Chemical Biology Approaches for Nuclear Receptors

- Molecular and Structural Insights -

PROEFSCHRIFT

ter verkrijging van de graad van doctor aan de Technische Universiteit Eindhoven, op gezag van de rector magnificus, prof.dr.ir. C.J. van Duijn, voor een commissie aangewezen door het College voor Promoties in het openbaar te verdedigen op woensdag 27 februari 2013 om 16.00 uur

door

Sascha Fuchs

geboren te Gütersloh, Duitsland
Dit proefschrift is goedgekeurd door de promotor:

prof.dr.ir. L. Brunsveld

A catalogue record is available from the Eindhoven University of Technology Library
ISBN: 978-94-6203-278-1

Cover Design: Maximilian Tomasoni & Sascha Fuchs
Printing: Wöhrmann Print Service, Zutphen, the Netherlands
to my family
# Table of Content

## CHAPTER ONE

*Molecular Insights into Nuclear Receptor Structure and Function*  

1.1. Nuclear Receptor Superfamily  
1.2. Nuclear Receptors in Chemical Biology  
1.3. Nuclear Receptors as Drug Targets  
1.4. Aim and outline of the thesis  
1.5. References  

## CHAPTER TWO

*Investigation into Posttranslational Modifications of the Estrogen Receptor Hinge Region*  

2.1. Introduction  
2.2. Predicted helical character of model 12-mer peptides  
2.3. Solid-phase peptide synthesis of 12-mers  
2.4. Structural analysis by far-UV circular dichroism  
2.5. Structural analysis using Nuclear Magnetic Resonance  
2.6. Acetylation by lysine acetyltransferase p300  
2.7. Discussion  
2.8. Conclusion  
2.9. Experimental  
2.10. References  

## CHAPTER THREE

*Studying the Estrogen Receptor Hinge Region Using Expressed Protein Ligation*  

3.1. Introduction  
3.2. Results and discussion  
3.2.1. Generating the recombinant protein for EPL  
3.2.2. Generating chemically synthesized thioester peptide for EPL  
3.2.3. Test ligations using EPL with the model protein C-YFP  
3.2.4. N’-terminal elongation of the ERα-LBD towards the ERα-HR  
3.2.5. Influence of hinge elongation on protein stability and coactivator binding  
3.3. Conclusion  
3.4. Experimental  
3.5. References
CHAPTER ONE

Molecular Insights into Nuclear Receptor
Structure and Function

Abstract. Nuclear receptors (NRs) are multi-domain transcription factors, typically under small molecule control. All proteins of the NR superfamily have a similar overall conformation and conserved domains. The ligand binding domain (LBD) receives small lipophilic molecules via its buried ligand binding pocket (LBP). The LBD and the partly involved hinge region (HR) are responsible for the subsequent dimerization of the receptor while the interaction with NR response elements is accomplished by the zinc fingers of the DNA binding domain (DBD) in cooperation with the HR. Rearrangement of the conformation of the LBD leads to the establishment of the activation function 2 (AF-2). Although this AF-2 is a main driving force for coactivator recruitment, transactivation of the receptor is supported by activation function 1 (AF-1), located on the amino-terminal domain (NTD). The available crystal structures of NRs have improved the molecular understanding of this complex ligand-mediated mechanism of signal transduction. Structural and functional analysis has revealed that receptor specificity can be introduced at different points, such as the size and shape of the LBP, size and structure of varying domains, including NTD and C-terminal F domain, and also by the appearance of the coregulators presenting motifs that are selectively accommodated by the AF-2. A consequence of the inherent conformational flexibility of the AF-2 is also the ability of NRs to respond differently to the presence of an agonist, a partial agonist, an antagonist or a reverse agonist leading to the differentiated recruitment of coactivators, corepressors or coactivator binding inhibitors that ultimately determine the transcriptional activation or repression of NRs.
Chapter One

1.1. Nuclear Receptor Superfamily

Members of the nuclear receptor (NR) superfamily are transcription factors that directly interact with DNA. Upon binding of small lipophilic molecules the structure of NRs undergoes a dynamic and conformational change, which leads to an enhanced ability to recruit other proteins, such as coregulators. This signal transduction is involved in several physiological functions across the metazoan kingdom, including development, homeostasis, reproduction, and metabolism. \(^{1,2}\) Although the term ‘hormones’ was first used in the early stage of the 20\(^{th}\) century \(^{3}\) not much was known then about their function until ‘a specific high-affinity receptor’ was identified as the target for estradiol in the 1960s. \(^{4,6}\) This was the first biochemical evidence for the existence of NRs. However, the first cloned NR was the glucocorticoid receptor (GR, \(NR3C1\)), followed by the estrogen receptor \(\alpha\) (ER\(\alpha\, NR3A1\)) in the mid eighties. \(^{7,8}\) The full sequencing of the human genome finally identified 48 different NRs, \(^{9,10}\) which can be divided into six evolutionary groups (subfamilies and exemplary members): \(^{11,12}\)

- **Thyroid Receptors** (\(NR1xx\), thyroid hormone receptor-like)
  - thyroid hormone receptors (TRs, \(NR1Ax\))
  - retinoic acid receptors (RARs, \(NR1Bx\))
  - peroxisome proliferator-activated receptors (PPARs, \(NR1Cx\))
  - RAR-related orphan receptor (RORs, \(NR1Fx\))

- **Retinoid Receptors** (\(NR2xx\), retinoid X receptor-like)
  - hepatocyte nuclear factor 4 (HNF4s, \(NR2Ax\))
  - retinoid X receptor (RXRs, \(NR2Bx\))

- **Steroid Receptors** (\(NR3xx\), estrogen receptor-like)
  - estrogen receptors (ERs, \(NR3Ax\))
  - glucocorticoid receptor (GR, \(NR3C1\))
  - androgen receptor (AR, \(NR3C4\))

- **Nerve Growth Factor IB Receptors** (\(NR4xx\), nerve growth factor IB-like)
  - nerve growth factor IB (NGFIB, \(NR4A1\))

- **Steroidogenic Factor Receptors** (\(NR5xx\), steroidogenic factor-like)
  - steroidogenic factor 1 (SF1, \(NR5A1\))

- **Germ Cell Nuclear Factor Receptor** (\(NR6A1\))
All members of the NR superfamily share a common protein structure with high structural homology and conserved domains: the variable amino-terminal domain (NTD) including ligand-independent activation function 1 (AF-1); the central highly conserved DNA binding domain (DBD) that can distinguish between nuclear receptor response elements; the poorly conserved connecting hinge region (HR) featuring different functionalities, including regulation of nuclear translocation; and the moderately conserved ligand binding domain (LBD) that consists of 12 α-helices interacting with both ligands via the ligand binding pocket (LBP), and coregulators via activation function 2 (AF-2). Additionally, a carboxy-terminal F-domain may or may not be present, and is attributed to play a role in protein dimerization and the regulation of coactivator binding (Figure 1-1).

![Diagram of nuclear receptor domain structure](image)

**Figure 1-1 | Domain structure of the nuclear receptor superfamily**
Schematic representation of the nuclear receptor superfamily: amino-terminal domain (NTD), including activation function 1 (AF-1); DNA binding domain (DBD), including zinc finger structure; hinge region (HR); ligand binding domain (LBD), including ligand binding pocket (LBP), activation function 2 (AF-2) and binding function 3 (BF-3); and F domain. Left: representative domain structures of MR, AR, ERα, RXRα and VDR. Right: Crystal structure of estradiol-bound ERβ-LBD interacting with proline-based peptide including helix 1 – helix 12 (H1-H12).
1.1.1. Amino-terminal domain and activation function 1 (NTD/AF-1)

The amino-terminal domain (NTD) is the most varying region of NRs in terms of both size and sequence. While the mineralocorticoid receptor (MR, NR3C2) exhibits a 602 residue long NTD, the vitamin D receptor (VDR, NR3I1) is only 23 amino acids long. The progesterone receptor (PR, NR3C3) and GR with NTDs of similar size have only 15% sequence homology. This trend is widespread across the entire NR superfamily and even within subgroups such as the steroid receptors.\(^{18}\) Consequently the capacity of the ligand-independent activation function 1 (AF-1) to contribute to the constitutive activity of the receptor differs significantly.\(^ {19}\) Additionally, alternative splicing resulting in truncated receptor isoforms occurs preferential to the exclusion of the \(N'\)-terminus affecting the NTD.\(^ {20\text{-}22}\) The determination of the tertiary structure of entire NRs is very complicated due to difficulties in full-length NR expression and purification. While the structure of the intact PPAR\(\gamma\):RXR\(\alpha\) dimeric complex could be solved (Figure 1-1), the NTD was not visible due to its high mobility. This view is coherent with studies into the dynamics of NR domains, which revealed the NTD to be intrinsically flexible.\(^ {23}\) The conformation of the NTD in the context of the whole NR is currently described as a disordered domain with a partial, variable degree of secondary structure that converts to a stabilized structure as a consequence of both DNA and protein-protein interactions.\(^ {24\text{-}26}\) Indeed, studies discovered structural changes in the NTD in presence of the DBD or as a function of DNA binding for several NRs.\(^ {27\text{-}30}\) Although this intramolecular interaction does not lead to globular structure formation, circular dichroism (CD) and nuclear magnetic resonance (NMR) spectroscopy of the AF-1 (e.g. of GR) - in the presence of the helix-stabilizing trifluoroethanol - revealed the \(\alpha\)-helix forming propensity of the NTD.\(^ {31\text{-}32}\) Thus, the required secondary structure for AF-1-mediated activation is gained upon interaction with other proteins, such as coactivators.\(^ {33}\) This induced fit strategy of increased folding upon interaction with coactivator motifs could also be discovered within other transcription factors, including p53 and VP16.\(^ {34\text{-}35}\) Noticeably high numbers of posttranslational modifications (PTMs) - and their known influence on domain conformation and activity - also suggests an important role for NTDs in NR function.\(^ {18\text{-}36}\)

The androgen receptor (AR, NR3C4) features a very unique transcriptional activation mechanism. In contrast to other NRs, where AF-1 makes only a minor contribution to its overall transcriptional activity, in AR the main part is facilitated by its AF-1.\(^{37}\) In line with other NRs, the secondary structure is also intrinsically unstructured and stabilized upon interaction with other proteins.\(^{32}\) However, this interaction is not limited to intermolecular binding partners, such as coactivators, but also includes binding to the C'-terminal LBD or, more specifically, the AF-2 via a binding motif based on bulky aromatic residues.\(^ {38}\) This interplay between AF-1 and AF-2 (N/C interaction) in AR is ligand-dependent and is necessary for AR-mediated transactivation \textit{in vivo} (Figure 1-2).\(^ {39\text{-}40}\) To conclude, the active
conformation of AF-1 is induced by stabilization of the NTD by both intra- (DNA, AF-2) and intermolecular (coregulator) interactions.

1.1.2. DNA binding domain (DBD)

The first tertiary structure of a DNA binding domain (DBD) in complex with DNA was solved for the GR by nuclear magnetic resonance (NMR) and X-ray crystallography in the early nineties. Later, other DBDs were described with and without the DNA response element. This domain of approximately 70 residues is highly conserved across NRs and expresses a globular fold containing two zinc ions (zinc finger structure), which are necessary for the conformation and activity of the domain. Two α-helices are present: one (helix-1) interacts base-specific with the major groove of the DNA and the other (helix-2) stabilizes this complex. Upon binding of one monomer a conformational change in the DBD facilitates the dimerization of a second monomer. These two monomers can be either the same type of NR, resulting in homodimers that bind to palindromic inverted repeats on the DNA (steroid receptors), or RXR forming heterodimers with other NRs binding to direct repeats on the DNA (PPAR, VDR among others). However, RXR can also form homodimers, with the conformation of the less conserved C'-terminal extension (CTE) of the DBD -
extending into the hinge region (HR) - playing an important role in the decision to form hetero- or homodimers. In that sense, the CTE can either provide an additional dimerization interface to enhance interaction with RXR, or play an additional role in the interaction of the minor groove of the DNA (compare Figure 1-7, PPARγ-HR), similar to orphan receptors that bind only as monomers to extended response elements. Recently, the structure of intact PPARγ-RXRα (both full-length) in complex with DNA could be solved (Figure 1-7), which revealed the induction of DNA interaction of both DBDs by the PPARγ-LBD and is in accordance with findings that demonstrated reduced DNA binding upon mutations in the LBD of PPARγ.

1.1.3. Hinge Region (HR)

The hinge region (HR) of NRs is much less highly conserved compared to the flanking DBD and LBD. Apart from its flexibility, which enables different orientations of the domains of NRs with respect to each other the presence of the nuclear localization signal (NLS) was also discovered early on to be at least partially located in the HR of NRs. Indeed, the orientation between the NTD, including ligand-independent AF-1, and LBD, which includes the ligand-dependent AF-2, is based on the composition and length of the HR and therefore influences transactivation. Further studies revealed the involvment of HR as a multifunctional region in dimerization, nuclear translocation, DNA binding, transactivation, and receptor mobility.

Similar to the NTD, the HR is highly decorated with residues accessible for PTMs. Modifications such as acetylation are reported to be connected with the folding of AR, resulting in the regulation of transcriptional activity, subcellular distribution, and coactivator and corepressor binding, respectively. Similar discoveries were made for ER. Both small modifications - such as phosphorylation, acetylation or methylation at various positions - and the interaction of coregulators are known to stabilize the receptor leading to enhanced transcriptional activity. Simultaneously, the HR is a target for significant changes, such as mono- or polyubiquitination, resulting in proteosomal degradation. Interestingly, some phosphorylation sites, including S305 and T311, are connected to the failure of tamoxifen treatment in ER-positive breast cancer. However, S282, among others, is predicted to improve tamoxifen responsiveness.

The mobility of RXR-HR enables receptor interaction with various response elements as heterodimers. However, the interaction with these heterodimer partners is influenced by the HR of the interaction partners but not by the RXR-HR itself. The interaction with NR corepressors was also discovered for the RXR-heterodimer-partner TR, but not for RXR itself. The intramolecular interplay between HR and LBD is confirmed by the release of these corepressors upon ligand binding to the LBD.
While the HRs of both members of the PPAR-RXR heterodimer pair are rather flexible, the structurally similar HRs of two other RXR interaction partners - TR and VDR, respectively - are unique due to their α-helical structure as an extension of the DBD. While the HR was not or only partially visible in crystallization analysis - even in intact PPARγ-RXRα in complex with DNA - cryo electron microscopy of VDR-RXR allowed for the positioning of crystal structures of LBD and DBD, and for the first time the visualization of the HRs of both receptors. The crystal structure identifies that a conformationally well defined VDR-HR is important for the stabilization of the open conformation that does not require any interaction between LBD and DBD. Mutational studies revealed that the conserved HR length - rather than the sequence - of VDR is important, which does not apply for RXR. This validates the hypothesis for the HR of VDR and TR that upon DNA interaction the HR is repositioned, thereby modulating the dimer interface and consequently transactivation.

Taken together, the HR is poorly conserved across all members of the NR superfamily. This then provides a platform for possible specificity between proteins that are highly or moderately conserved at important regions including DNA, ligand, and coregulator interaction.

### 1.1.4. Ligand binding domain and activation function 2 (LBD/AF-2)

In the mid nineties the first structures of NR ligand binding domains (LBDs) were solved in the ligand-free apo-form (RXRa-LBD) and the more compact ligand-bound holo-form (RARγ-LBD and TR-LBD), respectively. To date, structures are available for nearly all subfamilies of NRs that feature a canonical fold including 12 α-helices (H1-H12) orientated in a three-layered anti-parallel fashion. Mutation and deletion studies of various LBDs discovered the most C'-terminal helix 12 (H12) being of particular importance for ligand-dependent activation of the receptor. Indeed, the apo-form identified H12 projecting away from the structure, however, upon ligand binding rearrangement of H12 occurs in the form of a folding over of the core structure of the domain and closing of the LBP. In this holo-form, H12 and residues from H3-H5 generate the activation function 2 (AF-2) which consists of a hydrophobic groove limited in size by two charged amino acids in H3 and H12 (charge clamp). The AF-2 was identified as a target for several coregulators which interact via a leucine-rich LXXLL binding motif (L, leucine; X, any amino acid). Further investigations revealed this motif - embedded in a well-defined two-turn-α-helix - to be necessary, but also sufficient for NR-coactivator interaction.

Ligands that bind to the LBP of the receptor regulate this protein-protein interaction by means of a repositioning of the H12. Stabilization of this helix can be reached by direct contact with the ligand, a stabilizing effect on helices close to the LBP by the ligand, or through long-range interactions between LBP and AF-2. Ligand selectivity is achieved through cooperative function of H3, H5, and H11, which feature residues capable of
hydrophobic interactions and specific hydrogen bonding networks. The size of the LBP buried in the LBD can vary between different members of NRs between 450 and 1,600 Å². Apart from receptor-activating ligands (agonists), receptor-inhibiting ligands (antagonists) can also target the LBP. Upon binding of an antagonist ligand to the LBP, H12 is either sterically hindered to transition into the holo-form, or instead covers the AF-2 resulting in an active blockage of the binding of coactivators. Another way to silencing receptor activity is through the recruitment of corepressors, which takes place in ligand-free nonsteroidal receptors, such as thyroid and retinoid receptors. This protein-protein interaction is mediated via an extended LXXLL-like motif that binds to the AF-2. In this way, the lack of H12 results in the ability to accommodate the one-turn-enlarged α-helices – including sequences such as LXXI or HIXXXL/I, respectively - exposed by corepressors. This NR-corepressor complex can be stabilized by antagonists since they can provide an even more open AF-2 than is present in the apo-form. Indeed, even steroid receptors – except AR, which naturally does not bind to corepressors – were discovered in a DNA-corepressor-complex repressing transcriptional activity. In addition to these H12-related means of modulating NR activity, RXR regulation can be accomplished by steric hindrance of AF-2. The formation of tetramers renders coactivator binding impossible, whereas upon ligand binding the receptor exists in the active dimer state.

As previously discussed, AR possesses a unique mechanism through which transcriptional activity can be modulated. The intramolecular interaction between the unstructured AF-1 in the NTD and the ligand-dependent AF-2 in the LBD (N/C interaction) is required for the recruitment of coactivators (Figure 1-2). A phenylalanine-rich motif (FXXLF) on the NTD facilitates the interaction with the AR-AF-2 by providing more space on the protein surface – coated with less bulky residues - compared to other NRs. Further studies focusing on the atypical NR binding groove revealed that other bulky aromatic residues are accommodated, while the NR-binding motif LXXLL showed less binding affinity compared to other receptors.

Recently, another interaction platform on the LBD was discovered in AR: the binding function 3 (BF-3), which is located adjacent to AF-2 and involves principally H9 and partly H1. Structural and functional studies suggested a possible allosteric function for BF-3 (modulating AF-2) transmitted through amino acid residues adjacent to AF-2. While no natural target has yet to be directly identified, studies on FKB52-induced AR activity revealed the BF-3 surface as a putative FKB52 interaction and regulatory surface. Superimposition of the solved crystal structures of other NR-LBDs identified similar sequence and structural identity, presumably in steroid receptors but also in other subclasses.

Apart from ligand and coregulator binding and thus signal transduction, the LBD of NRs also features a dimerization surface, with H9 and H10 being the main helices involved.
Hydrophobic amino acids with flanking charged residues constitute a cluster for interaction with other LBDs. Although the general principle is the same, certain differences could be discovered between heterodimer and homodimers formation, for instance the involvement of H11 and H12 in the case of the PR.\textsuperscript{125}

1.1.5. F domain

The F domain is located at the very C’-terminus of NRs, but is not present in all receptors.\textsuperscript{24,126} Where present, the F domain varies in size - 19 to 80 residues - and is poorly conserved. Indeed, between the ER subtypes $\alpha$ and $\beta$ the F domain varies by ten amino acids in length and with less than 25\% homology.\textsuperscript{127,128} However, this variation and its involvement in both receptor dimerization and coregulator recruitment makes it partially responsible for the different activities discovered for the ER subtypes.\textsuperscript{17,129-131} Although the domain is required for the agonistic effect of tamoxifen on the ER element,\textsuperscript{132,133} it is not necessary for ligand-dependent transactivation, and can even enhance it.\textsuperscript{24} While there is currently no structural data available for ER, other NRs have given (partial) structural insights: visible as an extended $\beta$-strand in PR,\textsuperscript{134} involved with dimer partners in RAR.\textsuperscript{135,136} Investigations on the hepatocyte nuclear factor 4$\alpha$ (HNF4$\alpha$) also confirmed the importance of this domain. With 60 residues, it is the longest F domain known among NRs and is responsible for different transcription activities (four fold) in two HNF4$\alpha$ isoforms.\textsuperscript{137-140}

1.1.6. Posttranslational Modifications

Posttranslational modifications (PTMs) play an important role in NR function.\textsuperscript{141} These receptor modifications can be achieved by either the transfer of a functional chemical group (e.g. phosphate and acetyl) or the coupling of other polypeptides (e.g. sumoylation and ubiquitination). Studies have connected various PTMs to several diseases, with phosphorylation, acetylation and sumoylation of AR, ER$\alpha$, GR and PPAR$\gamma$ being the most reported kinds.

For decades now, advanced prostate cancer has been successfully treated through hormone deprivation therapy. However, increased death rates have been linked to hormone-refractory prostate cancer (HRPC) as a function of AR-phosphorylation.\textsuperscript{142} Interestingly, the same modification on the same residue leads to reduced toxicity of the elongated polyglutamine in Kennedy’s Disease.\textsuperscript{143} Acetylation of AR is also connected to both an increase in cell growth in prostate cancer tumor models and antagonist-resistance.\textsuperscript{58} Furthermore, sumoylation of AR is directly linked to the onset of prostate cancer.\textsuperscript{144}

In the case of ER$\alpha$-positive breast cancer, some posttranslational phosphorylation events, including S118 and S167, are linked with an improved response to tamoxifen and aromatase inhibitor treatment,\textsuperscript{145-147} while others, such as S305, T311 and S559, are associated with a poorer response.\textsuperscript{69} Analysis of breast cancer samples identified a predominantly high lysine-to-arginine mutations rate in ER$\alpha$ leading to a reduced susceptibility to acetylation. Since this
mutation also results in enhanced estradiol sensitivity, acetylations were attributed to the suppression of this ligand sensitivity and thus a lowering of the risk of breast cancer development.148–151

Glucocorticoid resistance in the treatment of various inflammatory diseases is a result of both GR phosphorylation and acetylation.152–155 While posttranslational phosphorylation of PPARγ is also linked to insulin resistance and possibly to obesity156 Furthermore, PPARγ has anti-inflammatory effects by sumoylation-mediated recruitment of NR corepressor, resulting in the silencing macrophage inflammatory genes.157–159 In summary, PTMs are another component in the regulation of NR function, which can be used as to measure disease development as biomarkers that provide information about the patient responsiveness to NR-mediated drug therapy.141
1.2. Nuclear Receptors in Chemical Biology

The basic function of NRs are closely related to their conserved structure, especially of the two domains DBD and LBD. These are mainly involved in communicating with the surrounding environment, such as small molecules interacting with the buried ligand binding pocket, DNA binding to Zn finger structure, and coactivator proteins binding to hydrophobic groove. This phenomenon of cellular response controlled by small lipophilic molecules is the ideal platform for chemical biology approaches using the receptor interfaces with ligand, DNA, and coregulators as targets for the design of biochemical tools.

1.2.1. Protein degradation

Induced protein degradation is one possibility to inhibit NR function. The design of the proteolysis targeting chimeric molecule (PROTAC) is a posttranslational chemical genetic technique to induce protein degradation by means of the ubiquitin proteasome system. The PROTAC consists of both the E3 ubiquitin ligase recognition motif and a recognition motif for the target protein. Until now, PROTACS targeting AR and ERα have been developed through chemical labeling of the endogenous ligands dihydroxytestosterone and 17β-estradiol (Figure 1-3). Another approach uses a three-hybrid small-molecule-protein interaction system in which split-ubiquitin functions as a sensor. In a model system using the two ligand-receptor pairs GR/dexamethasone (Dex) and dihydrofolate/methotrexate (Mtx), two halves of ubiquitin could be successful reconstituted. In this case, a small bivalent hybrid molecule Dex-Mtx functioned as the dimerizer for dihydrofolate and GR resulting in ubiquitin specific protease cleavage and subsequent degradation.

![Figure 1-3](NR-specific PROTAC used for the induced degradation of NRs)

1.2.2. Intein-mediated protein engineering

Protein splicing is a self catalyzed posttranslational process resulting in the excision of an internal protein segment (intein). This intein-mediated self-cleavage activity is a powerful tool to activate the protein of interest in a controlled fashion. One approach is expressed protein ligation (EPL) in which a peptide featuring a N’-terminal cysteine is ligated to a recombinantly expressed protein bearing a C’-terminal thioester. This thioester can be generated by the fusion protein consisting of protein and C’-terminal intein. EPL was used to investigate the dynamic mechanism of NR activation by incorporation of a fluorescent
Chapter One

... label in the relevant H12 of the PPARγ-LBD. Anisotropy measurements revealed a reduced mobility of H12 upon ligand binding. Thus, these findings were the first direct evidence for the molecular switch of H12 being essential for transcriptional activity. The same approach was used in a study concerning the influence of tyrosine-phosphorylation on H12 of ERβ. Crystal structure analysis of the site-specific posttranslationally modified protein gave insights into structural conformation as a function of H12 phosphorylation.

Another way of modulating protein activity is through the integration of inactive split inteins that are activated by small molecules. Expression of both a split intein and a split green fluorescent protein (GFP) separated by the VDR-LBD resulted in an intein-based vitamin D fluorescent biosensor leading to detection of α,25-hydroxyvitamin D₃ in the pM-range. In this case, ligand-binding induces a conformational change in the LBD, which brings the two intein halves together with subsequent protein splicing and reconstitution of GFP the read out for the binding event (Figure 1-4). A similar approach for the detection of NR agonists used a modified β-galactosidase assay based on both ERα-LBD and a split intein. Upon agonist binding the protein hydrolyzes the β-glycosidic bond resulting in yellow O-nitrophenol.

1.2.3. Photo-caged ligands

Photo-caged bioactive substrates, which are biologically inert small molecules that gain their function upon exposure to light, can function in a unique spatial and temporal manner. The introduction of photo-caged ligands for NRs allows for light-activated gene expression (LAGE), such as via 3’-hydroxyl inactivated estradiol, which activates ER upon irradiation with UV light (Figure 1-5). In a reverse approach, selective receptor modulators (SERMs) have also been photo-caged through the introduction of the same nitroveratryl group at the hydroxyl group of the SERM. A similar design has resulted in the light-dependent targeting of other NRs, such as RARγ, TRβ, and VDR.
1.2.4. Genetically encoded sensors

The generation of genetically encoded sensors is another possibility for the investigation of cellular processes.\textsuperscript{178} The introduction of fluorescent proteins allows the use of techniques such as Förster resonance energy transfer (FRET), fluorescence recovery after photobleaching (FRAP) or Fluorescence-lifetime imaging microscopy (FLIM).\textsuperscript{179-181} In the NR field these fluorescence-based approaches have been widely used. Umezawa et al. used a fluorescent indicator to discriminate between estrogen receptor agonists and antagonist in living cells. A construct consisting of the FRET-donor-acceptor-pair cyan/yellow fluorescent protein (CFP/YFP) was separated by a sequence based on the NR coactivator 1 (NCOA1/LXXLL) fused to the N’-terminus of ERα-LBD. Agonist binding led to recruitment of the LXXLL motif to the AF-2, leading to close proximity of the FRET-pair and thus FRET. However, in the antagonist-bound conformation both domains are held apart and no FRET can occur (Figure 1-6).\textsuperscript{182,183} A reverse approach was used in the development of a bile acid sensor for real time intracellular imaging in single living cells, consisting of bile acid-binding FXR an LXXLL-based peptide and a FRET-pair. Mutations in the domains of the cerulean/citrine FRET-pair led to intramolecular association and thus efficient FRET in the ligand-free state. However, bile acid-induced FXR-LXXLL-interaction resulted in reduced FRET signal as a function of increased distance between the FRET-pair.\textsuperscript{184} In a similar way the conformational change of AR and its N/C-interaction was used to monitor subcellular distribution and ligand-induced conformational change. However, in this particular case the fusion protein consists only of full-length AR with one component of the CFP-YFP FRET-pair on each side. Ligand-dependent N/C-interaction and the induced proximity of FRET pairs resulted in FRET.\textsuperscript{184,185} Houtsmuller and coworker took this approach one step further and performed simultaneous FRET and FRAP measurements to study the spatiotemporal organization of AR in living cells. They discovered that the N/C interaction is a function of AR mobility. While present in solution, upon binding to DNA, this interaction is abolished to provide space for coactivator interaction.\textsuperscript{186}

![FRET-based biosensing of NR agonists](image-url)

**Figure 1-6** | FRET-based biosensing of NR agonists
1.3. Nuclear Receptors as Drug Targets

The superfamily of nuclear receptors (NRs) is involved in a wide range of physiologic but also pathophysiologic processes.\textsuperscript{11,187} Already in the early stages of the 20\textsuperscript{th} century it was discovered that both ovariectomy decreases chance of mammary cancer and extracts from adrenal gland can be used in the treatment of Addison’s disease based on glucocorticoid deficiency.\textsuperscript{188,189} Purified cortisone - utilized against inflammatory diseases - and its subsequent total synthesis (1948) was pioneering work for subsequent syntheses of steroid compounds.\textsuperscript{190,191} In the 1970s tamoxifen was identified as an anti-estrogen in the screen of nonsteroidal compound libraries and became a leading drug, mainly in the treatment of early breast cancer, but also in ductal carcinoma, and breast cancer chemo-prevention.\textsuperscript{189,192-195} The ability to control NR function through the binding of small lipophilic molecules has made NRs one of the most targeted classes of proteins in the field of drug discovery. Indeed, statistical analysis of marketed drugs revealed that over 50\% interfere with only four protein classes, including G-protein-coupled receptors, NRs (13\%), ligand-gated ion channels or voltage-gated ion channels, with GR and the histamine H\textsubscript{1} receptor being the top targets.\textsuperscript{196,197} Diabetes (ER, PPAR), asthma (GR), atherosclerosis (LXR), osteoporosis (ER, VDR), cancer (ER, AR, GR, RXR), arthritis (GR) and heart failure (MR) are only some examples of many diseases currently treated with drugs that mainly target the ligand binding pocket (LBP) of NRs.\textsuperscript{197,198} However, aside from the LBP, other alternative sites of NR modulation are under investigation, including the activation function 2 (AF-2), binding function 3 (BF-3), the zinc finger motif, and the NR response element (Figure 1-7).\textsuperscript{123}
1.3.1. Targeting the Ligand Binding Pocket

The interaction of ligands at the NR ligand binding pocket (LBP) results in a conformational change of the LBD, which produces an interface for coactivator binding (AF-2). This activation or agonism of NRs by small lipophilic molecules has inspired the design of compounds that block this function (Figure 1-8, small molecule). For the first time antagonism was structurally confirmed with the crystal structure of ERα-LBD bound to either 4-hydroxytamoxifen or raloxifen. Although the antagonist-induced overall structure is similar to the agonist-bound form (H1-H11), either bulky side chains of the chemical molecules lead to incomplete/alternative folding of H12 (blocking AF-2) or it is not at all part of the AF-2 (incomplete AF-2). Either way, coactivators are not recruited to AF-2. However, since corepressor and coactivator share parts of the binding site, antagonist can also lead to reduced NR-corepressor interaction as discovered for RARα bound to BMS614.
Chapter One

Total inhibition of NR function, including their basal transcriptional activity, is a consequence of the interaction of inverse agonists. In this case H12 is stabilized in a different position that allows the binding of corepressor.\textsuperscript{101} Inverse agonists of RAR that can naturally recruit corepressors significantly induces these silencing interactions.\textsuperscript{200,201} However, even NRs that do not recruit corepressors innately, are capable of changing their conformation - upon inverse agonist binding - in a way that allows corepressors to bind.\textsuperscript{110,202}

Partial agonists, another class of ligand, are poorly understood, as their NR binding mechanism differs from those of agonist and antagonist. The functionality of partial agonists depends on the promoter context and cellular concentration, as shown for PPARγ.\textsuperscript{203} Similar to agonists, they induce a confirmation that allows the recruitment of coactivators, however, poor H12 stabilization leads to a receptor conformation that is still capable to interact with corepressors. Studies on LXR confirmed the hypothesis that the tissue-specific coactivator - corepressor ratio determines the activity of partial agonists.\textsuperscript{204}

Selective NR modulators are similar to partial agonists in terms of their different functionalities in various tissues. However, they not only show different levels of agonism, but even antagonistic effects. The first discovered selective ER modulator (SERM) was 4-hydroxytamoxifen, initially thought to be an exclusive antagonist.\textsuperscript{109} Functioning as an antagonist in mammary cells, it functions as a partial agonist in endometrial cells.\textsuperscript{205} Like partial agonists the presence of coactivators and corepressors are responsible for the SERM agonistic or antagonistic activity, respectively.\textsuperscript{206} Apart from other ER-targeting SERMs similar molecules for AR, PPAR and PR have also been discovered.\textsuperscript{207–210}

1.3.2. Targeting the activation function 2

An alternative approach to classical ligands is the direct targeting of the NR-coactivator interaction (e.g. AF-2). Due to the large number of crystal structures available for nearly all nuclear receptors and the detailed characterization of this protein-protein interaction, it has gained strong attention in the field of drug discovery.\textsuperscript{94} Overcoming the limitations of LBP-binding therapeutics, including insufficient selectivity and drug resistance due to long-term treatment, were the initial driving force to explore other druggable interactions on the NR surface.\textsuperscript{123} Due to the limited length of the α-helix of NR coactivators - two distinct residues on both sides of the AF-2 – it is unique among protein-protein interaction interfaces that are preferably greater in size.\textsuperscript{211,212} Thus, the design of coactivator binding inhibitors (CBIs) is a promising strategy to block NR transactivation.
The design of small molecules that mimic the well-structured α-helix of coactivators (Figure 1-8, peptidomimetics) combines the high selectivity of natural peptides with the bioavailability of small lipophilic drug-like compounds. Another advantage compared to peptide-based approaches is the high structural diversity since there are no limitations placed on the choice of side chain residues. However, the optimal arrangement of the side chain mimics - achieved in nature through the rigidity of the α-helix - is challenging to imitate. In a de novo design based on a pyrimidine core the leucine side chains of the LXXLL motif were successful mimicked using branched alkyl substituents, which resulted in an ER-selective coactivator binding inhibitor (CBI). The introduction of aromatic benzylic groups as a mimic for the phenylalanine-rich motif (FXXLF) involved in AR-AF-2 binding led to AR selective CBIs. Other attempts were based on biaryl scaffolds enabling out-of-plane projection of side-chain residues mimicking residues i, i + 3 to prevent steric repulsion. A similar core structure was used to specifically target the charge clamp, which emerges from two charged residues at both ends of the AF-2. In addition to the branched alkyl side chains, substituents were introduced to address electronic interactions with the charged amino acids. Another way to identify lead compounds that interact with a certain target is the high-throughput screening (HTS) of large compound libraries. The combination of an automated fluorescence-based assay with X-ray soaking studies discovered several lead compounds targeting the AR. The associated structural analysis directly gave insights into the location and mechanism of receptor interaction and revealed a BF-3 binding site adjacent to the AF-2. Another FRET-based approach involved the screening of an 86,000-strong library with diverse compounds. Subsequent structure-activity-relationship (SAR) studies and cell-based assays identified several CBIs. Modeling studies, however, revealed a much reduced interaction with the receptor surface compared to model peptides based on coactivator sequences. These findings suggest a possible non-perfect fit of these CBIs,
resulting in an entropically unfavored state and might provide a structural explanation for the difficulty of optimization μM lead structures towards higher nM potency.\textsuperscript{219} Covalent inhibitors of TR have also been developed which are small-molecule-based but mechanistically different from the classical CBIs discussed here. In this case, a cysteine in the AF-2 of the NR reacts with the β-aminoketone of the inhibitor to form an irreversible Michael adduct resulting in inhibition of coactivator recruitment.\textsuperscript{221-223}

Difficulties in the optimization of small molecule as CBIs due to an imperfect fit with the receptor AF-2 has (re)stimulated the design of peptide-based CBIs (Figure 1-8, peptide-based inhibitor). Phage display is a powerful technology to identify the optimal peptide sequence for a specific target.\textsuperscript{224-226} Combinatorial phage display against ER\textalpha revealed three different classes of peptides with CBI-functionality. In some cases NR selectivity could be achieved by the context of the flanking residues of the common LXXLL motif.\textsuperscript{97} Screening the AR surface with phage display discovered, aside from the FXLXF motif, other aromatic residues capable of binding the AR-AF-2. Structural analysis showed a perfect match of the identified hits and the AR surface. Additionally it explained the preferential binding of aromatic-based peptides compared to LXXLL, which forms only a single hydrogen bond between the backbone and the AR charge clamp instead of two in other NRs.\textsuperscript{227} Targeting other NRs in a similar way revealed different enlarged consensus motifs, such as HPLLXXLL in the case of PPAR\gamma and MPXLXXXLL for MR.\textsuperscript{228,229} Another study also used phage display screening to obtain information about the influence of LXXLL-flanking residues on ER binding. Due to the localization on well-defined α-helices on miniproteins, this way of screening identifies residues that are optimal for protein binding, however, disconnecting their effect on helix stability.\textsuperscript{230,231}

Next to low cell permeability and weak cellular stability, stabilization of the isolated α-helix is one of the major challenges on the way to drug-like structures.\textsuperscript{232} Thus, stabilizing the α-helix, which contains residues targeting the AF-2 (LXXLL/FXXLF), has been the goal of various studies. The connection of two side-chains by both a \(i, i + 4\) lactam-bridge or \(i, i + 3\) disulfide cyclization showed helical character, effective binding, and receptor selectivity of these small LXXLL peptides.\textsuperscript{233} In a similar approach using computational design and combinatorial chemistry both ERs and TR\β could be selectively targeted. Based on the docking studies, the leucines were sequentially substituted with hydrophobic non-natural side chains and additionally stabilized by a lactam bridge.\textsuperscript{234,235} In a separate study, the rigidity of the peptide backbones was enhanced by the introduction of a hydrocarbon staple at \(i, i + 3, i, i + 4,\) or \(i, i + 7\). Apart from hydrophobic interactions of the leucine side chain, structural analysis identified additional van der Waals contact of the staple contributing to the ability of the peptide to bind ER.\textsuperscript{236}
1.3.3. Targeting alternative sites

Crystallographic HTS against AR identified - aside from CBIs targeting the AF-2 - compounds targeting an alternative site on the LBD.22 This binding site, called binding function 3 (BF-3), is located in close proximity to the AF-2, and is clearly distinguishable. In total seven molecules in the mid-μM-range were co-crystallized with the AR-LBD, which revealed the recognition of residues that form part of both BF-3 and AF-2 and are therefore allosterically disruptive of coactivator recruitment (Figure 1-7, BF-3).237 Another approach focused on the identification of inhibitors for the positive AR-regulator FKBP52. The modified receptor-mediated reporter assay was used to screen a natural compound library, out of which two compound candidates were discovered that inhibit FKBP52-enhanced mutated AR-P723S function. Interestingly the responsible binding motif on the AR surface overlaps with the recently discovered BF-3 function.244 Inspired by this finding, an in silico screen of AR-BF-3 identified several BF-3 targeting ligands. Structural analysis frequently identified dual binding properties of the active compounds, namely binding to both BF-3 and AF-2.238

Other work focused on targeting of the nuclear response element of NRs. The high affinity molecular recognition of the minor groove by pyrrole-imidazole polyamides led to a blocking of both ER and ERR (Figure 1-7, response element). A similar strategy was used to target the AR element. Hairpin and cyclic polyamides showed a high ability to reduce prostate-specific antigen (PSA) mRNA levels in androgen-sensitive cancer cells (LNCaP).242,243 Another study concerning the treatment of ER-positive human breast cancer also focused on DNA, however, targeting directly the zinc finger of the DBD. The electrophilic disulfide benzamide (DIBA) inhibits cell growth of tamoxifen resistant breast cancer in vivo (Figure 1-7, zinc finger).244,245 Recent studies focused on targeting the NTD of AR. Both a small biaryl-based molecule and chlorinated peptides could be discovered to inhibit transactivation of the N'-terminus of AR in prostate cancer cells.246-250

HTS also identified compounds with high binding affinities for AR. The molecular interaction mechanism, however, is unclear. Two hits were identified that were synergistic with each other - suggesting different binding sites - and were neither competing with the endogenous ligand nor binding to the AF-2. Furthermore, one hit showed improved inhibition of LNCaP cancer cell proliferation and, in addition to its anti-androgen function, caused a reduction in prostate weight (Figure 1-7, unknown). In a mammalian two-hybrid assay, two other compounds were discovered that showed a similar profile. No recruitment of AR to the PSA enhancer in combination with a lack of influence on subcellular localization suggested the interference at the level of DNA or coactivator binding.246,247
1.4. Aim and outline of the thesis

The transcriptional activity of nuclear receptors (NRs) involves a complex interplay between, on the one hand the conserved DNA binding domain (DBD) and the ligand binding domain (LBD), and on the other hand the less well conserved amino-terminal domain (NTD), the hinge region (HR) and the carboxy-terminal F domain. Major structural elements involved in this activation event are activation function 1 (AF-1) at the NTD, the zinc finger motif on the DBD, and ligand binding pocket (LBP) and activation function 2 (AF-2) both located at the LBD. Perturbations in this highly sensitive regulatory machinery can lead to various disease states. Modulation of this permanently activated process can be achieved at different positions, such as the LBP or AF-2.

The aim of the thesis is to gain some molecular and structural insights into NR function with a main focus on both the hinge region (HR) of the estrogen receptor (ER) and the activation function-2 (AF-2) of ER, androgen receptor (AR) and retinoic X receptor (RXR). These structural domains and their role in protein-protein interactions are also investigated as possible drug targets for small lipophilic molecules and peptide-based binders, respectively.

In chapter 2 the structural influence of the acetylation of lysines 266 and 268 located in the estrogen receptor α (ERα) hinge region (HR) are investigated. Molecular dynamics (MD) simulations, circular dichroism (CD) studies and nuclear magnetic resonance (NMR) spectroscopy provided information about the change in secondary structure. Furthermore the ability of the acetyltransferase p300 to acetylate these lysines was characterized.

Chapter 3 deals with the semi-synthesis of ERα LBD combined with the HR. The site-specific introduction of posttranslational modifications (PTMs) was achieved by means of expressed protein ligation (EPL). Generating both recombinant ERα-LBD with an N’-terminal cysteine and peptides with C’-terminal thioester were the requirements to perform the native chemical ligation (NCL). Furthermore the progressive extension of the ERα-LBD in the direction of the HR forms part of a second study to acquire some insights into the function of this short and poorly conserved region of the receptor.

In chapter 4 ribosome display was used to effectively screen the surface of the estrogen receptor (ER) ligand binding domain (LBD) for novel natural peptide binders. Validated by in vitro and in vivo methods, structural insights were achieved by molecular dynamics (MD) simulation and X-ray crystallography studies. Optimized proline-based peptides were tested and compared to natural binding sequences in subsequent binding polarization and cellular assays. Co-crystallization with ERβ-LBD revealed a proline-mediated binding mechanism based on helix-stabilization.
Chapter 5 describes an *in silico* design approach to target the AR. Design of the peptides was based on separate results from ER screening. MD simulation studies were performed to gain information about the propensities of the designed peptides to form α-helices. These modeling studies were used to direct the solid-phase synthesis of selected peptides. CD studies were performed to confirm the predicted α-helical character. The AR binding affinities of the peptides were determined by means of a competitive polarization assay and cellular studies.

In chapter 6 NR-coregulator interaction profiling was performed to test a small compound library of biaryl natural products (NPs) for potential RXR binding. This microarray-based technique allowed for a rapid and simultaneous screen of these molecules against known coactivator and corepressor proteins important for NRs. Structure-activity relationships trends were assigned and the best binders verified in a polarization assay format.

Gaining molecular and structural insights to understand function and regulation of the structural elements of NRs is not only important to understand receptor pharmacology, but also to provide novel entries to target these proteins.

![Figure 1-9 | Outline of the thesis](image-url)
1.5. References

24. V. Kumar, S. Green, G. Stack, M. Berry, J.-R. Jin, and P. Chambon, Cell, 1987, 51, 941–951.
Chapter One


Chapter One

CHAPTER TWO

Investigation into Posttranslational Modifications of the Estrogen Receptor Hinge Region

Abstract. Posttranslational modifications are known to be important for protein stability and activity. Especially in less structured regions they can make a significant impact on structural changes and thus potentially modulate the activity of certain protein domains. In this chapter the structural influence of the acetylation of lysines 266 and 268 in estrogen receptor α (ERα) was investigated. These lysines form part of a linker domain between the DNA and ligand binding domain – namely the ERα hinge region – which is known to be mainly unstructured. Molecular dynamics (MD) simulation studies predicted structural changes upon lysine acetylation, with each mono-acetylated isoform (K*K and KK*) showing different amounts of enhancement in terms of their propensity to form α-helices. Double lysine-acetylation exerted the greatest effect on helix formation. Both CD and NMR studies of 12-mer peptides synthesized by solid phase peptide synthesis confirmed these results. Additionally, CD data suggested the presence of a transient polyproline II (PPII) helix that is less pronounced for K*K* than for KK. This structural element could be important in both the exploration of molecular and physiological mechanisms and specific targeting. The acetyl transferase p300 was capable of acetylating both the unacetylated peptide KK and the mono-acetylated peptides K*K and KK* without any additional auxiliary protein. The amount of acetylation observed for both mono-acetylated peptides K*K and KK* was however significantly lower compared with the unacetylated peptide KK accessible for double acetylation. The acetylation efficiency of the second lysine is thus impaired by the first acetylation event.
2.1. Introduction

The human genome consists of only 20,000 – 25,000 protein-coding genes. Thus, the initially expected or hypothetical number of 2,000,000 genes was far off the mark. An explanation for this is that nature introduces diversity at the protein level of an organism in different ways, resulting in two to three orders of magnitude higher complexity compared to the corresponding genome. Modifications during transcription are one way to introduce diversity. Alternative promoters, mRNA editing, or alternate splicing expands the genome four fold, resulting in a transcriptome of around 100,000 mRNAs. Further diversification is reached by covalent post-translational modifications (PTMs) of proteins. This leads to a set of approximately 1,000,000 human proteins which is then closer to the number of expected genes. These PTMs can be divided into two classes: peptide backbone cleavage, including controlled proteolysis and protein splicing, and the enzymatic introduction of chemical groups resulting in their covalent binding to residue side chains. From the 20 proteinogenic amino acids, 15 can be diversified with PTMs, with phosphorylation, acetylation, glycosylation, amidation and hydroxylation being the most common reaction types.

Nuclear receptors (NRs) are transcription factors that play an important role in various cellular processes. PTMs influence both the stability and activity of NRs, and thereby provide a platform for their in addition to small molecule and protein-protein interaction (PPI)-based regulatory mechanisms. Recently, PTMs could be connected with the onset and progression of several NR-based human diseases. For example, a direct link between PTMs and disease could be made for sumoylation, acetylation and phosphorylation of various NRs, including estrogen receptor (ER) and androgen receptor (AR). Hormone-refractory prostate cancer (HRPC) has been related to AR phosphorylation (p). Upregulation of the PI3K/Akt pathway results in AR pT308 and pS473 and HRPC development. Reduced AR ligand binding could be correlated to pS215 and pS792, resulting in decreased toxicity of the elongated polyQ stretch in Kennedy’s Disease. Enhanced prostate cancer cell growth has been linked to both AR acetylation (K630) and desumoylation (K386 and K520). In case of ERα, tamoxifen therapy in breast cancer treatment was more successful in the case of serine phosphorylation pS118 and pS167, but resulted in tamoxifen resistance for pS305. Sumoylation of certain positions in ERα, however, might be useful in the prediction of breast cancer response to endocrine therapy. Further connections to PTMs could be made in the case of insulin resistance in adiposity and inflammatory diseases. In conclusion, PTMs are an important constituent of NR function and show potential as both disease progression biomarkers and predictive patient response markers in terms of NR-directed treatments. Acetylation, or more specifically ε-N-acetyl-lysine (Figure 2-1), could be identified for all of the different sub classes of NRs, including thyroid hormone receptor-like, retinoid X receptor-like (e.g. HNF-4α), estrogen receptor-like (e.g. ERα/AR), and steroidogenic factor-
like. They are involved in the modulation of various cellular functions, such as stability and activity of the receptor, sensitivity of hormone response, DNA interaction, and subcellular distribution.\textsuperscript{24}

The acetylation sequences of NRs are predominantly localized in the flexible hinge region (HR). This region of NRs bridges the DNA binding domain (DBD) with the ligand binding domain (LBD) and plays an important role in DNA binding, NR dimerization, and nuclear translocation (see chapter 3 for more details).\textsuperscript{25-28} In the case of ER\textsubscript{a}, four lysines (K) are known to be acetylated: K266, K268, K299, K302 and K303 (Table 2-1, ER\textsubscript{a}).\textsuperscript{29,30} The last two were studied more intensively since they are part of a R/KXKK motif conserved across all NRs members, and because of their proximity to the serine at position 305 (S305), which is important for ER activity. Sequence homology of these lysine-rich areas adds weight to the idea of a functional importance of NR acetylation (Table 2-1). For example, the direct connection between subcellular localization and acetylation could be shown for HNF-4 \textalpha and AR.\textsuperscript{31,32}

Little is still known about NR structural change and the biological outcome upon introduction of PTMs and thus there is an urgent need to address this issue with chemical biology approaches. Motivated by this current state of affairs, this study is focused on the influence of lysine acetylation at positions 266 and 268 (K266 and K268, respectively) on the secondary structure of the mostly flexible HR (\textsuperscript{162}HR\textsuperscript{273}). Furthermore, the ability of the enzyme p300 to acetylate these small 12-mer peptide sequences of the HR - synthesized via solid phase peptide synthesis (SPPS) – was studied in the absence of any auxiliary peptide or protein. Kim \textit{et al.} could identify these positions as targets for p300 acetylation by sequential deletions of certain domains of the full length ER. Additionally, mutational studies, specifically mutating K266 or K268 reduced protein acetylation by 30-50\%\textsuperscript{30}. Interestingly, these lysine residues are conserved across multiple species, indicating a degree of relevance for NR function.\textsuperscript{30}
Chapter Two

Table 2-1 | Reported PTM acetylation of NRs

<table>
<thead>
<tr>
<th>NRs (human)</th>
<th>position</th>
<th>peptide sequence (R/KKKK consensus)</th>
<th>peptide sequence (other site)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ERα</td>
<td>295, 262</td>
<td>PLMIKRSKKNSLAL</td>
<td>GRMLKHKRQRDD</td>
</tr>
<tr>
<td>AR</td>
<td>626</td>
<td>TLGARKLLKGLNLK</td>
<td></td>
</tr>
<tr>
<td>GR</td>
<td>487</td>
<td>NLEARKTKKIKGI</td>
<td></td>
</tr>
<tr>
<td>HNF4α</td>
<td>111, 172</td>
<td>QCRYCRLKCFRAG</td>
<td>INDIRAKKIASIA</td>
</tr>
<tr>
<td>SF1</td>
<td>102</td>
<td>DRALKQQKKAQIRA</td>
<td></td>
</tr>
<tr>
<td>RXRα</td>
<td>197</td>
<td>GPGAGKRLCAI</td>
<td></td>
</tr>
<tr>
<td>TRα</td>
<td>106, 125</td>
<td>QCQLCRFKKCIAVG</td>
<td>DSKRVAKRKL</td>
</tr>
<tr>
<td>TRβ</td>
<td>160</td>
<td>QCQECRFKKCIVG</td>
<td></td>
</tr>
<tr>
<td>LXRα</td>
<td>78, 170, 424</td>
<td>EIRPOKRKKGPAPK</td>
<td>LRLQDKKLPLL</td>
</tr>
<tr>
<td>LXRβ</td>
<td>68, 161</td>
<td>EPEPERKRKKGPAPK</td>
<td>SEEQIRKKKIRKQ</td>
</tr>
<tr>
<td>FXR</td>
<td>208</td>
<td>LRKNVKQHADQ</td>
<td></td>
</tr>
</tbody>
</table>

Alignment of NR acetylation motif R/KKKK; bold NR are reported to be acetylated at the R/KKKK motif; bold residues, acetylation recognition motif or acetylation site

2.2. Predicted helical character of model 12-mer peptides

Molecular dynamic (MD) simulations were initially performed to gain information about the native structure of the peptides. For the calculations, an implicit solvent method was used to simulate the properties on the protein surface getting insights in the structure of protein domains upon binding to target proteins. These studies showed a moderate overall helicity (α- and 3₁₀ helix) of the 262HR²⁷₃ of Erα. (Figure 2-2B). The introduction of a side-chain acetyl group at lysines 266 and/or 268, however, significantly changed the helical character (KK < KK* < K*K < K*K*, Figure 2-2A). Compared to the natural non-acetylated sequence KK, the mono-acetylated peptide KK* showed an increase in helical content; especially in the C-terminal region of the peptide. K*K also increased the helical content, but, in contrast to KK*, did so in the N-terminal region. While the doubly-acetylated 12-mer peptide K*K* - lacking both positive charged side chain lysine residues - revealed the highest degree of helical secondary structure during the 20 ns MD simulation (Figure 2-2D).
Investigation into Posttranslational Modifications of the Estrogen Receptor Hinge Region

Figure 2-2 | Degree of helical content per residue obtained by MD simulation
Molecular dynamics (MD) simulations of the sequence of $^{26}$HR$^{273}$ of ERα. Starting from an extended initial conformation, simulations were fully unrestrained and all trajectories were generated. Each peptide was simulated for a total of 20 ns - 20,000 trajectory snapshots spaced every 1 ps - at 51.9 °C (325 K). The degree of helical content per residue was obtained using the ptraj module of AMBER omitting the first 5 ns of each simulation: (A) helical content of natural non-acetylated KK, 266-mono-N-acetylated K*K, 268-mono-N-acetylated K*K, and 266/268-doubly-N-acetylated K*K*. (B) Contribution of different helical motifs ($\alpha$-helix, $\beta$-sheet) to total helicity. More details in Experimental (1.2.9), MD Simulation and Figure 2-10.

2.3. Solid-phase peptide synthesis of 12-mers
Peptides of 12 amino acids lengths and representing the $^{26}$HR$^{273}$ sequence of the Hinge Region in different acetylated forms were prepared on solid phase (Rink Amide MBHA resin), using the fluorenylmethyloxycarbonyl (Fmoc) strategy.$^{36,37}$ The synthesis was performed on a 400 μmol scale for the first five amino acids of the peptide sequence. After confirmation of peptide sequence and purity by LC/MS, the resin was then divided into four 100 μmol batches, to allow for the synthesis of the non-, the two mono-, and the doubly acetylated peptides. The final yields of the peptides after purification were in the range of 15 to 25%. The purity and integrity of the final target 12-mer peptides were confirmed by LC/MS (Figure 2-3). The peptides were subsequently subjected to structural and enzymatic studies using far-UV circular dichroism (CD) (see 2.4), nuclear magnetic resonance (NMR) spectroscopy (see 2.5), and an acetyltransferase assay (see 2.6).
Figure 2-3 | Representative LC/MS analysis of K*K* peptide made by SPPS
Ion chromatogram (top, TIC), photodiode array chromatogram (bottom, PDA), and the corresponding mass spectrum of double-acetylated K*K* peptide analyzed by liquid chromatography–mass spectrometry (LC/MS). Other peptides in Experimental (1.2.9), SPPS and Figure 2-9.

2.4. Structural analysis by far-UV circular dichroism (CD)

Far-UV circular dichroism (CD) studies were performed to obtain information about the secondary structure of the 12-mer peptides and the influence of acetylation in this case. The amide peptide bonds possess two electronic transition states that are both CD active: the $\pi \rightarrow \pi^*$ transition is primarily responsible for the negative bands at 222 nm, while the $\pi_0 \rightarrow \pi^*$ transition results in both a positive band at ~190 nm and the negative band at 208 nm, all of which are characteristic of an $\alpha$-helix secondary structure. 38

In phosphate buffer (5 mM, pH 7.5) the peptides (50 $\mu$M) displayed a combination of weak $\alpha$-helical character (negative CD signal at <200 nm, no signal at 208 nm or 222 nm), and a random coiled polypeptide. It is known that the addition of 2,2,2-trifluoroethanol (TFE) increases the propensity of peptides to form secondary structures. 39 Indeed, as a consequence of the step-wise addition of TFE (10% - 50%), a shift of the minimum at <200 nm to higher wavelength (towards 208 nm) and the emergence of a minimum around 222 nm was observed (Figure 2-4A). Calculation of the fractional helicity $f_i$ showed an increase from 5.45% to 16.71%, with the greatest increase witnessed between 10% and 30% (v/v) TFE. Thus, a constant concentration of 30% (v/v) TFE was used to study the different peptides. The apparent helicity of the non-acetylated peptides was found to increase as a function of acetylation, with the doubly acetylated peptide showing the highest helicity ($f_i$K*K* < K*K < KK* < K*K*, Figure 2-4).
A B

Figure 2-4 | Influence of TFE and acetylation on α-helical character of the 12-mer peptides
(A) Comparison of circular dichroism (CD) data for peptide KK at 50 μM measured in 5 mM phosphate (P) buffer (pH 7.5, 20 °C) on varying the amount of additive 2,2,2-trifluorethanol (TFE). (B) Peptides KK, KK*, K*K and K*K* without TFE (condition a) and 30% (v/v) TFE (condition b). Mean residue ellipticity was calculated using equation 1 (Experimental (1.2.9), CD).

The impact of the intermolecular structuring on thermal stability was evaluated by monitoring the CD effect as a function of temperature. Unexpectedly, these studies revealed a decrease of ellipticity at 222 nm - normally correlated with greater α-helicity – upon an increase in temperature (Figure 2-5, grey).

A B

Figure 2-5 | Influence of temperature on the α-helicity of the 12-mer peptides
Circular dichroism (CD) measurements (A) tracking the mean residue ellipticity at 222 nm (MRE222nm) of 200 μM of KK (●) and K*K* (●), respectively, over a temperature range of 5°C - 90°C; and (B) the full MRE spectrum of at 5°C and 90°C. Mean residue ellipticity (MRE) was calculated using equation 1 (Experimental (1.2.9), CD).

2.5. Structural analysis using Nuclear Magnetic Resonance (NMR)
The NMR chemical shift perturbation studies were in line with the findings from the CD studies. Analysis of 1H-13C correlation spectra identified a specific pattern of chemical shift perturbation for α-helix stabilization, especially at the two lysine positions. Compared with peptides with a random coil conformation, β-sheet peptides exhibit a Hα (proton of backbone carbon next to the carbonyl carbon) downfield shift and a Cα (backbone carbon next to the...
Chapter Two

carbonyl carbon) upfield shift. By contrast, α-helical conformation results in the opposite effect, namely an upfield shift of $\text{H}^\alpha$ and a downfield shift of $\text{C}^\alpha$. While not all single residues show these characteristic effects of enhanced α-helicity, there is a significant overall trend in this direction. Interestingly, this trend is most pronounced at the acetylated lysine positions (Figure 2-6).

![Figure 2-6](image)

**Figure 2-6 | $\text{H}^\text{13C}$ chemical shift perturbation studies**

(A) Chemical shifts for both the $\text{H}^\alpha$ and the $\text{C}^\alpha$ nuclei upon acetylation of lysines (K) for each residue; (B) an expanded view of the superposition of the $\text{H}^\text{13C}$ heteronuclear single quantum coherence (HSQC) spectra of the free KK (solid thin contours) and K*K* (dashed contours). Arrows assigned to chemical shifts. HA = $\text{H}^\alpha$; CA = $\text{C}^\alpha$; e.g. K5 = lysine (K) at position 5.

### 2.6. Acetylation by lysine acetyltransferase p300

Protein side-chain acetylation by lysine acetyltransferases (KATs) occurs specifically at lysine residues (Ks). Histones are proteins whose activity is strongly regulated via lysine acetylation. The specific KATs for histones are otherwise referred to as histone acetyltransferases (HATs), although HATs can also target proteins other than histones. A versatile method to probe the substrate characteristics of HATs is the usage of enzyme-based assays. In this, the acetyl-group-transfer is directly coupled to a read-out signal, such as an increase in fluorescence intensity. In the HAT assay used in this present study, the minimum components for peptide acetylation are provided: namely (1) *acetyl coenzyme A* (acetyl-CoA), as a source of transferrable acetyl group and (2) the *acetyltransferase p300*, which catalyzes the transfer of an acetyl group from acetyl-CoA to form $\varepsilon$-N-acetyl-lysine. As a consequence of the transferred acetyl groups the sulphydryl group of the hydrolyzed CoA-SH reacts with 7-diethylamino-3-(4’-maleimidophenyl)-4-methylcoumarin (CPM) to yield a fluorescent product, which can be detected by excitation at 390 nm and emission at 469 nm (Figure 2-7).
The acetyl transferase p300 catalyzes the transfer reaction of the acetyl group from acetyl-CoA to a lysine residue, simultaneously releasing CoASH. This reacts with 7-diethylamino-3-(4’-maleimidophenyl)-4-methylcoumarin (CPM) to form a highly fluorescent product, which can be detected by excitation at 390 nm and emission at 469 nm. Thus the amount of transferred acetyl group is determined by the fluorescent signal.46

The non-acetylated peptide resulted in the highest fluorescent signal, indicating that most acetyl groups were transferred to KK. Interestingly, KK even overpowered the positive control histone H4 peptide, H4, a known substrate for p300.47 As expected, the double-acetylated K*K* peptide showed no signs of further acetylation, with a fluorescent signal similar to the control experiment (Figure 2-7). The mono-acetylated constructs K*K and KK* could be acetylated by p300, however, the sum of the transferred acetyl-groups to both K*K and KK* (each with one free lysine residue) did not reach the total amount transferred to KK (two free lysines). Furthermore, there was a difference in the amount of mono-acetylation of KK* compared to K*K.
2.7. Discussion

In a separate study - investigating the periodicity of lysines in histones - the \( i, i + 4 \) appearance of lysines in histones was reminiscent of that found in an \( \alpha \)-helix in which backbone hydrogen bonds occur between \( i, i + 4 \) residues.\(^{48}\) This insight provides some indication of a coherent relationship between lysine acetylation and its contribution to secondary structure, such as \( \alpha \)-helix formation. Similar to histones, tubulins possess unstructured domains in crystal structures that are known to be important for the recognition of microtubule-associated proteins. Interestingly, these domains are decorated with diverse posttranslational modifications (PTMs), such as acetylated lysine residues.\(^{49}\) Furthermore, acylation, including acetylation, has a stabilizing function in secondary structures.\(^{50}\) Another study revealed that the \( \alpha \)-helical content of histone H4 increases upon acetylation at the \( \epsilon-N \)-position of lysines.\(^{51}\) In this present work, the helical content of KK, a 12 residue long peptide based on the sequence of the hinge region (HR) of ERO, increased upon lysine acetylation. Mono-acetylated K*K (AcLys266) and KK* (AcLys268) as well as doubly acetylated K*K* (AcLys266/268) showed a higher propensity to form an \( \alpha \)-helix in MD simulations. Synthesis of the corresponding peptides (SPPS) and subsequent structural analysis using CD and NMR confirmed these helix-stabilizing findings: upon acetylation of Lys266 and Lys268 both the fractional helicity increased (CD) and the chemical shift perturbation studies (NMR) showed \( \alpha \)-helix stabilization characteristics in the form of an upfield shift of \( H^a \) and concurrent downfield shift of \( C^a \) nuclei.\(^{40-42}\)

In nature, acetylation of lysine residues is performed by acetyltransferases, such as histone acetyltransferase p300 (p300). Different crystal structures are available showing p300 in complex with its substrates, such as p53,\(^{52}\) hypoxia-inducible factor-1 (HIF-1)\(^{53}\) or nuclear receptor coactivator 3 (NCoA-3).\(^{54}\) In another study concerning the location of the involved components, the p300 HAT domain was co-crystallized with a bi-substrate consisting of acetyl coenzyme A (acetyl-CoA) and lysine (Lys-CoA). This crystal structure discovered in proximity to the pocket that facilitates the lysine acetyltransferase reaction an additional highly electronegative region.\(^{55}\) That signified that the properties of the amino acids flanking the target lysine residue play an important role. Sequence alignment of non-histone p300 substrates verified the hypothesis that positive charged residues are required in proximity (up to three to four residues up- and downstream) to the lysine acetylated by p300.\(^{55}\) In the case of KK, three additional positively charged amino acids (3xArg) are present in the 12-mer peptide next to the two lysines. Thus, from the viewpoint of Lys266, the two arginine residues, Arg263 and Arg269 are present three residues away, a position that is conserved with positive amino acids in p300 substrates.\(^{56}\) Furthermore, the other lysine residue (Lys268) also meets this requirement with Arg271 that is capable of forming a salt bridge with residues in the additional electronegative region of p300. Mutational studies on p300, which removed these negatively charged residues (E1505, D1628, D1625) resulted in a
decrease in acetylation ability, underlining the importance of the positively charged flanking residues in effective peptide acetylation.\textsuperscript{58} This need for a charge-dependent interaction to regulate the modification capabilities of the enzyme could also explain the low degree of acetylation in the case of K*K and KK*, respectively, compared to KK, as determined by the acetylation assay (HAT). The total amount of transferred acetyl-groups from acetyl-CoA to both peptides with one free lysine (K*K, KK*) was less compared with KK (two free lysines). In both cases a positive charge had been removed through the introduction of an acetyl lysine (AcLys) during peptide synthesis. The consequent neutralization – in each case two residues away from the lysine - could negatively influence the substrate binding capability and therefore the acetylation efficiency. In general, the positive effect of PTMs on the α-helical character of protein domains is also directly connected to the protein stability. In the case of hypoxia-inducible factor-1 alpha (HIF1α) protein, for instance, protein stabilization is a function of acetylation implemented by p300.\textsuperscript{57}

Kim et al. also studied acetylation of Lys266 and Lys268 of ERα using a variety of biochemical and cell-based approaches. In an acetylation assay consisting of a radiolabeled acetyl-CoA they could demonstrate acetylation by p300. However, efficient acetylation required SRC2\textsubscript{RID/PID} as an auxiliary protein, containing both the receptor interaction domain (RID) and p300 interaction domain (PID).\textsuperscript{58} Furthermore, they identified Lys268 as the major acetylation site of single acetylation. Although in the HAT assay the distinction between the two acetylation sites in the non-acetylated KK was not possible, single acetylation of K*K (free Lys268) was not favored over KK* (free Lys266).

The temperature-induced structural change, observed by CD spectroscopy at high peptide concentrations – expressed in the form of a decrease in ellipticity at 222 nm (Figure 2-5) - is an effect that has been reported for peptides,\textsuperscript{59} denatured proteins,\textsuperscript{60} as well as for intrinsically disordered proteins (IDPs) and natively unfolded domains, such as the N-terminal domain of p53,\textsuperscript{61} microtubule-associated protein Tau,\textsuperscript{62} or the extracellular domain of the nerve growth factor.\textsuperscript{63} This unusual temperature-dependent behavior - compared to the typical unfolding of globular proteins - was initially explained by the formation of α-helices inducing again partial folding of intrinsically unstructured proteins.\textsuperscript{64} However, multidimensional NMR spectroscopy studies of different IDPs and peptides revealed that transiently formed helices exhibited the highest helical character at low temperatures, indicating that the negative CD effect is not a function of α-helix formation. Whereas, a redistribution of the statistical coil involving a general loss of polyproline II (PP\textsubscript{i}) helix content is responsible for the change in the CD signal at 222 nm.\textsuperscript{65} Studies show that this left-handed helix \textbf{(9.3 Å/turn, 3 residues/turn)} occurs in a frequency of around 5%, resulting in one third of the coil state of proteins.\textsuperscript{66} Apart from proline-rich peptides, other amino acids including lysines can also adopt PP\textsubscript{i} helical character due to the electrostatic repulsion between the side chains.\textsuperscript{67} In this present work, the 12-mer peptides harbour two
lysine residues and an additional three arginines, resulting in five positive charges at physiological conditions. Thus a PPII helix conformation is very likely. Interestingly, the temperature-effect on the ellipticity at 222 nm is lower for K*K* compared to KK (Figure 2-5A). However, the curves were observed to level off at a similar value. This could be explained by the lower degree of initial PPII helical secondary structure of K*K* resulting from a double-acetylation of the lysines and thus less electrostatic repulsion. The possible transient presence of a PPII helix provides further opportunities for potential interaction with other proteins. Kinase activity, for example, can be modulated through the binding of the Src homology 3 (SH3) domain via PPII helices both intramolecular and in target proteins. Furthermore, PTMs, including acetylations, are both involved in the regulation of different cellular processes and are thus connected to various diseases. As an example, acetylation was shown to be important for the regulation of RAS oncogenicity as a stabilizer of certain domains. Counterparts of acetyl-lysines are bromodomains (BRDs), which form a specific hydrogen-bond with the oxygen of the acetyl carbonyl group. Thus, BRDs have recently become interesting as therapeutic targets blocking their interaction with acetyl lysine-containing domains. Thus, the elucidation of the structure of these domains and how they are modulated is not only important for the understanding of NR function, but also provides structural insights that are necessary for successful targeting.
2.8. Conclusion

In conclusion, this study of the linker domain between the DBD and the LBD of estrogen receptor $\alpha$, namely the ER$\alpha$ hinge region (HR), and the influence of posttranslational modifications (PTMs) has provided structural insights that may help to understand the function of what is intrinsically a poorly structured region. Initial MD simulation studies predicted structural changes upon acetylation of lysines at position 266 and 268. While the mono-acetylated $K^*K$ and $KK^*$ each showed different degrees of enhanced propensity to form $\alpha$-helices, the doublyacetylated peptides ($K^*K^*$) displayed the greatest structural change compared to the non-acetylated peptide $KK$.

Based on studies of 12-mer peptides synthesized by solid phase peptide synthesis (SPPS), CD (fractional helicity) and NMR studies (characteristic chemical shifts) confirmed these results: namely, a higher $\alpha$-helical character as a function of lysine acetylation. The CD data additionally hinted at the presence of a transient P$\beta$ helix, which was less pronounced for $K^*K^*$ than for $KK$. This potentially novel structural insight is not only important in the exploration for molecular and physiological mechanisms but could also be an interesting platform in specific targeting.

Finally, we could also show the acetylation of both mono-acetylated ($K^*K$, $KK^*$) and non-acetylated peptide ($KK$) by the acetyl transferases p300 in the absence of an additional auxiliary protein. Interestingly, the sum of both acetylations of the mono-acetylated peptides ($K^*K$, $KK^*$) were significantly lower compared to the double acetylation of the non-acetylated $KK$. This means that acetylation of Lys266 and the consequent charge-reduction diminishes the acetylation capacity of the adjacent Lys268 and vice versa. This charge-based acetylation capability of p300 taken together with their dissimilar degree of influence on the domain secondary structure suggests that these acetylations are important for the sensitive regulation of cellular processes through a change in the structure of what is otherwise a poorly structured part of the hinge region.
2.9. Experimental

General
MD simulations and NMR measurements and their interpretations were made by Dr. Lidia Nieto.

If not stated otherwise, chemicals were ordered from SIGMA-ALDRICH.

Solid Phase Peptide Synthesis & Purification
The synthesis of all peptides was performed on solid support using the Fmoc strategy on a Prelude peptide synthesizer (PROTEIN TECHNOLOGIES INC.). If not stated otherwise, 200 µmol (1 eq.) of Rink Amide MBHA resin (0.59 mmol/g, NOVABIOCHEM) was swollen for 30 min in 5 ml N-methylpyrrolidone (NMP). Washing was performed with NMP or dichloromethane (DCM), deprotection of the amino acids (AA, 0.2 M solution in NMP) was performed with piperidine (Dep., 20% (v/v) solution in NMP), activated with 2-(1H-Benzotriazole-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate (Act., HBTU, 0.4 M solution in NMP) and alkaliized with N,N-Diisopropylethylamine (Base, DIPEA, 1.6 M in NMP). The amino acids were double-coupled with a subsequent capping step (Cap., 20% (v/v) pyridine, 20% (v/v) acetic anhydride in NMP): 6 ml NMP, 4 ml Dep. (5 min), 2x[6 ml NMP, 4 ml AA, 2 ml Act., 2 ml Base, 20 min], 6 ml NMP, 3 ml Cap. (5 min), 3 ml NMP. After finishing the desired peptide sequence, the resin was washed 3 times alternately with diethylether (DEE) and dichloromethane (DCM) and dried under vacuum for 10 min. The side chain protection group and the cleavage of the peptide sequence from the resin were simultaneously achieved by the usage of mixture of trifluoroacetic acid, triisopropylsilane and water (TFA/TIS/H2O, 95/2.5/2.5) for 3 h. The free peptide was precipitated by adding it drop-wise to ice cold DEE. A final centrifugation step (2,000 rpm, 10 min) with subsequent washing (ice cold DEE), uptake in water and lyophilization results in the crude peptide. Purification was accomplished by reverse-phase HPLC on a Alltima™ HP C18 column (125x20 mm, Alltech). Water (0.1% TFA) was used as a polar phase, adding different amounts of an apolar acetonitrile (ACN, 0.1% TFA) phase. A linear gradient (20 ml/min) was optimized for each peptide to a 5% range, for instance 35-40% ACN. The purity of the peptide was determinate at an analytical LC-MS using GraceSmart RP18 (50x2.1mm, Grace, 3u, 120A) column. Finally, the peptides were lyophilized and stored at -80 °C.
Figure 2-9 | LC/MS-analysis of peptides made by SPPS
Ion chromatogram (top, TIC), photodiode array (bottom, PDA), and corresponding mass spectrum of unacetylated KK (A), mono-acetylated K*K (B) and KK* peptide (C), and double-acetylated K*K* (D) peptide analyzed by liquid chromatography–mass spectrometry (LC/MS).

Circular Dichroism spectroscopy\textsuperscript{73,74}
Far-UV Circular Dichroism (CD) spectroscopy measurements were performed under a constant nitrogen flow at 20 °C using 50 μM of the peptide in the presence of 30% (v/v) trifluoroethanol (TFE), if not stated otherwise. As a baseline a peptide-free buffer (30% TFE) was measured under identical conditions. The spectrum was recorded from 250 nm to 185 nm using a JASCO-815 spectrometer. Quartz cuvettes with path lengths of 1 mm or 0.2 mm (HELLMA ANALYTICS) were employed with a data pitch of 0.5 in a continuous mode, a scan speed of 20 nm/min with a response time of 2 s and a bandwidth of 0.5 nm. The graphs are representing an average of five scans. The observed ellipticity $\theta_1$ (degrees, in [mdeg]) was converted to mean residue ellipticity $\theta_{mrw,\lambda}$ (in [deg cm$^2$/dmol]) using equation 1.

$$\theta_{mrw,\lambda} = \frac{M}{(N-1)} \cdot \frac{\theta_1}{10^{-d-c}}$$  \text{equation 1}

Where $M$ is the molecular mass of the peptide (in [Da]), $N$ is the number of amino acids in the peptide, $d$ is the path length (in [cm]), and $c$ is the peptide concentration (in [g/ml]). The fractional helicity ($f_H$) was calculated using equation 2.
\[ f_H = \frac{\theta_{\text{mrw,222nm}} - (2220 - 53T)}{(2507 - 44000)(1 - \frac{n}{n}) - (2220 - 53T)} \]  

Where \( \theta_{\text{mrw,222nm}} \) is the measured mean residue ellipticity at 222 nm, \( T \) is temperature in °C, and \( n \) is the number of amide bonds in the peptide.\(^73\)

**Molecular Dynamics Simulation**

All MD simulations were carried out using the AMBER suite of programs, and the ff03 force field.\(^24\) An implicit solvent was used via the General Born solvation method (IGB 5), as implemented in AMBER, providing an approximation of the electrostatic free energy when a charge distribution is created in a cavity of low dielectric (protein, \( \varepsilon = 1 \)) embedded in a uniform dielectric medium (water, \( \varepsilon_w = 80 \)).\(^{33-35}\)

Starting with an extended initial conformation, built by the LEaP module of AMBER for each peptide, all MD simulations were fully unrestrained and all trajectories were generated using the sander program in the AMBER 9 package. Each peptide was simulated for a total of 20 ns at 51.9 °C, following a similar approach as previously reported to predict the conformation of a miniprotein.\(^76\) Analysis was performed on the 20 ns using 20,000 trajectory snapshots spaced every 1 ps. A snapshot with the lowest potential energy across the simulation was chosen as representative structure for each peptide. The secondary structure of each residue as a function of time was subsequently analyzed utilizing the STRIDE secondary structure assignment algorithms as implemented in VMD.\(^77,78\)

![Diagram](image)

**Figure 2-10 | Degree of helical content per residue based on MD simulation**

Starting from an extended initial conformation, MD simulations were fully unrestrained and all trajectories were generated. Each peptide was simulated for a total of 20 ns - 20,000 trajectory snapshots spaced every 1 ps - at 51.9 °C (325 K). Grayscale is used to represent the propensity of forming secondary structures: white, unordered; light grey, \( \alpha \)-helix; grey, turn; black, \( 3_{10} \) helix. Dashed line indicate the start/end of a helix. A snapshot of the lowest potential energy across the simulation was chosen as representative structure for each peptide: (A) 266/268-double-N-acetylated ERα-HR262-273 (K*K*). (B) The degree of helical content per residue (for K*K*) was obtained using the ptraj module of AMBER omitting the first 5 ns of each simulation.
Nuclear Magnetic Resonance spectroscopy

Samples of peptides for NMR experiments were dissolved in 5mM PBS buffer (pH 7.4) with 30% TFE (v/v) (trifluoroethanol-d3, Cambridge Isotope Laboratories) to give a final peptide concentration of 1 mM. Spectra were recorded at 283 K on a Bruker 600 spectrometer (H frequency of 600 MHz) by using a 5 mm triple-resonance Z gradient probe and processed by using Topspin software. One- and two-dimensional spectra were acquired by using standard pulse sequences and WATERGATE-based solvent suppression sequences. A standard suite of homonuclear two-dimensional experiments total correlation spectroscopy (TOCSY) and nuclear overhauser effect spectroscopy (NOESY) with mixing times of 70 ms and 150 ms, respectively, were acquired for chemical shift assignments. To complete the heteronuclear carbon chemical shift assignments, natural abundance H-13C heteronuclear single quantum coherence (HSQC) spectra were acquired. The data were processed in Topspin 2.1 from Bruker Biospin and analyzed using CARA1.5.

Acetylation Assay

The Histone AcetylTransferase (HAT) assay kit (ACTIVE MOTIF) was used, including acetyl-CoA, as a source for the acetyl-group and histone acetyltransferase p300 that catalyze enzymatically the transfer of the acetyl group from acetyl-CoA to form ε-N-acetyl-lysine. The amount of transferred acetyl groups is directly correlated to the fluorescence signal, since the sulphydryl groups of the acetyl-CoA (CoA-SH) can react with a provided substrate to a fluorescent product.

The assay was performed in 96-well-plates with a final volume of 50 µl containing 50 µM Acetyl-CoA, 22.2 nM p300 and 50 µM of peptide. After 30 min of incubation at room temperature 50 µl of stop solution and 100 µl developing solution was added and incubated for 15 min in the dark. Subsequently fluorescence was measured with 375 nm excitation and 460 nm emission.

2.10. References

CHAPTER THREE

Studying the Estrogen Receptor Hinge Region

Using Expressed Protein Ligation

Abstract. Protein engineering has long been limited to mutational studies focusing on the 20 proteinogenic amino acids. The introduction of protein (semi-)synthesis techniques has enabled access to a much broader set of proteins with chemical probes and modifications such as posttranslational modifications (PTMs). In this present study, expressed protein ligation (EPL) was used to investigate the hinge region (HR) of the estrogen receptor α (ERα) and its PTMs in the context of their impact on the ligand binding domain (LBD). Using the amino-terminal methionine excision approach, it was possible to generate recombinant ERα-LBD with an N’-terminal cysteine, required for the EPL reaction. Thioester peptides - as counterparts to the N’-terminal cysteine in the EPL - were synthesized by solid-phase peptide synthesis using the Dawson method, which uses a fluorenylmethyloxy-carbonyl-based coupling strategy in combination with a diaminobenzoyl linker attached to a Rink amide resin. While ligation of the posttranslationally modified HR peptides to the model protein C-YFP could be successfully demonstrated, the transfer of this method to ERα was not successful. Protein instability under ligation conditions proved to be the major issue in this case. Studies focusing on the extension of the ERα-LBD in the direction of the HR revealed that protein expression allowed the generation of ERα-LBD constructs including the HR up to position 251. Binding studies identified the HR sequence before position 271 to significantly decrease the ability of ERα to bind coactivator peptides, thus hinting at the existence of two regions of the HR, which differ in terms of their structure and function.
3.1. Introduction

Protein engineering plays a crucial role in biochemistry, biomedicine and biotechnology.\(^1\) Classical mutational studies have one essential limitation, however: namely, the introduction of solely 20 genetically encoded amino acids. This means that natural ribosomal synthesis is incapable of introducing unnatural amino acids, biophysical probes, or posttranslational modifications (PTMs). However, in the last decades different techniques have been developed for the specific modification of proteins, including unnatural amino acid mutagenesis and native chemical ligation (NCL). These are two powerful techniques to circumvent the limitations of natural protein synthesis and have multiple applications.\(^2,3\) While modifications are introduced at the stage of translation in the case of the mutagenesis method, NCL combines two fragments of peptides or proteins in a chemoselective manner.\(^4\) This requires a carboxy(C′)-terminal thioester group and an amino(N′)-terminal cysteine (N′-Cys), resulting in a native peptide bond bearing a cysteine residue at the junction (Figure 3-1). Main advantages of this chemical peptide synthesis technique are that the peptide can be fully unprotected and that the reaction can be performed under neutral conditions. In the 1950s the synthesis of small peptides had already been performed using thioester chemistry and thus Wieland et al. can be viewed as pioneers in the field of NCL.\(^5\) The introduction of this specific ligation reaction resulted in various chemical syntheses of chemically modified peptides of medium size. Standard peptide synthesis on solid-phase (SPPS) is typically limited to peptide sequences of no greater than 60 residues. NCL, however, can be used to couple large peptide fragments to generate even larger proteins.\(^6\) During this type of protein semi-synthesis, also named expressed protein ligation (EPL), a recombinantly expressed protein containing an thioester is ligated via NCL to a chemically synthesized peptide containing an N′-Cys or vice versa (with thioester on the chemically synthesized peptide and N′-Cys on the recombinant expressed protein)\(^6,7\) A wide range of chemical modifications have been introduced using EPL on different classes of proteins, including antibodies, ion channels, polymerases, phosphatases, kinases, signaling proteins and transcription factors like nuclear receptors (NRs).\(^8,9\)
The following six areas of application have in particular benefited from this new technique:

- site-specific solid-support-immobilization
  - e.g. covalent attachment of proteins onto glass surfaces\(^{10}\)

- polypeptide backbone-cyclization
  - e.g. biosynthesis of circular peptides and proteins using intramolecular EPL reaction \textit{in vitro} and \textit{in vivo}\(^{11}\)

- isotope-edited spectroscopy
  - e.g. decreasing spectral complexity via selective residue-labeling with NMR active nuclei\(^{12}\)

- incorporation of non-natural amino acids
  - e.g. such as the semi-synthesis of an ion channel using D-amino acids to modify selectivity\(^{13}\)

- incorporation of optical probes
  - e.g. synthesis of doubly fluorescently labeled protein to study stability in living cells by microscopic imaging\(^{14}\)

- introduction of PTMs for structural and biological studies
  - e.g. site-specific protein prenylation, \(\epsilon\)-acetylation\(^{15}\) or phosphorylation\(^{17}\)

The preparation of the starting products for the EPL reaction is an essential aspect of this approach for which a number of different methods are currently available. The generation of a recombinant protein with an N’-terminal cysteine residue (N’-Cys) can be achieved using different techniques. The introduction of cleavage sites of exogenous proteases (Factor Xa, TEV or thrombin) directly prior to a cysteine enables purification via an N’-terminal tag and subsequent cleavage with the exogenous proteases, resulting in an N’-Cys.\(^{18-20}\) Another option is to generate an N’-Cys by using endogenous leader peptidases in the periplasmic space to which an N’-terminal pelB leader sequence is directing.\(^{21-23}\) Furthermore, the approach of N’-terminal methionine excision (NME) can be used. In this case the endogenous methionyl aminopeptidases removes the methionine during expression in \textit{E. coli}, generating an N’-Cys.\(^{24,25}\) Finally, protein splicing and inteins can be used to produce N’-Cys. Several inteins such as \textit{Mxe} GyrA, \textit{Mth} RIR1 and \textit{Ssp} DnaB have been modified so that splicing can be accomplished in a pH- and temperature-dependent fashion to provide an N’-Cys.\(^{26-28}\)

The counterpart to the N’-Cys bearing protein in EPL, a thioester-featuring peptide, can be synthesized by solid phase peptide synthesis (SPPS) using a wide range of approaches, including mercaptopropionamide-, sulfonamide-, backbone amide-, S-protected oxazolidinone-, aryl hydrazine- and mercaptoarboxyethylene linkers,\(^{7,29-33}\) as well as Lewis
acid-activated cleavage and the introduction of an N-acyl-benzimidazolinone (Nbz) leaving group.\textsuperscript{34,35}

In this study EPL was used to investigate the hinge region (HR) of the estrogen receptor α (ER\textsubscript{α}) and its PTMs – in particular acetylation - on the lysine residues at position 266 and 268 (Lys\textsubscript{266}, Lys\textsubscript{268}) in the context of the entire ligand binding domain (Figure 3-1). The N’-Cys of the recombinant construct was generated using the NME approach, while the peptides were synthesized using fluorenylmethyloxycarbonyl (Fmoc)-based SPPS on the Dawson resin with a diaminobenzoyl (Dbz) linker.\textsuperscript{35} Activation of the peptide led to the corresponding N-acyl-benzimidazolinone (Nbz) leaving group, which was then converted to a reactive thioester during NCL. But aforetime, the ligand binding domain (LBD) of ER\textsubscript{α} was extended progressively in the N’-terminal direction towards the HR to gain information about its influence on protein stability. Furthermore, the variant expressing the entire HR can be seen as the recombinant counterpart to the semi-synthesized expressed LBD ligated to the chemically synthesized HR bearing posttranslational modifications.

**Figure 3-1 | Semi-synthesis of ER\textsubscript{α}-LBD and PTM-modified HR by the EPL reaction**
Expressed protein ligation (EPL) of the recombinantly expressed LBD of ER\textsubscript{α} (\textsuperscript{283}ER\textsubscript{α}-LBD\textsuperscript{553}) to peptides (\textsuperscript{262}HR\textsuperscript{281}) synthesized by solid phase peptide synthesis (SPPS), including PTMs (acytlation, Ac). N’-terminal methionine is cleaved off by endogenous methionyl aminopeptidases (NME) during expression in \textit{E. coli}. Release of N-acyl-benzimidazolinone (Nbz) leads to the C’-terminal peptide thioester. Transthioesterification and subsequent \textit{S,N}-acyl shift produces a native peptide bond between the chemically synthesized peptide and the recombinantly expressed protein.
3.2. Results and discussion

3.2.1. Generating the recombinant protein for EPL

The EPL at the N’-terminus of the LBD of ERα (around position 300) requires an N’-terminal cysteine (N’-Cys), one of the least abundant residues in proteins. The cysteine closest to the LBD, however, is located at position 381 and is thus impractical for this purpose. Site-specific mutation of a suitably located amino acid into a cysteine - preferably with similar properties - is an alternative option. Serine is ideally suited for this purpose since it differs from cysteine only in terms of one atom (oxygen versus sulfur) and is biosynthetically related. Serine occurs at a higher frequency than cysteine and in the case of the N’-terminus of ERα-LBD it can be found at positions 282, 294, 301, 305, and 309. For this work - the introduction of PTMs at position 266 and 268 - serine at position 282 (Ser282) is most convenient. Site-directed mutagenesis resulted in the following construct: ERα-\( ^{281}(S282C)HR-LBD^{553}\)-FXa-His\(_6\). This plasmid was transformed into E. coli BL21 expression cells and protein was expressed for 20 h at 15 °C. During expression endogenous methionyl aminopeptidases was envisaged to remove Met281, resulting in a free N’-Cys (C-ERα\(^{S282C}\), Figure 3-1, right). Purification was performed by means of the C’-terminal His\(_6\)-tag via Ni-NTA affinity chromatography. SDS-analysis of the protein showed a high purity after purification (Figure 3-2), however, the final yield of pure protein was low (0.5 mg/L).

![Figure 3-2](image1)

**Figure 3-2 | SDS-PAGE of expression and purification of ERα-LBD construct for EPL**

Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) of ERα\(^{281}(S282C)HR-LBD^{553}\)-FXa-His\(_6\) expression in E. coli (TB medium, 15 °C, 15 h, 10 μM estradiol, E2) and purification on a nitrilotriacetic acid-based nickel (Ni-NTA) column leads to pure C-ERα\(^{S282C}\) in the 250 mM imidazole fraction. Other fractions: 100 mM and 500 mM imidazole. Yield: 0.5 mg/L.

Subsequent LC-MS analysis revealed successful cleavage of the N’-terminal methionine. However, a concomitant increase in mass of 12-26 Da (e.g. +Δ12) was observed. This mass increase is the result of thiazolidine ring formation (five- or six-membered) through condensation of the cysteine sulffhydryl group and the N’-terminal amino group with aldehydes (Figure 3-3A) such as formaldehyde (+Δ12). This reaction of N’-Cys is well characterized and has also found use in various applications, such as prodrugs, peptide ligation, peptide purification, and N’-terminal protection of proteins. In the latter case,
the protective group is normally cleaved with methoxyamine, reverting the N’-Cys back into its unprotected form as required for EPL (Figure 3-3A). This deprotection of the N’-terminus - opening of thiazolidine ring - was also successful in case of the recombinantly expressed protein for NCL (Figure 3-3B).

![Figure 3-3](image)

**Figure 3-3 | Protection N’-Cys as the thiazolidine ring and subsequent deprotection using methoxyamine**

(A) Reaction scheme for the protection and deprotection of N’-Cys; (B) liquid chromatography–mass spectrometry (LC-MS) identifies the formation of the N’-terminal thiazolidine ring (A, +Δ12); (C) incubation with 125 mM methoxyamine hydrochloride in 500 mM acetate buffer at 4 °C for 12 h (pH 5.5) reverts back to the free N’-terminus.

3.2.2. Generating chemically synthesized thioester peptide for EPL

The peptide for the EPL was synthesized by solid phase peptide synthesis (SPPS) using Fmoc-protected amino acids on a Prelude peptide synthesizer. The coupling of the first glycine to the linker on the resin with 2-(6-Chloro-1H-benzotriazole-1-yl)-1,1,3,3-tetramethylammonium hexafluorophosphinate (HCTU) was very slow. To prevent incomplete coupling, the loading step of the first amino acid was repeated four times with the more reactive coupling reagent 2-(7-Aza-1H-benzotriazole-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate (HATU). The subsequent residues, including ε-N-acetyl-L-lysine, were doubly-coupled with a subsequent N-acetyl capping step. The last amino acid was introduced with an N’-terminal Boc-protective group (Figure 3-4). Further details can be found in 3.4 Experimental: Solid Phase Peptide Synthesis & Purification.
Figure 3-4 | Synthesis of the thioester peptide using the Dawson method
Peptide is elongated on a Dawson resin - featuring a diaminobenzoyl (Dbz) linker - using standard fluorenlymethyloxycarbonyl-based synthesis (Fmoc). The last residue was coupled with a tert-butyloxycarbonyl-(Boc)-protected N-terminus. Activation of the Dbz group led to the corresponding N-acyl benzimidazolizinone (Nbz) group, which was replaced by a thiol resulting in a C’-terminal thioester.\(^{30}\) R, amino acid side chains; R’S\(^\prime\), thiol additive. Further details can be found in 3.4 Experimental: Solid Phase Peptide Synthesis & Purification.

LC-MS analysis of the cleaved crude peptides confirmed that the nonacetylated (KK-Nbz), the lysine mono-acetylated (K*K-Nbz and KK*-Nbz) and the lysine doubly-acetylated peptide (K*K*-Nbz) had been successfully formed (Figure 3-5A, crude). The peptides were then purified by reverse phase HPLC. Due to poor separation of the highly polar peptides via standard C18 reversed-phase-based purification methods, the purification was instead performed using a hydrophilic interaction liquid chromatography (HILIC) column. Although the reasoning behind peptide retention on the HILIC column is not yet fully understood, it is thought to be due to a combination of hydrophilic interactions and ion exchange. The interaction with hydrophilic compounds is facilitated by the aqueous layer which is formed on the silica surface covered with cross-linked diol functional groups. In contrast to standard reverse phase chromatography, the strongest mobile phase has the highest water concentration, while the weakest has a higher concentration of the hydrophobic organic solvent, acetonitrile (ACN).\(^{42}\) The purity of the peptides obtained after HILIC purification was significantly higher compared to standard RP chromatography. As an example, the peptide KKK\(^*\) was isolated in sufficient purity for use in the optimization of the EPL reaction conditions (Figure 3-5).
3.2.3. Test ligation using EPL with the model protein C-YFP

In light of the low expression yield of C-ERαS282C the ability of the chemically synthesized peptides to undergo protein ligation was first explored using a yellow fluorescent protein with an N’-Cys (C-YFP) as a model protein. This protein showed improved stability compared to the ERα mutant (C-ERαS282C) intended for NCL. A portion of the test ligations focused on the choice and concentration of the thiol additive, which was used to catalyze in situ transthioesterification during the ligation reaction. Furthermore, both the protein-peptide ratio and the total concentration needed optimization, since they are known to influence ligation efficiency. As thiol catalyst, 4-mercaptophenylacetic acid (MPAA), thiophenol and 2-mercaptoethanesulfonate (MESNA) were used with protein:peptide ratios ranging from 1:10 to 1:33. The protein concentration varied between 5 - 116 μM, while the peptide varied from 0.1 - 1.7 mM (Table 3-1). Tris(2-carboxyethyl)phosphine) (TCEP) was used as reducing agent (Table 3-1).
Table 3-1 | Optimization of EPL of KK* with C-YFP

<table>
<thead>
<tr>
<th>reaction</th>
<th>protein [µM]</th>
<th>peptide [mM]</th>
<th>pro:pep ratio</th>
<th>thiol catalyst</th>
<th>TCEP</th>
<th>volume [µL]</th>
</tr>
</thead>
<tbody>
<tr>
<td>MPPA</td>
<td>30</td>
<td>1.0</td>
<td>1:33</td>
<td>I. 300 mM MPAA II. 200 mM MPAA</td>
<td>-</td>
<td>150</td>
</tr>
<tr>
<td>MPPA LOW</td>
<td>5</td>
<td>0.1</td>
<td>1:20</td>
<td>I. 100 mM MPAA II. 50 mM MPAA</td>
<td>-</td>
<td>150</td>
</tr>
<tr>
<td>MESNA</td>
<td>30</td>
<td>1.0</td>
<td>1:33</td>
<td>I. 500 mM MESNA II. 250 mM MESNA</td>
<td>-</td>
<td>150</td>
</tr>
<tr>
<td>MESNA HIGH</td>
<td>116</td>
<td>1.7</td>
<td>1:10</td>
<td>- II. 131 mM MESNA I. 3.4 mM II. 1.7 mM</td>
<td>190</td>
<td></td>
</tr>
</tbody>
</table>

All reactions are performed in 0.2 M phosphate buffer, 10% (v/v) glycerol, pH 7.0 with C-YFP and KK* thioester peptide. Ligation reaction: (I) 1 h preincubation and (II) main ligation reaction of 12 and 24 h, respectively.

The use of 2% (v/v) thiophenol resulted in a high degree of protein precipitation. However, in all other reactions protein instability was not an issue. A moderate level of ligation efficiency was observed, with a maximum conversion of 30% at high concentration of protein and peptide (1:33 ratio) in combination with a high MPPA concentration (Table 3-1, reaction: MPPA) after 24 h reaction time (Figure 3-6A).

Figure 3-6 | Ligation efficiency of EPL of KK* and C-YFP under different reaction conditions
MPPA and MESNA (protein:peptide: 1:33) both revealed moderate ligation, though marginally more for MPAA. SDS-PAGE analysis identified both unligated and ligated protein. Ligation: efficiency of ligation reaction is determined by integration of UV peak (photodiode array, FDA) of LC/MS spectrum.

Unfortunately, transfer of these ligation conditions to the recombinantly expressed C-ERαs252c construct resulted in protein precipitation and no evidence of successful ligation. In general, the EPL reaction has a number of limitations. Apart from the fact that the site-specific incorporation of chemically synthesized peptides can only be achieved close to the termini of the protein, the ligation site must also be solvent accessible; thus denaturing conditions are in some cases required.44 However, in this study, the site of ligation occurs at the terminus of the 30-residue-long HR. Thus accessibility of the cysteine residue should not be an issue in this case. Most EPLs performed in separate literature studies were performed at the C'-terminus, meaning a synthetic peptide, possessing a N'-Cys, was ligated to the C'-terminus of an expressed protein exhibiting a thioester.45 N'-terminal ligation,
However, requires a thioester peptide, which is sensitive to hydrolysis at pH > 7 (the optimal ligation condition is pH 7). Thus, although the ligation rate of this bimolecular reaction is increased by higher concentrations of the components, the ligation is still slow, and consequently hydrolysis almost always accompanies the desired reaction. Analysis of \textit{KK}⁺ by means of LC-MS analysis also identified hydrolyzed product over time. This confirms that the thioester - required for NCL - is not very stable.

Additionally, the identity of the C’-terminal amino acid of the thioester peptide can influence the rate of the EPL reaction. Although, the compatibility and efficiency of all proteinogenic amino acids could be shown, less hindered residues, including glycine and alanine resulted in a faster ligation. By contrast, valine, isoleucine, and proline were discovered to react at a significantly slower rate, while aspartic acid and glutamic acid were less effective due to side product formation. In our case, however, a glycine is present in the Xaa-Cys position which means that the reduced ligation efficiency cannot be explained by unfavored residues at the ligation site.

The nature of the thiol leaving group is another factor that influences the reaction rate. In the NCL the transthioesterification with the sulfhydryl group of the side chain of the \textit{N’}-Cys is the rate-limiting step. Therefore, exogenous thiols are added to NCL reactions to increase ligation kinetics by facilitating the \textit{in situ} formation of more active thioesters. Common thiol additives are thiophenol or a benzyl mercaptan/thiophenol mix for peptide thioesters or MESNA for recombinant protein-thioesters. A more recent, intense study investigating a set of 14 different sulfhydryl catalysts for NCL discovered aryl thiols, such as 4-mercaptophenylacetic acid (MPAA), as highly effective catalysts. However, in this present work the use of MPAA instead of MESNA did not improve the ligation efficiency significantly, but instead resulted in a maximal ligation of 30% in case of the model protein \textit{C-YFP} and no evidence of successful ligation in the case of \textit{N’-ERαGSSG}.

Another potential problem during the NCL reaction can be the formation of an unreactive disulfide linked dimer via two \textit{N’}-Cys. For this system, this oxidation reaction would lead to a covalent linkage of the \textit{N’}-termini of two proteins, which would make the EPL reaction impossible and could favor protein aggregation. During the EPL reaction, however, TCEP is added as a reducing reagent preventing the thiol of the \textit{N’}-Cys from oxidation. Disulfide bond formation might also occur after \textit{N’}-terminal methionine cleavage in the cell. While possible even under reducing conditions in the cytoplasm – by lowering the pH of cysteine due to charge interactions with adjacent residues – and close proximity of cysteines due to extension of the \textit{N’}-terminus, it is less unlikely to happen in the cell. However, the fact that \textit{N’}-Cys is located at the terminus of a 25 residue long linker might be connected with the low yield of the expression and high propensity of precipitation observed for this protein construct. To study the impact of the length of the linker in more detail, constructs with step-wise elongation towards the HR were generated (see 3.2.4).
3.2.4. N’-terminal elongation of the ERα-LBD towards the ERα-HR

The influence of extending the N’-terminus of the ERα LBD towards the HR was another part of the investigation of the HR. This 60 residue long region connects the DNA binding domain (DBD) with the LBD and is considered to be flexible and unstructured.\textsuperscript{55,56} However, PTMs in this region of the receptor modulate transcriptional activity,\textsuperscript{57} receptor degradation,\textsuperscript{58} hormone sensitivity,\textsuperscript{59} functional synergy between AF-1 and AF-2,\textsuperscript{55} and DNA binding.\textsuperscript{60} To gain more insight into the structure of the HR and its influence on the LBD, a step-wise elongation of the LBD was performed. Thus, the LBD construct ERα-\textsuperscript{309}LBD\textsuperscript{353}-FXa-His\textsubscript{6} in pET44a was extended towards the HR by insertional mutagenesis. Five different constructs were thus obtained (Table 3-2):

<table>
<thead>
<tr>
<th>construct</th>
<th>1st position</th>
<th>extension LBD (aa)</th>
<th>domains</th>
<th>total size (Da)</th>
</tr>
</thead>
<tbody>
<tr>
<td>LBD309</td>
<td>309</td>
<td>0</td>
<td>full length LBD</td>
<td>29,530</td>
</tr>
<tr>
<td>INT294</td>
<td>294</td>
<td>15</td>
<td>intermediate HR and LBD</td>
<td>31,330</td>
</tr>
<tr>
<td>INT272</td>
<td>272</td>
<td>37</td>
<td>intermediate HR and LBD</td>
<td>33,572</td>
</tr>
<tr>
<td>INT262</td>
<td>262</td>
<td>47</td>
<td>intermediate HR and LBD</td>
<td>34,863</td>
</tr>
<tr>
<td>HR252</td>
<td>252</td>
<td>57</td>
<td>full length HR and LBD</td>
<td>35,988</td>
</tr>
</tbody>
</table>

According to the classification system used by the protein knowledgebase UniProtKB – which attributes 60 amino acids from 251 to 310 to the human ERα-HR (NR3A1)\textsuperscript{61} – LBD\textsuperscript{309} (29,530 Da) is the full length LBD, HR\textsuperscript{252} (35,988 Da) is the full length LBD and HR and INT\textsuperscript{294}, INT\textsuperscript{272}, and INT\textsuperscript{262} are intermediate constructs (INT). Expression of these constructs in the presence of the natural ER full agonist estradiol (E2) was successful, resulting in good yields of protein (> 20 mg/mL) in the case of the LBD\textsuperscript{309} and INT\textsuperscript{294}, and and moderate yields (around 5 mg/mL) in the case of INT\textsuperscript{272}, INT\textsuperscript{262}, and HR\textsuperscript{252}, respectively (Figure 3-7).

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{fig3-7.png}
\caption{Protein expression and purification of HR252, INT262/272/294, and LBD309}
\end{figure}
The protein expression without E2, however, resulted in low yields and a high degree of impurities, of which the shortest constructs – LBD309 and INT294 – yielded the purest products. Nevertheless, the yield was worse than had been initially expected compared to protein expression in the presence of E2. (Figure 3-8).

![Image](image.png)

**Figure 3-8 | Ligand-free protein expression and purification of HR252, INT’s, and LBD309** SDS-PAGE gel with 100 mM (100), 250 mM (250), and 500 mM imidazole (500) elution fractions of protein expression and subsequent His-tag purification with Ni-NTA of ERα-LBD with different HR-extensions in the absence of the endogenous ligand, estradiol (E2).

3.2.5. Influence of hinge elongation on protein stability and coactivator binding

The stability of elongated proteins LBD309, INT294, INT262, and HR252 was investigated by circular dichroism (CD) spectroscopy. Only minor differences were observed for the melting temperatures of the four constructs (Tm: HR252 < LBD309 < INT294 < INT262, Figure 3-9A). In agreement with this finding, the spectrum of the mean residue ellipticity for the four constructs afforded a similar pattern with similar 222 nm to 208 nm ratios (>1.0, Figure 3-9B). However, the results also clearly showed differences in mean residue ellipticity (MRE), despite all samples having been measured at 7.5 μM concentrations. Interestingly, the lowest MRE were discovered for the longer protein constructs, while LBD309 had the highest (Figure 3-9B). The only explanation for this occurrence was due to precipitation during the preparation procedure. Thus, it seems that large extensions towards the HR leads to a higher tendency of the protein to precipitate. However, unfolding studies did not confirm this trend in the form of varying temperature-dependent stabilities for the different protein constructs (Figure 3-9).
The ability of the elongated constructs to recruit coactivators (evaluated via interaction with fluorescein-labeled SRC1-Box2) was tested in a polarization assay. While the extension from LBD309 via INT294 through to INT272 did not influence the binding properties ($K_d = 0.36$-$0.41$ µM) the 10-residue-step from INT272 to INT262 had a significant impact: namely, an increase in the dissociation constant $K_d$ by one order of magnitude ($K_d = 0.41$ µM → $5.5$ µM). Although binding affinity improved for the HR252 construct ($K_d = 1.2$ µM) it was still significantly lower than the three shorter constructs (Figure 3-10).
In summary, the extension of the LBD up to half way into the HR (\textsuperscript{272}HR-LBD\textsuperscript{553}: \textbf{INT272, INT294, LBD309}) did not significantly influence receptor stability and its binding affinity to LXXLL motif. However, further extension into the HR (\textsuperscript{252}HR-LBD\textsuperscript{553}) lowered the ability of the protein to recruit coactivator peptides. Although the CD spectrum (mean residue ellipticity) of the long proteins HR\textsuperscript{252} and \textbf{INT262} suggested lower receptor stability (through an increased propensity to precipitate) due to lowered concentration compared to \textbf{INT294} and LBD\textsuperscript{309}, temperature-induced denaturation studies did not confirm this. Interestingly, studies on peptides based on the sequence of the HR revealed a higher propensity to form an \(\alpha\)-helix in the region close to the LBD (\textsuperscript{272}HR\textsuperscript{310}) compared to the region close to the DBD (\textsuperscript{252}HR\textsuperscript{271}), including Lys\textsuperscript{266} and Lys\textsuperscript{268} (chapter 2 and unpublished data in the group). In both cases, however, the \(\alpha\)-helical character could be improved upon the introduction of PTMs. Thus, it seems to be that the ER\(\alpha\)-HR can be divided into two regions, namely from residue 310 to 372 (\textsuperscript{272}HR\textsuperscript{310}) located close to the LBD, featuring moderate \(\alpha\)-helicity and a conserved RXKK motif, and from 271 to 252 (\textsuperscript{252}HR\textsuperscript{271}) close to the DBD, being less helical and with possible PP\(\alpha\) helix and two posttranslational modified lysines (chapter 2). Interestingly, in a reporter-gene-assay-based study, acetylation of Lys\textsuperscript{266} and Lys\textsuperscript{268} resulted in the stimulation of ligand-dependent activity, whereas acetylation of Lys\textsuperscript{302} and Lys\textsuperscript{303} led to a diminishment of agonist response.\textsuperscript{63} This again supports the hypothesis of the presence of two major regions in the HR with different – or even opposing – outcomes upon the same PTM, such as lysine acetylation.

3.3. Conclusion

For EPL, both ligation partners were successfully synthesized. The recombinantly expressed protein construct C-\textbf{ER}\(\alpha\)\textsuperscript{assoc} could be made. However, its synthesis required an additional ring opening step to enable access to the protein with the N\textsuperscript{'}-Cys. The peptides (KK-Nbz, K\textsuperscript{*}K-Nbz, KK\textsuperscript{*}-Nbz, K\textsuperscript{*}K\textsuperscript{*}-Nbz) possessing both site-specific acetylations of lysines and a C\textsuperscript{'}-terminal thioester were synthesized by solid phase peptide synthesis. While test ligations of the peptides to a model protein could be optimized as far as 30% ligation efficiency, transfer of the same ligation conditions to ER\(\alpha\) was not successful. Possible peptide hydrolysis and mainly protein instability (manifest as an increased tendency for the protein to precipitate) were the major issues in this case. The studies concerning the enlargement of the ER\(\alpha\)-LBD towards the HR revealed that expression as far as position 272 in the presence of E2 gave good expression yields (\textbf{INT272, INT294, LBD309}). The melting point, reflecting the thermal stability of the protein, was not influenced upon further extension. However, the expression yields diminished and binding studies with coactivator peptides identified a significant decrease in affinity (\textbf{INT252, INT262}). These results, together with studies on peptides based on the HR sequence, suggested the presence of two distinct regions of the HR with different structure and function upon posttranslational modification.
3.4. Experimental

General
If not stated otherwise the following contents, descriptions, protocols, and distributer are valid: chemicals were ordered from SIGMA-ALDRICH. Proteins were handled at 4 °C. Standard culturing medium is lysogeny broth (LB) medium (10 g peptone, 10 g NaCl, 5 g yeast, 11 water, autoclaved). SRCI-Box2 or short SIB2 is the 20 amino acid long sequence of the nuclear receptor coactivator 1 (Ncoa1.692-71) with the sequence SLTERHKILHRLQEGPSD, including a LxxLL motif.

Protein Expression & Purification

Table 3-3 | Used constructs for protein expression

<table>
<thead>
<tr>
<th>construct</th>
<th>vector</th>
<th>tag</th>
<th>protein</th>
<th>origin</th>
</tr>
</thead>
<tbody>
<tr>
<td>ERα-full-length</td>
<td>pCDNA3</td>
<td>-</td>
<td>hERα-1-395</td>
<td>GIFT FROM ROB MICHALIDES</td>
</tr>
<tr>
<td>ERα-LBD309</td>
<td>pET44a</td>
<td>FXa-His (C-ter.)</td>
<td>hERα-LBD309-555</td>
<td>CLONING FROM ERα-full-length</td>
</tr>
<tr>
<td>ERα-INT294</td>
<td>pET44a</td>
<td>FXa-His (C-ter.)</td>
<td>hERα-HR-LBD294-555</td>
<td>CLONING FROM ERα-full-length</td>
</tr>
<tr>
<td>ERα-INT272</td>
<td>pET44a</td>
<td>FXa-His (C-ter.)</td>
<td>hERα-HR-LBD272-555</td>
<td>CLONING FROM ERα-full-length</td>
</tr>
<tr>
<td>ERα-INT262</td>
<td>pET44a</td>
<td>FXa-His (C-ter.)</td>
<td>hERα-HR-LBD262-555</td>
<td>CLONING FROM ERα-full-length</td>
</tr>
<tr>
<td>ERα-HR252</td>
<td>pET44a</td>
<td>FXa-His (C-ter.)</td>
<td>hERα-HR-LBD252-555</td>
<td>CLONING FROM ERα-full-length</td>
</tr>
</tbody>
</table>

Plasmids with the desired protein construct (Table 3-3) were transformed into E. coli BL21 cells. Emerging colonies were cultured over night in 25 ml LB medium with 100 μg/ml ampicillin at 37 °C. This preculture was added to 21 phosphate buffered, content-rich terrific broth (TB) medium (12 g peptone, 24 g yeast, 4 ml glycerol, 0.17 M KH2PO4, 0.72 M K2HPO4, 11 water, autoclaved) with ampicillin and incubated at 37 °C until an optical density at 600 nm (OD600) of 0.6 - 1.0. Addition of 100 μM Isopropyl β-D-1-thiogalactopyranoside (IPTG) induced expression of the protein of interest. For stabilization 10 μM of ligand (ER, estradiol) was added. After incubation for 20 h at 15 °C the cells were harvested by centrifugation (15 min at 7,000 rpm) and the resulting pellet was stored at -80 °C until further usage.

If not stated otherwise, the protein purification was performed at 4 °C. The pellet was resuspended in lysis buffer (LyB-1: 1x PBS tablet, CALBIOCHEM, 370 mM NaCl, 10% (v/v) glycerol, 0.1 mM TCEP, 1 mM PMSF, 1 μg/ml DNAsel, 10 μM ligand, pH 8.0, His6-tag: additional 40 mM imidazole) and lysed with the Emulsi Flex-C3 homogenizer (2 passes of 150,000 kPa, AVESTIN INC.). After another centrifugation step (40 min, 20,000 rpm) the supernatant was provided on the respective affinity chromatography column (His6-tag: His-Bind Resin, 3 ml, NOVAGEN), washed with wash buffer (1x PBS tablet, CALBIOCHEM, 370 mM NaCl, 40 mM imidazole, 10% (v/v) glycerol, 0.1 mM TCEP, 10 μM ligand, pH 8.0) and eluted with competing molecules (Wb-1 plus 100 mM / 250 mM / 500 mM imidazole). Fractions were analyzed with SDS-PAGE and optional with LC-MS.

The N’-terminal thiazolidine ring that was formed between the cysteine side chain and the free N’-terminus of the protein for EPL (+A11-21) could be opened with treatment of 125 mM methoxyamine in 500 mM acetate buffer at 4 °C for 24 h (pH 5.5). LC-MS confirmed the successful reaction.
CHAPTER THREE

Solid Phase Peptide Synthesis & Purification\textsuperscript{35}

The synthesis of all peptides was performed on solid support using the Fmoc strategy on a Prelude peptide synthesizer (PROTEIN TECHNOLOGIES INC.). If not stated otherwise, 200 µmol (1 eq.) of Dawson Dbz AM resin (0.43 mmol/g, 100-200 mesh, NOVABIOCHEM) was swollen for 30 min in 5 ml \(N\)-methylpyrrolidone (NMP). Washing was performed with NMP or dichloromethane (DCM), deprotection of the amino acids (AA, 0.2 M solution in NMP) was performed with piperidine (Dep., 20\% (v/v) solution in NMP), activated with 2-(7-Aza-1H-benzotriazole-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate (\textit{Act}A\textsubscript{AAA}, HATU, 0.4 M solution in NMP) or 2-(6-Chloro-1H-benzotriazole-1-yl)-1,1,3,3-tetramethy laminium hexafluorophosphate (\textit{Act}., HCTU, 0.4 M solution in NMP) and alkalized with \(N,N\)-diisopropylethylamine (\textit{Base}, DIPEA, 1.6 M in NMP). The loading of the first amino acid was performed in a four times coupling: 1 ml Dep. (5 min), 3x[2 ml NMP, 2 ml AA, 1 ml \textit{Act}A\textsubscript{AAA}, 1 ml \textit{Base}, 20 min], 1 ml NMP, 2 ml DCM. The following residues are double-coupled with a subsequent capping step (\textit{Cap}., 20\% (v/v) pyridine, 20\% (v/v) acetic anhydride in NMP): 6 ml NMP, 4 ml Dep. (5 min), 2x[6 ml NMP, 4 ml AA, 2 ml \textit{Act}., 2 ml \textit{Base}, 20 min], 6 ml NMP, 3 ml \textit{Cap}. (5 min), 3 ml NMP. The last amino acid was coupled BOC-protected. After finishing the desired peptide sequence, the resin was washed 3 times alternately with dimethylformamid (DMF) and dichloromethane (DCM) and the diaminobenzoyl (Dbz) group was activated 1 h with 0.5 mmol \(p\)-nitrophenyl chloroformate (\(p\)-NPCF) resulting in the corresponding \(N\)-acyl benzimidazolinone (Nbz). After washing with DMF, 0.5 M DIPEA in DMF was added for 30 min. After another washing step with DMF and DCM, the side chain protection group and the cleavage of the peptide sequence from the resin were simultaneously achieved by the usage of mixture of trifluoroacetic acid, triisopropylsilane and water (TFA/TIS/H\textsubscript{2}O, 95/2.5/2.5) for 3 h. The free peptide was precipitated by adding it drop-wise to ice cold DEE. A final centrifugation step (2,000 rpm, 10 min) with subsequent washing (ice cold DEE), uptake in water and lyophilization results in the crude peptide. Purification was accomplished by HPLC on a Luna\textsuperscript{®} 5 \textmu m HILIC 200 Å, LC Column 150 x 21.2 mm, AXIA\textsuperscript{TM} Packed (PHENOMENEX). Acetonitrile (ACN) was used as a apolar phase, adding different amounts of an polar ammonium formate (10 mM, pH3.5) phase. A linear gradient (20 ml/min) was optimized for each peptide to a 10\% range, for instance 60-50\% ACN. The purity of the peptide was determine at an analytical LC-MS using Luna\textsuperscript{®} 5 \textmu m HILIC 200 Å, LC Column 100 x 2 mm (PHENOMENEX). Finally, the peptides were lyophilized and stored at -80 °C.
Fluorescence Polarization Assay\textsuperscript{64,65}

Fluorescence polarization (FP) experiments were measured in black 384-well-plates (\textsc{Perkin Elmer} 384 F) on a Safire\textsuperscript{2} (\textsc{TECAN}) plate reader. The final volume of one well was 30 \(\mu\)l containing constant concentrations of ligand (5 \(\mu\)M) and Fluorescein-labeled peptide (0.1 \(\mu\)M, ER: NCoA-1/SRC1\textsubscript{153-703}, FL-LTERHKIHRLLQEGSPSD). The protein was sequentially diluted (24 steps) in TR FRET coregulator buffer E (\textsc{Invitrogen}). Briefly, in 96-well-plates dilution series were prepared in 55 \(\mu\)l and filled up with to 110 \(\mu\)l with a 2X master mix solution containing all non-varying components. Finally three times 30 \(\mu\)l was transferred to 384-well-plates, centrifuged (1,000 rpm, 2 min) and incubated for 1 h at 4 \(^\circ\)C. Plates were measured 50 times at 30 \(^\circ\)C (excitation 470 nm, emission 519 nm). Polarization \(y\) (in \([\text{mP}]\)) was plotted against the concentration \(x\) of either the protein (FP), whereby each data point represents an average of 3 experiments. The dissociation constant \(K_d\) of the protein-peptide complex (FP) was calculated using equation 1:

\[
y = A_1 + \frac{A_2 - A_1}{1 + 10^{(\log(x_0) - x)p}} \quad K_d = 10^{\log(x_0)} \tag{1}
\]

where \(A_1\) is the bottom asymptote, \(A_2\) is the top asymptote, and \(p\) the hill slope (steepness of the curve).

Circular Dichroism spectroscopy\textsuperscript{66,67}

Far-UV Circular Dichroism (CD) spectroscopy measurements were performed under a constant nitrogen flow at 20 \(^\circ\)C using 7.5 \(\mu\)M of the protein in 10 mM phosphate buffer (pH 7.5), if not stated otherwise. As a baseline a protein-free buffer was measured under identical conditions. The spectrum was recorded from 250 nm to 185 nm using a \textsc{Jasco}-815 spectrometer. Quartz cuvettes with path lengths of 1 mm (\textsc{Hellma Analytic}s) were employed with a data pitch of 0.5 in a continuous mode, a scan speed of 20 nm/min with a response time of 2 s and a bandwidth of 0.5 nm. The graphs are representing an average of five scans. The observed ellipticity \(\theta_x\) (degrees, in [mdeg]) was converted to mean residue ellipticity \(\theta_{mrw,\lambda}\) (in [deg \(\text{cm}^2/\text{dmol}\)]) using \textsc{Dichroweb}.\textsuperscript{68} Information about helical propensity in semi-quantitative manner could be shown by the ratio of the characteristic wavelengths at 208 nm and 222 nm: \(\theta_{mrw,222\text{nm}}/\theta_{mrw,208\text{nm}}\).

For the unfolding studies the mean residue ellipticity at 222 nm was used to detect unfolding of the protein. The bottom asymptote at low temperature was set as folded protein, meaning 0\% unfolding. By contrast, the top asymptote at high temperatures – no further change of mean residue ellipticity at 222 nm upon increase of temperature – was assumed as unfolded protein, thus 100\% unfolding. Plotting temperature \(x\) against unfolding \(y\) led to corresponding melting temperature \(Tm\) of each protein.

Expressed Protein Ligation

All optimization were performed with recombinantly expressed CYFP (provided by Ralph Bosmann) and KK*-Nbz peptide in 0.2 M phosphate buffer, including 10\% (v/v) glycerol at pH 7.0 at room temperature for 12h and 24h, respectively. Protein-, peptide-, thiol additive-, and TCEP concentration can be seen in Table 3-1. The ER ligation trials were performed with \textsc{MPPA} and \textsc{MESNA} conditions (Table 3-1, \textsc{MPPA/MESNA}).
3.5. References

Studying the Estrogen Receptor Hinge Region Using Expressed Protein Ligation

CHAPTER FOUR

Screening the Estrogen Receptor Coactivator Interaction

Abstract. Ribosome display was used to effectively screen the surface of the estrogen receptor (ER) ligand binding domain (LBD) for novel natural peptide binders. Whereas earlier rounds of ribosomal enrichment witnessed the expected emergence of the hallmark LXXLL motif subsequent rounds led to the identification of a more mature PXLXXLLXXP consensus, which could be validated by both biochemical and cellular methods. Molecular modeling (MD) and X-ray crystallography studies clearly defined a specific role for the flanking prolines as helix breakers, which prime the helix length for optimal interaction with the surface charge clamp. Furthermore, the conformational constraints imposed by the prolines on adjacent immediate flanking amino acid residues are believed to determine receptor subtype selectivity and binding affinity through the precise orientation of side chain functionality at the ER surface. These findings resulted in potent peptide inhibitors of the ER-coactivator interaction based alone on natural sequences and provide a structural rationale for the function of prolines in natural coactivator proteins. Finally, this work represents a fundamental re-evaluation of the NR-coactivator interaction and thus sets new, minimal, structural parameters based on linear sequences of proteinogenic amino acids. Moreover it also provides insights for peptide-derived tools and more druggable peptide agents.

4.1. Introduction

The standard approach to treat diseases caused by the misregulation of nuclear receptors (NR) is the use of small lipophilic drug molecules which modulate NR-mediated gene transcription at the ligand binding pocket (LBP) of the ligand binding domain (LBD). Recently, in the case of the estrogen receptor (ER; α, NR3A1; β, NR3A2), compounds with reduced toxicity have been developed, namely selective estrogen receptor modulators (SERMs). One example is tamoxifen which achieves improved cell and tissue selectivity through a more distinct recruitment of coactivator proteins. While clearly of significant benefit, classical approaches targeting the LBP suffer from a number of drawbacks, including endocrine resistance, and a general lack of selectivity at the level of coactivator recruitment, resulting in an urgent need for alternative approaches to modulate ER-mediated gene transcription.

A promising alternative strategy is direct inhibition of the protein-protein interaction between ER and its coactivators. In general, agonist binding - such as the endogenous ligand estradiol (E2) in case of ER - at the LBP of the LBD induces a conformational change, most prominent in the helix 12 (H12) region, which favors binding of an α-helical LXXLL consensus motif present in various NR-coactivator. A defining feature of this interaction is the charge clamp (with a positive charged lysine and negative charged glutamic acid), which is formed at the interface between H12 and other parts of the activation function 2 (AF-2). This clamp defines the maximal length of an α-helical peptide for optimal coactivator binding and offers therefore a unique opportunity for the development of selective inhibitors. Seminal work involving the screening of large LXXLL-derived peptide libraries by phage display led to the identification of potent and selective ER-coactivator binding inhibitors (CBI). However, issues of poor cell permeability and low metabolic stability typically complicate the translation of peptide hits into useful drug therapies. This has motivated to recent efforts to develop small lipophilic drug-like molecules which target the charge clamp or mimic the leucine side-chains of the coactivator α-helix, via computational design or using high throughput screening (HTS). Modified peptide inhibitors have also been investigated with improved binding affinity, ER subtype selectivity and proteolytic stability.

In recent years peptide drug discovery has matured and a number of different strategies to stabilize peptides have emerged, including macrocyclization via disulfide bond formation, macrolactamization, hydrocarbon stapling, or the incorporation of hydrogen bonding surrogates, which preorganize the α-helix through stabilization of its structure in the surface bound state. Nonetheless, the drawbacks of such approaches are the necessity for synthetic operations to upgrade amino acid building blocks to facilitate macrocyclization and the unpredictable effects that such modifications might have on binding. Furthermore, advances in this field have built only on prior knowledge of natural coactivators or peptide
sequences derived from phage display screening, and are therefore restricted to the inherent diversity of the systems at hand. In light of this, we wondered whether a consensus had yet to be reached on the minimal structural requirements for potent inhibition of the ER-coactivator interaction using only natural amino acids. The acquisition of such knowledge would greatly depend on the evaluation of vast, more structurally diverse peptide libraries than had previously been tested. Indeed the NR-coactivator interaction was first evaluated using phage display, and it was thanks to the efficiency of this method that the LXXLL motif was discovered as a preferential binder to ER and other NR surfaces. Nonetheless, the usage of an alternative screening approach, which samples from a more diverse peptide library, might reveal optimized extended LXXLL consensus motifs or even alternative sequences with improved binding profiles. Next to the well established phage display, also the yeast two-hybrid system, plasmid and cell surface display are available. But since they all are limited by their intrinsic transformation efficiency, the complexity of these libraries are at most $10^9$ unique peptide entities. Cell-free in vitro methodologies, including ribosome display (Figure 4-1), however, are resolving this limitation resulting in a random peptide library with high complexity and up to $10^{14}$ independent members. This advantage of that technique, first described by Mattheakis et al. and further developed by Plückthun et al., could enable a closer examination of the flanking regions of the LXXLL region than had previously been possible.
4.2. Ribosome display screening and preliminary validation

Initial rounds of ribosomal screening resulted in the emergence of the LXXLL consensus motif for both ERα and ERβ subtypes. Further rounds of enrichment, however, resulted in a more refined PXLXXLLXXP motif, which was particularly prominent in the case of ERβ (Figure 4-2).
<table>
<thead>
<tr>
<th>Round 4</th>
<th>ER ALPHA</th>
<th></th>
<th>ER BETA</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Round 6</th>
<th>ER ALPHA</th>
<th></th>
<th>ER BETA</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Round 8</th>
<th>ER ALPHA</th>
<th></th>
<th>ER BETA</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Figure 4-2 | Cluster analysis of different rounds of ribosome display screening**

Height of letters is representative of the abundance of an amino acid at certain positions of the amino acid sequence after ribosome display screening against the LBD of ERα (left) and ERβ (right), respectively. Alignment and position labeling was made relative to the emerging LXXLL-motifs.38

Hit sequences were selected for a preliminary validation with SNAP-tag-based cellular competitive inhibition studies. These binding tests performed in mammalian U2OS cell lines identified proline-based peptide sequences 4 (PXLXXLLXXP), 5 (PXLXXHL) and 10 (PXLXXLLXXXP) as the most efficacious binders compared with the natural SRC1-Box2 sequence (1).
Figure 4-3 | Validation of peptide sequences identified by ribosome display screening

(A) LXXLL aligned sequences of naturally derived NCoA1 Box 2 sequence, 1, and peptide sequences 2-10 identified after round eight of ribosome display. Comparison of binding properties of peptides 1-10 by $K_d$ values (mean ± s.e.m.) obtained as mEGFP-peptide fusions by a binding fluorescence polarization assay (B) or testing their luciferase activity (mean ± s.e.m.) in a cellular competitive inhibition assay against ERα (C) or ERβ (D), respectively.

Based on both the cluster analysis data (Figure 4-2) and results from the preliminary cellular screening (Figure 4-3), it was hypothesized that the presence of the prolines might be playing a significant role in determining the improved binding efficacy of sequences 4, 5, and 10 compared with the natural sequence (1). Therefore, 4, 5, and 10 were synthesized by solid phase peptide synthesis (SPPS) together with a series of rationally designed truncated analogues (Table 4-1) in order to probe the importance of the prolines in subsequent biochemical and structural studies.
4.3. Exploration of the PXLXXLLXXP motif

All peptides were first measured in a competition fluorescence polarization assay using fluorescein-labeled 1 as the competing peptide (Figure 4-4A). The main focus during further investigations was on the subtype ER\(\beta\), where the emergence of the prolines was discovered to be most prominent (Figure 4-2B). Overall, the peptides produced classical sigmoidal curves and behaved as competitive inhibitors with \(K_i\) values in the range 25 nM to 75 \(\mu\)M (Figure 4-4B-D, Table 4-1).

![Figure 4-4](attachment:fig4_4.png)

**Figure 4-4 | Fluorescence Depolarization Assay**
(A) General principle of the fluorescence depolarization assay based on the ER-SRC1-Box2 interaction. (B) Normalized sigmoid curves from the evaluation of 22 against other NRs. Fluorescein-labeled competing peptides: ER\(\beta\), SRC1-Box2 (1); AR, FLETT-1; RXR\(\alpha\), D22. Reference peptides: ER\(\beta\), 1; AR, AR-NTD (26); RXR\(\alpha\), phage display hit against ER\(\alpha\) and good binding properties for RXR\(\alpha\) (27). (C) Selected (1, 4, 5, 13, 22) sigmoid curves resulting from ER\(\alpha\) binding studies. (D) Selected (1, 4, 5, 13, 22) sigmoid curves resulting from ER\(\beta\) binding studies. IC\(_{50}\) and \(K_i\) values for both receptor subtypes in Table 4-1.

The ribosome sequences 4, 5, and 10, were found to bind more potently to ER\(\beta\) (\(K_i < 50 \text{ nM}\)) than the natural 1 sequence (\(K_i = 122 \text{ nM}\)). Truncation of peptide sequences 4 and 10 resulted in a decrease in binding affinity (Table 4-1). N’-terminal truncation of 19-mer 4 yielded 15-mer 11, which was 2.5 times less active than the parental peptide. By comparison,
Chapter Four

N’-terminal truncation of 19-mer 10 gave 15-mer 12, which was significantly less potent than 10. Further truncation of 12 at the C’-terminus, yielded 12-mer 13, which was similarly (in)active as 12 (Ki = 3.6 μM). This data would suggest that, at least for 4 and 10, the far-flanking regions of the LXXLL motif make an important contribution to binding, and that the presence of the proline residues alone is not sufficient for high potency binding.

Table 4-1 | Binding affinities of synthesized peptides

<table>
<thead>
<tr>
<th>peptide sequence</th>
<th>name</th>
<th>estrogen receptor β</th>
<th>estrogen receptor α</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>name</td>
<td>IC₅₀ ± s.e.m.b [μM]</td>
<td>Kᵢ (±)b [μM]</td>
</tr>
<tr>
<td>LTERHKILHRLLQEGSPSD</td>
<td>1</td>
<td>0.470 ± 0.059</td>
<td>0.122 ± 0.039</td>
</tr>
<tr>
<td>LTARSPILLRLLLQSPSD</td>
<td>4</td>
<td>0.171 ± 0.009</td>
<td>0.186 ± 0.007</td>
</tr>
<tr>
<td>LTARHPLLRLQLNQSPSD</td>
<td>5</td>
<td>0.184 ± 0.044</td>
<td>0.251 ± 0.009</td>
</tr>
<tr>
<td>LTARHPLLRLHHPSPSD</td>
<td>10</td>
<td>0.279 ± 0.028</td>
<td>0.756 ± 0.022</td>
</tr>
<tr>
<td>SPLITRLRLQQSPSD</td>
<td>12</td>
<td>0.466 ± 0.040</td>
<td>0.119 ± 0.026</td>
</tr>
<tr>
<td>HPLMRLLHPSPSD</td>
<td>11</td>
<td>7.150 ± 0.850</td>
<td>4.622 ± 0.245</td>
</tr>
<tr>
<td>HPLMRLLHHP</td>
<td>13</td>
<td>5.381 ± 0.611</td>
<td>3.640 ± 0.210</td>
</tr>
<tr>
<td>AcEPIHRLLQKP</td>
<td>14</td>
<td>38.35 ± 17.92</td>
<td>25.66 ± 12.09</td>
</tr>
<tr>
<td>AcSPIHRLLQEP</td>
<td>15</td>
<td>n.b. (b)</td>
<td>n.b. (b)</td>
</tr>
<tr>
<td>AcRhKPLHRLQEP</td>
<td>16</td>
<td>98.24 ± 12.52</td>
<td>66.04 ± 8.446</td>
</tr>
<tr>
<td>AcRHPILLRLQDEG</td>
<td>17</td>
<td>2.832 ± 0.118</td>
<td>1.711 ± 0.080</td>
</tr>
<tr>
<td>AcRHPILLRLQEPS</td>
<td>18</td>
<td>112.0 ± 17.64</td>
<td>75.35 ± 11.89</td>
</tr>
<tr>
<td>RHPKLHRRLQEP</td>
<td>19</td>
<td>1.460 ± 0.098</td>
<td>0.787 ± 0.066</td>
</tr>
<tr>
<td>AcPIHRLLQEP</td>
<td>20</td>
<td>0.680 ± 0.102</td>
<td>0.262 ± 0.068</td>
</tr>
<tr>
<td>AcChPILLRLQEP</td>
<td>21</td>
<td>0.395 ± 0.020</td>
<td>0.072 ± 0.013</td>
</tr>
<tr>
<td>AcHPLMRLLLSP</td>
<td>22</td>
<td>0.132 ± 0.021c</td>
<td>0.025 ± 0.015c</td>
</tr>
<tr>
<td>AcCMLRLLLS</td>
<td>28</td>
<td>0.301 ± 0.050</td>
<td>&lt; 0.050b</td>
</tr>
<tr>
<td>HPLLRLLPSP</td>
<td>23</td>
<td>0.289 ± 0.017</td>
<td>&lt; 0.050b</td>
</tr>
<tr>
<td>AcIKHLHRLLQEG</td>
<td>24</td>
<td>0.591 ± 0.120</td>
<td>0.203 ± 0.080</td>
</tr>
<tr>
<td>AcKiHLHRLLQEG</td>
<td>25</td>
<td>0.695 ± 0.144</td>
<td>0.272 ± 0.096</td>
</tr>
<tr>
<td>AcHPLMRLLLSP</td>
<td>22</td>
<td>against AR-LBD</td>
<td>16.61 ± 7.489</td>
</tr>
<tr>
<td>AcHPLMRLLLSP</td>
<td>22</td>
<td>against ARα-LBD</td>
<td>0.722 ± 0.196</td>
</tr>
</tbody>
</table>

LXXL motif highlighted in grey, flanking prolines underlined and bold; a, no exact value determinable due to assay limitation, b, standard error of the mean; c, non-standard conditions: 100 nM protein, d, no binding, e not determined; last two rows against AR/ARα-LBD, 22 measuring in presence of AR and RXRα-LBD, respectively, instead of ERα- or ERβ-LBD.

Interestingly though, insertion of the two proline residues into the 11-mer natural 1 sequence - at the same -2 and +3 positions relative to the LXXLL motif as observed in 4 and 10 (Table 4-1, 21) - resulted in a more potent binder (Ki = 72 nM) than the natural 11-mer sequence, 24 (Ki = 203 nM). Furthermore, this short proline-based peptide was more potent than the natural 19-mer sequence 1 (Ki = 122 nM). Interestingly, an 11-mer peptide consisting of the most frequent amino acids derived from the cluster analysis for ERβ (Figure 4-2 & Table 4-1, 22) resulted in high affinity binder (Ki = 25 nM). Modifications to negatively charged or polar residues adjacent to the proline residues (Table 1, 14 and 15) caused the peptide binding affinity to diminish. Subsequent investigations into the importance of the positioning of prolines relative to the LXXLL motif suggested that the proline at -2 was
optimally placed (Table 4-1, 16 and 18), while hinting a greater flexibility for the proline at +3 in the C’-terminal region (Table 4-1, 17 and 19). Taken together, these data suggest that the prolines make a significant and important contribution to peptide binding at the ER surface.

4.4. Structural analysis of potent binders

Circular dichroism (CD) measurements were performed to determine the α-helicity of the peptides (Table 4-2). Whereas in pure phosphate buffer all peptides measured were random coil, the addition of 30% (v/v) trifluoroethanol (TFE) – reported to stabilize intra-molecular hydrogen bonds and therefore mimic the environment at the protein surface – resulted in an increase in the percentage α-helicity for most of the peptides measured. This strongly suggests that helix nucleation and stabilization depend on binding at the NR surface.

Table 4-2 | Circular dichroism measurements of selected peptides

<table>
<thead>
<tr>
<th>peptide sequence</th>
<th>name</th>
<th>θ_{222}/θ_{310}</th>
<th>CD spectra</th>
</tr>
</thead>
</table>
| LTERHKILHRLQEGSPSD | 1^a | 0.85 | ![CD spectra image](image)
| LTARSPPLLRLQPSDP | 4^a | 0.56 |
| LTARSPPLLRLQPSDP | 4^- | 0.25^a |
| LTARHPPLLRLQPSDP | 5^a | 0.79 |
| LTARHPPLLRLQPSDP | 10^a | 0.75 |
| SPLTRLLQLPSDP | 12^a | 0.60 |
| HPLLRLLHPS | 11^a | 0.70 |
| AcRHPILHRLLQPEG | 17^a | 0.72 |
| AcRHPILHRLLQAPS | 18^a | 0.59 |
| RHKILHRLQPE | 19^a | 0.62 |
| AcPILHRLLQEP | 20^a | 0.69 |
| AcPILHRLLQEP | 21^a | 0.95 |
| AcPILHRLLQEP | 22^a | 0.86 |

CD spectroscopy measurements were performed under a constant nitrogen flow at 20 °C using 50 μM of the peptide in phosphate buffer (b) or in additional 30% (v/v) TFE (a). Information about helical propensity in semi-quantitative manner could be shown by measuring the ratio of the characteristic wavelengths at 208 nm and 222 nm: θ_{222}/θ_{310}.

This view was reinforced by data from molecular modeling (MD) studies, which provided the preferred folding of the peptides (Figure 4-5). In this case, the role of the flanking prolines as helix breaker, and by that their ability to determine the precise helix length, was clearly identified. Introduction of prolines in the natural binder 1, for instance, resulted in a reduction in the length of helical structure to the distance between the proline residues (Figure 4-5B). Furthermore, they predicted more influence concerning α-helicity in the region of the LXXLL motif upon N’-terminal truncation in the case of 4 (resulting in 11) compared to 10 (resulting in 12). Moreover, these MD simulations provide strong evidence that prolines apply significant torsional limitations on the peptide backbone (most notably proline -2).
To provide further structural insight into the relative significance of the flanking prolines within the complex, the X-ray structures of peptides 5 and 13 co-crystallized with the LBD of ERβ in the presence of the natural ligand 17β-estradiol (E2) were solved (Figure 4-6, 4-9 and 4-12). In the case of peptide 13, both prolines (located at -2 and +3 flanking positions) are situated above the charge clamp residues glutamic acid 493 (Glu493) and lysine 314 (Lys314) (Figure 4-6), respectively.
An overlay of the structures for 13 and 1 (PDB 3OLS) - both harboring a histidine residue at the -3 position (His-3) - revealed that for 13, the proline residue at position -2 (Pro-2) assists in directing the histidine side-chain residue toward glutamic acid 493 (Glu493) at ER surface, thereby enabling an additional stabilizing hydrogen-bonding interaction (next to Glu332-His-3). Similar structural insights are observed in the corresponding co-crystal structure of 5, which contains Pro-2. Moreover, atomistically detailed dynamics simulations showed these additional hydrogen bonds are extraordinary stable, along the 6-ns-simulation. This suggests that the proline with the histidine in front (HPX motif) might be stabilizing and might lead therefore to higher affinity binding. Indeed, scission of the histidine from 21 to 20, resulted in an approximate 3-fold reduction in binding affinity (Table 4-2, \( K_i = 262 \) nM). Additionally, cellular data highlighted the importance of the prolines resulting in clear enhanced binding in the case of ER\( \beta \) (Figure 4-7a). Mutational studies identified the proline-mediated hydrogen bond between the prior histidine and the glutamic acid (ER\( \alpha \): E542; ER\( \beta \): E493) in the charge clamp being essential, since charge clamp mutation to alanine leads to loss of activity (Figure 4-7b).
Figure 4-7 | Influence of the prolines on ER binding in the M2H assay
Mammalian two-hybrid studies on peptides 1, 21 and 22 comparing normalized luciferase activity (mean ± s.e.m.) for ERα vs. ERβ. The SRC1-Box2 mutants with one proline residue at either the -2 (I*) or +3 positions (I**) were also included: (A) with and without 17β-estradiol (E2), and (B) wild-type (wt) vs. ER charge clamp mutants E378A and E542A vs. AF-2 mutants E332A and E493A. (n=3)

The biochemical, molecular dynamics simulations and structural data all combine to suggest that peptides conforming to a PXXLXXLXXP consensus bind to ER more potently than conventional natural sequences. In this case, the prolines have evolved to function in two ways: firstly as a ‘molecular ruler’, which primes the helix length to enable an optimal fit between the two charge clamp residues, Glu493 and Lys314, and secondly to nucleate and stabilize α-helix formation through additional conformational constraints on adjacent amino acids.41

4.5. Discussion

Functional LXXLL motifs are not exclusive to NR coactivators, since they were also identified in several transcription factors, including the calcium response element binding protein (CBP), the acetyltransferase p300, and other mediator subunits.42 Indeed, nature makes frequent use of leucine rich α-helical structures to perform a variety of functions, such as the HIV-1 accessory protein Vpr,43 which is important for viral production, and the well characterized leucine zipper proteins such as c-jun/c-fos, which are necessary components of transcription.44 By contrast, the α-helix averse proline is frequently located at the termini of α-helices in water soluble proteins,45 while in contrast, α-helical transmembrane proteins such as GPCRs frequently make use of prolines to introduce kinks or perturbations into long helical structures, which engender functionality through increased conformational flexibility.46 According to Zimm-Bragg/Lifson-Roig helix-coil transition theory, whereas leucine functions both to nucleate helix formation and propagate helix length, proline in contrast serves as helix nucleator and terminator.47,48 Furthermore, in proteins, proline is thought to form capping motifs at the N’- and C’-terminus of α helices, respectively.49,50
which contributes towards helix stability and the specificity of protein folding through the correct positioning of adjacent side-chain residues. In this present work, cluster analysis of the ribosome hit sequences, however, could not identify any such known proline motifs; most likely due to the dependence of helix formation on NR surface binding. Indeed, the predominance for lipophilic amino acid residues C’-terminal to Pro-2 is dominated by the lipophilic nature of the NR box. Residues N’-terminal to this Pro-2, however, do not show any characteristic enrichment. The important role of the charge clamp in this unique case can be observed in the circular dichroism (CD) data, where the proline-derived peptides measured strong helicity only under conditions that mimic the hydrophobic environment of the charge clamp interaction (Table 4-2). The molecular modeling parameters used for this work have also been reported to mimic the effect of protein proximity on the peptide secondary structure. Under such conditions, the peptides showed strong helical character, and the prolines were consistently located at the -2 and +3, or corresponding N1 and C’ position (Figure 4-8).

<table>
<thead>
<tr>
<th>N’</th>
<th>N_cap</th>
<th>N1</th>
<th>N2</th>
<th>.....</th>
<th>C1</th>
<th>C_cap</th>
<th>C’</th>
</tr>
</thead>
<tbody>
<tr>
<td>-4</td>
<td>-3</td>
<td>-2</td>
<td>-1</td>
<td>LxxLL</td>
<td>+1</td>
<td>+2</td>
<td>+3</td>
</tr>
</tbody>
</table>

**Figure 4-8 | Designation of positions relative to α-helix or LXXLL motif.**

The top row shows labeling of amino acid positions with respect to the α-helix-formation. Positions Ni to Ci are part of the helix, while N/C’, N/C_cap are in the flanking regions. The second row shows labeling of the amino acid position with respect to the consensus sequence.

Interestingly, position-specific analysis of α-helices in globular proteins show prolines to be highly favored at position N1 and C’, respectively. Although not conforming to any previously identified motifs, the prolines are nonetheless postulated to nucleate helix formation and to stabilize the helix. Evidence for this can be found by careful examination of the crystallography data, where the precise anchoring of the prolines at their respective positions is thought to be critical in this case. The conformational constraints imposed by the prolines on adjacent amino acids results in a reorientation of side-chain residues and backbone carbonyl groups leading to more favorable interactions. At the N’-terminus this results in stabilization through additional interactions at the charge clamp, which can be evidenced by overlaying the co-crystal structures of 5 and 1 bound to ERβ (Figure 4-9). Both sequences possess a His-3 adjacent to the N’-terminal proline. In both peptides, the histidine residue is seen making an H-bonding interaction with Glu332. In the case of peptide 5, however, the presence of the proline in the peptide chain causes movement in the histidine side-chain toward the NR charge clamp resulting in an additional stabilizing H-bond interaction with Glu603. In a number of cases, this additional H-bonding interaction translates into a fine tuning of the binding in favor of greater affinity according to the fluorescence polarization data. For example, insertion of prolines into 1, and subsequent truncation (21)
leads to a statistically significant improvement in binding affinity while removal of the proline residues from 21 (peptide 24) or removal of the histidine residue (peptide 20) both lead to a loss of activity. Cellular data confirm these findings through enhanced activity upon introduction of Pro-2 into 1 (21 or 1*). Furthermore, the importance of the hydrogen bond between the prior histidine and the charge clamp glutamic acid (ERα: E542; ERβ: E493) could be confirmed due to loss of activity upon E → A mutation. Interestingly, the activity of 1 against ERα was reduced by only 50%, while 1 on ERβ and Pro-2-based peptides on both subtypes became completely inactive (Figure 4-7b). On the other hand, the E → A mutation of the rear side glutamic acid residue (ERα: E378; ERβ: E332) at the binding interface resulted in enhanced binding for Pro-2-based peptides, emphasizing the importance of the additional or alternative hydrogen bond with the charge clamp. Surprisingly, the peptide 23, a rationally designed peptide based on the most abundantly occurring amino acid sequence from cluster analysis of the ribosome display data, also featured significantly increased ligand-independent activity (Figure 4-7a).

The role of prolines as α-helix initiator has also been identified in nature. Position-specific analysis on a set of proteins identified a X-Pro-motif (with X often a His) at the beginning of α-helices, in which the proline at the N1 position facilitates the formation of hydrogen bonds involving the side chains of prior residues (Ncap). This directing role is comparable to the Pro-2 found in the most active sequence identified by ribosome display screening, with the difference being that the additional stabilizing hydrogen bond is formed intermolecular with the surface charge clamp and not within the peptide helix (Figure 4-9).

Additionally, this theoretical and experimental data is supported by sequence analysis of NR coactivator proteins, where prolines are rarely found within the LXXLL motif, but more frequently in the flanking regions. For instance, the two NR boxes of TRAP220/MED1 both contain Pro-2 (NPIIILSLLQ) or even HPX (HPMLMNLK), which is also the case for

![Figure 4-9](image-url)
NF-κ-B inhibitor beta (NPILARLLR)\(^{56}\) and NCoA-6/ASC-2 (SPLLVNLLQ).\(^{57}\) The proclivity of NRs to bind proline flanking peptide sequences is reflected in data from early phage display screening. Initially, the LXXLL coactivator motif was thought to function solely as a docking module.\(^{58}\) Later, phage screening of the flanking regions of the LXXLL motif against ER\(\alpha\) indentified three distinct extended recognition motifs, either dependent or independent of the surface charge clamp. Although PXLXXLL was one of those three peptide classes identified, only four analogues were cited in this case (one of which showed interesting selectivity for ER\(\beta\) over ER\(\alpha\)).\(^{56}\) In a separate piece of work, large diverse recombinant peptide libraries were simultaneously screened against ER\(\beta\) and TR\(\beta\). As for ER\(\alpha\) in the previous example, only a few of the ER\(\beta\)-specific hit peptides (three of nineteen) harbored a Pro-2. Somewhat intriguingly, however, is the fact that most of the peptide analogues reported for TR\(\beta\) (16 of 19) feature a PXLXXLL consensus, which is perhaps unsurprising given its natural propensity to bind TRAP220.\(^{27,59}\) These reports, combined with the data presented in this work hints at an evolutionary role played by proline residues for the fine-tuning of coactivator binding through control over helix length.

Despite the enlightening sequence homology data and the groundbreaking phage display work, there has been a lack of structural and biochemical evidence – especially co-crystal structure data – to explain the binding preference of proline flanking coactivator sequences at NR surfaces. Based on the evidence presented in this work, it could be speculated that the tendency for nature to introduce proline residues in the flanking regions of the LXXLL coactivator motif may function as a fine-tuning of binding properties. The resulting improved selectivity or binding affinity might even override other determining factors, by controlling the length of the \(\alpha\)-helix and increasing helix stability. The fact that the PXLXXLLXXP motif emerges early on during the ribosome display screening and is more pronounced based on the final cluster analysis for ER\(\beta\) compared to ER\(\alpha\) (Figure 4-2) supports the hypothesis that receptor preference is a function of the proline residue’s occurrence in the peptide sequence. Investigations of the binding affinity are in line with this, given the fact that the best hits indentified for ER\(\beta\) overpower the SRC1-Box2 interaction for ER\(\beta\), but not for ER\(\alpha\) (Table 4-1 and Figure 4-10). However, other factors such as charge and hydrophobicity are also certain to play a role in this protein-protein interaction. For drug discovery, the tactical employment of flanking prolines may therefore have important consequences for the future development of potent and selective peptide-based inhibitors of the NR-coactivator interaction.
Figure 4-10 | Comparison of $K_i$ values of the proline peptides
Comparison of $K_i$ values (mean ± s.e.m.) for ERα and ERβ determined in a competitive fluorescence polarization assay (values in Table 4-1). The dashed lines and arrows (dark = enhanced binding; light = decreased binding) facilitate comparison of peptide activities with the reference peptide 1.
4.6. Conclusion

In conclusion, a novel use of ribosome display has been described which effectively screens the ER surface for novel peptide inhibitors. In this way, a series of hit peptide sequences were discovered conforming to an advanced PXLXXLXXP consensus motif, which was most pronounced for ERβ. A selection of the hit sequences were prepared as protein fusions and as 19-mer peptides by SPPS and their binding to both ERα and ERβ determined by an initial cellular screen (SNAP-tag competition). According to circular dichroism measurements, while not intrinsically α-helical in pure phosphate buffer, the proline-derived peptides acquired significant helicity in 30% aq. TFE (v/v), suggesting that binding to the ER surface is a prerequisite for α-helix nucleation and stabilization. In this case, molecular modeling data highlighted the role of the prolines in determining the precise helix length, which in the optimal case is believed to match the distance measured between the two surface charge clamp residues Lys314 and Glu493. Furthermore, X-ray co-crystallography data has clearly demonstrated that the proline residues located at positions -2 and +3 (present in the most potent peptide binders) sit precisely above the surface charge clamp at the N1 and C′ positions of the α helix. Here, aside from their role as helix breakers, the conformational constraints imposed by the flanking prolines are postulated to restrict the orientation of H-bonding, carbonyl groups and adjacent amino acid residues, resulting in the improved stabilization of the α-helix, as well as more favorable interactions at the ER surface. These factors combined offer a structural rationale for the improved binding properties of the proline-derived peptides (compared with the natural sequence, SRC1-Box2), which has enabled the design of short (11-mer) and highly potent peptide inhibitors in the low nM-range. The initial higher occurrence of prolines in position -2 and +3 for ERβ compared to ERα is reflected in the ability to overpower SRC1-Box2 binding for ERβ, but not ERα. Therefore, the proline-derived peptide inhibitors reported in this work represent a set of minimal structural parameters for addressing the ER-coactivator interaction in preferential manner, based exclusively on natural amino acids and without recourse to additional chemical modifications. Results from this work have thus laid important foundations for the future development of peptide-derived tools and drug therapies with improved pharmacological profiles.
4.7. Experimental

General
Ribosome display experiments and initial validation experiments were performed by Dr. Hoang D. Nguyen, Dr. Trang T.P. Phan and Monireh Goodarzifard. Peptide truncation studies were performed together with Dr. Matthew F. Burton, Stijn M. Agen and Dr. Lech-Gustaw Milroy. Molecular dynamics simulations were conducted by Dr. Lidia Nieto. Solving the crystal structure was accomplished by Dr. Andrea Schmidt in cooperation with Dr. Rolf Rose and Dr. Christian Ottmann (MPI Dortmund, Germany). Cellular studies were performed by Dr. Ingrid J. de Vries-van Leeuwen.

If not stated otherwise the following contents, descriptions, protocols, and distributor are valid: chemicals were ordered from SIGMA-ALDRICH. Proteins were handled at 4 °C. Standard culturing medium is lysogeny broth (LB) medium (10 g peptone, 10 g NaCl, 5 g yeast, 1 l water, autoclaved). SRCl-Box2 or short 1 is the 20 amino acid long sequence of the nuclear receptor coactivator 1 (Ncoa1692-711) with the sequence SLTERHKILHRLLLQEGPSD, including a LXXLL motif.

Protein Expression & Purification

<table>
<thead>
<tr>
<th>construct</th>
<th>vector</th>
<th>tag</th>
<th>protein</th>
<th>origin</th>
</tr>
</thead>
<tbody>
<tr>
<td>ERα-LBD</td>
<td>pET15b</td>
<td>His6</td>
<td>hERα-LBD322-553</td>
<td>MSD, prior ORGANON</td>
</tr>
<tr>
<td>ERβ-LBD</td>
<td>pET15b</td>
<td>His6</td>
<td>hERβ-LBD263-502</td>
<td>BAYER HEALTHCARE PHARMACEUTICALS</td>
</tr>
<tr>
<td>AR-LBD</td>
<td>pGEX-KG</td>
<td>GST</td>
<td>hAR-LBD564-919</td>
<td>BAYER HEALTHCARE PHARMACEUTICALS</td>
</tr>
<tr>
<td>RXRα-LBD</td>
<td>pGEX-4T1</td>
<td>GST</td>
<td>hRXRα-LBD221-462</td>
<td>FOLKERTSMA et al.80</td>
</tr>
</tbody>
</table>

Plasmids with the desired protein construct (Table 4-3) were transformed into E. coli BL21 cells. Emerging colonies were cultured over night in 25 ml LB medium with 100 µg/ml ampicillin at 37 °C. This preculture was added to 21 phosphate buffered, content-rich terrific broth (TB) medium (12 g peptone, 24 g yeast, 4 ml glycerol, 0.17 M KH2PO4, 0.72 M K2HPO4, 11 water, autoclaved) with ampicillin and incubated at 37 °C until an optical density at 600 nm (OD600) of 0.6 - 1.0. Addition of 100 µM Isopropyl β-D-1-thiogalactopyranoside (IPTG) induced expression of the protein of interest. For stabilization 10 µM of ligand (ER, estradiol | AR, dihydrotestosterone | RXR, agonist LG) was added. After incubation for 20 h at 15 °C the cells were harvested by centrifugation (15 min at 7,000 rpm) and the resulting pellet was stored at -80 °C until further usage.

If not stated otherwise, the protein purification was performed at 4 °C. The pellet was resuspended in lysis buffer (LyoB-1: 1 x PBS tablet, CALBIOCHEM, 370 mM NaCl, 10% (v/v) glycerol, 0.1 mM TCEP, 1 mM PMSF, 1 µg/ml DNAseI, 10 µM ligand, pH 8.0, His6-tag: additional 40 mM imidazole) and lyzed with the Emulsi Flex-C3 homogenizer (2 passes of 150,000 kPa, AVESTIN INC.). After another centrifugation step (40 min, 20,000 rpm) the supernatant was provided on the respective affinity chromatography column (GST-tag: Protino GST/4B, 1 ml, MACHEREY-NAGEL | His6-tag: His-Bind Resin, 3 ml, NOVAGEN), washed with wash buffer (WaB-1: GST-tag: 1 x PBS tablet, CALBIOCHEM, pH 7.3 | His-tags: 1 x PBS tablet, CALBIOCHEM, 370 mM NaCl, 40 mM imidazole, 10% (v/v) glycerol, 0.1 mM TCEP, 10 µM ligand, pH 8.0) and eluted with competing molecules (GST-tag: 50 mM Tris, 10 mM glutathione | His-tag: WaB-1 plus 100 mM / 250 mM / 500 mM imidazole). Fractions were analyzed with SDS-PAGE and optional with LC-MS. The buffer of the combined fractions were
exchanged with Amicon Ultra-15 Centrifugal Filter Unit with Ultracel-10kDa membrane (MILLIPORE) to desalting buffer (20 mM Tris, 25 mM NaCl, 10% glycerol, pH 8.0) including 10 μM ligand and long-term stored at -80°C.

**Solid Phase Peptide Synthesis & Purification**

The synthesis of all peptides was performed on solid support, using the Fmoc strategy. If not stated otherwise, 200 μmol (1 eq.) of Rink Amide MBHA resin (0.59 mmol/g, Novabiochem) was swollen for 30 min in N-methylpyrrolidone (NMP). The protective Fmoc group was removed by incubation with 20% piperidine in NMP (2x5min). Subsequently the amino acids (AA, 4 eq.) were activated using 2-(1H-Benzotriazole-1-yl)-1,1,3,3-Tetramethyluronium hexafluorophosphate (HBTU, 4 eq.), and N,N-Diisopropylethylamine (DIPEA, 8 eq.). Single coupling was performed by shaking 45 min at room temperature. After finishing the desired peptide sequence, the resin was washed 3 times alternately with diethylether (DEE) and dichloromethane (DCM) and dried under vacuum for 10 min. The side chain protection group and the cleavage of the peptide sequence from the resin were simultaneously achieved by the usage of mixture of trifluoroacetic acid, triisopropylsilane and water (TFA/TIS/H2O, 95/2.5/2.5) for 3 h. The free peptide was precipitated by adding it drop-wise to ice cold DEE. A final centrifugation step (2.000 rpm, 10 min) with subsequent washing (ice cold DEE), uptake in water and lyophilization results in the crude peptide. Purification was accomplished by reverse-phase HPLC on a Alltima™ HP C18 column (125x20 mm, Alltech). Water (0.1% TFA) was used as a polar phase, adding different amounts of an apolar acetonitrile (ACN, 0.1% TFA) phase. A linear gradient (20 ml/min) was optimized for each peptide to a 5% range, for instance 35-40% ACN. The purity of the peptide was determine at an analytical LC-MS using GraceSmart RP18 (50x2.1 mm, Grace, 3u, 120A) column.

Finally, the peptides were lyophilized and stored at -80 °C.

**Fluorescence Polarization and Depolarization assay**

Both fluorescence polarization (FP) and fluorescence depolarization (FDP) experiments were measured in black 384-well-plates (PERKIN ELMER 384 F) on a Safire² (TECAN) plate reader. The final volume of one well was 30 μl containing constant concentrations of ligand (5 μM) and Fluorescein-labeled peptide (0.1 μM, ER: NCoA-1/SRC1δ-m, FL-LTERHKILHRLLLQEGPSD; AR: FLETT-I, CSSRFESLLAGEKESR, phage display hit AR-LBD;² D22, LPYGSLLKLRLRAPVVEV, INVITROGEN, phage display hit ERα-LBD, good binding properties RXRα²⁶). In case of FP assay the protein was sequentially diluted (24 steps) in TR FRET coregulator buffer E (INVITROGEN). On the other hand, during FDP assays the protein concentration was kept constant (ER: 0.4 μM; RXR, AR: 1 μM), varying the concentration of the potential peptide-based binder. Briefly, in 96-well-plates dilution series were prepared in 55 μl and filled up to with 110 μl with a 2X master mix solution containing all non-varying components. Finally three times 30 μl was transferred to 384-well-plates, centrifuged (1,000 rpm, 2 min) and incubated for 1 h at 4°C. Plates were measured 50 times at 30 °C (excitation 470 nm, emission 519 nm). Polarization y was plotted against the concentration x of either the protein (FP) or the test peptide (FDP), whereby each data point represents an average of 3 experiments.

The dissociation constant Kd of the protein-peptide complex (FP) was calculated using equation 1a,b:

\[
y = A_1 + \frac{A_2 - A_1}{1 + 10^{(\log(x_0) - x)p}} \\
K_d = 10^{\log(x_0)}
\]

where A₁ is the bottom asymptote, A₂ is the top asymptote, and p the hill slope (steepness of the curve).
In the competitive setup (FDP) the half maximal inhibitory concentration \( (IC_{50}) \) was calculated with the means of equation 1a and 2:

\[
IC_{50} = 10^{\log (x_0)}
\]

The inhibitory constant \( K_i \) of the different peptides could be determined with the \( K_D \) value of the protein (from equation 2) and equation 3a,b.

\[
K_i = \frac{IC_{50}}{1 + \frac{A_D(\gamma_0 + 2)}{K_D(\gamma_0 + 1)}} - K_D \cdot \left( \frac{y_0}{y_0 + 2} \right)
\]

\[
y_0 = \frac{[AB]}{[A]} = \frac{[B]}{K_D}
\]

where \( K_d \) is the dissociation constant of the fluorescein labeled peptide protein complex (equation 1), \( y_0 \) is the initial bound to free concentration ratio for labeled peptide, \([AB]\) is the concentration of protein-peptide complex, \([A]\) the concentration of unbound peptide, and \([B]\) the concentration of protein (ER).

**Ribosome Display**

If not stated otherwise, protein expression and purification were conducted like described above in *Protein Expression & Purification*. The following constructs were used to express the ER (α/β) LBD suitable for surface immobilization.

**Table 4-4 | Used construct for ribosome display**

<table>
<thead>
<tr>
<th>construct</th>
<th>vector</th>
<th>tag</th>
<th>protein</th>
<th>origin</th>
</tr>
</thead>
<tbody>
<tr>
<td>pHT503</td>
<td>pTriEx-4</td>
<td>Strep-His6</td>
<td>ERα-LBD302-553</td>
<td>CARRAZ et al.43</td>
</tr>
<tr>
<td>pHT504</td>
<td>pTriEx-4</td>
<td>Strep-His6</td>
<td>ERβ-LBD268-502</td>
<td>NGUYEN et al.44</td>
</tr>
</tbody>
</table>

Where His\textsubscript{6} tag was used for purification and Strep tag for immobilization. After induced protein expression in *E. coli* Rosetta 2 (DE3) pLacI cells (NOVAGEN), a 5 ml His-Trap HP 5 column (GE HEALTHCARE) was used for purification. SDS-PAGE and LC-ESI-MS analysis confirmed the ERα-LBD and ERβ-LBD containing fractions. Finally the purified proteins were immobilized on StrepTactin coated plate (IBA) and the activity was determined by its SRC1-Box2-eCFP binding ability. The DNA library was designed based on the known ER binder SRC1-Box2:

\[
gcc atg aat ttg ctg acc gcc cgc\]

\[
\text{NNK NNK NNK NNK NNK cgt NNK NNK NNK NNK NNK NNK NNK NNK NNK}
\]

\[
gcc ccc tgt gag aag ctt gcc ggc\]

Where N is any base and K is guanine (G) or thymidine (T). The plasmid with the randomized region, resulting in SLTARXXXXXXXXXPSD peptides (with X, any amino acid), was expressed *in vitro* using PureExpress *In Vitro* Protein Synthesis kit (NEB). In a 96-well plate format the library of ribosomal complexes, consisting of expressed peptide tethered to ribosome with unreleased mRNA, was screened against immobilized ERα-LBD and ERβ-LBD, respectively. The narrowed library with connected geno- and phenotype was used as basis for the following reversed transcription resulting in optimized DNA library with enhanced ER binding properties. After several of these selection rounds the library was processed and analyzed as followed: introduction in vector using In-Fusion TM Advantage PCR Cloning kit (CLONTECH), transformation in *E. coli* OmniMax, confirmation by colony PCR, sequencing, conversion to analogue amino acids and alignment by means of Weblogo.65,38
Selected hits:
2, LTARSSLRLRLLQPSD | 3, LTARMPQLTRLLSPSPSN
4, LTARSPRLRLLQPSD | 5, LTARHPRLRLLQPSND
6, LTARLLMSRLTLQPSD | 10, LTARHPLLMLHHPSPSD

Circular Dichroism spectroscopy

Far-UV Circular Dichroism (CD) spectroscopy measurements were performed under a constant nitrogen flow at 20 °C using 50 μM of the peptide in the presence of 30% (v/v) trifluoroethanol (TFE), if not stated otherwise. As a baseline a peptide-free buffer (30% TFE) was measured under identical conditions. The spectrum was recorded from 250 nm to 185 nm using a JASCO-815 spectrometer. Quartz cuvettes with path lengths of 1 mm or 0.2 mm (HELLMA ANALYTICS) were employed with a data pitch of 0.5 in a continuous mode, a scan speed of 20 nm/min with a response time of 2 s and a bandwidth of 0.5 nm. The graphs are representing an average of five scans. The observed ellipticity \( \theta_{\lambda} \) (degrees, in [mdeg]) was converted to mean residue ellipticity \( \theta_{mrw,\lambda} \) (in [deg cm²/dmol]) using equation 3.

\[
\theta_{mrw,\lambda} = \frac{M}{(N-1)} \cdot \frac{\theta_{\lambda}}{10d \cdot c}
\]

Where \( M \) is the molecular mass of the peptide (in [Da]), \( N \) is the number of amino acids in the peptide, \( d \) is the path length (in [cm]), and \( c \) is the peptide concentration (in [g/ml]). Information about helical propensity in semi-quantitative manner could be shown by the ratio of the characteristic wavelengths at 208 nm and 222 nm: \( \theta_{mrw,222nm}/\theta_{mrw,208nm} \).

Molecular Dynamics Simulation

All MD simulations were carried out using the AMBER suite of programs, and the ff03 force field. An implicit solvent was used via the General Born solvation method (IGB 5), as implemented in AMBER. Starting with an extended initial conformation, built by the LEaP module of AMBER for each peptide, all MD simulations were fully unrestrained and all trajectories were generated using the sander program in the AMBER 9 package. Each peptide was simulated for a total of 20 ns at 51.9 °C, following a similar approach as previously reported to predict the conformation of a miniprotein. Analysis was performed on the 20 ns using 20,000 trajectory snapshots spaced every 1 ps. A snapshot with the lowest potential energy across the simulation was chosen as representative structure for each peptide. The secondary structure of each residue as a function of time was subsequently analyzed utilizing the STRIDE secondary structure assignment algorithms as implemented in VMD. All MD simulations were performed by Dr. Lidia Nieto.
Figure 4-11 | Degree of helical content per residue and structure based on MD simulation
Starting from an extended initial conformation, MD simulations were fully unrestrained and all trajectories were generated. Each peptide was simulated for a total of 20 ns - 20,000 trajectory snapshots spaced every 1 ps - at 51.9 °C (325 K). (A, B) Grayscale is representing the propensity forming secondary structures: white, unordered; light grey, α-helix; dark grey, turn; black, 310 helix. Dashed line shows start/end of helix. A snapshot with the lowest potential energy across the simulation was chosen as representative structure for each peptide: (A) Natural binder SRC-Box2 (1), (B) after introduction of prolines in -2 and +3 position (1-Pro). (C) The degree of helical content per residue (for 1 and 1-Pro) was obtained using the ptraj module of AMBER omitting the first 5 ns of each simulation. (D) Predicted structure of peptides 5 and 13 based on MD simulation.

Co-crystallization
If not stated otherwise, protein expression and purification were conducted like described above in ‘Protein Expression & Purification’.

Table 4-5 | Used construct for co-crystallization

<table>
<thead>
<tr>
<th>construct</th>
<th>vector</th>
<th>tag</th>
<th>protein</th>
<th>origin</th>
</tr>
</thead>
<tbody>
<tr>
<td>ERα-LBD</td>
<td>pET16b</td>
<td>-</td>
<td>ERβ-LBD\textsuperscript{MDP1281-L500P}</td>
<td>Möcklinghoff et al.\textsuperscript{72}</td>
</tr>
</tbody>
</table>
After induced protein expression in E. coli cells, they were harvested, lysed (LyB-2: 20 mM Tris, 0.5 M NaCl, 5 mM DTT, 1 mM EDTA, pH 7.5), and subsequently the tag-free protein was purified with a 3-mL estradiol-sepharose column (PTI RESEARCH, INC.). After washing (250 mL WaB-2: 10 mM Tris, 0.5 M NaCl, 5 mM DTT, 1 mM EDTA, pH 7.5) cysteins of the protein were S-carboxymethylated using 5 mM iodoacetic acid (50 mL WaB-2, but 0.2 M NaCl, 1 h, 4 °C). After an additional washing step (400 mL WaB-2) the pure protein was eluted with step-wise increasing amounts of estradiol (WaB-2, 100 μM, 150 μM, 200 μM E2) and by means of a centrifugal 10 kDa-cut-off filter unit (AMICON) buffer-exchanged (WaB-2, but 0.2 M NaCl, 5 mM DTT) and concentrated to 8 - 12 mg/mL. Characterization was accomplished by SDS-PAGE and photometric determination of the concentration (NANODROP, 280 nm).

11 mg/mL (400 μM) of the protein was mixed with the peptide (5 or 13) in a 1:1.5 molecular ratio (600 μM) and incubated for 30 min at 4 °C. In order to crystallize this peptide-protein complex, it was combined with the respective well solution in a 2:1 ratio. Finally, a total volume of 4.5 μL was equilibrated at 4 °C against 1 mL reservoir solution on a 24-well crystallization plate (NEXTAL) using the hanging drop crystallization methodology. For initial screens the JCSG+ suite (QIAGEN), which consists of different premixed crystallization conditions, was used. Further optimization was performed with the suites JCSG Core I, Core II, Core III, and Core IV. Optimal conditions for crystal growing were 10% PEG6000, 1 M LiCl, 100 mM citric acid, and 20% glycerol at pH 5. For X-ray analysis single crystals were transferred into a Cryoloop (HAMPTON RESEARCH) and directly cryo-cooled in liquid nitrogen, since the optimal condition already consists of 20% glycerol.

In order to solve the crystal structure, molecular replacement of a ERβ LBD/ligand/peptide complex (PDB code: 1U9E7) using the program Phaser was performed. On a Nonius AXS MICRO A (MAR db detector, crystal-to-detector distance was 150 mm, oscillation range was 1.0°) a native data set from an optimized ERβ LBD peptide complex crystal was collected at -173 °C (100 K). After data indexing and integrating, the XDS package was used for scaling. The crystals diffracted to a maximal resolution of 2.1 Å (5) and 2.3 Å (13), respectively. Furthermore the crystals were assigned to space group C121 (5) and P3(2) (13) with 4 (5) and 2 (13) monomers per asymmetric unit. Statistics of both crystal structures are listed in TAB.
The program Coot was used to calculate the electron density maps.\textsuperscript{75} In order to perform the iterative structure refinement the program REFMAC (CCP4 suit) was used.\textsuperscript{76}
Figure 4-12 | Stereo image of a portion of the electron density map
ERβ-LBD in complex with peptide and peptide close up (A,B, 5; C,D, 13)
Mammalian Two Hybrid assay
All Mammalian Two-Hybrid assays (M2H) were performed in human osteosarcoma cells (U2-OS) seeding 30,000 cells/well in a 24-well plate (BD Biosciences). Plasmids are either from Mammalian Two-Hybrid Assay Kit (AGILENT) or cloned by Dr. Ingrid J. de Vries. Culturing conditions were 37 °C and 5% CO₂. Cells were transfected with following amounts of DNA/well using linear polyethylenimines (PEI, POLYSCIENCES): (1) 40 ng of the bait construct pCMV-BD-PEP containing peptide of interest (PEP) fused to the DNA-binding domain of the GAL4, (2) 40 ng of the target construct pCMV-AD-ERα/β, containing ER (α or β) fused to the transcriptional activation domain (AD) of the mouse protein NF-κB, (3) 200 ng of the reporter plasmid pFR-Luc with the synthetic promoter with five tandem repeats of the GAL4 binding sites that control expression of the Photinus pyralis firefly luciferase, and (4) 3.2 ng of control plasmid Renilla-Luc, expressing independently Renilla reniformis luciferase. After 16 h incubation cells were treated with 10 nM estradiol and after additional 24 h incubation the interaction was determined with a Dual-Luciferase Reporter Assay (PROMEGA), according to manufacturer’s instruction. The luminescent intensities were recorded on a Synergy HT Multi-Mode Microplate Reader (BioTEK). The interaction-dependent Firefly-luciferase signal was first normalized over the independent Renilla-luciferase signal and subsequently normalized over the ratio of untreated cells. Studies were performed by Dr. Ingrid J. de Vries-van Leeuwen.
Displacement assay (SNAP-tag) in cells were also based on M2H. Co-transfection of pCMV-AD-ERα/β and pCMV-BD-1 produce normal signal of luciferase representing for the binding of SRC1-Box2 to ER-LBD. Titration of a SNAP-PEP plasmid - carrying the sequence of the PEP fused to a SNAP-tag and a nuclear localization signal – and the consequent decrease of luciferase signal is a function of the ability of the PEP to compete with SRC1-Box2-containing target construct (pCMV-BD-1). Studies were performed by Dr. Hoang D. Nguyen.
4.8. References

Screening the Estrogen Receptor Coactivator Interaction

CHAPTER FIVE

In Silico Design of Androgen Receptor Coactivator Binding Inhibitors

Abstract. Molecular Dynamics (MD) Simulation was used to predict the propensity of short peptides to fold into an α-helix: one of the requirements for targeting the androgen receptor (AR) coactivator binding surface. The rational in silico design of the peptides was based on separate results from estrogen receptor (ER) screening, which identified optimized leucine-rich motif with flanking prolines as helix primers (Chapter 4). The scaffold sequence was based on a combination of the natural SRC1-Box2 sequence and the N’-terminal domain (NTD) of AR, later known to specifically interact with the AR ligand binding domain (LBD). Instead of a leucine-rich sequence (LXXLL) – required for binding to ER and several other nuclear receptors (NRs) – a phenylalanine-rich motif (FXXLF) was introduced to promote AR selective binding. Proline residues were inserted into the flanking regions of the central binding motif. Molecular dynamics (MD) simulation studies were performed to acquire information about the propensities of the designed peptides to form α-helices. Based on these results the solid-phase synthesis of selected peptides was performed. Circular dichroism studies of these peptides clearly confirmed the predicted α-helical character of the 11- and 12-mers. The AR binding affinities of the peptides was evaluated by varying the length of the helical segment of the peptides via the proline positioning and by elaborating the adjacent amino acid of the N’-terminal proline in terms of charge, size, and hydrophobicity. This conceptual approach resulted in short peptides that interact exclusively with AR-LBD with Kd/Ki values in the low μM-range, and that could even overpower known AR-specific binding in a cellular context with the fully intact AR. Thus, the identified sequences can be considered as lead structures for the future development of AR-selective coactivator binding inhibitors (CBIs).
5.1. Introduction

The androgen receptor (AR, NR3C4) is part of the subclass of steroid receptors of the nuclear receptor (NR) superfamily, which also includes glucocorticoid-, mineralocorticoid-, progesterone-, and estrogen receptors. Misregulation of AR signaling has classically been treated by either reducing androgens at the steroidogenesis level or through blocking of the binding of the endogenous ligand, dihydrotestosterone (DHT), to the AR ligand binding pocket (LBP). While initial steroid-based anti-androgens produced off-target effects on other NRs, the non-steroidal small molecules hydroxyflutamide (Eulexin, SCHERING-PLOUGH), bicalutamide (Casodex, ASTRAZENECA) and nilutamide (Nilandron, AVENTIS PHARMACEUTICALS) are known to be selective androgen receptor modulators (SARMs). The first co-crystal structure of the AR-bicalutamide complex provided an entry point for the rational drug design of small-molecules antagonists of increased potency and efficiency. The drawback of such compounds, however, is that long term treatment can result in the loss of antagonistic effect due to amino acid mutations in the AR-Ligand binding domain (LBD), as well as reduced activity due to AR and AR-coactivator up-regulation or modification (post-translational modification), or improvement of ligand (DHT) production. Thus, alternative to the classical approach targeting the AR-ligand binding pocket (LBP), the protein-protein interaction (PPI) between the AR-LBD and its coactivator – via activation function 2 (AF-2) - is currently under investigation as a controllable and druggable interface.

The human proteome consists of more than 375,000 PPIs. The emergence of PPIs as drug targets for drug development has been aided by X-ray crystallography, NMR and electron microscopy, which has enabled the atomistic level determination of large multiprotein complexes. Additionally, next generation sequencing and high-throughput genomic mapping has helped to identify peptide sequences in certain PPIs that are involved in certain disease states. The use of biologics such as peptides for targeting PPIs has been validated through the increasing numbers of approved drugs in recent years.

The AR-LBD exhibits only moderate sequence homology compared with the LBD of other NRs (e.g. 50% compared to the mineralocorticoid- (MR) and progesterone receptor (PR) or around 20% in the case of the estrogen receptors subtypes α and β, ERα/β). However, overall, they share a similar structure. Additionally, AR also possesses some unique attributes, especially at the interface with coactivator proteins. While the common ability of NRs to bind leucine-rich motifs (LXXLL) also includes AR, binding studies using the full length AR has revealed the extraordinary influence of the receptor’s N-terminal domain (NTD). NR-positive interaction partners show limited binding properties for AR, suggesting that the activation function 2 (AF-2) of AR has dissimilar sequence-specificity compared with other NRs. Detailed AF-2 investigations discovered, indeed, an ability for AR to accommodate aromatic amino acid-based sequences, including the FXXLF binding pattern. This unique motif is also present in the AR-NTD, which is reported to interact
with the AR-LBD (AR N/C interaction). Furthermore, the charge clamp residues lysine 720 and glutamic acid 897 of the AR-AF-2 can interact with AR-specific motifs forming two hydrogen bonds with their backbone, compared to only one hydrogen bond with LXXLL-based peptides.\textsuperscript{15-17} In a separate piece of work, the importance of the FXXLF motifs in coactivator recruitment was demonstrated, including members of the AR-associated (ARA) coactivator family or the checkpoint protein hRAD9. The introduction of LXXLL motifs in these coactivators resulted in the complete loss of AR binding ability.\textsuperscript{18,19} \textit{Vice versa}, peptides featuring FXXLF motifs show no binding to other NRs, except for PR-LBD, which nevertheless shows preferential binding to leucine-based sequences.\textsuperscript{17} In general, structural analysis of the interface between NRs and their coactivators showed that the peptide exists as an \(\alpha\)-helix of well-defined length, which is determined by the charge clamp.\textsuperscript{15,20}

In a previous study of the AF-2, the AR-LBD was screened for novel peptide sequences resulting in phenylalanine-based motifs, but not the typical LXXLL NR motif.\textsuperscript{21,22} Thus, drug discovery is using this atypical surface pattern - featuring the ability to accommodate bulky aromatic residues - to target specifically AR. Small molecule design through a mimic of the aromatic phenyl-rings discovered CBIs with \(K_i\) values in the low \(\mu\)M-range.\textsuperscript{23-25} Other studies have focused on peptide-based coactivator binding inhibitors: for example the introduction of an FXXLF motif into \(\alpha\)-helical miniproteins stabilized by intramolecular disulfide bonds.\textsuperscript{26,27} In general, stabilization of peptide secondary structures (e.g. by macrolactamization,\textsuperscript{28} hydrocarbon stapling,\textsuperscript{29} or the incorporation of hydrogen bonding surrogates,\textsuperscript{30}) can lead to enhanced binding properties due to preorganization of the \(\alpha\)-helix in a surface-bound-like state. Nonetheless, the drawbacks of such approaches are the necessity for synthetic operations to upgrade amino acid building blocks to facilitate macrocyclization and the unpredictable effects that such modifications might have on binding.

In the present study, potential AR CBIs were designed \textit{in silico}, making recourse to both the well-characterized specificity of the unique FXXLF motif for AR-AF-2 binding and the influence of prolines on \(\alpha\)-helix formation and stability (PXLXXLXXXP) discovered for ER-AF-2 interaction (chapter 4). The required degree of \(\alpha\)-helicity was initially monitored by molecular dynamics (MD) studies and later confirmed by CD spectroscopy of peptides made by solid-phase peptide synthesis (SPPS). Finally, the potential CBIs were validated in polarization-based binding studies and using cell-based assays.
5.2. In silico peptide design

In a prior study, the LBDs of ERα and ERβ were screened for novel binding motifs using the ribosome display technique. The main advantage of ribosome display over other screening methods - in particular phage display - is the exceptionally high diversity of the DNA library, since the screening is performed entirely in vitro. Results from this protein surface screen resulted in a consensus motif featuring the hallmark LXXLL sequence – known to be important for NR binding – and additional prolines in the flanking regions of the core motif (chapter 4). Transferring these investigations to AR requires a detailed comparison of both coactivator binding surfaces (AF-2) with respect to size, hydrophobicity and charge, since the total sequence homology between ERβ and AR is relatively low (18.6%) compared to other NRs.
Figure 5-2 | AF-2 of AR compared to ERβ
Shown are amino acids of the ligand binding domain involved in the generation of the AF-2, the surface for coactivator interaction. (A) AR (PDB, 1T7R), in comparison to (B) ERβ (PDB, 2J7X), provides more space for interaction with bulky, aromatic residues reported to be selective AR binders. Charged amino acids considered to make important electrostatic interactions are observed in parentheses.

Ligand binding to the LBP of the NR-LBD, and the subsequent conformational change, including a repositioning of the mobile helix 12 (H12) region, results in coactivator binding to the NR AF-2. In the case of AR, this binding platform consists of nine amino acids, which are involved in forming the hydrophobic interaction with the coactivator protein (Figure 5-2a): V716 and K720 on helix 3 (h3), Q733, M734, I737, and Q738 on helix 4 (h4), and M894, E897, and I898 on H12. Although six of these nine residues are different compared with ERβ, the general hydrophobic environment does not change, since the amino acid residues remain predominantly hydrophobic. One exception is E287(ERβ)→Q738(AR), which results in a net loss of a negative charge. Furthermore, residues with reduced steric bulk, which accommodates the greater steric bulk of the aromatic phenylalanines as a unique feature of the FxxLF binding motif AR. In particular, residues I898 (in place of M449) and V716 (in place of I265) provide more room for interaction with the first phenylalanine (Figure 5-7, Phera). Furthermore, the distance between the charge clamp residues (AR: E897/K720 and ERβ: E448/K269) is 5% greater for AR. This AF-2 shape difference in combination with knowledge acquired from previous ER screening facilitated the design of a set of novel CBIs (Figure 5-1, In Silico Design Rationale).

5.3. Molecular Dynamics Simulations – identification of scaffold

The helical propensity of the designed, proline-derived peptides was verified using molecular dynamics (MD) simulations (Figure 5-1, MD simulation). Implicit solvent was used starting with an extended initial conformation. In fully unrestrained conditions, all trajectories were generated and each peptide was simulated for a total of 20 ns at 51.9 °C (325 K), in accordance with a similar approach as previously reported to predict the conformation of a miniprotein.32 Analysis was performed on the 20 ns time-scale using 20,000 trajectory snapshots spaced at 1 ps intervals. Secondary structure analysis of each residue as a function of time was subsequently performed.33,34
**Figure 5-3 | Initial MD simulations**

Molecular Dynamics (MD) simulations were performed to find the optimal scaffold with respect to α-helicity. Helicity vs. peptide sequence derived from a 20 ns molecular dynamics simulation: comparison of peptides 1, 2, 3 and 4. The degree of helical content per residue was obtained using the *ptraj* module of AMBER. Finally, 4 consists of an N’-terminal part of peptide 2 and 3 (dashed line) and both the FXXLF motif and C’-terminal part of 1 (black line), with an V→P mutation (grey line).

Initially, MD simulations were performed to find the ideal scaffold for introducing both the AR-specific FxxLF motif and the helix-modulating/controlling prolines. A logical starting point was the sequence of the N’-terminal domain of AR (1), an FXXLF bearing sequence capable of interacting with the AF-2. However, this peptide already showed limited α-helical character around Pheα, prior to introducing the proline residues (Figure 5-3A). An alternative approach used the natural coactivator SRC1-Box2 as a scaffold with which to introduce both the central FxxLF motif and the flanking prolines (2 and 3, respectively). But again the propensity to form α-helical structures in this case was not optimal according to the MD calculations. Finally, a combination of the N’-terminal motif of SRC1-Box2 (LTERHIP) bearing both the core motif and C-terminus from 1 and additional prolines (FQNLFQSPRE) suggested the best formation of a well-defined α-helix (Figure 5-3, 4).
5.4. First generation peptides – the second proline position

First, the optimal position of the second proline was investigated, because structural analysis had revealed that, compared to ER, the AF-2 of AR is not only deeper but also features a larger distance between the two charge clamp residues (Figure 5-1 and 5-2).

![Figure 5-1](image1.png)

**Figure 5-1** | Influence of the proline position on AR binding
Moving the second proline one position in the N'-terminal direction (P+3 \(\rightarrow\) P+4) by introducing a valine (V), naturally occurring in AR-NTD at that position, leads to enhanced \(\alpha\)-helicity (MD) and AR binding (polarization assay). (A) Polarization competition assay of 5 and 7 and V-inserted 6 and 8 competing with FLETT-1 (CSSRFESLFAGEKESR, phage display hit AR-LBD\(^{10}\)). (B) Helicity vs. peptide sequence derived from a 20 ns molecular dynamics simulation: comparison of peptides 5 and 6. The degree of helical content per residue was obtained using the * ptraj* module of AMBER.

Repositioning of the proline (\(-P\rightarrow-VP\)) in two different peptides (5 \(\rightarrow\) 6, 7 \(\rightarrow\) 8) resulted in both greater propensity for \(\alpha\)-helix formation and improved binding to AR-LBD (Figure 5-4). While 6 and 8 were inactive, the chain extended variants (by one amino acid) were capable of competing with the strong FXLF reference peptide (5, \(K_i = 77 \mu M\); 7, \(K_i = 85 \mu M\)). Together with the FXLF core motif, this proline at the +4 position – realized by introducing an additional valine at the +3 position - was kept constant in the second set of peptides.

5.5. Influence of first proline position and charge on structure

In this part of the study both the position of the first proline and the influence of the amino acid after the proline were simultaneously investigated. In a first set of 12-mer peptides, consisting of the sequence \(X_3-X_2-X_1-FQNLFSVP\) (\(X_Y = \) variable positions Y), the proline was relocated between position -1, -2 and -3 (PQL\(^{-}\), QPL\(^{-}\), QLP\(^{-}\)). A second set of peptides was prepared keeping the histidine and proline residue at positions -3 and -2, respectively, while varying the -1 position (HP-X\(^{-}\)). The choice of residues was as follows: glutamic acid (E) representing negatively charged amino acids; lysine (K) as a positive amino acid, glutamine (Q) with its polar but uncharged side chain properties; and leucine (L) representing
hydrophobic amino acids. All peptides were synthesized on solid phase as 12-mers to enable structure activity relationship studies.

Table 5-1 | Circular dichroism measurements of AR peptides.

<table>
<thead>
<tr>
<th>CD spectra of potential AR binders</th>
<th>peptide sequence</th>
<th>label</th>
<th>$\theta_{222}/\theta_{208}^c$</th>
</tr>
</thead>
<tbody>
<tr>
<td>first proline position</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>AcQLPFQNLFSVP</td>
<td>9c</td>
<td>0.73</td>
</tr>
<tr>
<td></td>
<td>AcQLPFQNLFSVP</td>
<td>10b</td>
<td>0.81</td>
</tr>
<tr>
<td></td>
<td>AcPQLFQNLFSVP</td>
<td>11b</td>
<td>0.89</td>
</tr>
<tr>
<td>influence of solvent (TFE)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>AcHPQFQNLFSVP</td>
<td>12b</td>
<td>0.28b</td>
</tr>
<tr>
<td></td>
<td>AcHPQFQNLFSVP</td>
<td>12b</td>
<td>0.82</td>
</tr>
<tr>
<td>influence of residue after first proline (-1 position)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>AcHPKFQNLFSVP</td>
<td>13b</td>
<td>0.83</td>
</tr>
<tr>
<td></td>
<td>AcHPKLFQNLFSVP</td>
<td>14b</td>
<td>1.14</td>
</tr>
<tr>
<td></td>
<td>AcHPEFQNLFSVP</td>
<td>15b</td>
<td>0.75</td>
</tr>
</tbody>
</table>

CD spectroscopy measurements were performed under a constant nitrogen flow at 20 °C at 50 µM peptide concentration in phosphate buffer (b) or in additional 30% (v/v) TFE (a, all except 12b). Information about helical propensity in a semi-quantitative manner could be shown by determining the ratio of the characteristic wavelengths at 208 nm and 222 nm: $\theta_{222}/\theta_{208}^c$.

The investigation of the secondary structure of the synthesized peptides was determined by CD spectroscopy. These structural data are consistent with the MD simulation data (Figure 5-5), which show, for instance, the propensity to form longer $\alpha$-helices upon gradual repositioning of the proline from position -1 to position -3 (Table 5-1 and Figure 5-5A, 9-11). Furthermore, the observed influence of the helix-promoting solvent 1,1,1-trifluoroethanol (TFE), which simulates the proximity of the protein surface, fits coherently with the theoretical and practical structural analysis. With no observable $\alpha$-helical character in water according to MD simulations, or a low $\theta_{222}/\theta_{208}$ ratio based on CD measurements in phosphate buffer (an indicator for $\alpha$-helicity), TFE was observed to cause an increase in the degree of secondary structure formation (Table 5-1, 12). The result for 14, having the most helical character, is coherent for both studies. (Table 5-1, and Figure 5-5B, 13-15).
5.6. Influence of the first proline position and charge on binding

After structural analysis, the peptides were tested for their ability to bind AR-LBD. In a competitive experimental set up, potential CBIs were made to compete with the fluorescein-labeled FLETT-1 peptide - identified from phage display ($K_d = 0.67 \pm 0.07 \mu M$) - for binding to the AR-AF-2. Overall, the peptides produced classical sigmoidal curves and behaved as competitive inhibitors with $K_i$ values in the range of $1 \mu M$ to $110 \mu M$ (compared to the FXXLF-featuring peptide with a sequence based on AR-NTD, 1, $K_i = 0.6 \mu M$). The investigations with respect to the proline position N'-terminal of the fixed FXXLF motif identified position -2 (10) as the optimal position ($K_i = 1.5 \mu M$). The peptide analogue with proline at -3 position (11) was equally active ($K_i = 1.8 \mu M$). However, location of the proline at the -1 position, directly prior the FXXLF motif (9) resulted in a significant decrease in binding affinity ($K_i = 64 \mu M$). Keeping the Pro-2 fixed and the simultaneous exploration of various amino acids residues at the position between Pro-2 and the FxxLF motif, identified both positively charged (13) and hydrophobic (14) residues to be beneficial for peptide binding ($K_i = 3.5 \mu M$ and $4.1 \mu M$, respectively). In contrast, polar (12) and negatively charged amino acids (15) resulted in poor binding affinities ($K_i = 42 \mu M$ and $110 \mu M$, respectively). Further investigations into the most active peptide (10) against other NRs, including estrogen receptor $\beta$ (ER$\beta$) and the retinoic X receptor $\alpha$ (RXR$\alpha$), confirmed the selectivity for AR. Cellular binding with full length AR showed enhanced binding for 10 compared to 1 and ARA54, a known AR-specific coactivator.
Figure 5-6 | Evaluation of proline-based peptides
Depolarization competition assays with fluorescein-labeled cofactor (AR: FLETT-1, ERβ: SRC1-Box2, RXRα: D22). (A) Competition assay of peptides with varying proline-position (9-11). (B) Competition assay of peptides with different residues at position -1 (12-15). (C) Binding properties of 10 against ERβ and RXRα. D22, LPYEGSLIKLKLAPVEEV, IC50 = 184 nM; S1B2, SRC1-Box2, SLTERHKILHRLQEGSPSD, IC50 = 289 nM. (D) Mammalian one-hybrid studies on peptides 10 and 13 comparing normalized luciferase activity with known AR binders, including AR-NTD (1) and ARA54. L and S are assigned to the linker between the sequence and bait domains. 13, 10, and 10* have corresponding linker to 15. 10* is peptide 10 with removal of Val+3 resulting in Pro+3. (E) IC50 resulting from a fit of the sigmoid curves. K values calculated using equation 3a,b (Experimental). FXXLF motif highlighted in light grey, flanking prolines underlined and bold; ± sign is the standard error of the mean (s.e.m.); b no binding.
5.7. Discussion

In this study the AR-coactivator interaction was investigated as a potential target to block AR function. Targeting this hydrophobic interaction instead of the ligand binding pocket (LBP) is a potentially advantageous approach, given that classical antagonists tend to suffer from undesired toxicity, in addition to long term treatment resistance caused by point mutations in the AR-LBP, enhanced ligand production, or up-regulation and modification of either AR itself or AR-coactivators and accessory proteins. For the design of novel AR peptide inhibitors, we applied knowledge about prolines as helix breakers, obtained from previous investigations into ribosome display screening of the ER surface, and combined this with the predicted α-helical propensity of the potential CBIs (a known requirement for efficient binding), as determined by molecular dynamics (MD) simulations. Our results demonstrate the possibility of knowledge transfer within the steroid receptor subfamily of the NR superfamily resulting in a proline derived 12-mer peptide endowed with strong and selective binding affinity for AR. The initial evaluation of in silico designed peptides in respect of their secondary structure eliminates streamlines the development process and in particular limits the possibility for negative hits at the early stages of the study.

The initial position of the prolines in the AR targeting peptides was based directly on the ER screening investigations. This study exposed flanking prolines at both the -2 and +3 position with respect to the central LXXLL motif (chapter 4). Structural analysis by co-crystallization and CD studies connected these positions to N1 and C', respectively. The first choice of scaffold structure for potential AR CBIs design was the N'-terminal AR domain (1), which interacts with the AR-AF-2 via an FXXLF motif. However, MD simulations predicted low α-helical character in this case, especially in the region of the first phenylalanine Pheα (Figure 5-3A, 7 Phe). The reason for the poor α-helix propensity could be the presence of a glycine residue N'-terminal of the FXXLF motif. Previous studies investigating the α-helical propensity of different peptides and proteins have shown that glycines - aside only from proline – possess the lowest helix propensity. The formation of α-helices is an interplay between the loss of side-chain configurational entropy and the gain of enthalpy through mainly hydrogen bond formation between i and i + 4 amino acids. Apart from the side-chain-based entropy difference between the helical and coiled state (helix-coil transition) there are also different enthalpic contributions from the residues. Thus, in the case of glycine, aside from the high entropic costs through fixing of the backbone dihedral angles in an α-helix, the enthalpic contribution turns out to be the lowest of all residues, which is again a combination of the poor solvent accessibility of the backbone and the lack of possible stabilizing van der Waals interaction between amino acid side chains.

The next step in designing proline derived AR inhibitor peptides focused on the introduction of both the specific FXXLF motif and the flanking prolines into the SRC1-Box2 coactivator sequence - also part of the aforementioned ER ribosome display study - which
resulted in a PXFXLFXXP motif (Figure 5-3C, 3). The introduction of the aromatic amino acid residues caused a destabilization of the α-helical secondary structure, compared to its equivalent LXXLL-bearing sequence (Figure 5-3B, 2). Compared to 1, however, the N’-terminal region between Pheα and Pro-2 showed improved α-helicity, while the total helicity, especially at the C-terminus (Pheε/Pro+3) was observed to decrease. Indeed, phenylalanines display a lower propensity to form α-helices compared to leucines.36 Nevertheless, this dramatic loss of overall helical character proved to be an unexpected outcome of these investigations.

A logical next step in the design, therefore, was to combine the most suitable characteristics of both sequences, since one (3) possessed a well-defined structure in the region surrounding Pheα, while the other (1) around Pheε. The 11-mer resulting from this new design consisted of the following sequence: H-P-I-E-Q-N-L-E-Q-S-P. MD simulation predicted in this case a good α-helical propensity with prolines in N1 (beginning of helix) and C’ position (two positions after the helix) (Figure 5-3).

With this scaffold sequence in hand, the ideal position of the second proline was investigated with respect to α-helicity and finally AR-binding. This first generation AR-CBIs were inactive with the second proline located at position +3 (Pro+3), however, introduction of a valine at position 3 (Val+3) and consequent repositioning of the proline (Pro+4) resulted in enhanced binding properties. This result would suggest that AR prefers longer helices for coactivator binding compared to ER. Indeed, analysis of the AF-2 of AR in comparison with ERβ reveals the slightly greater distance between the two charge clamp residues glutamic acid and lysine (Figure 5-7). This difference of 0.8 Å could provide sufficient space for an additional amino acid in the helical segment spacing in between the two charged amino acids (3.6 Residues are needed for one α-helix full turn of 5.4 Å, resulting in a 1.5 Å contribution per amino acid).41 The improved binding properties are also a function of α-helix stability. MD data showed a significant increase in helical character upon a repositioning of the proline from +3 to +4 (Figure 5-4). Thus, the introduction of the valine in position +3 (Val+3) plays a crucial role in stabilizing the α-helix.
Hydrophobic interactions are important in protein folding and for protein-protein interactions (PPIs). The driving force for the hydrophobic effect is the reduction of hydrophobic residues exposed to polar water. Since α-helices are mainly stabilized by electrostatic interactions, such as hydrogen bond formation between i and i + 4 residues, hydrophobic interactions of side chains receive less attention, and the degree with which they contribute to helix stability is the subject of ongoing discussions. However, studies on alanine-based peptides have shown that hydrophobic residues at the i and i + 4 positions exert a positive effect on the α-helical nature of the peptide. Even though the two side chains at i and i + 3 sit close to one another, the side chain orientations of the i, i + 4 residues are more suitable for hydrophobic interactions, including van der Waals and π/π stacking. Thus the significantly increased binding properties of peptides 5 and 7 over 6 and 8 are a function of enhanced helix stability through additional interactions between the hydrophobic residues Leuδ↔Val+3 (Figure 5-7B, -C-H⋯H-C-, 3.4 Å). This contribution to helix stability seems to be necessary due to the lower α-helix propensity of phenylalanines (Pheα/Pheε of FXXLF) compared to leucines (Leuα/Leuε of LXXLL), which are used in ER binding. Furthermore, the introduction of additional hydrophobic residues (Val+3) is again only possible due to the greater distance between the AR charge clamp residues E897/L720 (compared to E542/L362 in ER β), thus providing space for a more extended α-helix (Figure 5-7A).

In the second library of AR CBIs the core structure (FXXLF) and the C′-terminal Val-Pro motif remained unchanged, while the N′-terminal residues surrounding the first proline (positions -1, -2, and -3) were varied. First, the ideal position of the first proline was investigated. CBIs were designed with varying proline positions and two other neutral
amino acids of high helical propensity – namely the polar glutamine and hydrophobic leucine – which most likely exert little effect on the binding affinity of the peptide inhibitor. The circular dichroism (CD) data support the predictions made by MD simulations, in that the movement of the proline from position -1 through -2 to -3 resulted in a more extended α-helix ($\theta_{222}\text{nm}/\theta_{208}\text{nm}; 11 > 10 > 9$, Table 5-1). Competitive binding studies showed that Pro-2 has similar binding compared to Pro-3, both with strong inhibitory properties ($10, K_i = 1.1; 11, K_i = 1.8 \mu M$). Co-crystal structures with other coactivators peptide sequences have shown that the interaction with the backbone and the charge clamp does not occur at the terminus of the helix, but instead at the amide bond located between the first two amino acids (here, Pro-Gln, if Pro-2 = N1) in the case of the interaction with the charge clamp residue E897. At the opposite end of the charge clamp, K720 interacts with the amide bond situated between the fourth and third last amino acid residues (here, Leu-Phe, if Pro+4 = C'). Combining this knowledge with a helix contribution of 1.5 Å per residue and taking a typical hydrogen bond distance of 2.7 - 3.1 Å into account, results in a theoretical, optimal α-helix length of 9.8 amino acids. Peptides with proline at position -2 or -3 are matching this prediction with a helix length of 9 and 10 residues, respectively (assuming prolines in N1 and C' position) (Figure 4-8, Figure 5-7B). Thus, the results from the binding studies confirm the theoretical ideal length defined by the flanking prolines, and provide a point of entry to address NR subtype selectivity based on helix length recognition. Interestingly, cluster analysis of the AR-specific motifs (Figure 5-8B) showed a preference for proline at the -2 position.

After identifying the optimal position of the first proline, the position between this residue and the FXXLF motif was next investigated (position -1): H-P-(X)-E-Q-N-L-E-Q-S-V-P. Lysine and glutamic acid were chosen, representing positively (13) and negatively charged (15) residues, respectively. Furthermore, a hydrophobic (14) and polar (12) amino acid were introduced. Perhaps unsurprisingly, 15, bearing a glutamic acid at the -1 position, displayed the weakest binding properties. Crystal structure analysis of AR-LBD revealed that the charge clamp residues, E897 and K720, - conserved across all NRs - form hydrogen bonds with the backbone of coactivators, and in this way determine the size of the axial length. Together with the adjacent residues E893 and E709, E897 forms a cluster of negative charge on one side of the AF-2. Introduction of a negative amino acid at the positions adjacent to the first proline (15) resulted in electrostatic repulsion and thus in a weaker binding affinity. On the other hand, mutation to a positively charged lysine (13) resulted in significantly enhanced binding ($K_i = 3.5 \mu M$). These results are in line with the LBD screening of both AR and also ER, which predominantly identified positively charged amino acids amino-terminal to FXXLF or LXXLL, respectively. Furthermore, the cluster analysis of nuclear receptor coactivators, including the p160 family, and specifically androgen receptor associated (ARA) protein family, including ARA54/55/70, show the occurrence of positive residues amino-terminal to the LXXLL and FXXLF motif (Figure 5-8). Although, negative amino acids
residues are also found in this region, especially in the case of the ARA family, they are always in direct proximity to a neutralizing positive charged amino acids, while positive residues typically stand alone.  

The effects of introducing polar residues (12 and 5) compared to the hydrophobic leucine (14) were less dramatic. Yet the hydrophobic residue still invoked an order of magnitude greater inhibition (Ki = 4.1 vs. 38.6/42.6 μM). This is perhaps as expected, since this specific PPI is based on hydrophobic interactions, including leucines or phenylalanines. In general, hydrophobic residues are important for the interior of both protein monomers and protein-protein interfaces, having a large positive entropic effect, which is directly connected to the release of water molecules from the cavities of hydrophobic surfaces upon protein folding and PPI, respectively. Although the influence on PPIs is not as strong as for protein folding, it remains the principle driving force. However, salt bridge formation, including hydrogen bonding, is also contributing to these interactions because of the unfavorable, hydrophobic surfaces of unassociated monomers exposed to the polar environment and the greater possibilities for selectivity necessary in the case of PPIs.  

![Figure 5-8](image)

**Figure 5-8 | Cluster analysis of NR and AR coactivators**

(A) Cluster analysis of NR coactivators featuring the LXXLL motif (n = 40) and (B) coactivator interacting nearly exclusively with AR via the FXXLF motif (n=5). [Weblogo by Crooks et al.](#)  

In addition, mutation of the histidine at position -3 of the peptide (14, His-3), which played a major role in ERβ binding (chapter 4), to an uncharged glutamine (10, Gln-3) led to enhanced inhibition of AR-coactivator binding (Ki = 4.1 to 1.1 μM). Comparison of the AF-2 of both NRs reveals that ERβ has a negatively charged residue E287 to which the positively charged His-3 is pointing towards, while AR possesses a neutral amino acid (Q738) at the equivalent position. Thus the difference in residue preference between these two NRs could be explained by the subtraction of charge in this part of the AF-2. Histidine (positively charged and a hydrogen bond donor) and glutamic acid (negatively charged and a hydrogen bond acceptor) are good interaction partners in terms of salt bridge formation, through a combination of hydrogen bonding and electrostatic interactions. On the other hand, glutaminines are both hydrogen donor and acceptor, and thus well suited for hydrogen bond formation with each other.  

Subsequently, the most potent inhibitor from this study was tested for its selectivity. Thus, both another member of the subclass of steroid receptors (NR3) - AR - and a member of the
class of retinoid receptors (NR2) – RXRα - were tested for their binding properties to bind 10. In both cases no binding could be detected, suggesting that the peptides found during this line of investigation show exclusive inhibition for AR. Final cellular studies confirmed the good binding properties of 10 and 13. Interestingly, although both peptides showed weaker binding compared to 1 in the biochemical assay (polarization assay), in this mammalian one hybrid set-up 13 showed a similar and 10 even four-fold enhanced ability to activate luciferase expression (Figure 5-6). This can be explained by the fact that in this binding assay receptor activity can be obtained by the intact protein compared to only the LBD in the biochemical assays. And especially for AR it is known that the N/C interaction between NTD and LBD plays a major role in the overall transcriptional activity.18,50,51

5.8. Conclusion

In conclusion, a successful conceptual approach for developing selective AR CBIs has been described. The rational in silico design was based on knowledge from investigations into the PPI between the AR-related ER and its coactivators in which prolines were identified as helix regulators (breaking and stabilization) resulting in enhanced inhibitory properties. MD simulation studies revealed that phenylalanines - present in the AR-exclusive FxxFL binding motif - contribute less to the formation of a well-structured α-helix compared to leucines (LXXLL motif). Thus, for the introduction of both FXXLF motif and flanking prolines the scaffold sequence had still to be optimized. A combination of the NCoxA1-Box2 sequence, known to bind various NRs, and a sequence based on the N’-terminal domain of AR (AR-NTD) showed the best propensity to form an α-helix and was therefore used as the basic sequence. This combination of the unique AR-binding ability of phenylalanine-rich motifs with helix breaking prolines resulted in short peptides with moderate binding properties (Kd/ki values in the high μM-range) for the AR-LBD. Further optimization focusing on the second proline position (Pro+4 >> Pro+3), the first proline position (Pro-2 > Pro-3 >> Pro-1), the -1 position between Pro-2 and FXXLF motif (K > L >> Q > E), and position -3 prior to Pro-2 (Q > H) culminated most notably in a 12-mer peptide with promising CBI properties (10, Kᵢ = 1.1 μM). Cellular studies involving intact AR identified the ability of 10 to overpower known AR-AF-2 interaction partners, such as the AR-NTD or AR-associated protein 54 (ARA54). Thus, this peptide-based inhibitor can be seen as a lead compound targeting the AR AF-2 surface and thus AR activity in preferential manner, based exclusively on natural amino acids and without recourse to additional chemical modifications. The fact that it interferes one step downstream of classical ligand-competing drugs opens new possibilities in drug selectivity and treatment of both androgen-resistant prostate cancer and diseases based on ligand-independent misregulation.
5.9. Experimental

General
MD simulation studies were performed by Dr. Lidia Nieto. The final cellular investigations were performed by Dr. Ingrid J. de Vries-van Leeuwen.

If not stated otherwise the following contents, descriptions, protocols, and distribute are valid: chemicals were ordered from SIGMA-ALDRICH. Proteins were handled at 4 °C. Standard culturing medium is lysogeny broth (LB) medium (10 g peptone, 10 g NaCl, 5 g yeast, 11 water, autoclaved). Peptide sequences:
- SRC1-Box2, SLTERHKILHRLLQEGPSD, LXXLL motif, natural NR coactivator (Ncoa1,002-711)
- FLETT-1, CSSRFESLFAGEKESR, FXXLF motif, phase display hit AR-LBD15
- D22, LPYEGLIIKLLTRPVEEV, LXXLL motif phase display hit ERα-LBD14
- AR-NTD, KTYRGAFQNLFPQSVRE, motif FXXLF, NTD of AR, binding to AR-AF-2 (AR17-32)
- ARA54, NDGSPCENRLFYAVDVDD, motif FXXLF, specific AR-coactivator (RNF1447-463)

Molecular Dynamics Simulation
All MD simulations were carried out using the Assisted Model Building with Energy Refinement (AMBER) suite of programs, and the ff03 force field.52 An implicit solvent was used via the General Born solvation method (GB 5), as implemented in AMBER.53 Starting with an extended initial conformation, built by the LEaP module of AMBER for each peptide, all MD simulations were fully unrestrained and all trajectories were generated using the sander program in the AMBER 9 package. Each peptide was simulated for a total of 20 ns at 51.9 °C, following a similar approach as previously reported to predict the conformation of a miniprotein.52 Analysis was performed on the 20 ns using 20,000 trajectory snapshots spaced every 1 ps. A snapshot with the lowest potential energy across the simulation was chosen as representative structure for each peptide. The secondary structure of each residue as a function of time was subsequently analyzed utilizing the STRIDE secondary structure assignment algorithms as implemented in VMD.33,34 All MD simulations were performed by Dr. Lidia Nieto.

Protein Expression & Purification

<table>
<thead>
<tr>
<th>construct</th>
<th>vector</th>
<th>tag</th>
<th>protein</th>
<th>origin</th>
</tr>
</thead>
<tbody>
<tr>
<td>AR-LBD</td>
<td>pGEX-KG</td>
<td>GST</td>
<td>hAR-LBDb14-319</td>
<td>BAYER HEALTHCARE PHARMACEUTICALS</td>
</tr>
<tr>
<td>ERβ-LBD</td>
<td>pET15b</td>
<td>His6</td>
<td>hERβ-LBDa100-302</td>
<td>BAYER HEALTHCARE PHARMACEUTICALS</td>
</tr>
<tr>
<td>RXRα-LBD</td>
<td>pGEX-4T1</td>
<td>GST</td>
<td>hRXRα-LBD121-162</td>
<td>FOLKERSTMA et al.34</td>
</tr>
</tbody>
</table>

Plasmids with the desired protein construct (Table 5-2) were transformed into E. coli BL21 cells. Emerging colonies were cultured over night in 25 ml LB medium with 100 μg/ml ampicillin at 37 °C. This preculture was added to 21 phosphate buffered, content-rich terrific broth (TB) medium (12 g peptone, 24 g yeast, 4 ml glycerol, 0.17 M KH2PO4, 0.72 M K2HPO4, 11 water, autoclaved) with ampicillin and incubated at 37 °C until an optical density at 600 nm (OD600) of 0.6 - 1.0. Addition of 100 μM Isopropyl β-D-thiogalactopyranoside (IPTG) induced expression of the protein of interest. For stabilization 10 μM of ligand (ER, estradiol | AR, dihydrotestosterone | RXR, agonist LG) was
Chapter Five

added. After incubation for 20 h at 15 °C the cells were harvested by centrifugation (15 min at 7,000 rpm) and the resulting pellet was stored at -80 °C until further usage.

If not stated otherwise, the protein purification was performed at 4 °C. The pellet was resuspended in lysis buffer (LyB-1: 1x PBS tablet, CALIBIOCHEM, 370 mM NaCl, 10% (v/v) glycerol, 0.1 mM TCEP, 1 mM PMSF, 1 µg/ml DNaseI, 10 µM ligand, pH 8.0, His6-tag; additional 40 mM imidazole) and lyzed with the Emulsi Flex-C3 homogenizer (2 passes of 150,000 kPa, AVESTIN INC.). After another centrifugation step (40 min, 20,000 rpm) the supernatant was provided on the respective affinity chromatography column (GST-tag: Protino GST/4B, 1 ml, MACHERY-NAGEL l His6-tag: His-Bind Resin, 3 ml, NOVAGEN), washed with wash buffer (WaB-1: GST-tag: 1x PBS tablet, CALIBIOCHEM, pH 7.3 l His6-tag: 1x PBS tablet, CALIBIOCHEM, 370 mM NaCl, 40 mM imidazole, 10% (v/v) glycerol, 0.1 mM TCEP, 10 µM ligand, pH 8.0) and eluted with competing molecules (GST-tag: 50 mM Tris, 10 mM glutathione l His6-tag: WaB-1 plus 100 mM / 250 mM / 500 mM imidazole). Fractions were analyzed with SDS-PAGE and optional with LC-MS.

Solid Phase Peptide Synthesis & Purification
The synthesis of all peptides was performed on solid support, using the Fmoc strategy. If not stated otherwise, 200 µmol (1 eq.) of Rink Amide MBHA resin (0.59 mmol/g, Novabiochem) was swollen for 30 min in N-methylpyrrolidone (NMP). The protective Fmoc group was removed by incubation with 20% piperidine in NMP (2x5min). Subsequently the amino acids (AA, 4 eq.) were activated using 2-(1H-Benzotriazole-1-yl)-1,1,3,3-Tetramethyluronium hexafluorophosphate (HBTU, 4 eq.), and N,N-Diisopropylethylamine (DIPEA 8 eq.). Single coupling was performed by shaking 45 min at room temperature. After finishing the desired peptide sequence, the resin was washed 3 times alternately with diethyl ether (DEE) and dichloromethane (DCM) and dried under vacuum for 10 min. The side chain protection group and the cleavage of the peptide sequence from the resin were simultaneously achieved by the usage of mixture of trifluoroacetic acid, triisopropylsilane and water (TFA/TIS/H2O, 95/2.5/2.5) for 3 h. The free peptide was precipitated by adding it drop-wise to ice cold DEE. A final centrifugation step (2.000 rpm, 10 min) with subsequent washing (ice cold DEE), uptake in water and lyophilization results in the crude peptide. Purification was accomplished by reverse-phase HPLC on an Alltima™ HP C18 column (125x20 mm, Alltech). Water (0.1% TFA) was used as a polar phase, adding different amounts of an apolar acetonitril (ACN, 0.1% TFA) phase. A linear gradient (20 ml/min) was optimized for each peptide to a 5% range, for instance 35-40% ACN. The purity of the peptide was determinate at an analytical LC-MS using GraceSmart RP18 (50x2.1mm, Grace, 3u, 120A) column. Finally, the peptides were lyophilized and stored at -80 °C.

Fluorescence Polarization and Depolarization assay
Both fluorescence polarization (FP) and fluorescence depolarization (FDP) experiments were measured in black 384-well-plates (PERKIN ELMER 384 F) on a Safire2 (TECAN) plate reader. The final volume of one well was 30 µl containing constant concentrations of ligand (5 µM) and Fluorescein-labeled peptide (0.1 µM, ER: NCoA-1/SRC1, LTERHKIHHLLLQEGSPSD; AR: FLETT-1, CSSRFSFLAGEKESR, phage display hit AR-LBD; RXRα: D22, LPYESGLLKLLRAPVEEV, INVITROGEN, phage display hit ERa-LBD, good binding properties RXRα). In case of FP assay the protein was sequentially diluted (24 steps) in TR FRET coregulator buffer E (INVITROGEN). On the other hand, during FDP assays the protein concentration was kept constant (ER: 0.4 µM; RXR, AR: 1 µM), varying the concentration of the potential peptide-based binder. Briefly, in 96-well-plates dilution series were prepared in 55 µl and filled up with to 110 µl with a 2X master mix solution containing all non-varying components. Finally three times 30 µl was transferred to 384-well-plates,
centrifuged (1,000 rpm, 2 min) and incubated for 1 h at 4 °C. Plates were measured 50 times at 30 °C (excitation 470 nm, emission 519 nm). Polarization \( y \) (in [mP]) was plotted against the concentration \( x \) of either the protein (FP) or the test peptide (FDP), whereby each data point represents an average of 3 experiments.

The dissociation constant \( K_d \) of the protein-peptide complex (FP) was calculated using equation 1:

\[
P = P_0 + \frac{(P_{in}-P_0)(A_0+B_0+K_d[A_0+B_0+K_d]^{2-4A_0B_0})}{2A_0}
\]

where \( P \) is the measured polarization value (in mP), \( P_0 \) is the polarization of the free fluorescent ligand, \( P_{in} \) is the polarization of the bound ligand, \( A_0 \) is the total concentration of fluorescent peptide, and \( B_0 \) is the protein concentration.

In the competitive setup the half maximal inhibitory concentration \( (IC_{50}) \) was calculated with the means of equation 2a and 2b:

\[
y = A_1 + A_2-A_1 \cdot \frac{10^{(\log (x_0)-x)p}}{1+10^{(\log (x_0)-x)p}} \quad IC_{50} = 10^{\log (x_0)} \quad \text{equation 2a,b}
\]

where \( A_1 \) is the bottom asymptote, \( A_2 \) is the top asymptote, and \( p \) the hill slope (steepness of the curve).

The inhibitory constant \( K_i \) of the different peptides could be determined with the \( K_d \) value of the protein and equation 3a and 3b.

\[
K_i = \frac{IC_{50}}{\frac{A_0}{2K_D(\frac{A_0}{K_D})}+y_0} - K_D \cdot \left( \frac{y_0}{y_0+y_2} \right) \quad y_0 = \frac{[AB]}{[A]} = \frac{[B]}{K_D} \quad \text{equation 3a,b}
\]

where \( K_d \) is the dissociation constant of the fluorescein labeled peptide protein complex (equation 1), \( y_0 \) is the initial bound to free concentration ratio for labeled peptide, \([AB]\) is the concentration of protein-peptide complex, \([A]\) the concentration of unbound peptide, and \([B]\) the concentration of protein (ER).

**Circular Dichroism spectroscopy**

Far-UV Circular Dichroism (CD) spectroscopy measurements were performed under a constant nitrogen flow at 20 °C using 50 μM of the peptide in the presence of 30% (v/v) trifluoroethanol (TFE), if not stated otherwise. As a baseline a peptide-free buffer (30% TFE) was measured under identical conditions. The spectrum was recorded from 250 nm to 185 nm using a JASCO-815 spectrometer. Quartz cuvettes with path lengths of 1 mm or 0.2 mm (HELLMA ANALYTICS) were employed with a data pitch of 0.5 in a continuous mode, a scan speed of 20 nm/min with a response time of 2 s and a bandwidth of 0.5 nm. The graphs are representing an average of five scans. The observed ellipticity \( \theta_\lambda \) (degrees, in [mdeg]) was converted to mean residue ellipticity \( \theta_{mrw,\lambda} \) (in [deg cm²/dmol]) using equation 3.

\[
\theta_{mrw,\lambda} = \frac{M}{(N-1)} \cdot \frac{\theta_\lambda}{10^d\cdot c} \quad \text{equation 3}
\]
Where $M$ is the molecular mass of the peptide (in [Da]), $N$ is the number of amino acids in the peptide, $d$ is the path length (in [cm]), and $c$ is the peptide concentration (in [g/ml]). Information about helical propensity in semi-quantitative manner could be shown by the ratio of the characteristic wavelengths at 208 nm and 222 nm: $\frac{\theta_{\text{mrw,222nm}}}{\theta_{\text{mrw,208nm}}}$.

**Mammalian One Hybrid assay**

All Mammalian One-Hybrid assays (M2H) were performed in human osteosarcoma cells (U2-OS\textsuperscript{37}) seeding 30,000 cells/well in a 24-well plate (BD Biosciences). Plasmids are either from Mammalian Two-Hybrid Assay Kit (AGILENT) or cloned by Dr. Ingrid J. de Vries-van Leeuwen. Culturing conditions were 37 °C and 5% CO\textsubscript{2}. Cells were transfected with following amounts of DNA/well using linear polyethylenimines (PEI, POLYSCIENCES): (1) 40 ng of the bait construct pCMV-BD-PEP containing peptide of interest (PEP) fused to the DNA-binding domain of the GAL4, (2) 40 ng of the target construct pDNA3.1-AR, (3) 200 ng of the reporter plasmid pFR-Luc with the synthetic promoter with five tandem repeats of the GAL4 binding sites that control expression of the *Photinus pyralis* firefly luciferase, and (4) 3.2 ng of control plasmid *Renilla*-Luc, expressing independently *Renilla reniformis* luciferase. After 16 h incubation cells were treated with 10 nM estradiol and after additional 24 h incubation the interaction was determined with a Dual-Luciferase Reporter Assay (PROMEGA), according to manufacturer’s instruction. The luminescent intensities were recorded on a Synergy HT Multi-Mode Microplate Reader (BiOTEK). The interaction-dependent Firefly-luciferase signal was first normalized over the independent *Renilla*-luciferase signal and subsequently normalized over the ratio of untreated cells.
5.10. References
CHAPTER SIX

Modulation of retinoid X receptor activity by biaryl natural products

Abstract. The retinoid X receptor (RXR) is an important transcription factor involved in the regulation of gene networks connected to cell growth, cell differentiation, and cell death. Misregulation of RXR or pathways that are based on RXR heterodimer formation with other nuclear receptors (NRs) can lead to various malignancies, such as cancer, cardiovascular and inflammatory diseases. Small molecules targeting the receptor are used in the treatment of those diseases. Natural products (NPs) are privileged structures for the discovery of new drugs since they are intrinsically favored to bind to enzymes or protein receptors. Furthermore, NPs typically have biological relevance and cell permeability, which are advantageous properties for the development of novel orally active drug therapies. Thus, microarray-based nuclear receptor-coregulator interaction profiling was performed to screen a small compound library of biaryl NPs originating from the Magnolia species for potential RXR binding. Magnolol and 4’-O-methyl-honokiol featured (partial) agonism functionality. The microarray profiling assay additionally identified honokiol itself to feature non-competitive antagonism through allosteric RXR binding. Both the lipophilic side chain groups and the hydroxyl functionality of these NPs were shown to be important aspects of their activity. Modifications to the two lipophilic groups of honokiol, towards an improved mimic of the structure of leucine side-chains of coactivator peptides, resulted in an improved ability to inhibit the RXR–coactivator interaction in the profiling assay. This data in combination with the structural similarity with known NR-CBIs identifies honokiol as the first NP-CBI.
6.1. Introduction

Nuclear receptors (NR) are ligand-activated gene transcription factors that determine a host of fundamental (patho)-physiological processes including embryonic development, diabetes and cancer.\(^1\) Ligand binding to the lipophilic NR ligand binding pocket (LBP) induces a protein conformational change, which leads to the recruitment of co-activator proteins as a prerequisite for the orchestrated assembly of the transcriptional machinery.\(^2\) Therefore, NRs represent important drug targets.\(^3\) Screening natural products (NPs) is one possibility to discover new compounds that can function as lead structures in the development of drugs for the treatment of human diseases.\(^4\) NPs offer a rich source of structurally diverse, biologically relevant and cell permeable small molecules or scaffold structures, which achieve high hit rates across a broad range of in vitro and cell-based screens compared with scaffolds based on combinatorial chemistry.\(^5\) Indeed, the success of NPs translates at the clinical level, where a high percentage of marketed drugs are either whole natural products, natural product analogues, or natural-product-like.\(^6\) The privileged status of NPs has even been attributed to the notion that, as metabolites of conservative biosynthetic processes, they are evolutionarily pre-disposed to re-interact with protein-based targets.\(^6\)\(^-\)\(^11\)

In the field of NRs several NPs have been discovered to have NR interacting properties, aside from the native NR ligands such as steroids. One of the best characterized NR-targeting NP is the isoflavone genistein. It was isolated from soybeans, directly interacts with estrogen receptors (ER\(\alpha\), NR3AI; ER\(\beta\), NR3A2) and reminiscent of the endogenous ligand 17\(\beta\)-estradiol. It is used as dietary supplement to ease menopausal symptoms.\(^12,13\) Guggulsterone - isolated from the guggul tree\(^14,15\) - has antagonistic properties for the farnesoid X receptor (FXR, NR1H4), is similar to androstenedione, a precursor for estrogens and androgens in steroidogenesis;\(^16\) and is marketed in India as an antihyperlipoproteinemic drug.\(^17,18\) Interestingly, Swinhosterol B - isolated from the sponge Theomella swinhoei - features a dual function, vidental agonism in terms of PXR (NR1I2) and antagonistic properties for FXR (NR1H4).\(^19\) Recently, several NPs were discovered that block androgen receptor (AR, NR3C4) transcriptional activity in prostate cancer, namely epigallocatechin-3-gallate from green tea,\(^20\) atricar from plant Pygeum africanum,\(^21\) lupeol found in fruits vegetables and medicinal plants,\(^22\) and naphthenone B from the marine sponge Niphates digitalis.\(^23\) Interestingly, the latter binds to the activation function-1 (AF-1) of the N'-terminal domain of AR instead of the LBP of the ligand binding domain. Advantages of such alternative binding are both the circumvention of hormone resistance and the reduction in side effects due to combinatorial treatment involving lower doses of LBP-targeting antagonists.\(^24\)

These benefits have also stimulated investigations of activation function 2 (AF-2) – the binding platform for coactivators – as a target for small molecule intervention instead of the
LBP. A successful strategy for designing coactivator binding inhibitors (CBIs) has been to mimic the structural features of peptides commonly shared across all NR coactivators (peptidomimetics). Pyrimidine scaffolds mimicking the i, i + 3, i + 4 arrangement of LXXLL (ER) or FXL (AR) were developed to selectively target ER and AR, respectively (Figure 6-1, A and B).25-28 Other studies have indentified biaryl-based small molecules, such as pyridylpyridones (Figure 6-1C)29 or biphenyls (Figure 6-1D)30 targeting ER-AF-2 or more specifically the surface charge clamp. High-throughput screening (HTS) of large, 10,000 strong, compound libraries in combination with follow-up optimization studies (SAR/X-ray screen32) has provided an alternative approach to identifying inhibitors of AF-2 binding of ER (Figure 6-1E),33,31 AR,32 but also AR binding function 3 (BF-3), which is an allosteric binding site in proximity to AF-2 (Figure 6-1F).32

Figure 6-1  |  Nuclear receptor coactivator binding inhibitors
Small lipophilic drug-like molecules targeting activation function 2 (AF-2) or binding function 3 (BF-3) of NRs: (A, B) peptidomimetic pyrimidines targeting ER and AR, respectively,28 (C) pyridylpyridone binding to ER,29 (D) biphenyl proteomimetic interacting with ER,30 (E) quinazolinone derivatives,31 (F) pyrazolo pyrimidine targeting AR.32

NPs are also interesting as potential lead structures for the development of novel antagonists of the retinoid X receptors (RXRα, NR2B1; RXRβ, NR2B2). RXR is a type II NR, which on binding of the putative endogenous ligand, 9-cis-retinoic acid, can either homodimerize or form heterodimer pairs with other type II NRs such as the peroxisome proliferator-activated receptors (PPARs, NR1CX) and the vitamin D receptor (VDR, NR1I1).34-37 RXR plays a major role in the regulation of the gene networks connected to cell growth, cell differentiation, and cell death.37 Rexinoids, RXR selective ligands, are used in the treatment of cutaneous T-cell lymphoma and additional are in clinical trial for both breast and lung cancer.38-41 Furthermore, via modulation of pathways based on RXR heterodimer formation, including LXR and PPAR, such compounds could be useful in the treatment of
cardiovascular and inflammatory diseases, including diabetes, obesity, metabolic syndrome, and atherosclerosis.\textsuperscript{39,42-47} In two recent independent studies the biaryl NP honokiol (Figure 6-2, 1) has been discovered as a suspected partial RXR agonist.\textsuperscript{48,49} From a drug discovery perspective, this molecule is interesting due to its high molecular efficiency, its ready availability and ease of synthesis. Furthermore, biaryls are privileged structures for drug discovery, since they have shown a high capacity to bind a range of different proteins.\textsuperscript{50}

![Biaryl neolignan secondary metabolites isolated from Magnolia species](image)

Honokiol and its natural isomer magnolol (Figure 6-2, 2) are major components of the dried bark of the \textit{Magnolia (M.) officinalis} and \textit{M. obovata}, which for centuries has been used in traditional eastern medicine for the treatment of a variety of ailments, including indigestion, anxiety and depression.\textsuperscript{51} More recently, 1 and 2 have been isolated along with related biaryl metabolites, such as 4'-O-methyl-honokiol and 2'-O-methyl- honokiol (Figure 6-2, 3 and 4, respectively) and have since exhibited a range of beneficial biological effects, including the induction of neurite growth, neuronal protection and anti-angiogensis.\textsuperscript{52-54} The more complex neolignan sesquiterpene analogues, clovanemagnolol and caryolanemagnolol (Figure 6-2, 5 and 6, respectively), also display intriguing neuronal properties,\textsuperscript{55} which has stimulated recent synthetic efforts and subsequent phenotype studies.\textsuperscript{56-62}

### 6.2. Initial screen of biaryl natural products isolated from Magnolia species

First insights into the ability of RXR\textsubscript{\beta} to interact with downstream proteins upon interaction with Magnolia-derived biaryl natural products were obtained via a microarray assay for nuclear receptor-coregulator interaction profiling. This assay measures the ability of NRs to recruit a range of biologically relevant co-activator peptides to the receptor surface (Figure 6-3).\textsuperscript{63} The small compound library featured alongside commercially available 1 and
2, the known natural products mono-O-methyl analogues of 1 (3 and 4, respectively) and the more complex neolignan sesquiterpenes 5 and 6 (Figure 6-2, compounds 3-6 were synthesized by LGM, LHP, CVL, MS). \(^{55}\)

Figure 6-3 | Preliminary screen of biaryl natural products isolated from the Magnolia species. (A) Representation of the principle of the microarray-based profiling assay. Details and immobilized peptides sequences: 6.9 Experimental, Nuclear receptor-coregulator interaction profiling, Table 6-2.\(^{63}\) (B) Initial nuclear receptor-coregulator interaction profiling at 200 \(\mu\)M compound concentration identifies three different functionalities for biaryl compounds from Magnolia: honokiol (1) shows full inhibition of basal RXR-activity, magnolol (2) and 4’-O-methyl-honokiol (3) show either partial activation or did not reach their maximum, and the others (4-6) show no effect. As positive control: synthetic RXR agonist LG100268 10.\(^{64,65}\)

Each of the NPs showed a unique NR-coregulator interaction profile. While 2 showed enhanced overall binding compared to a DMSO control, 1 in comparison showed nearly total suppression of all interactions (Figure 6-3, 1 and 2, respectively). The neolignan sesquiterpene analogues 5 and 6, however were found not to influence the RXR\(\beta\) coregulator interaction. Interestingly, methylation of the 4’-hydroxyl group of 1 (resulting in 3) corresponded with an enhancement of coregulator binding, similar to 2. Methylation of the
second 2′-hydroxyl group (resulting in 4), however, led to a compound silent on coactivator binding (Figure 6-3, 3 and 4, respectively).

6.3. Concentration dependent effects of honokiol with and without agonist

The suppressive function of 1 on RXRβ cofactor binding was subsequently tested at different concentrations (0.01 - 100 µM). While lower concentrations between 0.01 µM and 0.1 µM showed only low or no inhibition of the RXRβ coregulator interaction, higher concentrations, however, lowered the binding affinity significantly in a dose dependent manner (Figure 6-4A). The introduction of the potent RXR agonist LG100268 (10) at 100 nM did not prevent 1 from binding to RXRβ. Although receptor activity was initially enhanced by 10 resulting in a positive effect (>intrinsic receptor activity), this was reversed due to the function of 1 to block cofactor interaction at higher concentrations. It is important to mention at this stage that the final negative effect of 1 - the ability of 1 to inhibit coactivator binding - was not affected by the presence of 10 (Figure 6-4B). This indicates that 1 and 10 are not competing for a binding site on the receptor surface.

![Figure 6-4](image)

**Figure 6-4 | Concentration-dependent effect of honokiol on the RXR-coregulator interaction**
Influence of increasing concentrations of 1 on RXR-coregulator interaction. (A) Without and (B) in competition with 100 nM agonist 10.

6.4. Insights into the honokiol RXRβ interaction

A more indepth characterization of the effect of RXR on the concentration-dependent effects of 1 was performed using a polarization assay. Agonist-activated RXRβ-LBD (3.2 µM of 10) bound to the fluorescein-labeled peptide D22 - indentified from phage display screening against ERα - afforded good binding affinity to RXR. Addition of increasing concentrations of 1 to this agonist activated receptor resulted in the disruption of the RXRβ – peptide interaction at high concentrations of 1. Importantly, this inhibitory function of 1 was insensitive to the concentration of agonist 10 (Figure 6-5A), supporting again the notion of different receptor interaction sites. In a separate experiment, increasing concentrations of 10
produced a gradual activation of RXRβ, which leveled off at saturating concentrations of the agonist. The addition of fixed concentrations of 1 to the receptor resulted in a lowering of the maximal effect achievable by 10. In contrast, the values for the half maximal effective concentration remained constant upon addition of 1 (EC₅₀, Figure 6-5B). Taken together these results thus identified 1 as a non-competitive antagonist - most probably a coactivator binding inhibitor (CBI) - which does not target the ligand binding pocket of RXRβ.

![Figure 6-5](image)

**Figure 6-5 | Polarization assay of RXRβ with honokiol and agonist 10**

(A) Increasing concentrations of 1 led to a decrease in the ability of RXRβ - activated by 10 – to recruit D22, which binds to the coactivator binding surface. Higher concentrations of 10 did not change the inhibitory property of 1. (B) Increasing concentrations of 10 led to activation of RXRβ, which resulted in an increase in polarization. The presence of 1 resulted in a lowering of the maximal effect, but did not change the half-maximal effective concentration (EC₅₀).

### 6.5. Studies on the pharmacological profile of honokiol

Getting insights in the biological relevance of honokiol as possible CBI both cellular and selectivity studies were performed. In a mammalian two-hybrid assay format 1 showed agonistic properties at a concentration of between 10-25 μM. This is in line with other studies that reported a partial agonistic effect of 1 for RXR. A further increase in concentration, however, led to a reduced luciferase activity. This reduced signal at 50 μM could also be observed for the same experiment in the presence of 10 nM of 10. However, cytotoxicity of 1 rather than CBI-activity might be an explanation for this decrease in signal.

The selectivity of 1 for RXR was evaluated on chip performing a NR-coregulator interaction profiling for the LBDs of RXRα, RXRβ, ERβ and AR in presence of an respective agonist. Receptor cofactor binding was evaluated in the presence of 100 nM agonist alone (RXR: 10, ER: estradiol, AR: dihydrotestosterone) and in combination with 100 μM of 1 as an intended non-competitive antagonist. Both RXRα and RXRβ showed a significant decrease in their ability to recruit cofactors upon addition of 1 compared to the receptor treated with the agonist alone. However, neither ERβ nor AR were effected by 1 (Figure 6-6). Honokiol, 1,
thus has exclusive inhibitory properties for RXR at the tested concentration, but does not discriminate between different RXR subtypes.

---

**Figure 6-6 | Biological relevance of honokiol (cellular activity/selectivity)**

(A) In a mammalian two-hybrid assay the influence of 1 on RXR activity was tested with and without 10 μM of agonist 10. (B) The ability to recruit cofactors in the presence of an agonist is blocked by 1 (A) in the case of RXR subtypes α and β, (C) but not in the case of ERβ or AR. Compounds: honokiol (1), RXR agonist LG100268 (10), ER agonist estradiol (E2) and AR agonist dihydrotestosterone (DHT).
6.6. Modification of honokiol towards exclusive targeting of the RXR AF-2

Jung et al. and Kotani et al. discovered partial agonistic effects at low concentrations of 1 that is based on binding to the LBP.49,49 Our study revealed additional non-competitive antagonistic effects of 1 at higher concentrations which requires binding to an alternative binding site, possibly the coactivator binding surface (activation function 2, AF-2). To discriminate this dual effect in favor of AF-2 binding, chemical modifications were introduced into 1 to mimic the LXXLL motif of coactivator peptides. The two allyl groups present in 1 were replaced by propyl-, epoxymethyl-, or cyclopropylmethyl side chains as leucine side chain mimics with varying degrees of lipophilicity and steric bulk (Figure 6-7, 7, 8, 11). Furthermore, allyl side-chains were conveniently installed in a speculative manner ortho to the hydroxyl group on the aromatic rings to extend lipophilicity in this region of the molecule in search of improved binding at the lipophilic RXR surface (Figure 6-7, 9, 12).

![Figure 6-7](image)

**Figure 6-7 | Design strategy to enhance binding properties at the alternative binding site**
Modification of 1 to optimize leucine mimic: either by modification of two allyl groups by propyl- (7), 2,3-epoxypropyl- (8), and cyclopropylmethyl substitution (11) or through the introduction of one (9) or two (12) additional allyl groups.

Initial screening identified 7, 8, and 11 to have the best inhibitory properties (data not shown). In a second screen in the presence of 100 μM agonist 10, both the epoxypropyl-substitution (8) and the introduction of one additional allyl group (11) corresponded with a sustained ability to inhibit RXR-coactivator recruitment. However, the activity was slightly reduced compared to 1. Adding more flexibility through substitution of the two allyl groups with propyl side chains (7) even enhanced the inhibitory properties (Figure 6-8).
6.7. Discussion

Profiling the retinoid X receptor β (RXRβ)-coregulator interaction against a set of biaryl natural products (NPs) resulted in unique coregulator binding profiles (Figure 6-2, 1-6). At 200 μM treatment with 2 led to an activation of the receptor regarding cofactor binding, while treatment with 1 on the other hand induced a loss in RXRβ cofactor recruitment ability. Interestingly, modification of 1 to give mono-O-methylated analogues 3 resulted in partial activation of coregulator recruitment (relative to 2), while the isomeric mono-O-methylated analogue 4 was effectively inactive. This would appear to suggest that at 200 μM, 1 functions as an inverse agonist or coactivator binding inhibitor (CBI), inhibiting basal coregulator binding ability of RXR, while analogues 2 and 3 induce coactivator recruitment in a partial manner (partial agonist). It is important to emphasize here how the small differences between the isomers 1 and 2 produces opposite effects. And the minor change from 1 to 3, namely the introduction of a methyl group, seems to switch activity from inhibition to partial agonist behaviour. The effects of clovanemagnolol (5) and caryolanemagnolol (6) in this assay were not significant, though the possibility that these two molecules bind to the LBP as antagonists and block co-activator could not be ruled out at this stage. At low concentrations of 1 no noticeable change in the coactivator profile was observed (Figure 6-4, 10 - 100 nM), which would indicate that inhibition of basal coactivator binding does not occur within this concentration range. 1 might still bind to the LBP. However, no activation of coactivator binding could be detected in the microarray format. At higher concentrations (Figure 6-4, 1 - 100 μM), a dose-dependent decrease in the fluorescence signal was observed similar to what had been observed during the initial 200-μM-screen. In the presence of 100 nM of a known RXR agonist 10, a similar inhibitory pattern was observed. Furthermore, competitive fluorescence polarization studies of 1 revealed that coactivator inhibition was insensitive to
changes in the concentration of the agonist 10 (Figure 6-5). This is in line with the profiling data, and suggests that 1 inhibits coactivator binding via an alternative mechanism and does not function as an inverse agonist targeting the RXR-LBP. Investigating the exclusiveness of RXR binding, the effect of 1 on four different NRs was tested. In this case, the proteins were incubated in the presence of 100 µM of 1 and 100 nM of a known full agonist – E2 (ERβ), DHT (AR), 10 (RXRα and RXRβ) – and the effect on coregulator binding monitored in the microarray format. While only minor changes were observed in the case of type I NRs ERβ and AR, a clear decrease in signal was measured for both RXRα and RXRβ (Figure 6-6). This suggests that 1 behaves as an alternative NR-co-activator inhibitor with selectivity for RXR (albeit without sub-type selectivity), and thus represents a new molecular structure for the potential development of RXR-selective inhibitors.

The results obtained via mammalian two-hybrid assays (M2H, Figure 6-6A) are in line with data from TR-FRET and reporter gene assays reported in the literature. These experiments provide evidence that 1 binds to the RXR LBP at lower, micromolar concentrations (10 - 30 µM). However, as suggested in the M2H assay and confirmed during NR-coragulator profiling, 1 is also capable of displaying an inhibitory effect. Thus, 1 most likely binds the LBP at lower concentrations, leading to weak co-activator recruitment, while at higher concentrations, a second molecule 1 binds at an allosteric site - away from the LBP - resulting in a concentration dependent inhibition of coactivator binding (CBI). The ability of NRs to bind the same molecule at two distinct sites on the protein has already been reported in the case of 4-hydroxytamoxifen and its ability to bind LBP and AF-2 of ER, but would represent a novel finding for RXR. Furthermore, the partial agonist properties measured for 2 (Figure 6-3) are concordant with structural studies showing 2 co-crystallized with RXRα, and suggest that 1 might be equally suited for binding at the RXR LBP as 2.

The use of co-activator binding inhibitors (CBIs) has emerged as a highly promising alternative strategy to classical LBP agonism/antagonism, though RXR-selective CBIs have not been reported thus far. CBIs target the NR-coactivator interaction by mimicking the leucine side chains of the LXXLL peptide motif, which is conserved in several NR coactivators (peptidomimetic). This has typically been achieved through the incorporation of lipophilic side-chains to a rigid, heterocyclic scaffold, leading to some notable successes in the case of ER and AR. Despite this, though, novel CBIs are still needed in order to expand on the current repertoire; especially compounds with selectivity for specific NR protein classes. Interestingly, Katzenellenbogen et al. recently disclosed the development of biphenyl-derived inhibitors of the ER-co-activator interaction, with specific designs for the AF-2. Central to the design of this compound class was the incorporation of lipophilic groups that mimic the leucine side chains of the signature LXXLL motif of the coactivator recognition sequence. In light of the structural similarity between 1 and known NR-CBI’s (Figure 6-1, e.g. D) it could therefore be reasonably hypothesized that the two allyl groups
Chapter Six

featured in 1 might also mimic the same leucine side-chains. The relative orientation of the allyl groups is important in this case, as is the ability of the biaryl bond to rotate enabling an out-of-plane configuration of the allyl substituents. Since 1 is also based on a biaryl core structure it meets all these requirements important for NR-AF-2 targeting.29,30

Investigations into 1 focusing on the optimization of the allylic side chains should strengthen the hypothesis that 1 binds to the AF-2 as CBI and not an alternative site on the receptor surface.71 Thus additional allyl groups were introduced ortho to the hydroxyl groups of 1 (Figure 6-7A, 9 and 12, respectively) or the two existing allylic groups were substituted with propyl (7), 2,3-epoxypropyl (8), or cyclopropylmethyl (11) to produce a small library varying in lipophilicity and possibly featuring higher affinity for RXR AF-2. This approach of sidechain optimization in search of improved binding to AF-2 has been successfully performed for other NRs.72,73 While 9 showed inhibitory activity similar to 1, the analogue containing four allylic subsitutents (12) showed a significantly reduced activity; although activity was not entirely abolished. One explanation could be that the introduction of the fourth allyl group on 1 (or 9) leads to a steric hindrance and a consequent limited access of the hydroxyl group, which play an important role in RXR binding, for example through a hydrogen bond interaction with one of the charge clamp amino acid residues.30 Surprisingly, 11 was one of the weakest binders to emerge from the small library. In contrast, 8 showed a good ability to block RXRβ-cofactor interaction, similar to 1, and analogue 7 even exceeded the inhibitory capabilities of 1. This preliminary data suggests that 1 one is indeed targeting the AF-2 of RXR. However, further evaluation of these compounds required. Structure-activity relationship studies of this natural product 1 - with high molecular efficiency - reveal that minor modifications such as the introduction of a methyl group (1 to 3) or a repositioning of the hydroxyl group (1 to 2) can lead to a switch between partial agonism and coactivator binding inhibition. Furthermore, first trials in the modulation of the inhibitory function through modifications at the side chain positions show promising results.
6.8. Conclusion

Initial microarray-based RXR-coregulator interaction profiling of a focused natural product compound library of biaryl molecules isolated from Magnolia species has resulted in the discovery of three classes of molecule with distinct modes of RXR binding: no agonistic effect with possible antagonistic properties (4-6), partial agonism (2 and 3, respectively), and inhibitory function as an inverse agonist or coactivator binding inhibitor (1). Further characterization in polarization assay suggested that 1 does not target RXR via the LBP, but via an alternative binding site. Subsequent diversification of 1 with the goal to optimize the leucine mimics - known to be important for NR-AF-2 binding - revealed possible gradual modulation through modifications to the allylic side chains (7-9,11,12).

Thus the compounds isolated from Magnolia bark can either directly or via close structural analogues function as (partial) agonist and antagonists - targeting both the LBP - or as CBIs blocking the AF-2-coactivator interaction. Taken together, these results provide further convincing evidence of the power of natural products as a source of novel structures and novel activities for NR-based drug discovery.
6.9. Experimental

General
Chemical synthesis of compound library performed by Dr. Lech-Gustav Milroy, Leslie in Het Panhuis, Chan Vinh Lam, Nicky Hoek and Marcel Scheepstra.

If not stated otherwise the following contents, descriptions, protocols, and distributor are valid: chemicals were ordered from SIGMA-A LD RICH. Proteins were handled at 4 °C. Standard culturing medium is lysogeny broth (LB) medium (10 g peptone, 10 g NaCl, 5 g yeast, 1 l water, autoclaved).

Protein Expression & Purification

Table 6-1 | Used constructs for protein expression

<table>
<thead>
<tr>
<th>construct</th>
<th>vector</th>
<th>tag</th>
<th>protein</th>
<th>origin</th>
</tr>
</thead>
<tbody>
<tr>
<td>AR-LBD</td>
<td>pGEX-KG</td>
<td>GST</td>
<td>hAR-LBD&lt;sub&gt;344-419&lt;/sub&gt;</td>
<td>BAYER HEALTHCARE PHARMACEUTICALS</td>
</tr>
<tr>
<td>ER&lt;sup&gt;+&lt;/sup&gt;-LBD</td>
<td>pET15b</td>
<td>His6</td>
<td>hER&lt;sup&gt;+&lt;/sup&gt;-LBD&lt;sub&gt;220-302&lt;/sub&gt;</td>
<td>BAYER HEALTHCARE PHARMACEUTICALS</td>
</tr>
<tr>
<td>RXR&lt;sup&gt;+&lt;/sup&gt;-LBD</td>
<td>pGEX-4T1</td>
<td>GST</td>
<td>hRXR&lt;sup&gt;+&lt;/sup&gt;-LBD&lt;sub&gt;221-302&lt;/sub&gt;</td>
<td>FOLKERTSMA et al. &lt;sup&gt;74&lt;/sup&gt;</td>
</tr>
</tbody>
</table>

Plasmids with the desired protein construct (Table 6-1) were transformed into E. coli BL21 cells. Emerging colonies were cultured over night in 25 ml LB medium with 100 µg/ml ampicillin at 37 °C. This preculture was added to 21 phosphate buffered, content-rich terrific broth (TB) medium (12 g peptone, 24 g yeast, 4 ml glycerol, 0.17 M KH<sub>2</sub.PO<sub>4</sub>, 0.72 M K<sub>2</sub>PO<sub>4</sub>, 11 water, autoclaved) with ampicillin and incubated at 37 °C until an optical density at 600 nm (OD<sub>600</sub>) of 0.6 - 1.0. Addition of 100 µM Isopropyl β-D-1-thiogalactopyranoside (IPTG) induced expression of the protein of interest. For stabilization 10 µM of ligand (ER, estradiol | AR, dihydrotestosterone | RXR, agonist 10) was added. After incubation for 20 h at 15 °C the cells were harvested by centrifugation (15 min at 7,000 rpm) and the resulting pellet was stored at -80 °C until further usage.

If not stated otherwise, the protein purification was performed at 4 °C. The pellet was resuspended in lysis buffer (1x PBS tablet, CALBIOCHEM, 370 mM NaCl, 10% (v/v) glycerol, 0.1 mM TCEP, 1 mM PMSF, 1 µg/ml DNAseI, 10 µM ligand, pH 8.0, His-tag: additional 40 mM imidazole) and lyzed with the Emulsi Flex-C3 homogenizer (2 passes of 150,000 kPa, AVESTIN INC.). After another centrifugation step (40 min, 20,000 rpm) the supernatant was provided on the respective affinity chromatography column (GST-tag: Protino GST/4B, 1 ml, MACHEREY-NAGEL | His-tag: His-Bind Resin, 3 ml, NOVAGEN), washed with wash buffer (GST-tag: 1x PBS tablet, CALBIOCHEM, pH 7.3 | His-tag: 1x PBS tablet, CALBIOCHEM, 370 mM NaCl, 40 mM imidazole, 10% (v/v) glycerol, 0.1 mM TCEP, 10 µM ligand, pH 8.0) and eluted with competing molecules (GST-tag: 50 mM Tris, 10 mM glutathione | His-tag: wash buffer + 100 mM / 250 mM / 500 mM imidazole). Fractions were analyzed with SDS-PAGE and optional with LC-MS.

Nuclear receptor-coregulator interaction profiling<sup>65,75</sup>

Assay was prepared on ice in three different mastermixes with TR-FRET coregulator buffer (INVITROGEN) with 1 M DTT (RXR: buffer G, ER: buffer E, AR: buffer A): (1) 4X antibody mix, containing 100 nM antibody (GST: anti-GST Alexa488, INVITROGEN; His: Pentahis Alexa488, QIAGEN) (2) 4X protein mix, containing 10 or 40 nM protein (GST-RXR: 40 nM; His-ER, GST-AR: 10 nM), and 2X ligand mix, containing double concentration of final compound concentration in 2% (v/v)DMSO. Combination of them is resulting in 25 µl solution including 25 nM antibody and 2.5 nM or 10 nM
protein and the compound in desired concentration (10 nM – 200 μM). All assays were performed in PamStation®-4 controlled by EvolveHT software (PAMGENE INTERNATIONAL B. V., 's-HERTOGENBOSCH, THE NETHERLANDS) at 28 °C. The Nuclear Receptor PamChip® Arrays (PAMGENE) contain 53 peptides (Table 6-2) immobilized on a 3D-cylindric-structured material with branched pores with 100 μm in diameter and 60 μm in length. This resulted in a approximate 500 times enlarged surface area that is directly correlated to the read-out signal. 20 initial cycles with 25 μl blocking buffer (1% BSA, 0.01% Tween 20 in Tris-buffered saline) are followed by 25 μl of assay mixture. In the time period of 40 minutes the solution containing protein, fluorescent antibody and ligand were pumped up and down (81 times, 2 cycles/min) through the porous membrane with the immobilized peptides (all 53 in parallel). The fluorescent signal was detected every 20 cycles by a charge-coupled device camera-based optical system interacted in the instrument. Image analysis was performed by Bionavigator software (PAMGENE), including automated spot finding and quantitation of signal-minus-background values (100 ms).

Mammalian Two-Hybrid assay
All Mammalian Two-Hybrid assays (M2H) were performed in human osteosarcoma cells (U2-OS76) seeding 30,000 cells/well in a 24-well plate (BD Biosciences). Plasmids are either from Mammalian Two-Hybrid Assay Kit (AGILENT) or cloned by Dr. Ingrid J. de Vries-van Leeuwen. Culturing conditions were 37 °C and 5% CO2. Cells were transfected with following amounts of DNA/well using linear polyethylenimines (PEI, POLYSCIENCES): (1) 40 ng of peptide construct pCMV-LXXLL containing HPLLMLRLLSP peptide - discovered to bind effectively RXRβ - fused to DNA-binding domain of the GAL4, (2) 40 ng of the target construct pCMV-AD-RXRβ, containing RXRβ fused to the transcriptional activation domain (AD) of the mouse protein NF-κB, (3) 200 ng of the reporter plasmid pFR-Luc with the synthetic promoter with five tandem repeats of the GAL4 binding sites that control expression of the Photinus pyralis firefly luciferase, and (4) 3.2 ng of control plasmid Renilla-Luc, expressing independently Renilla reniformis luciferase. After 16 h incubation cells were treated with 10 nM agonist LG100268 with varying concentrations of honokiol (10 nM – 50 μM). After additional 24 h incubation the interaction was determined with a Dual-Luciferase Reporter Assay (PRIMEGA), according to manufacturer’s instruction. The luminescent intensities were recorded on a Synergy HT Multi-Mode Microplate Reader (BioTek). The interaction-dependent Firefly-luciferase signal was first normalized over the independent Renilla-luciferase signal and subsequently normalized over the ratio of untreated cells. All cellular studies were performed by Dr. Ingrid J. de Vries-van Leeuwen.

Nuclear receptor-coregulator interaction profiling83,75
Assay was prepared on ice in three different mastermixes with TR-FRET coregulator buffer (INVITROGEN) with 1 M DTT (RXR: buffer G, ER: buffer E, AR: buffer A): (1) 4X anibody mix, containing 100 nM antibody (GST: anti-GST Alexa488, INVITROGEN; His: PentaHis Alexa488, QIAGEN) (2) 4X protein mix, containing 10 or 40 nM protein (GST-RXR: 40 nM; His-ER, GST-AR: 10 nM), and 2X ligand mix, containing double concentration of final compound concentration in 2% (v/v) DMSO. Combination of them is resulting in 25 μl solution including 25 nM antibody and 2.5 nM or 10 nM protein and the compound in desired concentration (10 nM – 200 μM). All assays were performed in PamStation®-4 controlled by EvolveHT software (PAMGENE INTERNATIONAL B. V., 's-HERTOGENBOSCH, THE NETHERLANDS) at 28 °C. The Nuclear Receptor PamChip® Arrays (PAMGENE) contain 53 peptides (Table 6-2) immobilized on a 3D-cylindric-structured material with branched pores with 100 μm in diameter and 60 μm in length. This resulted in a approximate 500 times enlarged surface area that is directly correlated to the read-out signal. 20 initial cycles with 25 μl blocking buffer (1% BSA, 0.01%
Tween 20 in Tris-buffered saline) are followed by 25 μl of assay mixture. In the time period of 40 minutes the solution containing protein, fluorescent antibody and ligand were pumped up and down (81 times, 2 cycles/min) through the porous membrane with the immobilized peptides (all 53 in parallel). The fluorescent signal was detected every 20 cycles by a charge-coupled device camera-based optical system interacted in the instrument. Image analysis was performed by Bionavigator software (PAMGene), including automated spot finding and quantitation of signal-minus-background values (100 ms).

Fluorescence Polarization assay

Both fluorescence polarization (FP) experiments were measured in black 384-well-plates (PERKIN ELMER 384 F) on a Safire² (TECAN) plate reader. The final volume of one well was 30 μl containing constant concentrations of ligand (5 μM) and Fluorescein-labeled peptide (D22, LPYEGSLI[LK]L1(RAPVEE,V INVITROGEN, phage display hit ERα-LBD, good binding properties RXRα[6]). The protein was sequentially diluted (24 steps) in TR FRET coregulator buffer E (INVITROGEN). Briefly, in 96-well-plates dilution series were prepared in 55 μl and filled up with to 110 μl with a 2X master mix solution containing all non-varying components. Finally three times 30 μl was transferred to 384-well-plates, centrifuged (1,000 rpm, 2 min) and incubated for 1 h at 4 °C. Plates were measured 50 times at 30 °C (excitation 470 nm, emission 519 nm). Polarization y (in [mP]) was plotted against the concentration x of either the protein, whereby each data point represents an average of 3 experiments. The dissociation constant \( K_d \) of the protein-peptide complex (FP) was calculated using equation 1:

\[
y = A_1 + \frac{A_2 - A_1}{1 + 10^{(lo g(x_0) - x)p}} \quad \text{EC}_{50} = K_d = 10^{lo g(x_0)} \quad \text{equation 2a,b}
\]

where \( A_1 \) is the bottom asymptote, \( A_2 \) is the top asymptote, and \( p \) the hill slope (steepness of the curve).
<table>
<thead>
<tr>
<th>position</th>
<th>protein</th>
<th>sequence</th>
<th>motif</th>
<th>uniprot</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>CBP</td>
<td>SVQQPPRSISPSALQDLLRTLKSP</td>
<td>LxxLL2067</td>
<td>Q92793</td>
</tr>
<tr>
<td>2</td>
<td>CBP</td>
<td>TADPEKKRLIQQLQLVLLHHAHKSQ</td>
<td>LxxLL358</td>
<td>Q92793</td>
</tr>
<tr>
<td>3</td>
<td>CBP</td>
<td>GNLVPDAASKHQLQMSELRRGGSQ</td>
<td>LxxLL70</td>
<td>Q92793</td>
</tr>
<tr>
<td>4</td>
<td>DAX1</td>
<td>GEDHPQGSGSLYRTSSKQTHVA</td>
<td>LxxLL146</td>
<td>P51843</td>
</tr>
<tr>
<td>5</td>
<td>DAX1</td>
<td>MAGENHQWQSGILYNLMSAKQT</td>
<td>LxxML13</td>
<td>P51843</td>
</tr>
<tr>
<td>6</td>
<td>DAX1</td>
<td>FSGKDPHRQGSGSLYRTSSKQ</td>
<td>LxxML80</td>
<td>P51843</td>
</tr>
<tr>
<td>7</td>
<td>CBP</td>
<td>TADPEKKRLIQQLQLVLLHHAHKSQ</td>
<td>LxxLL358</td>
<td>Q92793</td>
</tr>
<tr>
<td>8</td>
<td>P300</td>
<td>SPLKPGTVSOGQLQLRTLRS</td>
<td>LxxLL2051</td>
<td>Q09472</td>
</tr>
<tr>
<td>9</td>
<td>P300</td>
<td>GMDQDAASKHQLQLLRSSGS</td>
<td>LxxLL81</td>
<td>Q09472</td>
</tr>
<tr>
<td>10</td>
<td>IKBB</td>
<td>PLGSAMRLPNILRLARRHAHGAP</td>
<td>LxxLL289</td>
<td>Q15653</td>
</tr>
<tr>
<td>11</td>
<td>IKBB</td>
<td>LHLAVHIHPFLPLDFLFGSAT</td>
<td>LxxLL74</td>
<td>Q15653</td>
</tr>
<tr>
<td>12</td>
<td>TRIP9</td>
<td>PLVSQNNQEQSTLRLDDTTAGK</td>
<td>LxxLL2066</td>
<td>Q15652</td>
</tr>
<tr>
<td>13</td>
<td>TRIP9</td>
<td>QWYSSLNEREQELINLVEVSQ</td>
<td>LxxLL844</td>
<td>P52732</td>
</tr>
<tr>
<td>14</td>
<td>SRC1</td>
<td>TSGIPPTQPAQQKSLQLLTE</td>
<td>LxxLL1435</td>
<td>Q57888</td>
</tr>
<tr>
<td>15</td>
<td>SRC1</td>
<td>SDGDSDKYSQTSIKLQLTTTAEQ</td>
<td>LxxLL633</td>
<td>Q57888</td>
</tr>
<tr>
<td>16</td>
<td>SRC1</td>
<td>PSSHSLTERHKHRLQEGPS</td>
<td>LxxLL690</td>
<td>Q57888</td>
</tr>
<tr>
<td>17</td>
<td>SRC1</td>
<td>AKSKKESKDQHILLVLDDKD</td>
<td>LxxLL749</td>
<td>Q57888</td>
</tr>
<tr>
<td>18</td>
<td>TIP2</td>
<td>QGSRHJDSQKQTSLQQLTTSDQ</td>
<td>LxxLL641</td>
<td>Q57596</td>
</tr>
<tr>
<td>19</td>
<td>TIP2</td>
<td>STHGTSLKEKHLRLQDJSS</td>
<td>LxxLL690</td>
<td>Q57596</td>
</tr>
<tr>
<td>20</td>
<td>TIP2</td>
<td>EPVSPPKKEADFALLRLYLDKDTK</td>
<td>LxxLL745</td>
<td>Q57596</td>
</tr>
<tr>
<td>21</td>
<td>PGC1</td>
<td>PQEVAEELSLKQLLAPANT</td>
<td>LxxLL144</td>
<td>Q9UBK2</td>
</tr>
<tr>
<td>22</td>
<td>TIP2</td>
<td>SQSTFPNPRPQGLRGLPNQNF</td>
<td>LxxLL878</td>
<td>Q57596</td>
</tr>
<tr>
<td>23</td>
<td>SRC3</td>
<td>KKKGQGVQKDLSLQPLQQALDG</td>
<td>LxxLL613</td>
<td>Q9Y6Q9</td>
</tr>
<tr>
<td>24</td>
<td>SRC3</td>
<td>QRGPLESKHKQILQHQLTCSDD</td>
<td>LxxLL621</td>
<td>Q9Y6Q9</td>
</tr>
<tr>
<td>25</td>
<td>SRC3</td>
<td>QRGPLESKHKQILQHQLTCSDD</td>
<td>LxxLL621</td>
<td>Q9Y6Q9</td>
</tr>
<tr>
<td>26</td>
<td>SRC3</td>
<td>MHSSLQKGRHILQHQLQNGSN</td>
<td>LxxLL685</td>
<td>Q9Y6Q9</td>
</tr>
<tr>
<td>27</td>
<td>mSRC3</td>
<td>HGSSQNRPLRNSDLDLGPPSN</td>
<td>LxxLL1041</td>
<td>O09000</td>
</tr>
<tr>
<td>28</td>
<td>PGC1</td>
<td>PAFEVDELSSQKQLLATSYP</td>
<td>LxxLL156</td>
<td>Q86YN6</td>
</tr>
<tr>
<td>29</td>
<td>PRIP</td>
<td>LVSPAQRVELTSLQDLNSAG</td>
<td>LxxLL1491</td>
<td>Q14686</td>
</tr>
<tr>
<td>30</td>
<td>PRIP</td>
<td>PVKRDVTSPLLVLVNQSDSA</td>
<td>LxxLL887</td>
<td>Q14686</td>
</tr>
<tr>
<td>31</td>
<td>NCO</td>
<td>MGQVPRTHLTILADHICQITQ</td>
<td>LxxLL2051</td>
<td>O75376</td>
</tr>
<tr>
<td>32</td>
<td>NCO</td>
<td>MGQVPRTHLTILADHICQITQ</td>
<td>LxxLL2051</td>
<td>O75376</td>
</tr>
<tr>
<td>33</td>
<td>TRIP140</td>
<td>GHSFADPASNLQGLEDIRKALMG</td>
<td>LxxLL2263</td>
<td>O75376</td>
</tr>
<tr>
<td>34</td>
<td>SMRT</td>
<td>AHPVGHQKVRVTVLTAHISVETQ</td>
<td>LxxLL2135</td>
<td>Q9Y618</td>
</tr>
<tr>
<td>35</td>
<td>PGC1</td>
<td>AEFSILRELLAQDVLCDVSK</td>
<td>LxxLL343</td>
<td>Q86YN6</td>
</tr>
<tr>
<td>36</td>
<td>SMRT</td>
<td>QAVQEHASTSNLQGLEDIRKALMG</td>
<td>LxxLL2342</td>
<td>Q9Y618</td>
</tr>
<tr>
<td>37</td>
<td>RIP140</td>
<td>DSVRKGKQDSTLSLQFSR</td>
<td>LxxLL133</td>
<td>P48552</td>
</tr>
<tr>
<td>38</td>
<td>RIP140</td>
<td>KDRLCYVASSHLLKQLKSKV</td>
<td>LxxLL185</td>
<td>P48552</td>
</tr>
<tr>
<td>39</td>
<td>RIP140</td>
<td>KDSRLYVASSHLLKQLKSKV</td>
<td>LxxLL185</td>
<td>P48552</td>
</tr>
<tr>
<td>40</td>
<td>RIP140</td>
<td>RNNIKQAANSLLLHQLKSQITP</td>
<td>LxxLL380</td>
<td>P48552</td>
</tr>
<tr>
<td>41</td>
<td>RIP140</td>
<td>KNSKLNSHQKVTQTLQLLGKNE</td>
<td>LxxLL501</td>
<td>P48552</td>
</tr>
<tr>
<td>42</td>
<td>RIP140</td>
<td>SEIENLERRTVLQLLQNGPTKG</td>
<td>LxxLL713</td>
<td>P48552</td>
</tr>
<tr>
<td>43</td>
<td>RIP140</td>
<td>PVSQPQDFSKGNSLQRLQRQDQSYL</td>
<td>LxxLL819</td>
<td>P48552</td>
</tr>
<tr>
<td>44</td>
<td>RIP140</td>
<td>RSWAREKSNVNLQKLILLSCNV</td>
<td>LxxLL936</td>
<td>P48552</td>
</tr>
<tr>
<td>45</td>
<td>RIP140</td>
<td>RSWAREKSNVNLQKLILLSENSV</td>
<td>LxxLL936</td>
<td>P48552</td>
</tr>
<tr>
<td>46</td>
<td>GCN5</td>
<td>EEDADTKQVYFYLFKLRRSKLQ</td>
<td>LxxLL190</td>
<td>Q29831</td>
</tr>
<tr>
<td>47</td>
<td>PBP1</td>
<td>HGEDFSVQSNPILTSLQGWTNG</td>
<td>LxxLL604</td>
<td>Q51648</td>
</tr>
<tr>
<td>48</td>
<td>PBP1</td>
<td>VSMAGTKNHMPMLMNLDKNDPAQ</td>
<td>LxxLL645</td>
<td>Q51648</td>
</tr>
<tr>
<td>49</td>
<td>PGC1</td>
<td>DGTPPPQAEELSSLLKQLALAPANT</td>
<td>LxxLL144</td>
<td>Q9UBK2</td>
</tr>
<tr>
<td>50</td>
<td>SHP</td>
<td>TEFVAEAVPPSLKILEEPS</td>
<td>LxxLL118</td>
<td>Q51646</td>
</tr>
<tr>
<td>51</td>
<td>SHP</td>
<td>SPQGGAASPRLAYLLSSLLKA</td>
<td>LxxLL21</td>
<td>Q51646</td>
</tr>
<tr>
<td>52</td>
<td>TRIP4</td>
<td>FVNLTRYRQRDLRAVLPGRHPS</td>
<td>LxxLL161</td>
<td>Q51650</td>
</tr>
<tr>
<td>53</td>
<td>TRIP5</td>
<td>LQNKLKNLGESAFLKSLLLPHLR</td>
<td>LxxLL101</td>
<td>Q51649</td>
</tr>
</tbody>
</table>
6.10. References

Chemical Biology Approaches for Nuclear Receptors
- Molecular and Structural Insights -

Nuclear receptors (NRs) are multi-domain transcription factors that are controlled by small lipophilic molecules. Proteins of this protein superfamily feature domains with high sequence homology and domains which are comparatively diverse. Nevertheless the overall conformation of NRs is very similar. Small ligands interact with the ligand binding domain (LBD) of NRs via their buried ligand binding pocket (LBP). Dimerization of either two identical monomers or two different NRs is accomplished by structural rearrangements in the receptor, mainly in the hinge region (HR) and the LBD. Subsequent interaction with NR response elements on DNA is controlled by the zinc fingers of the DNA binding domain (DBD) with partial involvement of the HR. Additional conformational change of the LBD leads to the establishment of the activation function 2 (AF-2). Together with activation function 1 (AF-1), located on the amino-terminal domain (NTD), AF-2 controls coactivator recruitment and consequently transcriptional activation. Structural and functional studies led to the finding that receptor specificity is connected to size and shape of the intermolecular interaction interfaces, such as LBP and AF-2, but also to size and structure of varying domains, including NTD and C-terminal F domain. The intrinsic structural flexibility of the AF-2 is one reason for the ability of NRs to respond differently to the presence of diverse small molecules, including agonists, partial agonists, antagonists or reverse agonists. Thus both transcriptional activation and repression via NRs are under functional control of ligand-dependent recruitment of coactivators and corepressors.

Protein stability and activity are directly associated to posttranslational modifications (PTMs) of NRs. These PTMs can have a considerable impact on conformational changes - especially in less structured regions - and thus potentially modulate the functionality of certain protein domains and overall protein activity. In CHAPTER TWO acetylations of lysines in the estrogen receptor α (ERα) HR were investigated concerning their structural influence on the mainly unstructured domain. Molecular dynamics (MD) simulation studies predicted structural changes upon lysine acetylation, with double lysine-acetylation exerting the greatest effect in terms of their propensity to form α-helices. Structural analysis of 12-mer peptides synthesized by solid phase peptide synthesis by means of both CD and NMR studies confirmed these results. The discovery of a possible presence of a transient polyproline II (PPII) helix could be important in both the exploration of the molecular mechanism and specific targeting. The acetyl transferase p300 was capable of acetylating the unacetylated peptide and the mono-acetylated peptides without any additional auxiliary proteins. Comparing the amount of transferred acetyl-groups revealed that the acetylation efficiency of the second lysine is down-regulated by the first acetylation event.
In **CHAPTER THREE** expressed protein ligation (EPL), including native chemical ligation (NCL), was used to investigate the HR of ERα and its PTMs in the context of their impact on the LBD. An N’-terminal cysteine in a LBD construct, necessary for EPL, was generated by the methionine excision approach, while the required thioester peptides were synthesized by solid-phase peptide synthesis. Ligation of the posttranslationally modified HR peptides to a model protein could be successfully demonstrated, however, the transfer of this method to ERα was not successful. The major issue in this case was protein instability under ligation conditions. Studies focusing on the extension of the ERα-LBD in the direction of the HR revealed that protein expression allowed the generation of ERα-LBD constructs including the HR up to position 251. However, binding studies identified the HR sequence before position 271 to significantly decrease the ability of ERα to bind coactivator peptides, thus hinting at the existence of two regions of the HR, which differ in terms of their structure and function.

**CHAPTER FOUR** deals with the ER-LBD screening for novel natural peptide binders by means of ribosomal display. Earlier rounds of ribosomal enrichment witnessed the expected emergence of the well-studied LXXLL motif. However, subsequent rounds led to the identification of a more mature PXLXXLLXXP consensus, which could be validated by both biochemical and cellular methods. MD and X-ray crystallography studies revealed a specific role for the flanking prolines as helix breakers, which prime the helix length for optimal interaction with the surface charge clamp. Furthermore, the conformational constraints imposed by the prolines on adjacent immediate flanking amino acid residues are believed to determine binding affinity through the precise orientation of side chain functionality at the ER surface. This work does not only represent a fundamental re-evaluation of the NR-coactivator interaction and thus sets new, minimal, structural parameters based on linear sequences of proteinogenic amino acids, but also provides insights for peptide-derived tools and more druggable peptide agents.

MD simulation was used to predict the propensity of short peptides to fold into an α-helix in **CHAPTER FIVE**. Based on results from ER screening (**CHAPTER FOUR**), peptides were designed *in silico* to potentially function as androgen receptor (AR) coactivator binding inhibitors (CBIs). The scaffold sequence was based on a combination of a natural NR coactivator sequence and the NTD of AR, known to specifically interact with the AR LBD via a phenylalanine-rich FXXLF motif. Proline residues were inserted into the flanking regions of the central binding motif to define the helix length. MD simulation studies were performed to acquire information about the propensities of the designed peptides to form α-helices. Based on these results the solid-phase synthesis of selected peptides was performed. CD studies of these peptides clearly confirmed the predicted α-helical character of the 11- and 12-mers. The AR binding affinities of the peptides was evaluated by varying the length of the helical segment of the peptides via the proline positioning and by elaborating the adjacent amino acid of the N’-terminal proline in terms of charge, size, and hydrophobicity. This
conceptual approach resulted in short peptides that interact exclusively with AR-LBD and overpower known AR-specific binding in a cellular context with the fully intact AR. Thus, the identified sequences can be considered as lead structures for the future development of AR-selective CBIs.

The retinoid X receptor (RXR) is an important transcription factor involved in the regulation of gene networks connected to cell growth, cell differentiation, and cell death. Misregulation of RXR or pathways that are based on RXR heterodimer formation with other NRs can lead to various malignancies, such as cancer, cardiovascular and inflammatory diseases. Thus in CHAPTER SIX, microarray-based nuclear receptor-coregulator interaction profiling was performed to screen a small compound library of biaryl natural products (NPs) for potential RXR binding. Two NPs featuring (partial) agonism functionality could be identified, while another NP showed non-competitive antagonism characteristics through allosteric RXR binding. Both the lipophilic side chain groups and the hydroxyl functionality of these NPs were shown to be important aspects for their activity. Further modifications to the two lipophilic groups present in the NPs, towards an improved mimic of the structure of leucine side-chains of coactivator peptides, resulted in an improved ability to inhibit the RXR-coactivator interaction in the profiling assay. This data in combination with the structural similarity with known NR-CBIs revealed the first NP-CBI, also with potential for optimization.

This thesis provides molecular and structural insights to understand function and regulation of the structural elements of NRs. This is not only important for the understanding of receptor pharmacology, but also provides novel entries to target these proteins in a specific manner.
**Anwendungen in der Chemischen Biologie für Kernrezeptoren**

- Molekulare und Strukturelle Einblicke -


Die Stabilität von Kernrezeptoren und die damit zusammenhängende Aktivität stehen im direkten Zusammenhang mit Modifikationen von Aminosäuren dieser Proteine. Diese Veränderungen - in Form von Hinzufügung verschiedenster Molekülgruppen - haben einen entscheidenden Einfluss auf die Struktur des Rezeptors, im Speziellen in Bereichen die normalerweise wenig strukturiert sind. Somit können Modifikationen dieser Art die Funktionalität einzelner Rezeptorbereiche regulieren, die wiederum die Gesamtaaktivität des Rezeptors beeinflussen. In **KAPITEL ZWEI** wurde eine natürlich vorkommende Modifikation (**acetylation**) der Aminosäure Lysin und ihr Einfluss auf den relativ unstrukturierten Bereich (**hinge region, HR**) des Estrogenrezeptors untersucht. Computersimulationen (**MD**) zeigten strukturelle Veränderungen als Folge der Lysinmodifikation, wobei doppelte Modifikation den größten Effekt hatte spiralförmige Strukturen (**α-helix**) auszubilden. Die Herstellung (**SPPS**) der simulierten Peptide und die anschließende Strukturanalyse (**NMR, CD**) bestätigten diese Erkenntnisse. Die Entdeckung einer unüblichen helikalen Struktur (**PPα-helix**) könnte wichtig sein, um den Mechanismus zu verstehen, aber auch, um gezielte Bindepartner zu entwickeln. Weitere Untersuchungen in dem Kapitel beschäftigten sich mit

Im DRITTEN KAPITEL wurde eine Methode verwendet, die ein chemisch synthetisiertes Peptid und ein biologisch hergestelltes Protein miteinander verbindet (expressed protein ligation, EPL). Mit dem resultierenden Hybrid soll der Einfluss des relativ unstrukturierten Bereichs (HR) und dessen Modifikationen (acetylation) auf den Teil des Receptors, der verantwortlich ist für die Wechselwirkungen mit kleinen Molekülen und anderen Proteinen (ligand binding domain, LBD), untersucht werden. Damit diese chemische Reaktion (native chemical ligation, NCL) stattfinden kann muss sowohl das Peptid als auch das Protein spezielle Charakteristika aufweisen (thioester/N'-Cys). Das hergestellte Peptid mit Lysinmodifikation konnte erfolgreich an ein Testprotein geknüpft werden, jedoch war die Übertragung auf Estrogenrezeptor nicht möglich. Hauptproblem war hierbei Instabilität des Proteins unter den erforderlichen Reaktionsbedingungen. Studien zu einer sukzessiven Verlängerung der LBD-Region in Richtung HR zeigten, dass eine Proteinherstellung für alle Konstrukte möglich war. Jedoch zeigten anschließende Bindungsstudien, dass ein Überschreiten der Position 271 (Mitte von HR) eine entscheidende Verringerung der Fähigkeit Coffaktoren zu binden zur Folge hatte. Dies legitimiert die Hypothese, dass sich die HR in zwei Abschnitte aufteilen lässt, die sich sowohl in Struktur als auch in Funktion unterscheiden.


Fehler in der Regulierung des Androgenreceptors wird auch mit verschiedenen Krankheiten in Verbindung gebracht, wie zum Beispiel Prostatakrebs. In KAPITEL FÜNF


Diese Doktorarbeit liefert molekulare und strukturelle Einblicke, die für das Verständnis von Funktion und Regulation von strukturellen Elementen von Kernrezeptoren wichtig sind. Dieses Wissen ist Voraussetzung, neue Wirkstoffe entwickeln zu können, die ein spezielles Ziel nicht nur stark, sondern auch exklusiv binden, um Nebenwirkungen zu verhindern.
List of Publications


(to be) submitted


Fuchs, S.; Nieto, L.; de Vries-van Leeuwen I. J.; Brunsveld, L., *In silico* design of androgen receptor coactivator binding inhibitors, *to be submitted.*
Curriculum Vitae

Sascha Fuchs was born on February 16th, 1984 in Gütersloh, Germany. After secondary education at the Werdener Gymnasium in Essen he started studying Chemical Biology at the TU Dortmund University (both Germany). He performed part of his master’s program in the Department of Biochemistry and Molecular Biology at the University College in London (United Kingdom). During his research project in the group Prof. Dr. Herbert Waldmann at the Max Planck Institute of Molecular Physiology in Dortmund in cooperation with Prof. Dr. Henning D. Mootz at the TU Dortmund University (both Germany), he investigated the in vivo modulation of estrogen receptor α activity by conditional protein splicing. In December 2008 he joined the Chemical Biology group in the Department of Biomedical Engineering at the Eindhoven University of Technology (The Netherlands) starting his PhD position with Prof. Dr. Ir. Luc Brunsveld. His research focused on gaining molecular and structural insights to develop deeper understanding of the function and regulation of the structural elements of nuclear receptors, including estrogen-, androgen- and retinoid X receptor. His findings were important for the understanding of receptor pharmacology and provided novel entries to target these proteins. In this thesis the most important results are summarized.
Acknowledgements

Now, I come to the end of my thesis. But not before thanking some people that had a direct or indirect influence on my life and my work over the last four years in Eindhoven, thus being eventually co-responsible for the thesis in its final version.

First of all, I want to thank Prof. Dr. Ir. Luc Brunsveld. Luc, thank you for giving me the opportunity to work on this interesting and challenging topic. I felt always welcome to enter your office and discuss the project if you had time (and even if you had not). I remember that sometimes even I had to remind you about subsequent appointments as we both lost ourselves in the depths of intense discussion. Thank you for all the fruitful discussions about different scientific topics over the last years and your many suggestions and advice in several moments. It was also nice to be part of a growing, interdisciplinary group. I learned a heck of a lot these last few years.

I want to express my gratitude to the members of my defense committee Prof. Dr. Constant A.A. van Boeckel (TU/e & University Leiden), Dr. Christian Ottmann (TU/e), Univ.-Prof. Dr. Christian F. W. Becker (University Vienna) and Prof. Dr. Carlie J.M. de Vries (University of Amsterdam) for the critical reading of my thesis and for making suggestions to improve on its quality. Furthermore, I want to thank Dr. Maarten Merkx and Dr. Lech-Gustav Milroy for completing the committee. Lech, I want to thank you additionally for all the help especially in the final stage of my PhD. I can remember funny peptide synthesis days, excited discussions during the writing phase of ‘the paper’ and critical but always fair comments during the correction phase of my thesis. You taught me a lot about scientific writing and without you and your suggestions I would not be able to deliver the thesis in its present state. Thanks a lot for that.

Additionally I want to thank different people for their collaboration and support during my thesis. Dr. Lidia Nieto: Lidia, thank you a lot for all the help with the structural analysis of the peptides, namely molecular simulations, circular dichroism and NMR measurements (CHAPTER TWO, FOUR & FIVE). Your knowledge certainly improved the quality of my thesis. ¡Gracias! Dr. Ingrid J. de Vries-van Leeuwen: Ingrid, thank you for all the cellular assays you performed to evaluate the peptides (CHAPTER FOUR & FIVE). You were always willing to help and it was nice to work together with you. Dank je wel! Dr. Matthew Burton and Stijn M. Agten: Matt and Stijn, I want to thank you for all the work you made on the proline peptides (CHAPTER FOUR). Matt, you I want to additionally thank for all other help and discussions during the middle phase of my PhD. I enjoyed working together with you. Thanks! Dr. Hoang D. Nguyen, Dr. Trang T.P. Phan and Parisa M. Goodarzifard: Hoang, Trang and Parisa, I want to thank you for all the ribosome display work you have performed (CHAPTER FOUR). Cảm ơn ban! Parisa, you I want to thank also for the work you did during your master’s thesis. You learned quick and worked always very independently. Remco Arts: Mijn koning, thank you for all the work you did on the Hinge Region project (CHAPTER TWO), but also for joining me to celebrate the BVB. Chan Vinh Lam: Chan Vinh, I want to thank you for the work you did during your stage (CHAPTER SIX). You were always
interested in understanding the big picture. Thank you, my students! Dr. Christian Ottmann, Dr. Sabine Möcklinghoff and Andrea Schmidt: Christian, Sabine and Andrea, thank you for all your help during protein crystallization and with solving the structure. Crystal fishing and patience are definitely highly connected. I also want to thank the different OGO students and Wouter Engelen and Maarten Bakker that were part of the androgen receptor project. I liked your enthusiasm and your research formed the basis for a lot of the work presented in CHAPTER FIVE. I also wish to thank the analytical lab with Joost L.J. van Dongen, Ralph A.A. Bovee and Xianwen Lou. Your accurate measurements were on different occasions undoubtedly necessary.

Furthermore I want to thank the people that were not scientifically working together with me, but no less important for my thesis. Only a nice environment makes it possible to work occasionally more than a standard eight-hour-day. Thus, for me it is very important to ‘turn’ some of my colleagues into friends, as some of you might have experienced. In the beginning were all these people very important that had made the German-Dutch-transition just some month before me. Dana, Katja, Maëlle und Dung, you prepared everything in Eindhoven for my belated arrival. Due to you I quickly felt like home in Eindhoven. Although I left you office-wise some time later, you were very important during my time in Eindhoven, but also during my master’s in Dortmund. Maëlle, I enjoyed working together with you in the first few months and we always had fun together. Merci beaucoup! Dung, we had a lot funny conversations about all kinds of topics, including football and ManU. Câm on bán! Dana, du warst zwar nicht immer meiner Meinung, hast diese aber immer respektiert und ich fand unsere Gespräche immer interessant. Katja, oder besser gesagt Mullik, du hast mich als einzige (neben Luc) von Anfang bis Ende begleitet. Als ich Anfang 2008 in die Gruppe gekommen bin noch als alleinerziehende Mullik, dann später in Eindhoven dann aber zusammen mit Pullik. Ihr seid gerade in der Endphase meiner Thesis wirklich fast wie Eltern für mich geworden. Ihr habt euch um mich gesorgt, ihr habt mir Essen gebracht und mich zur Vernunft gemahnt. Diese Kombination kannte ich zuvor nur von meinen Eltern. Спасибо, auch dafür, dass du mein Paranime bist. But during my four years there were numerous other people responsible for making my time at work so enjoyable. Melissa, thank you for both your scientific help and also for being there for other things especially in the early phase of my PhD. Thanks! Ewelina, you were also important especially in the third year. Our coffee breaks, the cooking but also your critical questions about the state of my research and the pressure you generated motivated me to work efficiently and to finalize parts of my projects. Dziękujemy! Micha, auch wenn nicht der identische Charakter haben wir uns immer gut verstanden. Wir hatten immer jèhrisch Spaß, ob beim Fußball oder in London, auch wenn du dich des Öfteren zu Tode erschrocken hast wenn ich dich angesprochen habe. Pauline, the coffee/water breaks were always needed especially during my writing phase. Merci beaucoup, also for the very last proof-reading.

But also all the other members of the group I would like to thank. Wencke, Inga, Ralph, Christian and all the students. Also I want to thank Abidin for all the support especially in the last year and for being my paranime. Also thank you for ‘sharing’ your friends, like Başar, Seda, Gökhan, Kamil and Alessandra. The barbecue in Breda, the football game Germany-Italy in Essen
(although we lost), the short meeting in Delirium in Brussels, the gypsy festival and the Christmas market in Dortmund were all so much fun! Teşekkürler and hope to meet you all in Istanbul. I also want to thank Daniele, Mindaugas, Benjamin, Brian, Luuk, Asish, Lorenzo, Elisa, Miguel and all the others for joining in the Champions League viewing in the Irish, food in De Bengel or Touch of India, and diverse parties and other activities in Eindhoven.

I also want to thank my office members: Mantas (Malisauskas), Lars (Röglín), Luc (Scheres), Ingrid (de Vries-van Leeuwen) and the current members Laurens, Parisa and Chan Vinh. I remember funny table football competitions and Lars, der Spruch des Monats hängt hier immer noch an der Wand „Es ist so warm heute! Ich habe mir überlegt... !“. Furthermore, I want to thank all the members of the bio lab and the whole cell lab family. Thanks for all the scientific discussions and, Björne, I hope the cell lab family tree will continue growing. Also thank you to all the members of MST/SWO. I enjoyed the interdisciplinary environment and also the social events including uitje and Nolte-Mijer-Cup, although we won football only in the year I did not join. I also want to thank my housemates: Oliver, Sapo and Štěpán. Danke, grazie, děkujeme, you for being so uncomplicated and helpful. I think all of you are very easy to live with. That is what you can tell your girl-friend or future partner. But I also want to thank people I met on conferences and meeting and are still in touch with me: Annie and Christian, Marije, Randy, Ulrike, James, tack, dank je wel, thanks, danke for the nice time in Dubrovnik, Spetses, Amsterdam and all kinds of support (including Movember) and scientific exchange.


Last but not least, möchte ich mich natürlich bei Sarah bedanken. Sarah, thank you for all your support. Du hast immer an mich geglaubt and you always motivated me. Du bist der wichtigste Mensch in meinem Leben. Danke!

Sascha