proteins expressed as inclusion bodies

Inclusion body purification was performed in parallel for both of the LXR LBD isoforms. Figures for $(His)_6$-LXRα LBD are not shown. The cell pellet from the normal expression was used for the inclusion body experiment. The inclusion bodies were extracted by washing the pellet in CelLytic B and dissolved in highly concentrated urea buffer. The SDS PAGE analysis of the samples taken during the purification is shown in Figure 3.2.1.1. The proteins were refolded by rapid dilution into buffer, without and with 50 μM cholesterol, and no denaturants. During the refolding of the proteins in presence of cholesterol, immediate precipitation was observed in the $(His)_6$-LXRβ LBD sample. $(His)_6$-LXRα LBD precipitated after stirring for one hour in presence of cholesterol. The refolded protein samples, from refolding without cholesterol, were purified by IMAC on a nickel-loaded column. A close up view of the chromatogram from the analysis is shown in Figure 3.2.1.3. Fractions J8-K4 were analyzed by SDS-PAGE (Figure 3.2.1.2). Coomassie staining did not give good enough sensitivity, so the gels were silver stained. The gel shows that the peak from the chromatogram contained $(His)_6$-LXRβ LBD and some impurities. The fractions containing the protein were pooled and the protein concentrations in the pooled samples were determined to be 0.1 mg/ml and 0.4 mg/ml for $(His)_6$-LXRα LBD and $(His)_6$-LXRβ LBD, respectively. The proteins were concentrated in a centrifugal concentrator unit with 10 kDa MWCO (Vivaspin 20 from Sartorius). During the concentration, the proteins precipitated.



Lanes:
α1: Supernatant of $(His)_6$-LXRα LBD.
***1**: Error in application.
β1: Supernatant of $(His)_6$-LXRβ LBD.
***2**: Error in application.
α2: Pellet of $(His)_6$-LXRα LBD.
β2: Pellet of $(His)_6$-LXRβ LBD

Figure 3.2.1.1. Coomassie stained SDS gel showing the inclusion body extraction of $(His)_6$-LXRα LBD and $(His)_6$-LXRβ LBD.

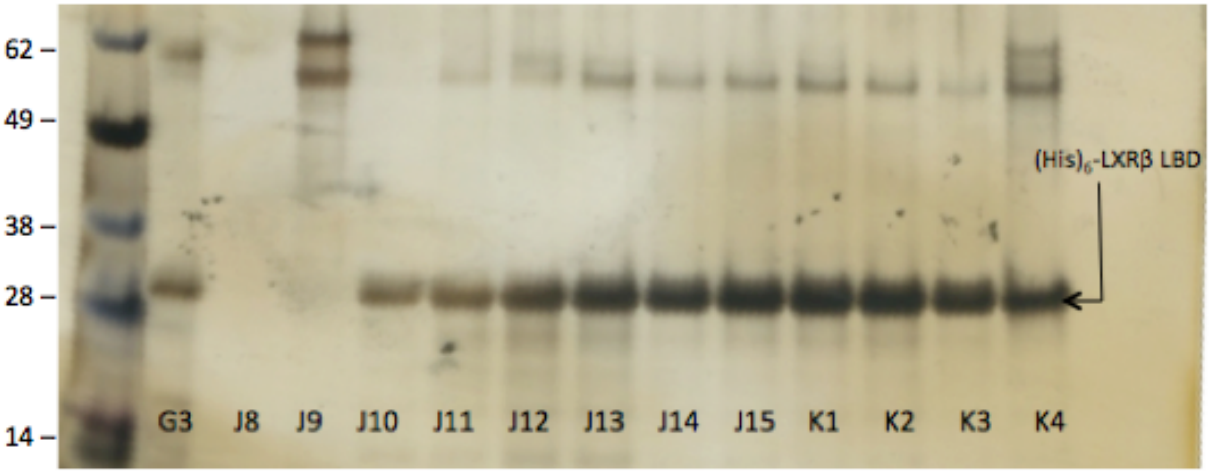


Figure 3.2.1.2. Silver staining SDS gel showing the analyzed fractions containing inclusion body isolated $(His)_6$-LXRβ LBD from the chromatogram in Figure 3.2.1.2. The fraction numbers from the chromatogram are shown in the figure. Molecular weight marker is in kDa.

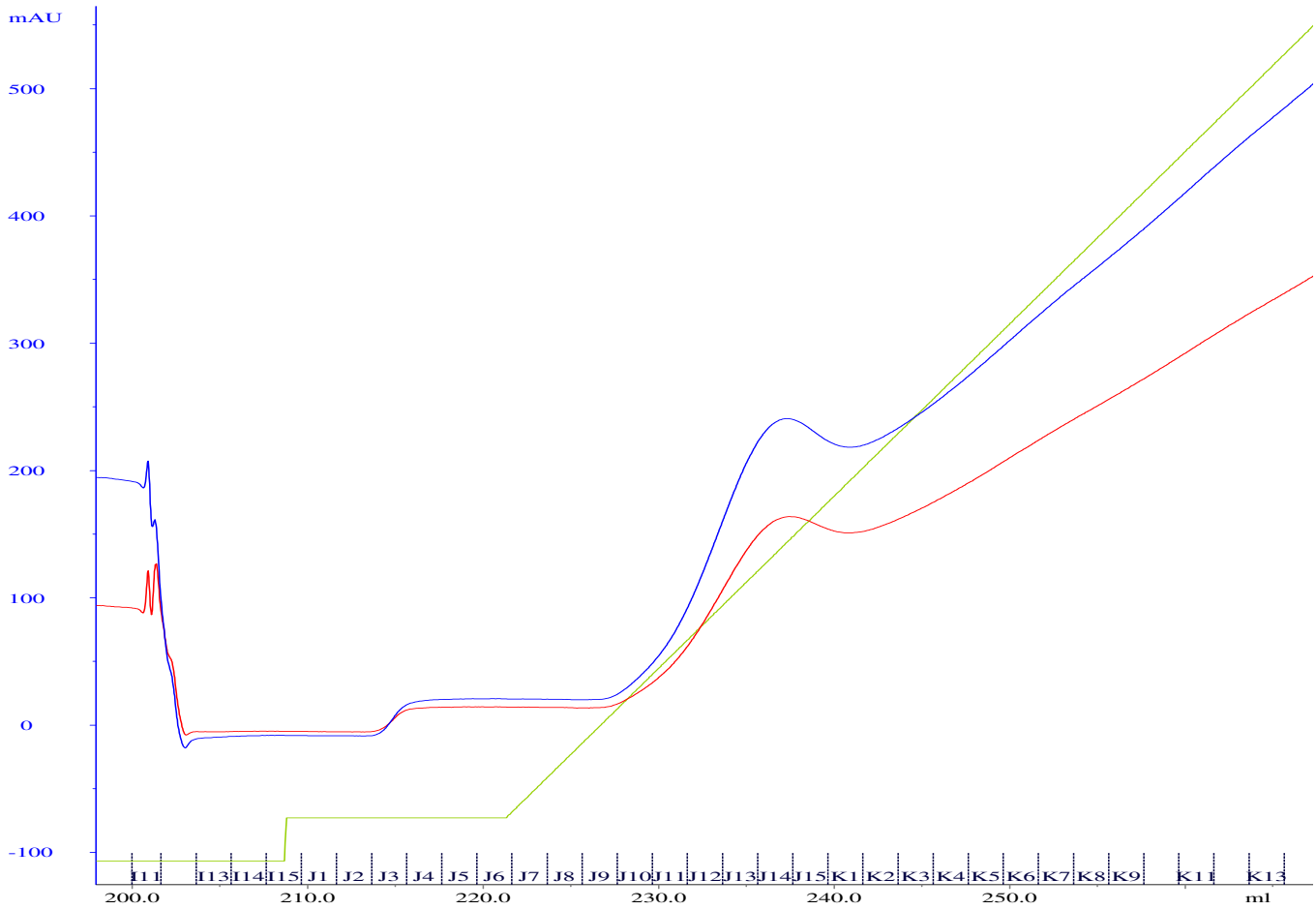


Figure 3.2.1.3. Zoomed chromatogram of elution of inclusion body isolated $(His)_6$-LXRβ LBD from the HiTrap Chelating HP column loaded with nickel. Blue graph: absorption at 280 nm. Red graph: absorption at 254 nm. Green graph: concentration of imidazole. Vertical-axis: absorbance displayed in milli absorbance units. Horizontal-axis: elution buffer volume measured in milliliters.

Conclusion of Strategy 1B

This experiment showed that it is possible to isolate the $(His)_6$-LXR LBDs from the inclusion bodies formed during expression, but the precipitation of the protein during concentration indicates that inclusion body purified protein is not stable in solution, even at low concentrations. The results from the inclusion body experiments indicate that the constructs are unstable and aggregate. Refolding of the constructs *in vitro* might be possible, but this requires searching for optimized conditions, which might be a relatively random process. The constructs might require a binding partner, a co-activator peptide or a ligand to be stabilized, which is hard to predict in advance of the experiments. We concluded that the optimization of the refolding would be to time consuming and economically unfavorable to pursue and decided to scale up the expression if the subcloning of a new cleavage site in the construct was successful.

3.3 Strategy 1C: Subcloning of Tobacco Etch Virus protease cleavage site into the pRSET construct

The sequences of the oligonucleotides for the TEVp cleavage site used for the subcloning were 5'-TCG AGC GAA AAC CTG TAT TTC CAG GGC-3' and 5'-TCG AGC CCT GGA AAT ACA GGT TTT CGC-3' for the forward and backward oligos respectively. The pRSET B plasmids with the $(His)_6$-LXR LBD genes were linearized by double digestion with *Xho*I and *Bam*HI and purified by agarose gel electrophoresis (Figure 3.3.1). Undigested plasmids were used as a reference.

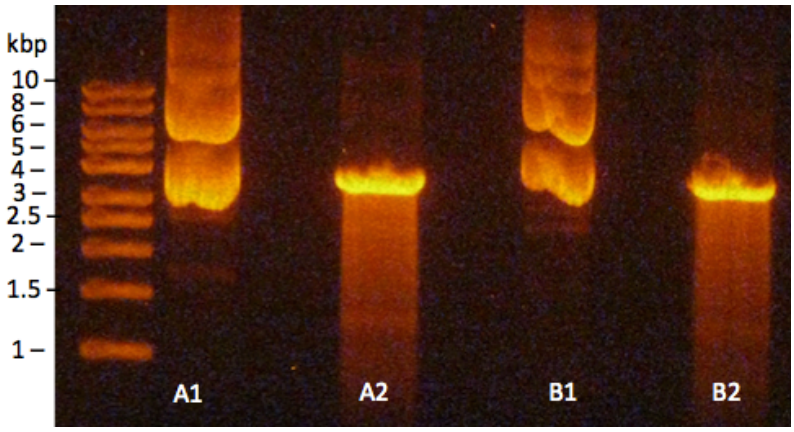


Figure 3.3.1. A 0.8% agarose gel showing supercoiled and linearized pRSET B plasmids. **A#**: with LXRα LBD gene. **B#**: with LXRβ LBD gene. **1**: undigested plasmid. **2**: double digested plasmid

The concentrations of the agarose gel-purified linearized plasmids were determined by spectrophotometry to be 14 ng/µl and 12 ng/µl for $(His)_6$-LXRα LBD and $(His)_6$-LXRβ LBD, respectively. The newly hybridized oligos were determined by spectrophotometry to have a concentration of 636 ng/µl. The linearized plasmids and double stranded oligos were ligated, transformed into Stellar™ cells (Clontech) and plated on LBamp plates. Promising colonies were picked and used to inoculate 5 ml cultures, which were grown overnight at 37 °C. The plasmids were purified from the cells using QIAprep Spin Miniprep Kit (Qiagen). Nineteen samples were sequenced by an in-house sequencing service. The results were compared with the original $(His)_6$-LXR LBD plasmids, and showed that none of the sequenced samples contained the TEVp cleavage site.

Conclusion of Strategy 1C

This experiment was conducted several times with the same negative result. None of the sequenced samples contained the TEVp cleavage site. The pRSET B vector contains an N-terminal fusion peptide fused to the LXR LBDs, as described in Strategy 1. As experiments have shown these constructs tend to aggregate and form insoluble inclusion bodies during expression with little or no soluble protein left in the supernatant. The addition of the TEVp cleavage site would lead to reliable cleavage of the tag, but as all of the experiments with pRSET B plasmids have been dead ends and required further optimization, we decided to try and circumvent all of the experienced problems and clone the genes into another expression vector coding for periplasmic expression, see Strategy 2.

3.4 Strategy 2: cloning of the LXR LBD genes into the pFKPEN periplasmic expression vector

The periplasmic expression vector was a kind gift from Geir Åge Løset (Department of Biosciences – The Faculty of Mathematics and Natural Sciences, University of Oslo) obtained via the hands of Master's student Hedda Johannesen, in XL1 Blue $CuCl_2$ cells (Stratagene). The In-Fusion HD cloning kit (Clontech) was used for the cloning reactions. This kit is designed for fast, directional cloning of one or more fragments of DNA into any vector by fusing PCR-generated fragments into linearized vectors using the In-Fusion enzyme premix that recognizes a 15 base pair (bp) overlap at the ends of the fragments and vectors. One set of primers was designed for amplifying the LXR LBD genes, and one set of primers was designed for adding a C-terminal hexahistidine tag on the LXR LBDs. All primers were designed to generate a 15 bp overlap with the pFKPEN vector on both termini of the LXR LBD genes. The hexahistidine tag was designed to be non-cleavable. The pRSET B plasmids were used together with the primers for PCR. To engineer the C-terminal hexahistidine tag the samples were amplified twice, first to add the His-tag, then to add the 15 bp overlap with pFKPEN. The pFKPEN plasmid was purified from the cells and linearized by double digestion with *Nco*I and *Not*I. The PCR products and linearized vector were purified by agarose gel electrophoresis (Figure 3.4.1 to Figure 3.4.3). Due to high amounts of sample the LXRα LBD-$(His)_6$ and LXRβ LBD-$(His)_6$ samples were loaded in two wells, respectively. The lengths of the genes and linearized plasmid are 814 bp, 763 bp and 3996 bp for LXRα LBD, LXRβ LBD and pFKPEN, respectively. The band at approximately 4 kbp in Figure 3.4.1 was purified and the concentration of the linearized plasmid was determined by spectrophotometry to be 6.5 ng/µl. The PCR products were determined to have DNA concentrations of 20 ng/µl, 77 ng/µl, 2.5 ng/µl and 10 ng/µl for LXRα LBD, LXRβ LBD, LXRα LBD-$(His)_6$ and LXRβ LBD-$(His)_6$, respectively.

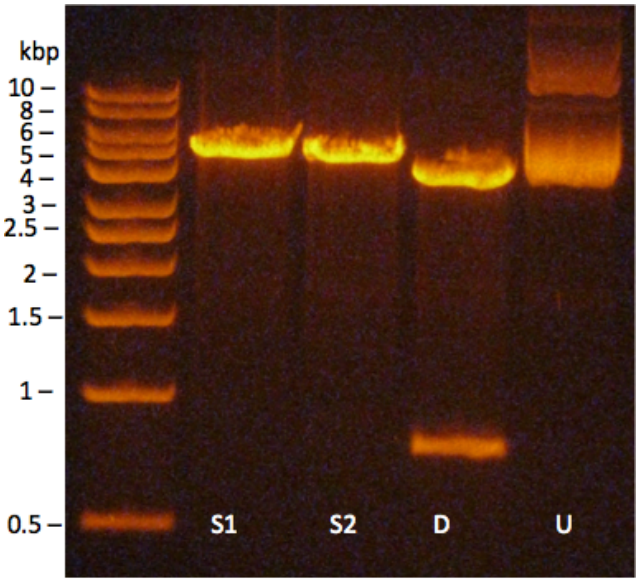


Figure 3.4.1. A 0.8% agarose gel showing pFKPEN undigested (U), single and double (D) digested with *Not*I (S1) and *Nco*I (S2)

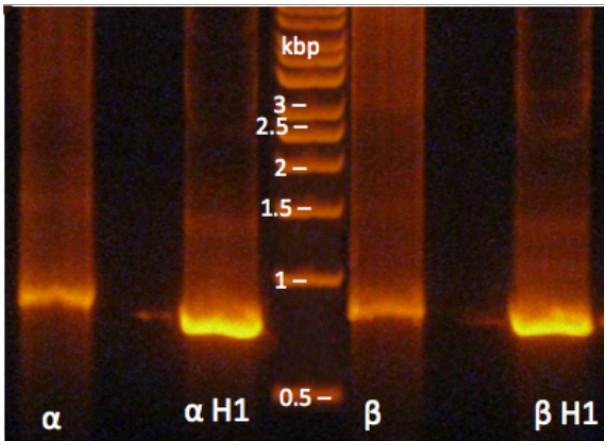


Figure 3.4.2. A 0.8% agarose gel showing the PCR products of LXR LBDs.

Fig.3.5.2.
α: LXRα LBD with 15 bp overlap to pFKPEN.
α H1: LXRα LBD with C-terminal hexahistidine tag.
β: LXRβ LBD with 15 bp overlap to pFKPEN.
β H1: LXRβ LBD with C-terminal hexahistidine tag.

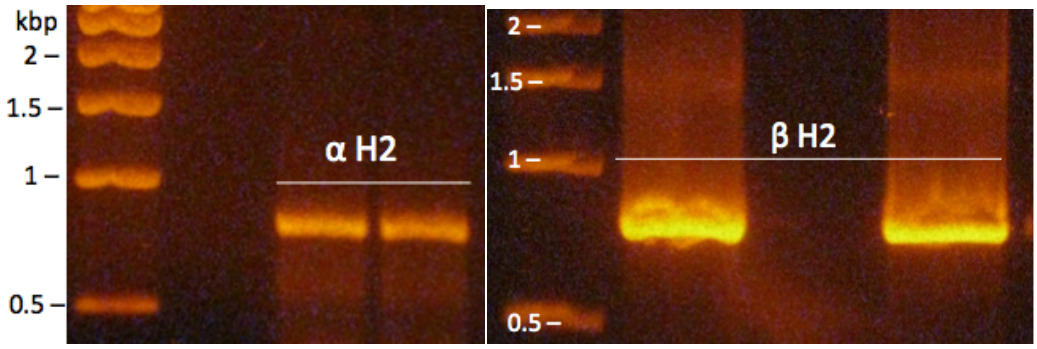


Figure 3.4.3. A 0.8% agarose gel showing the second PCR products of LXR LBDs with C-terminal hexahistidine tag. Due to high amounts of sample two wells were used, respectively. The his-tagged PCR products from Fig.3.5.2 were added a 15 bp overlap to pFKPEN in the second PCR.

The In-Fusion cloning and transformation procedure was performed according to protocols VI:B and VIII, respectively, of the In-Fusion® HD Cloning Kit User Manual (Clontech) (*51*). 50-100 ng of the PCR products and linearized plasmid were used for the cloning procedure. The controls were positive, i.e. the positive control showed more than hundred colonies on the LBamp plate and the negative control showed a few colonies. The LXRα LBD-$(His)_6$ plate was overgrown. 5 colonies were picked from LXRα LBD and LXRβ LBD, respectively. 24 colonies were picked from LXRβ LBD-$(His)_6$. The colonies were grown in LBamp and purified by miniprep. Diagnostic restriction digest was done to verify if the cloning procedure was successful. The restriction digest was done with the restriction enzymes *Acc*I and *Kpn*I. See Table 2.6.1 for the cleavage sites of the restriction enzymes in the different vectors. The samples of each cloned plasmid was single digested by the restriction enzymes, respectively. The original pFKPEN plasmid purified from the XL1 Blue cells was used as a reference and digested accordingly. Undigested samples were loaded on the gel as references. All samples were applied to the gel in the order of undigested, *Kpn*I-digested and *Acc*I-digested. One of the gels is shown in Figure 3.4.4. All of the gels were inconclusive and showed the same pattern as α1 and α2 the gel in Figure 3.4.4. The original, undigested, pFKPEN plasmid, with a size of 4.8 kbp, appears above the 10 kbp band of the weight marker in Figure 3.4.4, but at the expected size in Figure 3.4.1. However the cleavage pattern in Figure 3.4.4 is as expected for both the restriction enzymes. No conclusion could be drawn, as the reference samples are inconsistent in all of the gels. A diagnostic restriction digest gives only a verification expectation, but the samples would still have to be sequenced for a validation. Five random samples were picked for in-house sequencing. The samples were mixed with forward and backward sequencing primers, respectively, yielding ten samples. The sequencing primers used had the sequences 5'-CGGATAACAATTTCACACAG-3' and 5'-CTAGATTAGTGATGGTGATG-3', forward and backward respectively. The sequencing results did not contain any signal, but the background noise was present.

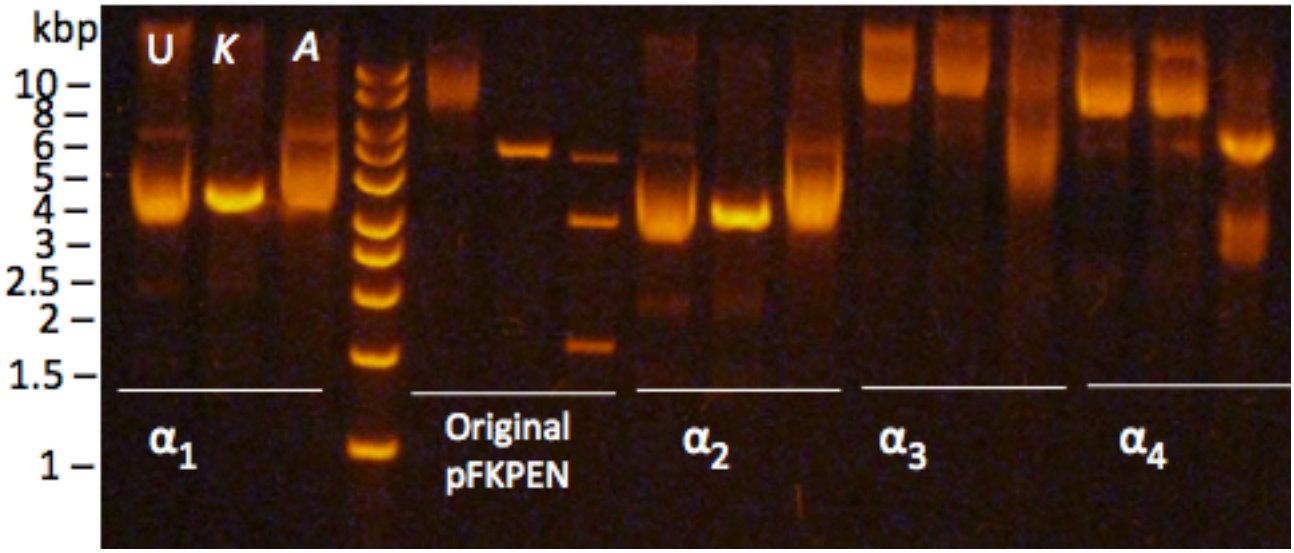


Figure 3.4.4. A 0.8% agarose gel showing the results of a diagnostic restriction digest of cloning experiments with LXR LBD genes, with and without C-terminal His-tag, in pFKPEN plasmid. All of the samples were applied in the same order. **U**: undigested. ***K****: Kpn*I single digested. ***A***: *Acc*I single digested.

Conclusion of Strategy 2

To circumvent all the issues experienced with expressing LXR LBD genes in the vector pRSET B we decided to clone the genes into pFKPEN periplasmic expression vector. pFKPEN is engineered to improve aggregation and proteolysis experienced with bacterial expression systems. The vector carries the gene encoding for a periplasmic chaperone shown to improve folding and prevent aggregation of recombinant proteins. In addition the genes cloned into the vector are N-terminally fused to a pelB signal sequence that directs protein translocation to the periplasm of the *E. coli* cell. The amplification of the genes and the linearization of pFKPEN were successful, but the cloning procedure yielded inconclusive results. The sequenced plasmid samples from the cloning procedure gave results without signal, but with background noise present. The in-house sequencing service that performed the sequencing of the samples were going out of business, and the sequencing of the samples was done on the day after the announced last day of the service. It is difficult to say if the inconclusive sequencing results are due to error in the service, the primers or the samples. Although a diagnostic restriction digestion test performed on the samples points to that the samples contained the wrong plasmid. No conclusions could be drawn from the digestion test due to unexpected and unexplainable patterns in the gels. The gels showed the known reference plasmid at double the expected size, but this anomaly was consistent throughout all gels. If we assume that something went wrong with the gels or the electrophoresis, the results of the digestion test indicates that all, except two, of the samples contained the original pRSET B plasmid and not the target pFKPEN plasmid. The expected digestion pattern of the newly generated plasmids would be undigested, and two bands for the restriction enzymes *Kpn*I and *Acc*I, respectively. Expected digestion pattern for the pRSET B plasmid would be a linearized DNA band, and undigested DNA band for *Kpn*I and *Acc*I, respectively. The gels showed the pRSET B expectation for all samples except two (α3 and α4 in Figure 3.4.4). These samples might show the pFKPEN expectation, but it is hard to say for sure because of smeary bands. The presence of pRSET B in the samples should not have happened as the PCR products were purified from the agarose gel, before being used for another PCR. The contamination might have come from the lab equipment, but it seems unlikely as the procedures of the experiments were conducted on different days. Unfortunately this was the last experiment in the project and there was no more time to conduct further analysis of the cloned samples.

4 Summary and future considerations

The aim of this Master's project was to express, purify and crystallize the ligand-binding domains of the Liver X Receptors, α and β, in their native forms and bound to a variety of ligands. The work from the Master's project would be used for structural studies and drug design, to find promising compounds against maladies related to lipid homeostasis. Several structures of the LXR LBDs bound to different ligands are reported in the PDB, but all the ligands are synthetic. Hence the ultimate goal of this Master's project was to obtain several crystals of the LXR LBDs bound to endogenous oxysterols and novel synthetic ligands, synthesized at the School of Pharmacy. Based on previous work on the LXR LBDs, it was expected that the biggest workload would lie in the crystallization and ligand soaking of the crystals, with the possibility of synthesis of potential ligands if there was enough time. This was, however, not the case. The purchased constructs of the LXR LBDs gave rise to more problems than we anticipated. Thus the main part of the Master's project became the optimization of expression and purification.

The purchased constructs expressed in high yield, but little of the protein was soluble. To avoid aggregation and inclusion body formation during the expression of the proteins, we tried to vary the temperature, expression time and add a ligand to all solutions when expressing the proteins in the standard expression strain *E. coli* BL21 DE3 and we tried to express the proteins in cold-adapted *E. coli* optimized for expression of proteins prone to aggregation. All of the experiments yielded little or no soluble protein. The soluble protein was successfully purified, but in low yields. Scaling up the expression could have increased the yield, but the proteins had a long disordered hexahistidine purification tag on the N-terminus separated form the LBDs by an enterokinase cleavage site. Cleavage experiments with enterokinase showed that it is highly unspecific at varying temperatures and with varying amounts of protease used. The replacement of amino acids adjacent to the cleavage site was reported to significantly increase the specificity of enterokinase (*56, 59*). However, this seems to be protein specific and may not be the solution for all fusion proteins (*55*). Removal of the tag could be unnecessary as a histidine-tag sometimes can be advantageous for crystallization by coordination of the proteins with nickel-ions in the mother liquor, so we tried to directly crystallize the purified tagged LXRα LBD protein, but the screen did not yield any protein crystals. The highly disordered purification tag still on the protein might have prevented crystal formation. We tried to circumvent the problems experienced with enterokinase by subcloning a novel cleavage site in the constructs encoding cleavage with Tobacco Etch Virus protease. In parallel with experiments of *in vitro* refolding of aggregated proteins. If these strategies had worked, they could be combined to obtain high yields of pure tag-less

protein. Unfortunately the subcloning did not go as expected, and the refolding yielded unstable protein that aggregated when concentrated to medium concentration. The experiments with optimization of the constructs had reached its limit, so we tried to clone the genes into a new expression vector encoding periplasmic translocation and a periplasmic chaperone to prevent aggregation. The expression system was developed by Gunnarsen *et al.* to improve expression of proteins prone to aggregation. The vector was a kind gift from Geir Åge Løset (Department of Biosciences – The Faculty of Mathematics and Natural Sciences, University of Oslo). The vector was successfully linearized by double restriction digestion and the genes were successfully amplified by PCR with ends overlapping with the linearized vector. However, the cloning procedure yielded ambiguous results. The samples from the cloning procedure were analyzed by restriction digestion and DNA-sequencing, and both results indicated that the cloned samples contained the old vector. Unfortunately the time had run out for the Master's project, so no more experiments were conducted on the cloned samples.

This Master's project shows that expression, purification and crystallization of recombinant proteins will not always go as expected. Although the ultimate goals of the project were not reached, a lot has been learned along the way. Considerations for future work in this project would be to do more experiments on the cloned samples, and if required repeat amplification of the genes to generate a short TEVp-cleavable N-terminal hexahistidine tag and repeat the cloning procedure. The periplasmic expression system was originally developed for T cell receptors, so it is unknown if the system will work for the LXR LBDs, but as the LXR LBDs also are prone to aggregation in bacterial systems it is likely that periplasmic expression with a chaperone would increase the yield of soluble protein. Co-expression with RXR could also be tried out to improve the solubility. Once pure active protein is obtained, it is strongly advised to crystallize the proteins in presence of the SRC-1 co-peptide and/or binding partner RXR and the ligand of choice.

References

1. N. I. Parikh, M. J. Pencina, T. J. Wang, K. J. Lanier, C. S. Fox, R. B. D'Agostino, R. S. Vasan, Increasing trends in incidence of overweight and obesity over 5 decades. *Am J Med* **120**, 242-U246 (2007); published online EpubMar (Doi 10.1016/J.Amjmed.2006.06.004).
2. K. H. Yoon, J. H. Lee, J. W. Kim, J. H. Cho, Y. H. Choi, S. H. Ko, P. Zimmet, H. Y. Son, Epidemic obesity and type 2 diabetes in Asia. *Lancet* **368**, 1681-1688 (2006); published online EpubNov 11 (Doi 10.1016/S0140-6736(06)69703-1).
3. R. Bender, H. Zeeb, M. Schwarz, K.-H. Jöckel, M. Berger, Causes of death in obesity: Relevant increase in cardiovascular but not in all-cancer mortality. *Journal of Clinical Epidemiology* **59**, 1064-1071 (2006)10.1016/j.jclinepi.2006.01.006).
4. N. H. L. a. B. Institute. (http://www.nhlbi.nih.gov/health/health-topics/topics/atherosclerosis/, July 01, 2011).
5. P. Greenwald, C. K. Clifford, J. A. Milner, Diet and cancer prevention. *Eur J Cancer* **37**, 948-965 (2001); published online EpubMay (Doi 10.1016/S0959-8049(01)00070-3).
6. M. B. S. Flores, G. Z. Rocha, D. M. Damas-Souza, F. Osorio-Costa, M. M. Dias, E. R. Ropelle, J. A. Camargo, R. B. de Carvalho, H. F. Carvalho, M. J. A. Saad, J. B. C. Carvalheira, Obesity-Induced Increase in Tumor Necrosis Factor-alpha Leads to Development of Colon Cancer in Mice. *Gastroenterology* **143**, 741-+ (2012); published online EpubSep (Doi 10.1053/J.Gastro.2012.05.045).
7. H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, P. E. Bourne, The Protein Data Bank. *Nucleic acids research* **28**, 235-242 (2000); published online EpubJan 1 (
8. RSCB Protein Data Bank. (RCSB Protein Data Bank, 2014), vol. 2014, pp. A graph that displays the number of searchable X-ray structures per year.
9. M. M. Babu, N. M. Luscombe, L. Aravind, M. Gerstein, S. A. Teichmann, Structure and evolution of transcriptional regulatory networks. *Curr Opin Struct Biol* **14**, 283-291 (2004); published online EpubJun (10.1016/j.sbi.2004.05.004).
10. D. S. Latchman, Transcription factors: an overview. *The international journal of biochemistry & cell biology* **29**, 1305-1312 (1997); published online EpubDec (
11. J. P. Overington, B. Al-Lazikani, A. L. Hopkins, How many drug targets are there? *Nat Rev Drug Discov* **5**, 993-996 (2006); published online Epub12//print (
12. F. M. Sladek, What are nuclear receptor ligands? *Molecular and Cellular Endocrinology* **334**, 3-13 (2011)10.1016/j.mce.2010.06.018).
13. Z. Zhang, P. E. Burch, A. J. Cooney, R. B. Lanz, F. A. Pereira, J. Wu, R. A. Gibbs, G. Weinstock, D. A. Wheeler, Genomic analysis of the nuclear receptor family: new insights into structure, regulation, and evolution from the rat genome. *Genome research* **14**, 580-590 (2004); published online EpubApr (10.1101/gr.2160004).
14. H. Gronemeyer, J. A. Gustafsson, V. Laudet, Principles for modulation of the nuclear receptor superfamily. *Nat Rev Drug Discov* **3**, 950-964 (2004); published online EpubNov (Doi 10.1038/Nrd1551).
15. W. Bourguet, V. Andry, C. Iltis, B. Klaholz, N. Potier, A. Van Dorsselaer, P. Chambon, H. Gronemeyer, D. Moras, Heterodimeric complex of RAR and RXR nuclear receptor ligand-binding domains: purification, crystallization, and preliminary X-ray diffraction analysis. *Protein Expr Purif* **19**, 284-288 (2000); published online EpubJul (10.1006/prep.2000.1248).

16. A. Chawla, J. J. Repa, R. M. Evans, D. J. Mangelsdorf, Nuclear receptors and lipid physiology: opening the X-files. *Science* **294**, 1866-1870 (2001); published online EpubNov 30 (10.1126/science.294.5548.1866).
17. E. W. Ottow, Hilmar, *Nuclear Receptors as Drug Targets*. (Wiley-VCH, ed. 1, 2008), pp. 498.
18. J. R. Schultz, H. Tu, A. Luk, J. J. Repa, J. C. Medina, L. Li, S. Schwendner, S. Wang, M. Thoolen, D. J. Mangelsdorf, K. D. Lustig, B. Shan, Role of LXRs in control of lipogenesis. *Genes Dev* **14**, 2831-2838 (2000); published online EpubNov 15 (
19. J. L. Collins, A. M. Fivush, M. A. Watson, C. M. Galardi, M. C. Lewis, L. B. Moore, D. J. Parks, J. G. Wilson, T. K. Tippin, J. G. Binz, K. D. Plunket, D. G. Morgan, E. J. Beaudet, K. D. Whitney, S. A. Kliewer, T. M. Willson, Identification of a nonsteroidal liver X receptor agonist through parallel array synthesis of tertiary amines. *Journal of medicinal chemistry* **45**, 1963-1966 (2002); published online EpubMay 9 (
20. G. A. Francis, E. Fayard, F. Picard, J. Auwerx, Nuclear receptors and the control of metabolism. *Annual review of physiology* **65**, 261-311 (2003)10.1146/annurev.physiol.65.092101.142528).
21. S. Alberti, K. R. Steffensen, J. A. Gustafsson, Structural characterisation of the mouse nuclear oxysterol receptor genes LXRalpha and LXRbeta. *Gene* **243**, 93-103 (2000); published online EpubFeb 8 (
22. E. J. Reschly, N. Ai, W. J. Welsh, S. Ekins, L. R. Hagey, M. D. Krasowski, Ligand specificity and evolution of liver X receptors. *The Journal of steroid biochemistry and molecular biology* **110**, 83-94 (2008); published online EpubMay (10.1016/j.jsbmb.2008.02.007).
23. M. H. Faulds, C. Y. Zhao, K. Dahlman-Wright, Molecular biology and functional genomics of liver X receptors (LXR) in relationship to metabolic diseases. *Curr Opin Pharmacol* **10**, 692-697 (2010); published online EpubDec (Doi 10.1016/J.Coph.2010.07.003).
24. F. Bray, J. Lortet-Tieulent, J. Ferlay, D. Forman, A. Auvinen, Prostate cancer incidence and mortality trends in 37 European countries: An overview. *Eur J Cancer* **46**, 3040-3052 (2010); published online EpubNov (Doi 10.1016/J.Ejca.2010.09.013).
25. J. J. Repa, G. S. Liang, J. F. Ou, Y. Bashmakov, J. M. A. Lobaccaro, I. Shimomura, B. Shan, M. S. Brown, J. L. Goldstein, D. J. Mangelsdorf, Regulation of mouse sterol regulatory element-binding protein-1c gene (SREBP-1c) by oxysterol receptors, LXR alpha and LXR beta. *Gene Dev* **14**, 2819-2830 (2000); published online EpubNov 15 (Doi 10.1101/Gad.844900).
26. E. G. Lund, L. B. Peterson, A. D. Adams, M. H. N. Lam, C. A. Burton, J. Chin, Q. Guo, S. Huang, M. Latham, J. C. Lopez, J. G. Menke, D. P. Milot, L. J. Mitnaul, S. E. Rex-Rabe, R. L. Rosa, J. Y. Tian, S. D. Wright, C. P. Sparrow, Different roles of liver X receptor alpha and beta in lipid metabolism: Effects of an alpha-selective and a dual agonist in mice deficient in each subtype. *Biochem Pharmacol* **71**, 453-463 (2006); published online EpubFeb 14 (Doi 10.1016/J.Bcp.2005.11.004).
27. N. Terasaka, A. Hiroshima, T. Koieyama, N. Ubukata, Y. Morikawa, D. Nakai, T. Inaba, T-0901317, a synthetic liver X receptor ligand, inhibits development of atherosclerosis in LDL receptor-deficient mice. *Febs Lett* **536**, 6-11 (2003); published online EpubFeb 11 (Pii S0014-5793(02)03578-0
Doi 10.1016/S0014-5793(02)03578-0).
28. J. Hardy, D. Allsop, Amyloid Deposition as the Central Event in the Etiology of Alzheimers-Disease. *Trends Pharmacol Sci* **12**, 383-388 (1991); published online EpubOct (Doi 10.1016/0165-6147(91)90609-V).
29. C. P. Chuu, Modulation of liver X receptor signaling as a prevention and therapy for colon cancer. *Med Hypotheses* **76**, 697-699 (2011); published online EpubMay (Doi 10.1016/J.Mehy.2011.01.037).
30. F.-Z. El-Hajjaji, A. Oumeddour, A. J. C. Pommier, A. Ouvrier, E. Viennois, J. Dufour, F. Caira, J. R. Drevet, D. H. Volle, S. Baron, F. Saez, J.-M. A. Lobaccaro, Liver X receptors, lipids and their reproductive secrets in the male. *Biochimica et Biophysica Acta (BBA) - Molecular Basis of*

Disease **1812**, 974-981 (2011); published online Epub8// (http://dx.doi.org/10.1016/j.bbadis.2011.02.004).
31. R. Koldamova, I. Lefterov, Role of LXR and ABCA1 in the pathogenesis of Alzheimer's disease implications for a new therapeutic approach. *Curr Alzheimer Res* **4**, 171-178 (2007); published online EpubApr (Doi 10.2174/156720507780362227).
32. D. H. Volle, K. Mouzat, R. Duggavathi, B. Siddeek, P. Dechelotte, B. Sion, G. Veyssiere, M. Benahmed, J. M. A. Lobaccaro, Multiple roles of the nuclear receptors for oxysterols liver X receptor to maintain male fertility. *Mol Endocrinol* **21**, 1014-1027 (2007); published online EpubMay (Doi 10.1210/Me.2006-0277).
33. M. Farnegardh, T. Bonn, S. Sun, J. Ljunggren, H. Ahola, A. Wilhelmsson, J. A. Gustafsson, M. Carlquist, The three-dimensional structure of the liver X receptor beta reveals a flexible ligand-binding pocket that can accommodate fundamentally different ligands. *J Biol Chem* **278**, 38821-38828 (2003); published online EpubOct 3 (10.1074/jbc.M304842200).
34. X. Fradera, D. Vu, O. Nimz, R. Skene, D. Hosfield, R. Wynands, A. J. Cooke, A. Haunso, A. King, D. J. Bennett, R. McGuire, J. C. Uitdehaag, X-ray structures of the LXRalpha LBD in its homodimeric form and implications for heterodimer signaling. *Journal of molecular biology* **399**, 120-132 (2010); published online EpubMay 28 (10.1016/j.jmb.2010.04.005).
35. S. Hoerer, A. Schmid, A. Heckel, R. M. Budzinski, H. Nar, Crystal structure of the human liver X receptor beta ligand-binding domain in complex with a synthetic agonist. *Journal of molecular biology* **334**, 853-861 (2003); published online EpubDec 12 (
36. S. Svensson, T. Ostberg, M. Jacobsson, C. Norstrom, K. Stefansson, D. Hallen, I. C. Johansson, K. Zachrisson, D. Ogg, L. Jendeberg, Crystal structure of the heterodimeric complex of LXRalpha and RXRbeta ligand-binding domains in a fully agonistic conformation. *The EMBO journal* **22**, 4625-4633 (2003); published online EpubSep 15 (10.1093/emboj/cdg456).
37. S. Williams, R. K. Bledsoe, J. L. Collins, S. Boggs, M. H. Lambert, A. B. Miller, J. Moore, D. D. McKee, L. Moore, J. Nichols, D. Parks, M. Watson, B. Wisely, T. M. Willson, X-ray crystal structure of the liver X receptor beta ligand binding domain: regulation by a histidine-tryptophan switch. *J Biol Chem* **278**, 27138-27143 (2003); published online EpubJul 18 (10.1074/jbc.M302260200).
38. D. Blow, *Outline of Crystallography for Biologists*. Oxford University Press (Oxford, Oxford, 2002), pp. 236.
39. D. Sheehan, *Physical Biochemistry: Principles and Applications*. (Wiley-Blackwell, ed. Second, 2009), pp. 374.
40. GE Healthcare, T. I. p. w. g. elution, Ed. (GE Healthcare, http://www.gelifesciences.com 2010).
41. EcoProDB. (*E. coli* Protein Database and KAIST - Institute for the BioCentury, *E. coli* Protein Database, 2006).
42. G. Toresson, G. U. Schuster, K. R. Steffensen, M. Bengtsson, J. Ljunggren, K. Dahlman-Wright, J. A. Gustafsson, Purification of functional full-length liver X receptor beta produced in Escherichia coli. *Protein Expr Purif* **35**, 190-198 (2004); published online EpubJun (10.1016/j.pep.2004.01.007).
43. Invitrogen. (LifeTechnologies, http://www.lifetechnologies.com/order/catalog/product/V35120 - citationsList, 2014).
44. M. K. Chow, A. A. Amin, K. F. Fulton, J. C. Whisstock, A. M. Buckle, S. P. Bottomley, REFOLD: an analytical database of protein refolding methods. *Protein Expr Purif* **46**, 166-171 (2006); published online EpubMar (10.1016/j.pep.2005.07.022).
45. J. Phan, A. Zdanov, A. G. Evdokimov, J. E. Tropea, H. K. Peters, 3rd, R. B. Kapust, M. Li, A. Wlodawer, D. S. Waugh, Structural basis for the substrate specificity of tobacco etch virus protease. *J Biol Chem* **277**, 50564-50572 (2002); published online EpubDec 27 (10.1074/jbc.M207224200).

46. K. S. Gunnarsen, E. Lunde, P. E. Kristiansen, B. Bogen, I. Sandlie, G. A. Loset, Periplasmic expression of soluble single chain T cell receptors is rescued by the chaperone FkpA. *Bmc Biotechnol* **10**, (2010); published online EpubFeb 3 (Artn 8 Doi 10.1186/1472-6750-10-8).
47. Agilent Technologies Inc. (Agilent Technologies Inc., http://www.genomics.agilent.com, 2009), vol. 2013, pp. Instruction manual for the ArcticExpress competent cells.
48. Invitrogen. (LifeTechnologies, http://www.lifetechnologies.com, 2006), vol. 2012, pp. User Manual for EnterokinaseMax™
49. R. G. Shipra Baluja, Nayan Vekariya, Mehul Bhatt, and Rah, u. Bhalodia, Solubility of Cholesterol in some alcohos from 293.15 to 318.15 K. *Archives of Applied Science Research* **1**, 263-270 (2009).
50. Thermo Scientific. (Thermo Scientific, http://www.thermoscientificbio.com, 2012), vol. 2013.
51. I. Clontech® Laboratories. (Clontech, http://www.clontech.com, 2014).
52. New England BioLabs Inc., vol. 2014.
53. X. Lou, G. Toresson, C. Benod, J. H. Suh, K. J. Philips, P. Webb, J. A. Gustafsson, Structure of the retinoid X receptor alpha-liver X receptor beta (RXRalpha-LXRbeta) heterodimer on DNA. *Nat Struct Mol Biol* **21**, 277-281 (2014); published online EpubMar (10.1038/nsmb.2778).
54. H. S. Moreno, F. and Farias, R.N., Influence of the incorporation of cholesterol on the doubling time and on the Arrhenius and Hill plots of Two membrane-bound enzymes of *Escerichia Coli* K12. *Current Microbiology* **8**, 215-220 (1983).
55. S. H. Shahravan, X. L. Qu, I. S. Chan, J. A. Shin, Enhancing the specificity of the enterokinase cleavage reaction to promote efficient cleavage of a fusion tag. *Protein Expression and Purification* **59**, 314-319 (2008); published online EpubJun (Doi 10.1016/J.Pep.2008.02.015).
56. T. Hosfield, Q. Lu, Influence of the amino acid residue downstream of (Asp)4Lys on enterokinase cleavage of a fusion protein. *Analytical biochemistry* **269**, 10-16 (1999); published online EpubApr 10 (10.1006/abio.1998.3084).
57. O. W. Liew, J. P. Ching Chong, T. G. Yandle, S. O. Brennan, Preparation of recombinant thioredoxin fused N-terminal proCNP: Analysis of enterokinase cleavage products reveals new enterokinase cleavage sites. *Protein Expression and Purification* **41**, 332-340 (2005); published online Epub6// (http://dx.doi.org/10.1016/j.pep.2005.03.006).
58. M. J. Mizianty, W. Stach, K. Chen, K. D. Kedarisetti, F. M. Disfani, L. Kurgan, Improved sequence-based prediction of disordered regions with multilayer fusion of multiple information sources. *Bioinformatics* **26**, i489-496 (2010); published online EpubSep 15 (10.1093/bioinformatics/btq373).
59. O. W. Liew, P. C. Jenny Chong, Y. Z. Lim, C. X. Ang, Y. C. Amy Lau, T. G. Yandle, S. O. Brennan, An SRLLR motif downstream of the scissile bond enhances enterokinase cleavage efficiency. *Biochimie* **89**, 21-29 (2007); published online EpubJan (10.1016/j.biochi.2006.10.005).

Appendix

Section A: sequences of LXR LBD genes and amino acids

DNA sequence of LXRα LBD

CAGGAGGAGGAGCAGGCCCACGCCACCAGCCTGCCCCCCAGGGCCAGCAGCCCCCCCCAG
ATCCTGCCCCAGCTGAGCCCCGAGCAGCTGGGCATGATCGAGAAGCTGGTGGCCGCCCAG
CAGCAGTGCAACAGGAGGAGCTTCAGCGACAGGCTGAGGGTGACCCCCTGGCCCATGGCC
CCCGACCCCCACAGCAGGGAGGCCAGGCAGCAGAGGTTCGCCCACTTCACCGAGCTGGCC
ATCGTGAGCGTGCAGGAGATCGTGGACTTCGCCAAGCAGCTGCCCGGCTTCCTGCAGCTG
AGCAGGGAGGACCAGATCGCCCTGCTGAAGACCAGCGCCATCGAGGTGATGCTGCTGGAG
ACCAGCAGGAGGTACAACCCCGGCAGCGAGAGCATCACCTTCCTGAAGGACTTCAGCTAC
AACAGGGAGGACTTCGCCAAGGCCGGCCTGCAGGTGGAGTTCATCAACCCCATCTTCGAG
TTCAGCAGGGCCATGAACGAGCTGCAGCTGAACGACGCCGAGTTCGCCCTGCTGATCGCC
ATCAGCATCTTCAGCGCCGACAGGCCCAACGTGCAGGACCAGCTGCAGGTGGAGAGGCTG
CAGCACACCTACGTGGAGGCCCTGCACGCCTACGTGAGCATCCACCACCCCCACGACAGG
CTGATGTTCCCCAGGATGCTGATGAAGCTGGTGAGCCTGAGGACCCTGAGCAGCGTGCAC
AGCGAGCAGGTGTTCGCCCTGAGGCTGCAGGACAAGAAGCTGCCCCCCCTGCTGAGCGAG
ATCTGGGACGTGCACGAG

Amino acid sequence of LXRα LBD

QEEEQAHATSLPPRASSPPQILPQLSPEQLGMIEKLVAAQQQCNRRSFSDR
LRVTPWPMAPDPHSREARQQRFAHFTELAIVSVQEIVDFAKQLPGFLQLSREDQIALLKT
SAIEVMLLETSRRYNPGSESITFLKDFSYNREDFAKAGLQVEFINPIFEFSRAMNELQLN
DAEFALLIAISIFSADRPNVQDQLQVERLQHTYVEALHAYVSIHHPHDRLMFPRMLMKLV
SLRTLSSVHSEQVFALRLQDKKLPPLLSEIWDVHE*

DNA sequence of LXRβ LBD

GGTGAAGGTGAAGGCGTTCAGCTGACCGCAGCACAAGAACTGATGATTCAGCAGCTGGTTGCAGCAC
AGCTGCAGTGTAATAAACGTAGCTTTAGCGATCAGCCGAAAGTTACCCCGTGGCCTCTGGGTGCAGA
TCCGCAGAGCCGTGATGCACGTCAGCAGCGTTTTGCACATTTTACCGAACTGGCAATTATTAGCGTGC
AAGAGATTGTGGATTTTGCCAAACAGGTTCCGGGTTTTCTGCAGCTGGGTCGTGAAGATCAGATTGCA
CTGCTGAAAGCAAGCACCATTGAAATTATGCTGCTGGAAACCGCACGTCGCTATAATCATGAAACCG
AATGTATCACCTTCCTGAAAGATTTCACCTATAGCAAAGATGATTTTCATCGTGCAGGTCTGCAGGTC
GAATTTATCAATCCGATTTTTGAATTTAGCCGTGCAATGCGTCGTCTGGGTCTGGATGATGCAGAATA
TGCCCTGCTGATTGCCATTAACATTTTTAGCGCAGATCGTCCGAATGTTCAAGAACCGGGTCGTGTTG
AAGCACTGCAGCAGCCGTATGTGGAAGCGCTGCTGAGCTATACCCGTATTAAACGTCCGCAGGATCA
GCTGCGTTTTCCGCGTATGCTGATGAAACTGGTTAGCCTGCGTACCCTGAGCAGCGTTCATAGCGAAC
AGGTTTTTGCACTGCGTCTGCAGGATAAAAAACTGCCTCCGCTGCTGTCAGAAATTTGGGATGTTCAT
GAATAA

Amino acid sequence of LXRβ LBD

GEGEGVQLTAAQELMIQQLVAAQLQCNKRSFSDQPKVTPWPLGADPQSRDARQQRFAHFT
ELAIISVQEIVDFAKQVPGFLQLGREDQIALLKASTIEIMLLETARRYNHETECITFLKD
FTYSKDDFHRAGLQVEFINPIFEFSRAMRRLGLDDAEYALLIAINIFSADRPNVQEPGRV
EALQQPYVEALLSYTRIKRPQDQLRFPRMLMKLVSLRTLSSVHSEQVFALRLQDKKLPPL LSEIWDVHE*

DNA sequence of pFKPEN vector obtained from Geir Åge Løset with the gene for 14F7 antibody in italics

CTCGAGAGCG GGCAGTGAGC GCAACGCAAT TAATGTGAGT TAGCTCACTC
51 ATTAGGCACC CCAGGCTTTA CACTTTATGC TCCCGGCTCG TATGTTGTGT
101 GGAATTGTGA GCGGATAACA ATTTCACACA GAATTCATTA AAGAGGAGAA
151 ATTAACCATG AAATACCTAT TGCCTACGGC AGCCGCTGGC TTGCTGCTGC
201 TGGCAGCTCA GCCGGCCATG G*CCCACCACC ACCACCACCA CGAAAACCTG*
TACTTCCAGG GTCAGGTGCA GCTGCAGCAG AGCGGCGCGG AACTGGCGAA
ACCGGGCGCG AGCATGAAAA TGAGCTGCCG CGCGAGCGGC TATAGCTTTA
CCAGCTATTG GATTCATTGG CTGAAACAGC GCCCGGATCA GGGCCTGGAA
TGGATTGGCT ATATTGATCC GGCGACCGCG TATACCGAAA GCAACCAGAA
ATTTAAAGAT AAAGCGATTC TGACCGCGGA TCGCAGCAGC AACACCGCGT
TTATGTATCT GAACAGCCTG ACCAGCGAAG ATAGCGCGGT GTATTATTGC
GCGCGCGAAA GCCCGCGCCT GCGCCGCGGC ATTTATTATT ATGCGATGGA
TTATTGGGGC CAGGGCACCA CCGTGACCGT GAGCAGCAAG CTTTCAGGGA
GTGCATCCGC CCCAAAACTT GAAGAAGGTG AATTTTCAGA AGCACGCGTA
GACATCCAGA TGACCCAGAC CCCGTCTTCT CTGTCTGCTT CTCTGGGTGA
CCGTGTTACC ATCTCTTGCC GTGCTTCTCA GGACATCTCT AACTACCTGA
ACTGGTACCA GCAGAAACCG GACGGTACCG TTAAACTGCT GATCTACTAC
ACCTCTCGTC TGCACTCTGG TGTTCCGTCT CGTTTCTCTG GTTCTGGTTC
TGGTACCGAC TACTCTCTGA CCATCTCTAA CCTGGAACAG GAAGACATCG
CTACCTACTT CTGCCAGCAG GGTAACACCC TGCCGCCGAC CTTCGGTGCT
GGTACCAAAC TGGAACTGAA ATAA
GCGGCC GCTGGATCCG AACAA
AAGCTGACTT CAGAAGAAGA CCTAAACTCA CATCACCATC ACCATCACTA
1051 ATCTAGAGGC CTGTGCTAAT GATCAGCTAG CTTGAGGCAT CAATAAAACG
1101 AAAGGCTCAG TCGAAAGACT GGGCCTTTCG TTTTATCTGT TGTTTGTCGG
1151 TTAACGTCGA CTCACTTGTC GTCATCGTCC TTGTAGTCTT TTTTAGCAGA
1201 ATCTGCGGCT TTCGCATCAG CTTCCGGCTT TGCATCAGCC TTCGGCGCTG
1251 GTTTCACATC CAGCAGCTCT ACGTCAAACA CCAGGGTAGA ATTCGGTGGG
1301 ATCCCCGGAA CACCCGCTTT GCCGTAAGCC AGTTCTGGTG GAATAACCAG
1351 TTTGATCTTA CCGCCTTTCT TGATGTTCTT CAGACCTTCT GTCCAACCCG
1401 GGATAACACC GTCCAGACGG AAAGAAAGCG GTTCACCACG GGTGTAAGAG
1451 TTGTCGAACT CTTTACCGTC GATCAGCGTA CCTTTGTAGT TCACTACAAC
1501 AGTATCGCTG TCTTTCGGTG CTTCGCCTTT ACCGGCTTCT ACTACCTGAT
1551 AAACCAGACC AGTTGAAGAG GTTTTCACAC CTTTCTCTTT GGCAAATTTC
1601 TCGCGGTACT CTTTACCTTT TGCTTCGTTA TCAGCCGCGT CTTTTTCCAT
1651 CTTCGCCTGA GCAGAAGACT TCACGCGAGC TTCGAATGCT TGTAGAGTCT
1701 GTTCGATCTC TTGGTCGGAG AGTTTGCTCT TATCAGCAAA TGCATCCTGA
1751 ACACCAGCGA TCAGCTGATC TTTATCCAGT TTGATGCCCA GTTTTTCTTG
1801 TTCTTTTAGA GAGTTTTCCA TGTAACGACC CAGCGAGGCA CCCAGTGCAT
1851 AAGCTGATTT CTGATCGTCA TTTTTGAACG CTGCTTTGCT GTCAGCAGCT
1901 GTAGCAGGTT TTGCAGCTTC AGCAGCAAAA GTGATTGGTG CATGCAGGGC
1951 AACGGCCATT GTGGTCGCCA GCAGCGTTAC TTTAAACAGT GATTTCATCC
2001 ATATCTCCAG GATCGGGGCA TCTCACCCCA GGGTTAACTA TTATCAGAAG
2051 GGTACTATAA AGCGTTGTCG AACAAATCTA CATACAGACA CGCCCTATTA
2101 TCATCTATTT TCAGACTCTT TTTGTTTAAA TTAGTTTCGA TGACCGCGAA
2151 ATGAGTGCTG TCTCGGGCAG CAAAGTTAAG TAGAATCCGC GGCGACCATT
2201 CGACAAAAGA GGTGAAGTCG ACCTGGCGTA ATAGCGAAGA GGCCCGCACC
2251 GATCGCCCTT CCCAACAGTT GCGCAGCCTG AATGGCGAAT GGGACGCGCC
2301 CTGTAGCGGC GCATTAAGCG CGGCGGGTGT GGTGGTTACG CGCAGCGTGA
2351 CCGCTACACT TGCCAGCGCC CTAGCGCCCG CTCCTTTCGC TTTCTTCCCT
2401 TCCTTTCTCG CCACGTTCGC CGGCTTTCCC CGTCAAGCTC TAAATCGGGG
2451 GCTCCCTTTA GGGTTCCGAT TTAGTGCTTT ACGGCACCTC GACCCCAAAA
2501 AACTTGATTA GGGTGATGGT TCACGTAGTG GGCCATCGCC CTGATAGACG
2551 GTTTTTCGCC CTTTGACGTT GGAGTCCACG TTCTTTAATA GTGGACTCTT
2601 GTTCCAAACT GGAACAACAC TCAACCCTAT CTCGGTCTAT TCTTTTGATT
2651 TATAAGGGAT TTTGCCGATT TCGGCCTATT GGTTAAAAAA TGAGCTGATT

2701 TAACAAAAAT TTAACGCGAA TTTTAACAAA ATATTAACGC TTACAATTTA
2751 GGTGGCACTT TTCGGGGAAA TGTGCGCGGA ACCCCTATTT GTTTATTTTT
2801 CTAAATACAT TCAAATATGT ATCCGCTCAT GAGACAATAA CCCTGATAAA
2851 TGCTTCAATA ATATTGAAAA AGGAAGAGTA TGAGTATTCA ACATTTCCGT
2901 GTCGCCCTTA TTCCCTTTTT TGCGGCATTT TGCCTTCCTG TTTTTGCTCA
2951 CCCAGAAACG CTGGTGAAAG TAAAAGATGC TGAAGATCAG TTGGGTGCAC
3001 GAGTGGGTTA CATCGAACTG GATCTCAACA GCGGTAAGAT CCTTGAGAGT
3051 TTTCGCCCCG AAGAACGTTT TCCAATGATG AGCACTTTTA AAGTTCTGCT
3101 ATGTGGCGCG GTATTATCCC GTATTGACGC CGGGCAAGAG CAACTCGGTC
3151 GCCGCATACA CTATTCTCAG AATGACTTGG TTGAGTACTC ACCAGTCACA
3201 GAAAAGCATC TTACGGATGG CATGACAGTA AGAGAATTAT GCAGTGCTGC
3251 CATAACCATG AGTGATAACA CTGCGGCCAA CTTACTTCTG ACAACGATCG
3301 GAGGACCGAA GGAGCTAACC GCTTTTTTGC ACAACATGGG GGATCATGTA
3351 ACTCGCCTTG ATCGTTGGGA ACCGGAGCTG AATGAAGCCA TACCAAACGA
3401 CGAGCGTGAC ACCACGATGC CTGTAGCAAT GGCAACAACG TTGCGCAAAC
3451 TATTAACTGG CGAACTACTT ACTCTAGCTT CCCGGCAACA ATTAATAGAC
3501 TGGATGGAGG CGGATAAAGT TGCAGGACCA CTTCTGCGCT CGGCCCTTCC
3551 GGCTGGCTGG TTTATTGCTG ATAAATCTGG AGCCGGTGAG CGTGGGTCTC
3601 GCGGTATCAT TGCAGCACTG GGGCCAGATG GTAAGCCCTC CCGTATCGTA
3651 GTTATCTACA CGACGGGGAG TCAGGCAACT ATGGATGAAC GAAATAGACA
3701 GATCGCTGAG ATAGGTGCCT CACTGATTAA GCATTGGTAA CTGTCAGACC
3751 AAGTTTACTC ATATATACTT TAGATTGATT TAAAACTTCA TTTTTAATTT
3801 AAAAGGATCT AGGTGAAGAT CCTTTTTGAT AATCTCATGA CCAAAATCCC
3851 TTAACGTGAG TTTTCGTTCC ACTGAGCGTC AGACCCCGTA GAAAAGATCA
3901 AAGGATCTTC TTGAGATCCT TTTTTTCTGC GCGTAATCTG CTGCTTGCAA
3951 ACAAAAAAAC CACCGCTACC AGCGGTGGTT TGTTTGCCGG ATCAAGAGCT
4001 ACCAACTCTT TTTCCGAAGG TAACTGGCTT CAGCAGAGCG CAGATACCAA
4051 ATACTGTCCT TCTAGTGTAG CCGTAGTTAG GCCACCACTT CAAGAACTCT
4101 GTAGCACCGC CTACATACCT CGCTCTGCTA ATCCTGTTAC CAGTGGCTGC
4151 TGCCAGTGGC GATAAGTCGT GTCTTACCGG GTTGGACTCA AGACGATAGT
4201 TACCGGATAA GGCGCAGCGG TCGGGCTGAA CGGGGGGTTC GTGCACACAG
4251 CCCAGCTTGG AGCGAACGAC CTACACCGAA CTGAGATACC TACAGCGTGA
4301 GCTATGAGAA AGCGCCACGC TTCCCGAAGG GAGAAAGGCG GACAGGTATC
4351 CGGTAAGCGG CAGGGTCGGA ACAGGAGAGC GCACGAGGGA GCTTCCAGGG
4401 GGAAACGCCT GGTATCTTTA TAGTCCTGTC GGGTTTCGCC ACCTCTGACT
4451 TGAGCGTCGA TTTTTGTGAT GCTCGTCAGG GGGGCGGAGC CTATGGAAAA
4501 ACGCCAGCAA CGCGGCCTTT TTACGGTTCC TGGCCTTTTG CTGGCCTTTT
4551 GCTCACATGT TCTTTCCTGC GTTATCCCCT GATTCTGTGG ATAACCGTAT
4601 TACCGCCTTT GAGTGAGCTG ATACCGCTCG CCGCAGCCGA ACGACCGAGC
4651 GCAGCGAGTC AGTGAGCGAG GAAGCGGAAG AGCGCCCAAT ACGCAAACCG
4701 CCTCTCCCCG CGCGTTGGCC GATTCATTAA TGCAGGTATC ACGAGGCCCT
4751 TTCGTCTTCA C

Section B: Chemicals, hardware and software

Table B1: Reagents

Chemical	Vendor
1 kbp DNA ladder	New England BioLabs Inc.
*Acc*I (*Xmi*I) FastDigest	Fermentas
Acetic Acid	Merck
Agar-agar	Merck
Agarose, low melting point	Sigma
Alkaline Phosphatase, Calf Intestinal, (CIP)	New England BioLabs Inc.
Ampicillin	AppliChem
*Bam*HI FastDigest	Fermentas
Bovine serum albumin (BSA, 100x)	New England BioLabs Inc.
Buffer #3, 10x	New England BioLabs Inc.
Buffer #3.1, 10x	New England BioLabs Inc.
CelLytic™ B 10x	Sigma
Cholesterol BioChemica	AppliChem
c*O*mplete inhibitor	Roche
Coomassie Brilliant Blue G250	Amersham Biosciences
Crystallisation Plate, 96 MRC	SWISSCI
Cuvettes	Sarstedt
Dibasic sodium phosphate (Na_2HPO_4)	G-Biosciences
Dithiothreitol (DTT)	Bio-Rad Laboratories
DNA Loading Dye Solution (6x)	Lonza
DNase I	AppliChem
DNase I	Invitrogen
dNTP mix	Fermentas
DreamTaq buffer	Fermentas
DreamTaq DNA polymerase	Fermentas
Duck Crystal Clear Sealing Tape	Henkel
E. coli ArcticExpress competent cells	Agilent Technologies
E. coli BL21DE3 cells	
E. coli XL1-Blue $CuCl_2$ cells	Stratagene
*Ecor*I FastDigest	Fermentas
EKMax™ 10x reaction buffer	Invitrogen
EnterokinaseMax™	Invitrogen
Eppendorf tubes, 1.5 ml and 2 ml	Eppendorf
Ethanol Absolut	VWR Chemicals
Ethylenediaminetetraacetic acid (EDTA)	Fluka
FastDigest 10x buffer	Fermentas

FastDigest 10x Green buffer	Fermentas
Gentamycin	AppliChem
Glycerol	Prolabo
HiTrap Chelating HP	GE Healthcare
Hydrogen-Chloride (HCl)	Merck
Imidazole	Sigma
In-Fusion® HD cloning kit	ClonTech
Isopropyl-β-D-thiogalactopyranoside	Sigma
*Kpn*I FastDigest	Fermentas
L-Arginine HCl	AppliChem
Ligase buffer, 10x	New England BioLabs Inc.
Lysozyme	Sigma
MES SDS Running buffer, 20x	Invitrogen
Milli-Q H_2O (MQ- H_2O)	Millipore
Mono Sodium phosphate ($NaH_2PO_{4)}$	Fluka
*Nco*I FastDigest	Fermentas
NEBbuffer	New England BioLabs Inc.
Nickel(II) sulphate ($NiSO_4$)	Sigma
*Not*I FastDigest	Fermentas
NuPAGE Bis-Tris 4-12% gel	LifeTechnologies
NuPAGE LDS Sample buffer	Invitrogen
NuPAGE loading buffer (4x)	Invitrogen
NuPAGE® LDS Sample Buffer (4X)	Invitrogen
Optifit Refill Tips	Sartorius
PCR Cuvette	Eppendorf
Peptone from casein	Merck
Petman pipettes	Gilson
pFKPEN vector	Gift from Geir Åge Løset – University of Oslo
Phenylmethanesulfonyl fluoride	Fluka BioChemica
Primers, PCR	Eurofins Genomics
pRSET B expression vector with LXRα LBD	LifeTechnologies
pRSET B expression vector with LXRβ LBD	LifeTechnologies
SeeBlue® Plus2 standard, molecular weight marker	Invitrogen
SH 10x buffer	Sigma
Sodium carbonate (Na_2CO_3)	Sigma
Sodium chloride (NaCl)	Prolabo
Sodium hydroxide (NaOH)	Kebo Lab
Sodium thiosulfate ($Na_2S_2O_3$)	Sigma
SuperDex200 10/300 GL	GE Healthcare
T4 DNA ligase buffer	Fermentas
Urea	Merck
Vivaspin® 20 Centrifugal concentrators	Sartorious

VWR® Syringe filter	VWR
*Xho*I	New England BioLabs Inc.
Yeast Extract, granulated	Merck
β-mercaptoethanol	Sigma

Table B2: Equipment

Crystallization kits	**Vendor**
JSCG+ suite	Molecular Dimensions
Morpheus	Molecular Dimensions
Kit	**Vendor**
PCR Purification Kit (250)	QIAGEN
QIAprep Spin Miniprep Kit	QIAGEN
QIAquick gel extraction kit	QIAGEN
SilverQuest™ staining kit	Invitrogen
Instruments and lab equipment	**Vendor**
ÄKTApurifier-900	GE Healthcare
Avanti Centrifuge J-26 XP	Beckman Coulter
Bench pH meter, 3510	Jenway
Benchtop UV M-20 Transilluminator	UVP
Biofuge Fresco, Heraeus	Thermo Scientific
BX41 microscope	Olympus
Canon PowerShot A480	Canon
Capsulefuge PMC-06	Tomy
Centrifuge 5810 R	Eppendorf
CP224s Analytical Balance	Sartorius
Dri-BLOCK® DB-2A	Techne
Electrophoresis power supply-EPS 601	GE healthcare
Eppendorf Thermomixer comfort	Eppendorf
IKA-WERK VF2 vortex mixer	Janke & Kunkel
Jet 900W	Whirlpool
Kelvitron® t	Heraeus instruments
LKB Model: GPS 200/400	Pharmacia
Mini Horizontal Submarine Unit	Amersham
MQ-H2O, Direct Q	Millipore
Multitron Standard	Infors HT
NanoPhotometer	IMPLEN
Oryx4 robot	Douglas Instruments
PCB Precision Balance	Kern
PhotoDoc-It™ 60 Imaging system	UVP

Spectrafuge™ Mini Centrifuge	Labnet International Inc.
SSM1 Miniorbital shaker	Stuart
SterREO Discovery V12	Zeiss
TC-3000 PCR Thermal Cycler	Techne
Thermomixer comfort	Eppendorf
Transilluminator 2000	Bio-Rad
Tuttnauer 3870-ML, autoclave	Tuttnauer
UltraRocker Rocking Platform	Bio-Rad
Ultrasonic cleaner	VWR
Ultraspec III KKB	Pharmacia
Unichromat 1500	UniEquip
Vacuubrand®	VWR
XCell SureLock™ Mini Cell Electrophoresis system	Invitrogen

Table B3: Software

Software	**Vendor**
4Peaks	Nucleobytes Inc.
Clustal2W	EMBL-EBI
DeltaPix LE	DeltaPix
Multilayered Fusion-based Disorder Predictor	Biomine Software
NEBcutter V2.0	New England BioLabs Inc.
ProtParam, ExPASy	Swiss Institute of Bioinformatics
Translate tool, ExPASy	Swiss Institute of Bioinformatics
Unicorn 5.11	GE Healthcare
WaspRun Screening	Douglas Instruments

Section C: PCR mixtures and programs, and PCR primers

PCR of LXR LBDs to generate genes with and without a C-terminal hexahistidine tag, with 15 bp overlap with pFKPEN

PCR mixture for amplification

5	µl	10 mM dNTP
1	µl	10 µM forward primer (Eurofins)
1	µl	10 µM reverse primer (Eurofins)
2	µl	pRSET B with LXR LBD
5	µl	10x standard *Taq* reaction buffer
0.25	µl	DreamTaq DNA polymerase
39.75	µl	MQ – H_2O
50	µl	

PCR program for amplification of linear hEndo-FL

95	°C	30 seconds	
95	°C	15 seconds	x 10
45-68	°C	60 seconds	
68	°C	15-60 seconds	x 20
68	°C	5 minutes	

PCR-primers for amplifying LXR LBD and generating hexahistidine tag and 15bp overlap with pFKPEN vector. Some of the primers are common for both genes.

α With and without tag: Forward (generating overlap):
5’ – GCTCAGCCGGCCATGG CCCAAGAAGAAGAACAGGCA – 3’

β With and without tag: Forward (generating overlap):
5’ – GCTCAGCCGGCCATGGCAGGTGAAGGTGAAGGCGTT – 3’

α and β Without tag: Reverse (generating overlap):
5’ – TCGGATCCAGCGGCCGCTTATTCATGAACATCCCAAATTTC– 3’

α With histidine tag: Forward (generating random)
5’ – CCAGGCTTTACACTTTATGCTCCAAGAAGAAGAACAGGC – 3’ @

β With histidine tag: Forward (generating random)
5’ – CCTATCTTTACAATTTATGCT CGGTGAAGGTGAAGGCGT – 3’

α and β: With histidine tag: Reverse (generating his-tag)
5’ – ATTTGGGATGTTCATGAACATCACCACCACCACCACTAA – 3’

α and β: With histidine tag: Second Reverse (generating overlap)
5’ – CATCACCACCACCACCACTAA GCGGCCGCTGGATCCGA – 3’