

**Development of an orthogonal ligand-receptor pair
based on synthetic estrogen analogs and engineered
estrogen receptor for transcriptional regulation**

Dissertation
zur Erlangung des Doktorgrades
der Mathematisch-Naturwissenschaftlichen Fakultäten
der Georg-August-Universität zu Göttingen

vorgelegt von
Kazi Mohammed Didarul Islam
aus Gopalganj, Bangladesh

Göttingen 2007

D7

Referent:

Prof. Dr. Christiane Gatz

Korreferent:

PD Dr. Wolfgang Dröge-Laser

Tag der mündlichen Prüfung:

03. Mai 2007

Contents

Abbreviations

1	Introduction	1
1.1	<i>Regulation of Gene Expression.....</i>	<i>1</i>
1.1.1	Level of Regulation.....	1
1.1.1.1	Transcriptional regulation	2
1.1.1.2	Translational control	2
1.1.1.3	Direct regulation of protein function.....	3
1.1.2	Regulation at transcriptional level.....	3
1.1.2.1	Regulation of transcription factor expression	4
1.1.2.2	Regulation of transcription factor activity	4
1.1.3	Trends in inducible control of gene expression	5
1.1.4	Orthogonal combination.....	6
1.1.4.1	Criteria for orthogonal ligand-receptor pair.....	7
1.1.5	Nuclear hormone receptor (NHR) and estrogen receptor (ER).....	7
1.1.5.1	The Estrogen Receptor	8
1.1.6	Mechanism of E2 action.....	14
1.1.7	Orthogonal ligand-receptor pair: chemical synthesis vs. protein engineering	15
1.1.7.1	Engineering approach.....	16
1.2	<i>Aim of the study.....</i>	<i>19</i>
2	Materials and Methods	20
2.1	<i>Materials</i>	<i>20</i>
2.1.1	Organisms	20
2.1.1.1	Bacteria.....	20
2.1.1.2	Yeast (<i>Saccharomyces cerevisiae</i>).....	20
2.1.2	Media and additives	20
2.1.2.1	Bacterial media.....	20
2.1.2.2	Yeast media	21
2.1.2.3	Plant Media	21
2.1.2.4	Additives	21
2.1.3	Nucleic Acids	21
2.1.3.1	Plasmid	21
2.1.3.2	Primer	22
2.1.4	Enzymes	24
2.1.5	Synthetic ligands	25
2.1.6	Chemicals	25

2.1.7	Kits	27
2.1.8	Buffers and solutions.....	27
2.1.9	Consumables	29
2.1.10	Software	29
2.1.11	Equipment	29
2.2	<i>Methods</i>	31
2.2.1	Culture and storage of microorganisms	31
2.2.1.1	<i>E. coli</i> strains	31
2.2.1.2	Yeast <i>S. cerevisiae</i> strains	31
2.2.2	Standard molecular biology methods.....	32
2.2.2.1	Isolation of plasmid DNA from <i>E. coli</i>	32
2.2.2.1.1	Alkaline lysis:.....	32
2.2.2.1.2	High-purity plasmid DNA isolation.....	32
2.2.2.1.3	Isolation of plasmid DNA from yeast	33
2.2.2.1.4	Estimation of nucleic acids concentration and purity	33
2.2.2.2	Nucleic acids gel electrophoresis	34
2.2.2.2.1	Separation of DNA on Agarose Gels	34
2.2.2.2.2	Elution of DNA fragment from agarose gel.....	34
2.2.2.3	Restriction digestion of DNA molecules	34
2.2.2.4	Ligation of DNA fragments	36
2.2.2.5	Homologous recombination in yeast.....	36
2.2.2.6	Polymerase chain reaction (PCR)	36
	Random point mutagenesis using error-prone PCR.....	38
	Saturation Mutagenesis	38
2.2.2.7	DNA sequencing	39
2.2.2.8	Gene transfer in bacteria	40
2.2.2.9	Gene transfer in yeast.....	41
2.2.2.9.1	Frozen yeast transformation for screening.....	41
2.2.2.9.2	High efficiency transformation for selection	41
2.2.2.10	Construction of plasmids.....	42
2.2.3	Standard protein biochemical methods	43
2.2.3.1	Expression of 6x His-fusion protein expression in <i>E. coli</i>	43
2.2.3.2	Purification of recombinant 6x His-fusion proteins.....	44
2.2.3.3	Protein concentrations determination.....	44
2.2.3.4	Denaturing sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE)	45
2.2.3.5	Coomassie staining of proteins separated on SDS-PAGE	45
2.2.3.6	Western blot analysis	45
2.2.3.7	β -galactosidase activity (ONPG assay).....	46
2.2.3.8	<i>In vitro</i> ligand binding assay	47
2.2.3.9	Chemical complementation based library selection.....	48
2.2.3.9.1	One hybrid system.....	48

2.2.3.9.2	Two hybrid system	48
2.2.3.10	Chemical complementation based screening	49
2.2.3.10.1	One hybrid system	49
2.2.3.10.2	Ligand dose-response/titration assay	49
2.2.4	Toxicity test	49
2.2.4.1	Germination of seed	49
2.2.4.2	Adventitious shoot induction test	50
2.2.5	Molecular modeling	50
3	Results	52
3.1	<i>Validation of test system</i>	52
3.1.1	Yeast one hybrid system:	52
3.1.1.1	On plate assay	53
3.1.1.2	β -galactosidase assay	54
3.1.2	Yeast two hybrid system	55
3.2	<i>Toxicity test</i>	56
3.2.1	Germination Test	56
3.2.2	Adventitious shoot induction	57
3.3	<i>Rational design: Site-directed mutagenesis at M388 and L428</i>	58
3.4	<i>Irrational design using random mutagenesis</i>	60
3.4.1	Characterization of V400A mutant	62
3.4.1.1	3-D Model of LBD with V400	62
3.4.1.2	Ligand-dependent activation of V400A	63
3.4.1.3	<i>In vitro</i> binding assay	65
3.5	<i>Directed Evolution</i>	68
3.5.1	Ligand classification in yeast two hybrid system	68
3.5.2	Identification of ligand for the saturation mutagenesis	70
3.5.3	Identifying ligand-contacting residues in hER LBD	71
3.5.4	Saturation mutagenesis	72
3.5.5	Random Mutagenesis	75
3.5.6	Additive effect of mutations for improved affinity and better transactivation	77
3.6	<i>Combinatorial improvement of transactivation: combining V400A with L384F L387M Y537S</i>	79
4	Discussion	81
4.1	<i>Choice of ligand</i>	81
4.2	<i>Yeast-based test system</i>	83
4.3	<i>Chemical complementation based one and two hybrid test system</i>	85
4.3.1	One hybrid system	86

4.3.2	Two hybrid system.....	87
4.4	<i>Rational design</i>	89
4.5	<i>Irrational design</i>	90
4.6	<i>Directed evolution</i>	93
4.7	<i>Evolution of novel ligand-receptor combination through systematic protein engineering</i>	95
4.8	<i>Functional mutations identified in the evolved hERα variants: L384F L387M Y537S</i>	96
4.8.1	L384F.....	97
4.8.2	L387M.....	97
4.8.3	Y537S.....	99
4.8.4	Additivity of the functional mutations.....	100
4.9	<i>Orthogonal ligand-receptor pair with unique specificity is useful in controlling gene expression</i>	102
4.10	<i>Outlook</i>	102
5	Summary	104
6	Bibliography	106
7	Appendix	119
7.1	<i>List of ligands</i>	119
7.2	<i>The amino acid code</i>	122

Abbreviations

3-AT	3-amino triazole
°C	Degree celcius
μM	Micromolar
Å	Ångström
β-gal	β-galactosidase
A ₆₀₀	Absorbance at 600 nm wavelength
AD	Activation domain
AF	Activation Function
Amp	Ampicillin
ATP	Adenosine-triphosphate
bp	Base pairs
BSA	Bovine serum albumin
Cam	Chloramphenicol
cDNA	Complementary deoxyribonucleic acid
DBD	DNA binding domain
ddH ₂ O	Double-distilled water
DES	Diethylstilbestrol
DHB	4,4'-dihydroxybenzil
DNA	Deoxyribonucleic Acid
dNTP	Deoxynucleosidtriphosphate (dATP, dGTP, dCTP, dTTP)
DPN	Diarylpropionitrile
<i>E. coli</i>	<i>Escherichia coli</i>
E2	Estradiol
EA	Enzyme acceptor
EC ₅₀	Effective concentration of half-saturation (50%) = half-saturation constant
ED	Enzyme donor
ER	Estrogen receptor

ERE	Estrogen response element
EtOH	Ethanol
Gal4	Transcription factor required for the activation of the GAL genes in response to galactose in <i>S. cerevisiae</i>
GLV	Gal4 DBD-hER α LBD-VP16 AD
GR	Glucocorticoid receptor
HEGO	Human estrogen receptor, Gly at residue 400 variant
HEO	Human estrogen receptor, Val at residue 400 variant
hER α	Human estrogen receptor α
hER β	Human estrogen receptor β
Hrp	Horse radish peroxidase
hsp	Heat shock protein
IPTG	Isopropyl β -D-1-thiogalactopyranoside
Kn	Kanamycin
<i>lacZ</i>	β -galactosidase gene
LB	Luria Broth
LBD	Ligand binding domain
LBP	Ligand binding pocket
MeO	Methoxy-group
MOE	Molecular operating environment
MeOH	Methanol
mRNA	Messenger Ribonucleic Acid
MS medium	Murashige and Skoog medium
MU	Miller unit
NHR	Nuclear hormone receptor
NLS	Nuclear localization signal
nm	nanometer
nM	Nanomolar
NR	Nuclear receptor
OD ₆₀₀	Optical density at 600 nm wavelength
OECD	Organization of economic co-operation and development
OH	Hydroxyl-group

ONPG	<i>o</i> -nitrophenyl- β -D-galactopyranoside
PBS	Phosphate buffer saline
PCR	Polymerase chain reaction
PDB	Protein Data Bank (Brookhaven)
PEG	Polyethylene glycol
pH	Negative decadic logarithm of H ⁺ -ion concentration
PPT	Propyl pyrazole triol
PVDF	Polyvinylidene fluoride
RAR	Retinoid acid receptor
RMSD	Root mean square deviation
RNA	Ribonucleic Acid
rpm	Rotation per minute
RXR	Retinoid x receptor
<i>S. cerevisiae</i>	<i>Saccharomyces cerevisiae</i>
SD-medium	synthetic drop-out medium
SDS	Sodium dodecyl sulfate
SDS-PAGE	Sodium dodecyl sulfate-polyacrylamide gel electrophoresis
SE	Polysorbate: ethanol (10:1)
SERM	Selective estrogen receptor modulator
SH3	Src homology 3
SRC-1	Steroid receptor coactivator 1
<i>Taq</i>	<i>Thermus aquaticus</i>
TE	Tris-EDTA
TF	Transcription Factor
TIF-2	Transcription intermediary factor 2
TSR	Template suppression reagent
UTR	Untranslated region
v/v	volume/volume
w/v	weight/volume
VP16	Herpes simplex virus (HSV) regulatory protein
WT	Wild type
Y2H	Yeast two hybrid

1 Introduction

1.1 Regulation of Gene Expression

Transgenic manipulation has become a powerful tool for investigating gene function for both fundamental research and applications. In most cases, foreign genes under the control of a constitutive promoter are introduced into pro- and eukaryotes. Constitutive promoters have been used successfully to express a spectrum of genes in transgenic organisms. However, such promoters have not been useful in situations where expression of a (trans)gene needs to be restricted to specific tissues or at a particular time. Constitutive expression of foreign genes may interfere with physiological processes and on occasion, may be deleterious or even lethal. Understanding developmental and physiological processes in plants often requires the conditional or targeted expression/suppression of either transgenes or native genes. For example, systems that allow researchers to control gene regulation precisely facilitate the over- or underexpression of genes that have deleterious effects on plant growth and development. These regulated gene expression systems also have possible applications in biotechnology that are not provided by constitutive gene expression. Such applications include providing conditional male sterility, altering the timing of defense and reproduction, controlling the production of novel material in plants and microorganisms for industrial or pharmaceutical uses and in gene therapy, and construction of molecular sensors or “smart” materials [1-4].

1.1.1 Level of Regulation

The synthesis of proteins, which are usually the functional end-products of the information stored in genes, involves multiple enzymatic reactions carried out by large protein complexes and all of these processes appear to be potential regulatory targets. Regulation of gene expression can be achieved in either one of these three broad stages: transcription, translation and protein function. Transcriptional regulation offers the potential to effect the widest change in protein levels. Translational control with small

molecules was discovered only recently, but appears to have intermediate properties of speed and breadth of regulation. Direct regulation of protein function has the highest precision, but it suffers from a lack of generality. Regulation can be achieved by turning a gene on (induce) or off (repress). While turning off a gene function is desirable in some cases, the majority of regulation application is benefitted from the induction of expression. Again if a gene remains silent and can be induced while the function is needed, the necessity of repression is theoretically alleviated in most of the cases. Because of its application potential, we will only focus on inducible control of gene expression.

1.1.1.1 Transcriptional regulation

Most organisms regulate gene expression primarily through altering mRNA levels. An important characteristic of transcriptional regulation is that the regulatory signal is amplified. Binding of transcription factors to one or several copies of a gene can generate many mRNAs and thousands of active protein molecules. Another added advantage is the level of control being at the beginning of central dogma, thus reducing the metabolic burden of the cell. Of the three levels of regulation, transcriptional control typically provides the widest range of response.

1.1.1.2 Translational control

Recent work by Breaker and coworkers has led to the discovery of RNA-mediated small molecule regulation of translation. Termed “riboswitches”, these natural translational switches are found in the 5'- untranslated region (5'-UTR) of mRNAs of many prokaryotes [5]. In one well-characterized riboswitch, the binding of thiamine pyrophosphate blocks translational initiation by up to 16-fold in several genes involved with thiamine biosynthesis in *E. coli* [6]. Researchers have shown that the principles of translational control seen in many natural riboswitches can be applied to the creation of artificial genetic switches. Switching activity achieved with most of the riboswitches is the down regulation of gene expression, which is not always desirable. Recent development in the upregulation of gene expression using riboswitches is promising, but the induction kinetics limit their applicability [7, 8].

1.1.1.3 Direct regulation of protein function

Post-translational regulation of protein function can provide high temporal precision and is frequently used in natural systems to rapidly modulate protein activities. One major drawback is the negative regulation, i.e. repression/inhibition of gene function. While binding of a small-molecule to a protein target can exert an immediate effect, this direct regulation of protein function suffers from a lack of generality [9]. Whereas a single genetic switch can regulate transcription or translation of an arbitrary gene of interest, small-molecule protein interactions are usually specific to a single protein target. The synthesis and identification of potent and selective inhibitors or activators for a target protein remain significant challenges despite improvement in synthesis and screening methods.

1.1.2 Regulation at transcriptional level

Since the transcriptional regulation provides the greatest potential for positive control of gene function, we will discuss briefly the mechanism of transcriptional regulation. The core promoter region of a eukaryotic gene shows by itself only a very weak, if any, transcriptional activity. The transcriptional activity is regulated, positively or negatively, through DNA regions (*cis*-acting sequences) that are located more distally to the core promoter and contain binding sites for regulatory proteins (*trans*-acting factors). The *cis*-elements do not have a direct effect on transcription by themselves. Their role is rather to serve as “landing-sites” for a diverse group of DNA-binding proteins, collectively called transcription factors (TFs) [10]. Binding of TFs to upstream *cis*-elements is a pre-requisite for transcription initiation. Today several hundred TFs have been identified, many of them belonging to one of several different protein families. Although members of different families can be entirely unrelated in their sequences and structures, they usually encompass two common features. The first is the presence of a DNA binding domain (DBD), which allows the sequence-specific binding of the factor to a short, defined DNA-element within upstream promoter regions or enhancers of target genes. The other function present in transcription factors is an activation domain (AD), involved in modulation of the transcriptional activity of the target promoter. The DNA binding and activation properties are frequently found to be located in separate domains of the

transcription factor [11, 12] as demonstrated by the fact that DBDs and TADs often can be interchanged between transcription factors in so called swapping experiments [13-15].

Transcription initiation is one of the most exciting stages of regulation and numerous studies have been conducted to understand the process in a deeper sense as well as to manipulate the regulation. Two major strategies have been forwarded for such regulation:

1.1.2.1 Regulation of transcription factor expression

The activities of TFs are controlled in different ways to coordinate cell- and developmental-specific gene expression. One strategy is to control the expression of the transcription factors themselves and it has been shown for several TFs that they are expressed predominantly in certain tissues and sometimes during special timeframes. Tissue-specific TFs have been intensely investigated in e.g. liver [16, 17]. For many tissues less is known about mechanisms involved in tissue-specific gene regulation and involvement of cofactors in initiation of transcription makes this strategy complicated.

1.1.2.2 Regulation of transcription factor activity

Whereas a restricted expression pattern of a transcription factor is an efficient way for tissue-specific control of target genes, this regulatory mechanism is not useful for mediating a rapid response to extracellular environmental and chemical signals. To enable this, several transcription factors are pre-existing within the cell in inactive forms that can rapidly be induced to modulate gene expression. This modulation is often mediated by covalent modification of the TF, e.g. by phosphorylation [18, 19] or by keeping the TF in an inactive state bound to an inhibitory protein [20-22]. The transcriptional activity of other TFs is mediated by the non-covalent, high affinity interaction with small signal substances, such as antibiotic, sugar or hormones. This type of transcriptional regulation is termed as inducible control of gene expression. The glucocorticoid receptor for example, is present in an inactive form in the absence of the cognate ligand. Binding of glucocorticoids activates the receptor, leading to the formation of a homodimer capable of DNA binding and transactivation [23].

1.1.3 Trends in inducible control of gene expression

Since the transcriptional regulation has several advantages like broader control range, positive and negative control as well as being a well-studied phenomenon, it is no wonder that there were numerous approaches to develop regulatory systems based on native or heterologous promoter elements [24]. In almost all the cases, the regulation or switching was achieved by various signals like environmental stimuli [25, 26], metal ions [27, 28] and chemicals [29-32]. Native promoters have been used for tissue-specific induction but they have limitation like activation during the process of regeneration, causing unwanted pleiotropic effects. In addition, they also restrict analysis to a few cell types or tissues. Promoters responding to environmental signals (temperature, light etc) and metal ions are also problematic as these stimuli are difficult to control in nature and induction of expression often results in unwanted effects.

To achieve effective regulation of transgene expression, a series of chemically regulated expression system have been developed, with which expression can be switched on or off, by adding or removing specific chemicals [1, 24]. Based on the purpose and type of control as well as on the organism, chemically inducible systems could vary significantly. However, when considering in general, the following criteria and absolute requirements for an ideal chemically inducible system are:

- Low basal level of expression
- High inducibility with rapid response
- Termination of expression after removal of inducer
- Low toxicity and high specificity of the inducer
- No pleiotropic effect
- Inducer should be easily applicable

Because the ideal inducer should affect expression only of the transgene, it favours the use of well-characterized regulatory elements from evolutionarily distant organisms (heterologous systems), which respond to chemical signals that are usually not encountered by higher plants [24]. Considering the above-mentioned criteria, various major chemically inducible systems developed in the past are summarized in Table 1.1:

System	Tested in	Strengths	Limitations
Tetracycline-inducible TetR	Tobacco, tomato, potato, BY2 cells, SR1 cells	Low amounts of inducer is sufficient for activation, inducer readily taken up by cells	Leaky expression, high concentrations of TetR are required, short half-life of inducer, does not work in <i>Arabidopsis</i>
Tetracycline-inactivatable tTA	<i>Arabidopsis</i> , tobacco, NT1 cells	Target promoter can be shut off efficiently, turnover of transgene can be assessed, lower basal levels compared to TetR	Plants have to be maintained with tetracycline to turn off gene expression, negatively controlled system
Dexamethasone-inducible GR fusions	<i>Arabidopsis</i> , tobacco	GR::transcription factor fusion proteins identify immediate target genes, post-transcriptional activation	Only suitable for transcription factors, inducer toxic in some cases, induction of defense-related genes in <i>Arabidopsis</i>
Dexamethasone-inducible GVG	<i>Arabidopsis</i> , tobacco	Dexamethasone easily permeates plants cells and can be applied by various routes	Inducer toxic in some cases, induction of defense-related genes in <i>Arabidopsis</i> , slow off rate
Estradiol-inducible ER-C1	Black Mexican sweet corn cells	Relatively low levels of estradiol is required, no apparent toxic effects	Not tested in transgenic plants, not suitable for plants with phytoestrogens
Estradiol-inducible XVE	<i>Arabidopsis</i> , tobacco, BY2 cells	No apparent toxic effects, low basal level and high inducible levels	Not suitable for plants with phytoestrogens (soybean), not suitable for field use
Dex-inducible and tet-repressible TGV	<i>Arabidopsis</i> , tobacco, BY2 cells	Dual control, quick shut off	Inducer toxic in some cases, defense-related genes may be induced
Tebufenozide-inducible GVHvEcR	Tobacco	Safe inducer, suitable for field application	High basal activity, foliar uptake of inducer is poor
Methoxyfenozide-inducible GVCfEcR	<i>Arabidopsis</i> , tobacco, BY2 cells	Low basal and high inducible levels, safe inducer, suitable for field application.	Foliar uptake of inducer is poor, turn off is slow
Methoxyfenozide-inducible GVOncEcR	Maize	Safe inducer, inducer moves systemically, suitable for field application.	Foliar uptake of inducer is poor
Ethanol-inducible AlcR	<i>Arabidopsis</i> , tobacco, potato	Inducer inexpensive and biodegradable, rapid reversible induction, suitable for field application	Volatile inducer, induction can triggered inadvertently, induction due to anoxia, inducer cannot be used for more than 2 d
Copper-inducible ACEI	<i>Arabidopsis</i> , tobacco, root nodules, BY2 cells	Ease of application of inducer, simple to use, inducer inexpensive	Induced levels are low, long exposure of inducer is toxic, does not work in BY2 or protoplasts derived from tobacco leaves, expression is variable in <i>Arabidopsis</i>
Benzothiadiazole-inducible PR-1a	Tobacco	Inducer not phytotoxic, long-lasting response, suitable for expression of disease resistance genes in field	Low induction levels, promoter also responds to oxidative stress and endogenous salicylic acid signals, inducer activates native genes
Safener-inducible In2-2	<i>Arabidopsis</i> , BY2 cells	Suitable for field application, inducer is an agrochemical	Inducer causes growth abnormalities, constitutive expression in roots, promoter responsive to other chemicals

Table 1.1: A comparison among different regulatory systems developed for plant gene expression based on chemically inducible promoter system [33].

1.1.4 Orthogonal combination

One of the major drawbacks of available inducible systems is that they are not truly orthogonal, i.e. putting genes under control of a native ligand binding domain requires the use of ligands that will also activate the endogenous pathways that are normally controlled by the receptor within a cell. This unwanted gene activation would make it impractical or impossible to accurately evaluate the physiological role of the intended target gene. For a receptor-ligand pair to overcome this obstacle they must be orthogonal to the contents of the cell [34], i.e., the receptor cannot be activated by endogenous small molecules and the ligand cannot activate endogenous receptors. One strategy to achieve this goal is to employ heterologous ligand-receptor pairs that are not found in target cell or organism, but this then significantly reduces the applicability. An alternate strategy that

does not depend on identifying heterologous ligand-receptor pairs from other organisms is to modify the receptor to be activated by compounds that do not activate the wild-type protein. This strategy has been used to selectively modulate the function of engineered immunophilins [35], kinases [36], and Src homology domain 3 (SH3) [37]. Peet and coworkers have previously observed that the nuclear hormone receptor retinoid X receptor (RXR) can be mutated to alter its activation in response to various ligands [38]. Koh and colleagues have shown that the related retinoic acid receptor (RAR) can be mutated to permit activation by neutral and basic ligands [39] and in recent work have demonstrated that the same strategy can be applied to the estrogen receptor [40]. Katzenellenbogen has also used mutagenesis of the estrogen receptor in combination with chemical synthesis to obtain new hormone-receptor combinations [41].

1.1.4.1 Criteria for orthogonal ligand-receptor pair

While creating a heterologous chemically inducible promoter system, knowledge of model system (e.g. plant or animal) to be used helps in deciding the orthogonal combination of ligand and receptor. The choice of receptor is based on the following criteria. First, the model organisms should not express any activity that is homologous or related to the exogenous transcription factor used. Second, the inducer of this transcription factor should be an inert signal for the cells used. Third, the binding site for this transcription factor should be complex and therefore unlikely to occur by chance in the control region of a mammalian gene. Good induction kinetics and structural knowledge are also important criteria to consider.

1.1.5 Nuclear hormone receptor (NHR) and estrogen receptor (ER)

One such receptor class fulfilling the above-mentioned criteria is nuclear hormone receptor (NHR) proteins which form a class of ligand activated proteins that, when bound to specific sequences of DNA serve as on-off switches for transcription within the cell nucleus. These switches control the development and differentiation of skin, bone and behavioral centers in the brain, as well as the continual regulation of reproductive tissues. NHRs function by binding to specific sites within the DNA genome and regulating the

expression of adjacent target genes in response to cognate ligand. Both negative and positive effects on transcription can occur [42-47]. Nuclear receptors (NRs) share a modular domain structure, which includes, from N-terminus to C-terminus, a modulatory A/B domain, the DNA-binding domain (DBD; C domain), the hinge D domain, the ligand-binding domain (LBD; E domain) and a variable C-terminal F domain that is absent in some NRs [48]. The nuclear receptor (NR) superfamily represents a large group of transcriptional regulators that encompass receptors for steroid and thyroid hormones, retinoids, vitamin D, peroxisome proliferator-activated receptors and orphan receptors [46].

Estrogen receptor (ER) is a well-characterized NHR that perfectly satisfies the above-mentioned criteria. Additionally ER shows rapid induction kinetics with its natural ligand estradiol (E2) and accepts diverse range of chemical structures as ligands where the induction varies significantly. The receptor shows dose-dependency which in turn can make the inducible system a quantitative one [49]. ER is a well studied NHR with significant structural data available [50]. The receptor has also good tolerance to modular exchange and to engineering approaches, making it an ideal candidate for orthogonal inducible system [51]. ER does not occur in plants, thus the regulatory system would not have any pleiotropic effect with host regulatory pathway. And once engineered for a synthetic chemical ligand without any hormonal properties, this system can also be used in mammals.

1.1.5.1 The Estrogen Receptor

Already at the beginning of the century, it was recognized that some substance(s) present in ovaries was important for e.g. uterine and vaginal functions. It was subsequently found, during the early thirties, that an estrous-inducing substance could be isolated from urine from pregnant animals, which led to the first successful isolation and purification of an estrogen, estrone. Somewhat later, a second estrogen, estradiol, was isolated from ovarian tissue. In the early sixties, a major breakthrough came with the work of Jensen [52, 53] reporting the presence of receptors for estrogens in target tissues. Intense research during the following years led to an improved understanding of the action of these receptors.

Another important step came forward during the mid-eighties when estrogen receptor-encoding cDNAs were cloned from several species [54-57] allowing detailed molecular investigations. Simultaneously, cDNAs coding for related proteins were cloned, indicating the presence of a large family of hormone receptors with similar mechanisms of action [46]. Thanks to the dramatic progress since then, a wealth of information has accumulated [58, 59]. We know today that the ER is a ligand-activated transcription factor that after activation binds to specific DNA elements (estrogen response elements, EREs) and modulates the transcription levels of target genes.

Like other nuclear receptors, ER exhibit a modular structure with different regions corresponding to autonomous functional domains (Fig. 1.1) that can be interchanged between related receptors without loss of function. ER has six domains named A through F from N- to C terminus as follows: Activation Function 1 (AF-1), a conserved DNA-binding domain (DBD), a linker region, a conserved region that contains the ligand binding domain (LBD) and the C-terminal Activation Function 2 (AF-2) region.

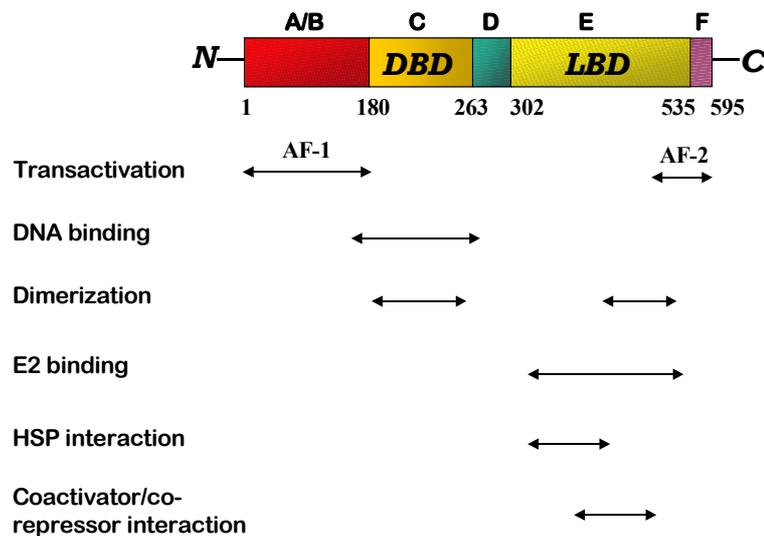


Figure 1.1: Graphical representation of structure and function of human ER α . The numbers indicate positioning of different domains within the receptor.

The **N-terminal A/B domain** contains the autonomous (ligand-independent) transactivation function AF-1 that modulates transcription in a gene and cell-specific manner. The A/B domain shows promoter and cell-specific activity, suggesting that it is

likely to contribute to the specificity of action among receptor isoforms and that it could interact with cell type-specific factors [60-63].

The **DNA binding domain** (DBD) is made of the C domain and plays an important role in receptor dimerization and in binding of receptors to specific DNA sequences. Several three-dimensional structures (nuclear magnetic resonance as well as X-ray investigations) are known for ER α DBD alone and in complex with DNA [64-66]. The DBD binds as a dimer with each monomer recognizing a six base pair sequence of DNA.

The **hinge region** or **D domain** is a 40 amino acid sequence separating the DBD and LBD which contains sequences for receptor dimerization and nuclear localization sequences (NLSs). Nuclear co-repressor proteins interact with the hinge region and increase the partial agonist activity of antagonist ligands. The D domain can be considered as a linker peptide between the DBD and the LBD [67].

The **Ligand Binding Domain** is a globular domain that harbors a hormone binding site, a dimerization interface (homo- and heterodimerization), and a coactivator and corepressor interaction function. In the three dimensional structure of ER α , the polypeptide chain is folded into the canonical α -helical sandwich topology associated with NR-LBDs [68]. This compact “wedge-shaped” domain is composed of 12 helices (H1- H12) that are arranged into 3 anti-parallel layers. The observed helical arrangements act as a molecular scaffold that maintains a sizable buried cavity of approximately 500 Å³ at one end of the LBD. While the overall conformation of the ER LBD is remarkably similar in the various ligand complexes, one aspect of the LBD, namely the orientation of the C-terminal transactivation helix (H12) is highly sensitive to the nature of the bound ligand [69]. The LBD also harbors activation function 2 (AF-2), which is a complex region whose structure and function are governed by the binding of ligands. Crystallographic studies with the LBDs of ER α and ER β revealed that the AF-2 interaction surface is composed of amino acids in helix 3, 4, 5, and 12 and that the position of helix 12 is altered by binding of ligands [70-72]. The LBD participates in several activities including hormone binding, homo- and/or heterodimerization, formation of the heat-shock protein complex and

transcriptional activation and repression. The binding of the hormone induces conformational changes that seem to control these properties and influence gene expression. The conformational changes that accompany the transition between the liganded and unliganded forms of the nuclear hormone receptors affect dramatically their affinity for other proteins.

Dimerization

All liganded ER-LBD complexes determined to date form homodimer within the crystal. The dimerization surface is extensive and encompasses about 14-16% of each monomer's accessible surface area. The interface is dominated by the H11 helices from each respective monomer which interact via a stretch of conserved hydrophobic residues at their N-terminal ends [73].

Ligand Binding

ER binds a wide repertoire of compounds with remarkable structural and chemical diversity [74]. Typically, ER ligands comprise two hydroxyl groups separated by a rigid hydrophobic linker region ranging from 10-12 Å. In addition, effective ligands possess a phenolic hydroxyl group. The ligand binding cavity is formed by residues from H3, H6, the loop region between H7 and H8, H8, H11 and H12. Ligand recognition is achieved through a combination of specific hydrogen bonds and the complementarity of the hydrophobic residues that line the cavity of the non-polar nature of ER ligands. The first crystal structure of an ER α -LBD provided the molecular basis of the interaction of the receptor with its natural ligand E2 [73]. The cavity is completely shielded from the external environment and buries the ligand in a highly hydrophobic environment mostly defined by 22 residues. Two polar regions located at opposite sides of the ligand binding pocket can be identified and they are involved in the anchoring of the E2 hydroxyl moiety at positions 3 and 17. The phenolic hydroxyl group of the A-ring (3-OH) is hydrogen bonded to E353 from H3, and to R394 from H5 and a water molecule. The hydroxyl group of the D-ring (17 β -OH) forms a single hydrogen bond with H524 (H11). The cavity delimited by the protein exhibits a probe accessible

volume of 450 Å³, which is much larger than the molecular volume of natural ligand (250 Å³) [75]. While the rigid protein architecture around the ‘A-ring’ pocket imposes an absolute requirement on effective ER ligands to contain a planar ring group, the remaining of the binding cavity is quite accommodating [69, 72, 73]. It is well known that ligand binding induces a conformational change in nuclear receptors. Protease digestion and antibody accessibility studies reveal that agonists and antagonists trigger distinct structural alterations of nuclear receptor LBDs [76]. A comparison between apo (ligand-free) and holo-(ligand-occupied) ER α LBD reveals several differences in the variable region. The most striking difference is the repositioning of the C-terminal helices, which is shown in Figure 1.2. In the apo LBD, helix H11 is almost perpendicular to H10 and points towards the LBP and some of the hydrophobic residues of H11 partially fill and stabilize the LBP. Helix H12, which contains the residues of the AF-2 activation domain core, extends away from the LBD. In the holo-LBD (bound E2), helix H11 is positioned in the continuity of H10, and H12 sits snugly over the ligand binding cavity and is packed against H3, H5/6 and H11. Although it makes no direct contact with E2, it forms the lid of the binding cavity and projects its inner hydrophobic surface towards the bound hormone. This precision positioning of H12, which is observed in all well-known structures of liganded forms of LBD, seems to be prerequisite for transcriptional activation as, by sealing the ligand-binding cavity, it generates a competent AF-2 that is capable of interacting with the coactivators. In contrast, the alignment of H12 over the cavity is prevented by antagonists (for example, raloxifene, tamoxifene etc), and instead the helix lies in a groove formed by H5 and the C- terminal end of H3 [73].

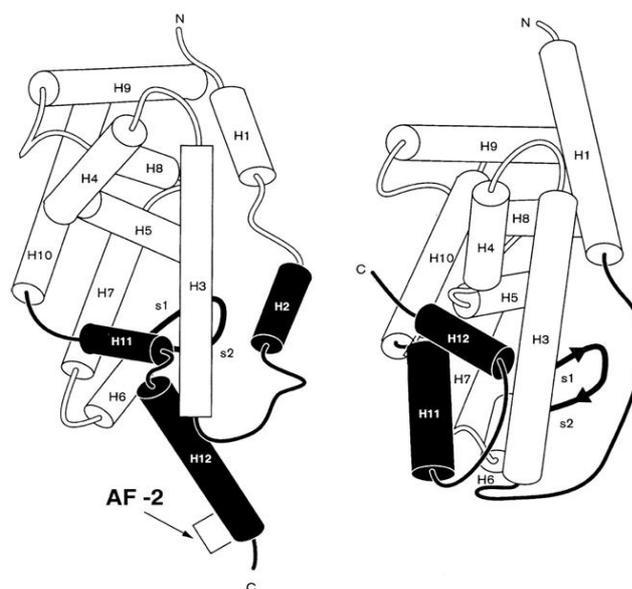


Figure 1.2: Schematic drawing of the ER LBD. On the *left*, the LBD from the crystal structure of the unliganded ER α is shown. On the *right*, the ligand-bound LBD of the ER α is shown. Cylinders represent α -helices that are numbered from 1 to 12. Note the different position of the C-terminal helix 12 that contains the core AF-2 domain in both situations.

Interaction with coactivators

A variety of nuclear factors termed coactivators serve as intermediaries between liganded NRs and the general transcription machinery [77]. p160 coactivators, such as steroid receptor coactivator 1 (SRC-1) and transcription intermediary factor 2 (TIF-2), are recruited by ER in a ligand- and AF-2-dependent manner through a distinctive common signature motif termed the NR-box which comprises the core consensus sequence LxxLL (where L is Leucine and x is any amino acid) [78]. The integrity of the coactivator binding groove is highly dependent on the orientation of H12. In turn, the alignment of H12 is highly sensitive to the nature of the bound ligand. Only those ER ligands that promote the positioning of H12 over the ligand binding cavity will act as full agonist. ER AF-2 antagonists, such as raloxifene and tamoxifene, are able to prevent the proper alignment of H12 through direct steric effects between their characteristic basic side chain substituents and H12 [72, 73]. Consequently, the recruitment site is incorrectly formed and such receptors are unable to interact with p160 coactivators.

1.1.6 Mechanism of E2 action

The dissection of the molecular mechanism by which estrogen activates its receptor and thereby modulates target gene transcription is central to the understanding of the complex biology governed by estrogen. In the absence of any ligand, the ER is generally believed to reside in the cytoplasm of target cells where it is associated with heat shock proteins (hsps) and remain transcriptionally inactive [79]. After diffusion into the cell, estradiol binds the estrogen receptor with high affinity. The binding of estradiol to ER induces an allosteric change which subsequently leads to dissociation of hsp from the ER followed by homodimerization of the receptors which allows the receptor–hormone complex to be transported through the nuclear membrane into the nucleus and bind to its specific DNA target, the ERE (Figure 1.3) [80]. These DNA sequences are found in the promoter region near the start site of transcription of many but not all genes directly regulated by estrogens. This ERE-bound, ligand-occupied ER complex can either activate or suppress transcription of a downstream target gene in a cell and promoter-specific manner [46, 81, 82]. In the case that the complex recruits coactivators and induces the transcription of a downstream target gene, this biological function is generalized as transactivation or transcriptional activity of ER.

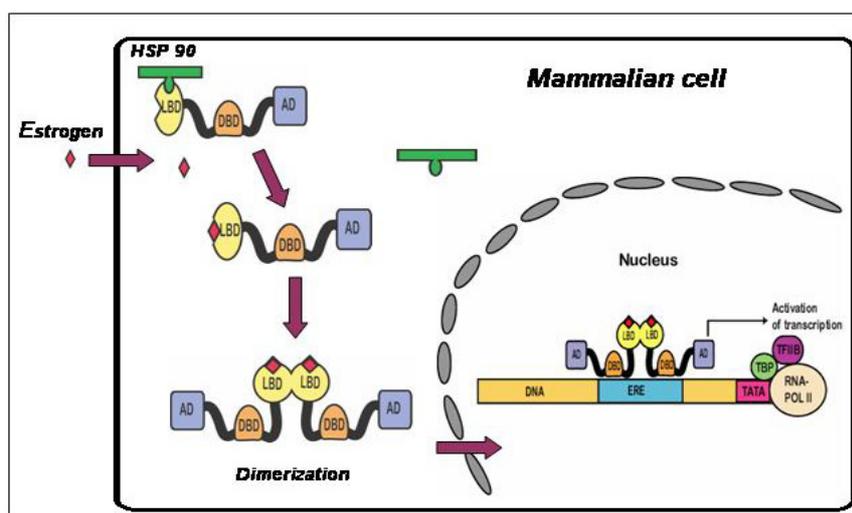


Figure 1.3: Schematic representation of E2 action in ER α .

1.1.7 Orthogonal ligand-receptor pair: chemical synthesis vs. protein engineering

The engineering of highly specific ligand–protein pairs that are orthogonal to existing natural ligand–protein pairs represents an important challenge that would allow us to selectively regulate gene expression for applications such as the study of gene function, gene therapy, tissue engineering and metabolic engineering [83, 84]. The ability to engineer new specific ligand–protein pairs would also be useful in the selective regulation of cellular processes for the study of phenomena such as apoptosis, genetic recombination, signal transduction and motor protein function [34].

Numerous strategies for engineering orthogonal ligand–receptor pairs have been developed with varying degrees of effectiveness in terms of generality of approach, as well as degree of specificity of interaction achieved. Two broad approaches have been applied (Figure 1.4): (i) identification of a mutant receptor with weakened response to the natural ligand and creation of various synthetic ligands to rematch the interaction with the target mutant receptor; and (ii) identification of a synthetic ligand that poorly activates the natural receptor and creation of variant receptors that are strongly activated by the target ligand.

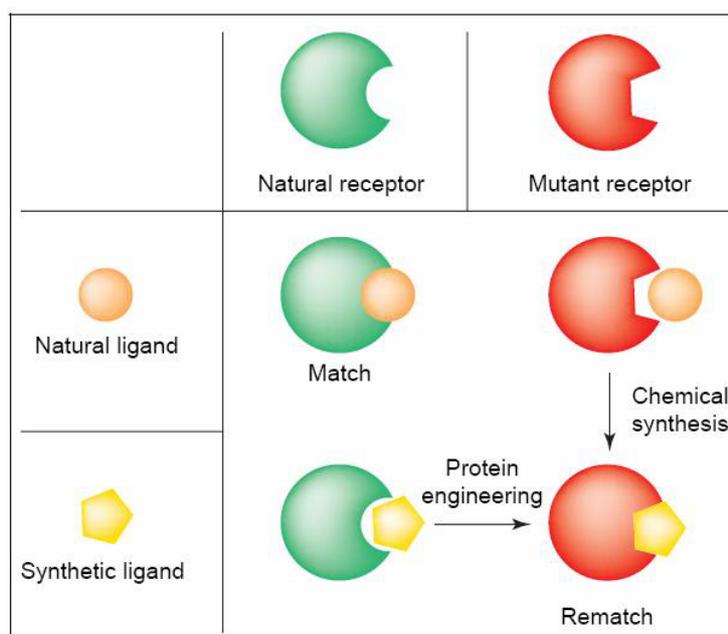


Figure 1.4: Scheme showing approaches to create orthogonal ligand-receptor pair. Two broad approaches can be used to achieve the desired rematch: (i) chemical synthesis, and (ii) protein engineering.

The chemical manipulation of ligands based on rational structure–function considerations to rematch interaction with a given mutant receptor of human estrogen receptors (hER α or hER β) has been demonstrated by several studies [40, 41, 85]. Despite these advances, the degree of specificity shift towards the target mutant receptor achieved in even the best cases was only moderate. In addition, the general application of this approach is hindered by two factors: (i) our ability to modify ligands to create optimized variants to rematch a mutant receptor is limited to the changes allowable by existing chemical synthesis methods; and (ii) the synthesized ligand modifications are mostly based on rational predictions of complementary interaction with the protein ligand binding pocket, and such predictions are often inaccurate.

Although proteins are far more structurally complex on a molecular level than small molecules, they are in some sense easier to manipulate structurally. For example, substitution of one or more protein residues with amino acids of other identities is routinely carried out using any of several simple procedures for DNA-based site-directed mutagenesis and large quantities of the protein with the substituted amino acid can be expressed and purified in a matter of days. Therefore, numerous attempts have been made to create protein variants based on NHRs that preferentially respond to a selected synthetic ligand over the natural ligand.

1.1.7.1 Engineering approach

Protein engineers now possess an unprecedented capability to design, modify and engineer naturally occurring proteins at the molecular level. Two different yet complementary approaches have been developed in the past two decades: rational design and directed evolution. Rational design involves the rational alterations of selected residues in a protein via site-directed mutagenesis, and requires detailed knowledge of protein folding, structure, function, and dynamics. In contrast, directed evolution mimics the process of natural evolution in the test-tube, involving repeated cycles of creating molecular diversity by random mutagenesis and/or gene recombination and screening/selecting the functionally improved variants. Both approaches have been used successfully to engineer a wide variety of protein functions such as stability, activity,

affinity, selectivity and pH profiles [86-90]. While rational design has been used successfully in protein design strategies for the past two decades, the difficulty of fully comprehending the structure-function relationship of proteins makes the success rate of this approach low. Despite recent advances in computational protein design and structural proteomics, identifying the molecular basis for the desired protein function by rational design is not straightforward and reliable [87]. In addition, recent findings that many protein functions are not confined to a small set of amino acids but are affected by residues far away from active sites have added another level of difficulty [86, 91-95]. It is therefore, that the second general approach for protein design- directed evolution is important. In directed evolution, changes are imposed on a protein on a DNA-level via some form of mutagenesis or DNA recombination, and the resulting library of protein variants is screened out using a screening system that assess the protein's performance in the function of interest. Protein variants showing improved function are then subjected to the same diversification-screening treatment in an iterative approach, until a protein variant with sufficiently improved function is achieved.

While directed evolution is undoubtedly a powerful approach for protein design, it also has some significant limitations. In order to fully consider all the possible variants that can be generated from a 300 amino acid protein, one would need to screen through a library of 20^{300} different protein variants. Not only is this library size impossible to generate with conventional mutagenesis techniques such as error-prone PCR [96], but even if this library size is possible to generate, it would be far too large to experimentally screen out (in a practical *E. coli* or yeast screening host, the limitation of DNA transformation means that a generated library size of $10^6 \sim 10^7$ is considered very good [97, 98]). Table 1.2 shows the theoretical library size for various numbers of simultaneous mutations generated in a protein consisting of 300 amino acids. It is evident from the analysis that a library where only one amino acid substitution is allowed per protein lends itself to comprehensive screening. If two simultaneous mutations were to be allowed per protein, the large library size would make it very difficult, if not impossible to screen through all the possible variants. If three simultaneous mutations were to be allowed per protein, no currently available techniques would be able to screen out all the library variants within a reasonable amount of time. To date, most directed evolution approaches

employ a mutation rate whereby one, sometimes two amino acid substitutions per protein (corresponding to ~2-3 base substitutions per gene) are allowed [99]. In rare cases, higher mutation rates (translating to ~10-30 base substitutions per gene) have been employed to successfully engineer improved mutants [100, 101], although the resulting library sizes from these cases were not by any means comprehensively screened out. These observations demonstrate that a directed evolution approach can only explore a tiny fraction of the total possible search space when designing a protein. Therefore, a purely directed evolution approach may not be the best way to optimize a protein's function.

Mutation(s)	Potential Library Size *
1	6.00×10^3
2	1.79×10^7
3	3.06×10^{10}
4	4.31×10^{13}
5	4.85×10^{16}

* Calculated by $N=20^m \times 300! / [(300-m)!m!]$ where m = number of simultaneous mutations in a protein.

Table 1.2: The potential mutant library size of a protein with 300 amino acids.

One answer to the dilemma that arises from using a purely directed evolution approach to design proteins is to use rational design to guide and focus the directed evolution approach [102, 103]. Such an integration of rational design and directed evolution may take the form of narrowing the region of DNA-based random mutagenesis down to a specific sub-fragment of the gene of interest, randomly mutating the base pairs corresponding to a given subset of important protein residues, and /or narrowing down the number of amino acid substitutions that can be made to particular residues (in total, 20 possible amino acid substitutions can be made per residue).

1.2 Aim of the study

The aim of this thesis was to use a two-pronged approach of rational design combined with directed evolution to create a transcriptional switch. Here we describe mutation of hER α LBD and identification of an orthogonal ligand-receptor pair. The nonsteroidal orthogonal ligands were chosen because of their relatively small structure, which in turn made them quite inactive in natural ER. While inactive, they still have the A and D ring which is typical to estrogen and thus serve as the platform for creating an orthogonal ligand receptor pair.

2 Materials and Methods

2.1 Materials

2.1.1 Organisms

2.1.1.1 Bacteria

Bacterial Strain	Description	Reference
<i>Escherichia coli</i> BL21 (DE3)	F-; <i>omp T hsdSB (rB mB) gal dcm</i>	[104]
<i>E. coli</i> DH5a	F-; <i>gyrA96 (Nal^r), recA1, relA1, endA1, thi-1, hsdR17 (rk-mk⁺), glnV44, deoR, D (lacZYA-argF) U169 [p80dD (lacZ) M15]</i>	[105]

2.1.1.2 Yeast (*Saccharomyces cerevisiae*)

Yeast strain	Description	Reference
PJ69 4A	MAT α ; <i>trp1-901, leu 2-3,112, ura3-52, his3-200, gal4Δ, gal80Δ, GAL2-ADE2, LYS2::GAL1-HIS3, met2::GAL7-lacZ</i>	[106]
YRG2	MAT α ; <i>ura3-52 his3-200 ade2-101 lys2-801 trp1-901 leu2-3 112 gal4-542 gal80-538 LYS2::UAS_{GAL1}-TATA_{GAL1}-HIS3 URA3::UAS_{GAL4} 17mers(x3)-TATA_{CYC1}-lacZ</i>	Stratagene

2.1.2 Media and additives

All media and heat-stable solutions were sterilized by autoclaving for 20 min at 121 °C. The heat sensitive solutions were sterilized by filtering through a 0.2 μ m membrane filter (Heinemann Labortechnik GmbH, Germany).

2.1.2.1 Bacterial media

Medium	Components and concentrations
dYT	16g/l tryptone, 10g/l yeast extract, 5g/l NaCl, pH 7.0 (NaOH)
LB	10g/l tryptone, 5g/l yeast extract, 10g/l NaCl, pH 7.0 (NaOH)
SOC	20g/l tryptone, 5g/l yeast extract, 0.5g/l NaCl, 186.38 mg/l KCl, 2.033g/l MgCl ₂ , 3.6g/l glucose, pH 7.0 (NaOH)

2.1.2.2 Yeast media

Medium	Components and concentrations
yPD	20g/l peptone, 10g/l select yeast extract, 15g/l agar (only for solid medium), 20g/l glucose pH adjusted to 5.8 with HCl
yPAD	yPD medium with 100mg/l adenine
SOS (always prepared fresh)	750µl yPD, 750µl 2M sorbitol and 9.75µl 1M CaCl ₂
SD drop-out medium	6.7g/l yeast nitrogen base without amino acids, 12-15g/l select agar (only for solid medium), 0.61 CSM-appropriate drop-out, pH 5.8 with KOH

2.1.2.3 Plant Media

Media	Components and concentrations
2MS	4.4g/l MS salt, 20g/l sucrose, 6.4g select agar, pH adjusted to 5.7 with KOH
Shoot Induction Media	2MS medium supplemented with 0.5mg/l BAP and 0.1mg/l NAA

2.1.2.4 Additives

Substance	Working concentration	Stock solution and solvent
100XAdenine	20mg/l	2g/l in H ₂ O
Ampicillin	100mg/l	100mg/ml in H ₂ O
Chloramphenicol	25µg/ml	25mg/ml in 100% EtOH
IPTG	0.5mM	100mM in H ₂ O
500XHistidine	50mg/l	2.5mg histidine in 100ml H ₂ O
Kanamycin	50mg/l	50mg/ml in H ₂ O
100XLeucine	20mg/l	2g/l in H ₂ O

2.1.3 Nucleic Acids

2.1.3.1 Plasmid

Plasmid	Description	Reference
pGBT9	Vector containing Gal4DBD for Y2H screening where the fusion protein is constitutively expressed from <i>ADHI</i> promoter; a shuttle vector with <i>amp^r</i> (<i>E. coli</i>) and <i>TRP1</i> (yeast) marker	Clontech
pGBT-LBD	hERα LBD (residue 282-595) fused to the Gal4DBD and VP16AD in the N- and C-terminus in pGBT9 vector	This work

Plasmid	Description	Reference
pBDGal4Cam	Vector containing Gal4DBD for Y2H screening where the bait fusion protein is constitutively expressed from <i>ADHI</i> promoter; a shuttle vector with <i>cam^r</i> (<i>E. coli</i>) and <i>TRP1</i> (yeast) marker	Stratagene
pBDGal4LBDCam	hER α LBD (residue 312-595) fused to the Gal4DBD the pBDGal4Cam vector	[107]
pAD-Gal4-2.1	Vector for GAL4AD fusion of insert DNA for Y2H screen; contains <i>amp^r</i> and <i>LEU2</i> marker	Stratagene
pAD-Gal4-2.1 SRC-1	Human SRC-1 fused to the Gal4AD in the pAD-Gal4-2.1 vector	[108]
pET15b	Vector for 6XHis-fusion protein expression under the control of IPTG-inducible <i>T7</i> promoter and lac-operator, <i>amp^r</i>	Novagene
pET15b-LBD	pET15b derivative containing the coding sequence for of hER α LBD (residue 302-553)	This work
pET28a	Vector for 6XHis-fusion protein expression under the control of IPTG-inducible <i>T7</i> promoter and lac-operator, <i>km^r</i>	Novagen
pET28a-LBD	pET28a derivative containing the coding sequence for of hER α LBD (residue 302-553) as a 6XHis fusion	This work

2.1.3.2 Primer

Name of the primer	Sequence (5'→3')
LBD Upper	TCCGCTGACTAGGGCACATCTGAC
LBD Lower	CGTCTAGCGCGTCGGCATG
M388A Upper	GAGATCCTCGCGATTGGTCTCGTCTGGCGCTC
M388A Lower	GACCAATCGCGAGGATCTCTAGCCAGGCACATTC
L428A Upper	CGACATGGCGCTGGCTACATCATCTCGGTTCC
L428A Lower	TAGCCAGCGCCATGTCTGAAGATCTCCACCA
Y537S Up	GCCCCTCTTGACCTGCTGCTG
Y537S Low	GTCAGAGAGGGGCACCACGTTC
LBD Sense	AATTCTCTAAGAAGAACAGCCTGGCCTTGTC
LBD Antisense	CTCGAGTAGTGGGCGCATGTAGGCG
Cam L ER	CGACATCATCATCGGAAGAG
Cam R ER	GCTTGGCTGCAGTAATACGA

Name of the primer	Sequence (5'→3')
343Sat5	GCTTCGATGNNSGGCTTACTG
343Sat3	CAGTAAGCCNNSCATCGAAGC
346Sat5	GATGGGCTTANNSACCAACCTG
346Sat3	CAGGTTGGTNNSTAAGCCCATC
347Sat5	GGCTTACTGNNSAACCTGGCAG
347Sat3	CTGCCAGGTTNNSCAGTAAGCC
349 Sat5	CTGACCAACNNSGCAGACAGG
349Sat3	CCTGTCTGCNNSGTTGGTCAG
350 Sat5	GACCAACCTGNNSGACAGGGAG
350 Sat3	CTCCCTGTCNNSCAGGTTGGTC
383 Sat5	GAATGTGCCNNSCTAGAGATC
383 Sat3	GATCTCTAGNNSGGCACATTC
384 Sat5	TGTGCCTGGNNSGAGATCCTG
384 Sat3	CAGGATCTCNNSCCAGGCACA
387 Sat5	CTAGAGATCNNSATGATTGGTC
387 Sat3	GACCAATCATNNSGATCTCTAG
388 Sat5	GAGATCCTGNNSATTGGTCTC
388 Sat3	GAGACCAATNNSCAGGATCTC
391 Sat5	GATGATTGGTNNSGTCTGGCGC
391 Sat3	GCGCCAGACNNSACCAATCATC
404 Sat5	GAAGCTACTGNNSGCTCCTAAC
404 Sat3	GTTAGGAGCNNSCAGTAGCTTC
421 Sat5	GTAGAGGGCNNSGTGGAGATC
421 Sat3	GATCTCCACNNSGCCCTCTAC
424 Sat5	CATGGTGGAGNNSSTTCGACATG
424 Sat3	CATGTCGAANNSCTCCACCATG
425 Sat5	GTGGAGATCNNSGACATGCTG
425 Sat3	CAGCATGTCNNSGATCTCCAC
428 Sat5	CTTCGACATGNNSCTGGCTACATC

Name of the primer	Sequence (5'→3')
428 Sat3	GATGTAGCCAGNNSCATGTCTGAAG
521 Sat5	GAGTAACAAANNSATGGAGCATC
521 Sat3	GATGCTCCATNNSTTTGTTACTC
524 Sat5	GGCATGGAGNNSCTGTACAGC
524 Sat3	GCTGTACAGNNSCTCCATGCC
525 Sat5	CATGGAGCATNNSTACAGCATG
525 Sat3	CATGCTGTANNSATGCTCCATG
528 Sat5	CTGTACAGCNNSAAGTGCAAG
528 Sat3	CTTGCACTTNNSGCTGTACAG
II-383Sat5	GAATGTGCCNNSTTCGAGATC
II-383Sat3	GATCTCGAANNSGGCACATTC
II-388Sat5	GAGATCATGNNSATTGGTCTC
II-388Sat3	GAGACCAATNNSCATGATCTC
pBDCam5N	TGCCGTCACAGATAGATTGG
pBDCam3C	AGAAATTCGCCCCGGAATTAG

2.1.4 Enzymes

Enzyme	Manufacturer
DNA Restriction enzymes type II	Fermentas New England Biolabs
DNase (RNase-free)	Promega
iProof High Fidelity DNA Polymerase	Biorad
Lysozyme	Serva
Lyticase	Sigma
Moltag	Molzym
Pfu DNA polymerase	Stratagene
Phusion High Fidelity DNA Polymerase	New England Biolabs
Reverse Transcriptase	Fermentas

Enzyme	Manufacturer
Ribonuclease Inhibitor	Fermentas
T4 DNA Ligase	Fermentas
Taq DNA Polymerase	Fermentas

2.1.5 Synthetic ligands

A library of 56 novel estron analogs were synthesized by Dr. Carsten Vock and Dr. Ilga Krimmelbein [109] in the Institute for Organic and Biomolecular Chemistry, Georg August University Göttingen, Germany. These analogs were synthesized using a palladium-catalyzed *Sonogashira*-coupling reaction. A broad range of substitution patterns were covered by these analogs. A complete list of ligands with structure is included in the Appendix.

2.1.6 Chemicals

Chemical	Manufacturer
Acrylamide: N,N-methylene-bisacrylamide (37.5:1) 30% (w/v)	Roth
Acrylamide: N,N-methylene-bisacrylamide (19:1) 40%(w/v)	Roth
Acetic acid	Roth
Adenine hemisulfate	Sigma
Agar bacteriological	Gibco BRL
Agarose Seakem LE	Biozym
Albumin fraction V	Applichem
3-Amino-1,2,4 Triazole (3-AT)	Sigma
Ammonium persulfate (APS)	Biometra
Ampicillin	AGS
β -mercaptoethanol	Roth
17 β -estradiol (E2)	Sigma
Boric Acid	Serva
Bradford Reagent	Roth
Bromophenol Blue	Roth
Calcium Carbonate	Merck
Calcium Chloride	Merck
Chloramphenicol (Cam)	Roth
Chloroform	Merck
Coomassie Brilliant Blue	Merck
Dimethylsulfoxide (DMSO)	Roth
Dithiothreitol (DTT)	Sigma

Chemical	Manufacturer
DNA Carrier	Sigma
dNTPs	Fermentas
Human Estrogen Receptor α , Full Length	Sigma
Ethanol (EtOH)	Merck
Ethidiumbromide (EtBr)	Roth
Ethylenediaminetetraacetic acid (EDTA)	Roth
Formaldehyde (37%)	Roth
Formamide	Fluka
Glucose	Sigma
Glycerol	Roth
Glycine	Roth
Hepes	Roth
Histidine	Sigma
Hydrochloric acid (HCl)	Roth
Imidazole	Sigma
Isoamylalchol	Roth
Isopropanol	Roth
Isopropylthiogalactoside (IPTG)	BioTech Grade
Leucine (Leu)	Sigma
Lithium Acetate	Sigma
Kanamycin (Km)	Sigma
Magnesium chloride	Applichem
Magnesium sulfate, heptahydrate	Applichem
Methanol (MeOH)	Roth
Methionine (Met)	Sigma
Non-fat dried milk power SUCOFIN	TSI
o-nitrophenyl- β -D-galactopyranoside (ONPG)	Sigma, Roth
Orange G	Sigma
Peptone	Roth
Phenol	Applichem
Phenylmethanesulfonylfluoride (PMSF)	Fluka
Polyethylene glycol of various concentrations	Roth
Potassium chloride	Roth
Select Agar	Life Technologies
Sodium acetate	Roth
Sodium Carbonate	Merck
Sodium Chloride (NaCl)	Roth
Sodium dodecyl sulfate (SDS)	Roth
Sodium hydroxide (NaOH)	Merck
Sodium Phosphate, dibasic (Na ₂ HPO ₄)	Merck
Sodium Phosphate, monobasic (NaH ₂ PO ₄)	Merck
Sorbitol	Sigma
Sucrose	Roth
TEMED	Roth
Tris	Roth

Chemical	Manufacturer
Tryptophan	Sigma
Tween-20	Sigma
X-ray film developer LX24	Kodak
X-ray film fixator AL4	Kodak
Xylenecyanol FF	Roth
Yeast extract	Gibco BRL
Yeast nitrogen base without amino acids	Difco

2.1.7 Kits

Kit	Manufacturer
BD Universal His Detection Kit	BD Biosciences
Big Dye Enhanced Terminator Cycle Sequencing Kit v3.1	Perkin-Elmer
Hithunter Estrogen Assay Kit	DiscoverRx
Nucleospin Extract II Kit	Macherey-Nagel
Nucleospin Plasmid	Macherey-Nagel
Nucleobond AX 500	Macherey-Nagel
Qiafilter Midi/Maxi Kit	Qiagen
Qiaprep Spin Miniprep Kit	Qiagen
Qiaquick Gel Extraction Kit	Qiagen
Zymoprep Kit	Zymo Research Corporation

2.1.8 Buffers and solutions

Buffer/Solution	Component and concentration
Buffer I for alkaline lysis	50mM Tris-HCl, pH 8.0 10mM EDTA 100µg/µl RNase A
Buffer II for alkaline lysis	0.2M NaOH 1% (w/v) SDS
Buffer III for alkaline lysis	29.4 g potassium acetate 5ml formic acid water till 100ml
Coomassie destain solution	10% (v/v) glacial acetic acid
Coomassie fixing solution	25% (v/v) isopropanol 10% (v/v) glacial acetic acid
Coomassie staining solution	0.01% w/v coomassie brilliant blue G-250 10% (v/v) glacial acetic acid
DNA loading buffer	67% (w/v) sucrose 50mM EDTA, pH 8.0 0.42% (w/v) bromophenolblue 0.42% (w/v) xylenecyanol 0.42% (w/v) orange G

Buffer/Solution	Component and concentration
H Buffer for ONPG	100mM HEPES/KOH, pH 7.0 150mM NaCl 2mM MgCl ₂ 1% (w/v) BSA
Ni-NTA binding buffer	300mM NaCl 50mM (Na ₂ HPO ₄ + NaH ₂ PO ₄), pH 8.0 10mM imidazole
Ni-NTA washing buffer	300mM NaCl 50mM (Na ₂ HPO ₄ + NaH ₂ PO ₄), pH 6.0 20mM imidazole
1XPBS	68mM NaCl 58mM Na ₂ HPO ₄ 17mM NaH ₂ PO ₄ Adjust pH to 7.4 with NaOH
PBST	1XPBS 0.1% (v/v) Tween 20
Ponceau S	2g Ponceau S 30g Trichloroacetic acid 30g Sulfosalicylic acid H ₂ O to 100 ml
10X SDS-PAGE running buffer	25mM Tris-HCl, pH 8.3 200mM Glycine 0.1% (w/v) SDS
3X SDS-PAGE sample stop buffer (without β-mercaptoethanol)	4.5g SDS 15g sucrose 18.8 ml 1M Tris pH 6.8 1 teaspoon of bromophenolblue
20XTAE	800mM Tris 20mM EDTA 2.3% (v/v) glacial acetic acid
100XTE	1M tris-HCl, pH 8.0 0.1M EDTA
Transfer Buffer for western blotting	39mM glycine 48mM Tris base 0.037% (w/v) SDS (electrophoretic grade) 20% (v/v) MeOH

2.1.9 Consumables

Product	Manufacturer
3MM paper	Whatman
96-well plate	Evergreen Scientific Rainin Instruments Sarstedt
384-well black plate	Greiner Nunc
Electroporation cuvette	Biorad
Glass material	Brand Schott
Ni-NTA resin	Qiagen
Nylon Membrane Hybond N ⁺	Amersham Pharmacia
Plastic ware	Sarstedt, Greiner, Eppendorf

2.1.10 Software

Program	Manufacturer
Acrobat Reader 6.0	Adobe
Chemwindow	SoftShell International
Chromas 1.55	Technelysium Pty Ltd
Clone v7	Scientific and Educational Software
Molecular Operating Environment	Chemical Computing Group, Montreal
Multalin	http://bioinfo.genopole-toulouse.prd.fr/multalin/
Office	Microsoft
Oligo 4.0	MedProbe
PyMOL vers. 0.99rc6	DeLano Scientific LLC
Rastop2.1	SourceForge.net

2.1.11 Equipment

Equipment	Model	Manufacturer
Autoclave	3870 ELV	Tuttnauer
Automatic pipette		Gilson
Balance	SP051, SAC62 1207 MP2	Scaltec Sartorius
Cooling Centrifuge	Sorvall RC 5B Plus	DuPont
Digital Camera	Canon Powershot A510	Canon
DNA/RNA Calculator	GeneQuant II	Pharmacia
Electroporator	GenePulser II	Biorad
Fluorometer	Cytofluor II Plate Reader	PerSeptive

Equipment	Model	Manufacturer
Gel Documentation Station		MWG Biotech
Heating Block		Unitek Boekel Scientific
Heating Stirrer	RCT Basic	IKA Labortechnik
Ice Machine	AF 20	Scotman
Incubation Chamber		WTC Binder, Memmert
Microtitre Plate Spectrophotometer	MRX Plate Reader	Dynex
PCR Thermocycler	MiniCycler PTC 150 PTC-200	MJ Research
pH-meter	HI 9321	Hanna Instruments
Polyacrylamide Gel Chamber		AGS
Shaker	ST5	M Zipperer GmbH
Scanner	Scanjet 4c	Hewlett Packard
Sequencer	ABI Prism 310	Perkin-Elmer
Spectrophotometer	Novaspec Biochrom Spectramax 340PC	LKB Molecular Devices
Sterile Bench	Microflow Biohazard	Nunc
Table-top Centrifuge	Biofuge Pico	Heraeus Christ
Table-top centrifuge with cooling	5403	Eppendorf
Freezer (-80 ⁰ C)	C54285	New Brunswick Scientific
Ultrasound Sonicator	Soniprep 150	MSE
Vortex	L46	Labinco BV
Water Deionization System	Option 4, Maxima	ELGA
Water Bath	1086	GFL

2.2 Methods

2.2.1 Culture and storage of microorganisms

2.2.1.1 *E. coli* strains

A single colony of *E. coli* cells was cultured at 37°C in LB or dYT liquid or solid media in the presence of selective antibiotic. The liquid culture was grown under continuous shaking at 250 rpm in a 37°C shaker, while solid culture was grown in a 37°C incubator. Number of cells in a liquid culture was determined by measuring the optical density at 600nm (OD₆₀₀). A liquid culture with OD₆₀₀ equal to 0.1 corresponds to 2X10⁷ cells/ml. For long-term storage, the overnight culture was supplemented with glycerol to 20% and stored at -70°C.

2.2.1.2 Yeast *S. cerevisiae* strains

A small portion of the frozen glycerol stock was streaked onto a yPD (or appropriate SD) agar plate and incubated at 30°C for 2-3 days so that the yeast colonies became ~2 mm in diameter. These colonies were used as the working stock. The addition of adenine to yPD was crucial to enhance the growth of yeast strains that contain the *ade2-101* mutation (strain YRG2). Transformed yeast strains were grown on SD drop-out medium (lacks specific nutrients) in order to keep selective pressure on transformed plasmids. The 3-amino-1,2,4-triazole (3-AT), a competitive inhibitor of the *HIS3* gene product, was added in order to suppress background growth on SD medium lacking histidine, at concentrations ranging from 10-50mM, when required. Cells number of the liquid culture was determined by measuring the OD₆₀₀. A liquid culture with OD₆₀₀ equal to 1 corresponds to 1 x 10⁶ cells/ml. Sealed plates with Parafilm were stored at 4°C for up to 3 weeks. Yeast strains can be stored indefinitely in YPD medium with 25% glycerol at -70°C. A fresh culture was derived from the short or long-term stored culture always before performing chemical complementation based selection or screening.

2.2.2 Standard molecular biology methods

2.2.2.1 Isolation of plasmid DNA from *E. coli*

2.2.2.1.1 Alkaline lysis:

Small amounts of plasmid DNA for analytical purposes were isolated from *E. coli* using a modification of the alkaline lysis method [110]. *E. coli* overnight culture (stationary phase) of 1.5 ml was collected by centrifugation at 13000 rpm for 1 minute. The supernatant was removed and the cells were resuspended in 100µl of buffer I for plasmid DNA isolation (50mM Tris-HCl, pH 8.0; 10mM EDTA; 100µg/µl RNase A). The cell suspension was lysed for 5 minutes on ice using 200 µl of buffer II [0.2M NaOH; 1% (w/v) SDS]. The suspension was neutralized with 150 µl of buffer III (29.4 g potassium acetate; 5ml formic acid and water till 100ml). The solution was mixed well by inverting 8 times and the suspension was centrifuged for 10 minutes at 13000 rpm at room temperature. The aqueous solution (~400 µl) was transferred into a new eppendorf tube containing 1 ml of 96% (v/v) ethanol. The DNA was left to precipitate for 20 min at -20°C. Plasmid DNA was collected by centrifugation for 10 minutes at 13000 rpm and 4°C. The pellet was washed with 70% (v/v) ethanol and air-dried for 10 minutes at 37°C. The DNA was dissolved in 20 µl of EB buffer (10mM Tris-HCl, pH 8.5).

2.2.2.1.2 High-purity plasmid DNA isolation

Spin Miniprep

For sequencing and yeast transformation purposes, high-purity plasmid DNA was isolated using QIAprep or Nucleospin Mini kit following the manufacturer's instructions. Optional steps were always followed according to the manufacturer's recommendation. A 4ml overnight culture was used to isolate plasmid and the isolated DNA was eluted with 50µl (high copy) or 30µl (low copy) EB buffer.

Preparative quantity (Midi and Maxiprep)

Larger quantities of plasmid DNA from *E. coli* with high purity were isolated using Qiagen or Macherey-Nagel Midi and Maxi kit depending upon the required end concentration. Manufacturer's protocol including the optional recommendations was followed and final elution volume depended on the plasmid copy number, size of the DNA pellet to be eluted and final concentration required.

2.2.2.1.3 Isolation of plasmid DNA from yeast

To isolate plasmid from yeast for plasmid rescue purpose, a single colony (less than 2 weeks on selection plate) was used to inoculate a 3 ml overnight SD liquid culture (lacking the nutritional marker to keep selective pressure on the plasmid) at 30°C. Next day, cells were vortexed briefly and a 1.5 ml aliquot was centrifuged for 5 minutes at 13000 rpm to harvest the cells. Pelleted cells were suspended in 30µl of Lyticase solution (5 U/µl in TE, pH 7.5). The cell suspension was incubated for 1.5 hour at 37°C for lyticase-mediated cell wall digestion. The suspension was used then to isolate the plasmid DNA using the QIAprep Spin Mini kit following the manufacturer's instructions. Alternatively, 800µl overnight yeast culture was used for isolating plasmid using a Zymoprep yeast plasmid miniprep kit according to the protocol provided by the manufacturer. Thirty-µl EB buffer (pre-incubated at 70°C) was used for elution purpose. The concentration and quality of isolated yeast DNA plasmid was relatively low and was used for retransformation in *E. coli* so that plasmid DNA with analytically suitable quantity and quality could be isolated.

2.2.2.1.4 Estimation of nucleic acids concentration and purity

The concentration of nucleic acids was estimated by measuring their absorption in a spectrophotometer at a wavelength of 260 nm (maximum nucleic acid absorption value; due to the π -electron systems of the heterocycles of the nucleotides). In a cuvette having 10mm path-length where OD₂₆₀ reading is 1 corresponds to 50 and 40 µg/ml double-stranded DNA and RNA, respectively. Absorption at 280 nm (for the presence of aromatic rings from amino acids and phenol compounds) was used to give information

about the purity of the DNA or RNA sample, where an optimal ratio OD₂₆₀/OD₂₈₀ is in the range of 1.9-2.0 for RNA and 1.8 for DNA. DNA concentrations lower than 100ng/ μ l were measured on an agarose gel using the Gene Ruler Ladder Mix.

2.2.2.2 Nucleic acids gel electrophoresis

2.2.2.2.1 Separation of DNA on Agarose Gels

The electrophoretic separation of DNA for analytical and preparative purpose was done in a horizontal agarose gel (10 cm X 7 cm X 0.3 cm, 16 lanes) with 1X TAE as running buffer. DNA fragments ranging between 500 bp and 14 kb were run in an agarose gel concentration of 1% where DNA fragments with lower size were run in a 2% agarose gel. DNA samples were mixed with 1/10 volume of 10x DNA loading buffer, loaded in separate lanes and run at 120 V for 40-45 min. Ethidium bromide solution (0.1% w/v) was used to stain the DNA fragments. The detection of DNA was done under UV light (260nm). When a preparative gel was run and particular band fragments were needed to cut out, detection was done using larger wavelength UV light (320nm). Before exposure to the UV light, the gel was rinsed briefly in H₂O to reduce background staining. In a gel-documentation station, gels were visualized on a UV-transilluminator and documented. The sizes and amount of the DNA fragments were determined using DNA standards.

2.2.2.2.2 Elution of DNA fragment from agarose gel

The elution of DNA fragments from agarose gel was done using the QIAquick or Nucleospin Extract II Gel Extraction kit following the manufacturer's instructions. The eluted fragments were verified by electrophoresis as described above.

2.2.2.3 Restriction digestion of DNA molecules

Type II endonucleases were used to digest a double stranded DNA molecule for analytical and cloning purposes. The enzymes cut the DNA either as 5' or 3' "sticky" overhangs or as blunt ends. The digestion reactions were incubated in a buffer system optimized for the

used enzyme and in the case of double digestion a universal buffer system was used. The activity of the restriction enzymes was estimated in “units” (U), where 1 U was defined as that amount of enzyme cutting completely 1 μg of λ DNA in 60 minutes at optimal conditions. The minimal amount of enzyme necessary for each restriction was determined according to the following formula:

$$U_{\text{min}} = \frac{\text{bp}[\lambda] \times \text{No. of restriction sites in target DNA} \times \text{incubation period (hour)}}{\text{No. of restriction sites in } [\lambda] \times \text{bp of target DNA}}$$

Where, $\lambda = 48500$ bp

The incubation temperature was 37°C unless otherwise mentioned for particular restriction enzyme. Due to the adverse effect of high glycerin concentration, we preferred the total volume of restriction enzymes less than 10% in the restriction mix.

2.2.2.4 Ligation of DNA fragments

The conventional cloning of a DNA fragment into a selected plasmid was performed using the T4-DNA ligase enzyme, which is able to catalyze the formation of a phosphodiester chemical bond between free 5'-phosphate and 3'-OH groups of double-stranded DNA fragments and vectors. The donor DNA fragment (10 x accesses to the vector) was incubated with the vector DNA, 2 μ l of ligation buffer and 1 μ l of T4-DNA ligase for 2 hours at room temperature. The ligation of DNA fragments with blunt ends was performed in the presence of 5% (w/v) PEG 4000 with the ligation mix described above. Ligase activity was destroyed by heating at 65°C for 10 min before using the ligated DNA for transformation.

2.2.2.5 Homologous recombination in yeast

A 10-bp fragment was removed from the multiple cloning site of pBD-Gal4-Cam by digestion with *EcoRI* and *SalI*. For individual site saturation mutagenesis or error-prone PCR libraries, this gapped expression vector was cotransformed with mutagenized hER α LBD PCR product into YRG2 yeast cells pretransformed with the pGAD424-SRC1 plasmid by using the lithium acetate/single-stranded DNA/polyethylene glycol protocol [98]. The two cotransformed linear DNA fragments share 40-60 bp of homology at their ends, allowing the yeast cells to recombine the linear fragments *in vivo*, giving rise to circular plasmids expressing the mutagenized fusion protein Gal4DBD-hER α LBD. All saturation mutagenesis library transformations were plated onto SC minimal media agar plates lacking leucine and tryptophan (for selection of the plasmids expressing the pGAD424-SRC1 and pBD-Gal4-Cam plasmids, respectively). Error-prone PCR library transformations were plated onto SC minimal media agar plates lacking leucine, tryptophan, and histidine and containing appropriately concentrated target ligand.

2.2.2.6 Polymerase chain reaction (PCR)

Depending upon aim of our amplification, we made four different polymerase chain reactions with variables like DNA polymerase, primer, buffer etc.

Standard amplification

Amplification of target DNA fragments was performed by polymerase chain reaction in thermal cycler according to the protocol from Mullis and Faloona [111] with required modification. The reaction started with the denaturation of two strands of a DNA template. The 5' complementary strands of the denatured DNA were recognized and hybridized with specific primers (annealing). A high fidelity polymerase (iProof or Phusion) was used which catalyzed elongation of a newly synthesized chain and the complementary polymerization of nucleotides to the free 3'-OH group of the primer. Repeating the denaturation, annealing and elongation for x cycles (usually from 25 to 35) exponentially enriched the reaction with the primer-flanked DNA sequence. In some cases suitable synthetic restriction sites were incorporated to the 5'-end of the primer for cloning purposes. The PCR reaction was carried out in a 50 μ l reaction volume with the following constituents: 10-50 ng template DNA, 10 pmol sense primer, 10 pmol antisense primer, 0.2 mM dNTPs, 2 μ l of 10x buffer, 2U polymerase and H₂O up to 50 μ l. The amplification reaction was done in a PCR thermocycler using program mentioned in Table 2.1.

PCR Site Directed Mutagenesis

The principle of the PCR site-directed mutagenesis is that a mismatched oligonucleotide is extended incorporating a "mutation" into a strand of a selected DNA that can be cloned [112]. In this study the point mutation was introduced by PCR overlap extension method using high fidelity polymerase like iProof (Biorad) or Phusion High Fidelity DNA Polymerase (New England Biolabs). In the overlap extension method two separate PCRs were conducted to amplify two PCR fragments. Each PCR reaction uses one flanking primer that hybridizes at one end of the target sequence and one internal primer that hybridizes at the site of the mutation and contains the mismatched nucleotide bases. Finally, a third PCR was conducted using flanking primers and the generated DNA fragments from previous reaction as template.

Random point mutagenesis using error-prone PCR

For generating randomly point-mutated (error-prone PCR) libraries, two flanking primers were used for a template DNA. The method introduces mutations in the gene of interest using a PCR reaction under conditions that induce an increased error-rate of the DNA-polymerase. For the manipulation of mis-incorporation, we used higher Mg^{+2} concentration, supplemented with Mn^{+2} , used unbalanced dNTP concentrations, decreased annealing temperature and increased number of cycles. We used the protocol from Cadwell and Joyce with slight modification where Moltaq (Molzym) or *Taq* DNA polymerase (New England Biolab) were used with $MnCl_2$ concentration of 0.15mM. $MnCl_2$ was added at the end in order to avoid precipitation.

	Standard amplification		Random mutagenesis		Site-directed mutagenesis	
	Temp	Time	Temp	Time	Temp	Time
Initial denaturation	98°C	1 min	94°C	2 min	98°C	1 min
Denaturation	98°C	15 sec	94°C	30 sec	98°C	15 sec
Annealing	58°C	30 sec	55°C	30 sec	55°C	30 sec
Extension	72°C	2 min	68°C	2 min	72°C	2 min
X time to 2	30		30		25	
Final extension	72°C	10 min	72°C	7 min	72°C	10 min

Table 2.1: Variable PCR program for standard amplification, random and site directed mutagenesis.

Saturation Mutagenesis

The procedure used for generating libraries whereby single residues were randomized to all 20 possible amino acids involves overlap extension coupled with PCR [112]. Briefly, four primers were used to generate an amplified gene library consisting of a saturation mutagenized residue: two primers flanking the region to be amplified and two exactly complementary degenerate primers incorporating the residue to be mutated. The two degenerate primers incorporating the randomized amino acids substituted the codon corresponding to the target residue with the sequence NNS and contained 9-10 additional bases on either side (5' and 3'). The choice of the substitution NNS allows the incorporation of all 20

amino acids while keeping the total number of codon possibilities low, at 32 instead of 64. For each gene library containing a randomized codon, four PCRs were performed. First, two separate PCRs were performed by using the vector harboring the appropriate region to be amplified as template and a primer pair consisting of a flanking and a degenerate primer to amplify a 5' portion and 3' portion of the hER α LBD gene containing the NNS substitution at the codon of interest. A high fidelity DNA polymerase (Phusion High Fidelity DNA Polymerase, New England Biolabs) was used in saturation mutagenesis purpose. The PCR components were: 30ng template DNA, 25pmol each of flanking and residue primer, 1X cloned *Pfu* reaction buffer and 2 U of Phusion polymerase in a total of 50 μ l reaction volume. The PCR program was similar to standard amplification PCR described above. Both PCR products from these reactions were gel purified and treated with the restriction enzyme *DpnI* to remove any residual methylated template from the products. Each PCR product was then combined in an overlap extension reaction without primers in 20 μ l volume. The reaction conditions of this overlap extension were identical to those used for the standard PCR described above except for the absence of primers with 15 cycles. Finally, this overlap extension reaction was used as the template for a standard amplification PCR for the amplification of the gene library incorporating a randomized codon by using the flanking primers.

2.2.2.7 DNA sequencing

The DNA sequencing was done using the BigDye Terminator RR Mix Cycle Sequencing kit. The principle of DNA sequencing is based on the chain-termination method described by Sanger *et al.* [113]. In the chain-termination method, dideoxynucleotides (terminators) are incorporated into a newly synthesized complementary chain that will lead to stop its elongation in a PCR reaction. Each of dideoxynucleotides is labeled with a specific fluorescent dye and the terminated chains can be specifically detected using an ABI Prism 310 Capillary Sequencer (Applied Biosystems). The PCR sequencing reaction was performed using 500-1000ng plasmid DNA, 5pmol primer, 2 μ l RR mix (ready reaction)

and H₂O up to 10µl. The samples were subjected to 25 cycles of: 10 seconds at 95°C, 5 seconds at 50°C, 4 minutes at 60°C in a thermocycler. The DNA product was precipitated using 9.5µl water and 30.5µl of absolute ethanol and left for 1 hour. The DNA was collected by centrifugation for 20 minutes at 13000 rpm. The pellet was washed using 125µl 70% ethanol and then centrifuged for 10 minutes at 13000 rpm. The pellet was dried at 95°C for one minute and resuspended in 15µl of template-suppression reagent (TSR, Perkin-Elmer). A denaturation step at 95°C for 2 minutes took place and the tube was immediately placed on ice. The reaction was loaded on an ABI-Prism 310 capillary electrophoresis sequencing station for analysis.

2.2.2.8 Gene transfer in bacteria

E. coli have no competence nature, *i.e.*, they are not able to accept naked DNA molecules from the environment. To enable the bacterial cells to take up circular vector DNA they have to be made competent using special treatments. Two transformation methods were used to transform bacteria cells: the heat shock and the electroporation. The heat shock method was used only to transform *E. coli* chemical competent cells. The transformation procedure was done after Hanahan [105]. In brief, 200µl competent *E. coli* cells were thawed on ice for 20 min, 50ng of plasmid DNA was added to the cells and mixed gently. The mixture was incubated on ice for 30 minutes. The cells were heat shocked for 90 seconds at 42°C and were placed immediately on ice for at least 3 minutes. Eight hundred- µl of SOC medium was added to the tube and the suspension was mixed on a roller for 45-60 min at 37°C depending on selectable antibiotic resistance marker. Different volumes of the culture were plated on plates containing LB medium supplemented with antibiotics. The plates were incubated overnight at 37°C.

The transformation using electroporation was done for *E. coli* cells according to a slight modification of Dower *et al.* [97]. The electroporation was done using a Gene Pulser II. Bacterial competent cells were thawed on ice slowly before adding 2µl of plasmid DNA. The mixture was transferred into an ice-cooled electroporation cuvette (2mm electrode distance). The cuvette was subjected to electroporation at 25µF, 2.5kV, 200Ω. The cells were suspended immediately with 1ml SOC medium and incubated for 45-60 min at

37°C. Different volumes of the culture were plated on LB media supplemented with antibiotic and incubated overnight (12-16 hours) at 37°C.

2.2.2.9 Gene transfer in yeast

Yeast cells were transformed with plasmid DNA using two different methods. The first method was using frozen competent yeast cells for screening a small variable population whereas the second method was using freshly prepared competent cells for large scale selection.

2.2.2.9.1 Frozen yeast transformation for screening

A slightly modified transformation method described by Dohmen *et al.* [114] was used to transform previously prepared competent yeast cells for screening a limited variable for reporter gene assays like β -galactosidase or ligand titration curve assay. Shortly, frozen competent yeast cells were thawed on ice for 15 min. A cocktail was prepared using 1 μ l plasmid DNA (0.1-1 μ g), 10 μ l single-stranded carrier DNA (50 μ g) and 40 μ l lithium-sorbitol solution [0.1M Li-acetate; 1mM EDTA; 10mM Tris-HCl, pH 8.0 and 1.82% (w/v) sorbitol]. This cocktail was added to the yeast cells and mixed well. Five hundred eighty- μ l of 40% poly ethylene glycol (PEG) 3350 (in Li-Acetate-TE) was added to the mix and vortexed until a homogenous suspension appeared and incubated for 30 min at 30°C. A heat shock was done at 42°C for 15 min followed by brief (30 sec) centrifugation at 6000 rpm in a table top centrifuge. The supernatant was discarded, the pellet was resuspended in 220 μ l SOS media and incubated again at 30°C for 30 min in a rotor. Finally, the cells were plated on a synthetic dropout (SD) plate lacking appropriate amino acid mediating selection for the plasmid. The transformation plate was incubated at 30°C for 2-3 days so that the yeast clones could appear.

2.2.2.9.2 High efficiency transformation for selection

For a large scale selection of libraries where high efficiency is desired, we used the following method described by Gietz and Woods [98]. Frozen yeast cells were

streaked on SD agar plate with appropriate selection and grown for 2 days at 30°C. A single colony was picked to inoculate 3ml of the similar liquid medium and incubated overnight (12-14 hours) in a rotary shaker. This overnight culture was used to inoculate 50ml yPAD rich medium (5×10^6 cells/ml) and incubated for 5 hours at 30°C in a rotary shaker (250 rpm). Cells were harvested by centrifuging at 3000g for 6 min. The pellet was washed twice with 25ml sterile ddH₂O and dissolved in 1ml sterile ddH₂O. A transformation cocktail was prepared from 240µl PEG 3500 50% (w/v), 36µl 1M Li-acetate and 50µl single-stranded carrier DNA. Fifty-µl aliquot of yeast cells (supernatant removed by centrifugation), 326µl of transformation cocktail and 34µl plasmid DNA were mixed together gently and incubated at room temperature for 10 minutes. Cells were heat shocked at 42°C for 10 min with a mixing in between. Ethanol was added to final concentration of 2.5% to the transformation mix and incubated for another 10 min with intermittent mixing. Cells were harvested by centrifugation (3000g, 30 sec), resuspended in sterile ddH₂O and plated in pre-warmed SD plates with appropriate selection. The transformation plate was incubated at 30°C for 2-3 days for appearance of colony.

2.2.2.10 Construction of plasmids

pGBT-LBD

Plasmid pHCA/Gal4ERVP16 coding for Gal DBD (1-93), hER α LBD (282-595) and VP16 AD (424-490) [115] was double-digested with two different sets of restriction enzymes as (a) XhoI-HindIII and (b) BamHI-HindIII. Two fragments of 272bp and a 1046bp length were cut out from the gel. A second plasmid pGBT-TGA2.2 [116] was digested with XhoI and BamHI to remove the TGA2.2 insert and a 5276bp vector backbone pGBT was gel purified. A 3-fragment ligation was done including the 272- and 1046bp fragments from pHCA plasmid containing the GalDBD-hER α LBD-Vp16AD cassette with the pGBT vector backbone to get pGBT-LBD.

pET28a-LBD

A PCR was made on pHCA/Gal4ERVP16 using primer pair LBD sense and LBD antisense to amplify the 756bp LBD sequence (amino acid residue 302-553). This LBD sequence was cloned into the pGEM-T vector (Promega) and named pGEM-T LBD. The newly created pGEM-T LBD and pET28a (Novagen) were digested with EcoRI and XhoI to get the LBD and vector backbone, respectively. The LBD fragment and linearized pET28a vector backbone was ligated to yield pET28a-LBD.

pBD-Gal4- LBD-Cam

pCMV5-ER α plasmid containing the wild type full-length hER α cDNA [117] was first digested with EagI and BamHI and then treated with T4 DNA polymerase to form blunt ends at both termini followed by ligation into the EcoRI site (also treated with T4 DNA polymerase) of pBD-GAL4-Cam vector (Stratagene) to form pBD-GAL4 hER α containing hER α amino acids 312–595.

pAD-Gal4-2.1-SRC-1:

pGAD424 SRC-1 containing the full-length SRC-1 was constructed as described elsewhere [108].

2.2.3 Standard protein biochemical methods

2.2.3.1 Expression of 6x His-fusion protein expression in *E. coli*

Expression and purification of WT and mutant hER α LBD was done according to Eiler *et al.* [118] and Chen *et al.* [92] with minor modification. *E. coli* BL21(DE3) cells transformed with the pET28a plasmid containing the wild type or V400A mutant hER α LBD fragments encoding residues 302–553 fused with 6X His fusion were grown at 37°C until A600 = 0.6. No ligand was added to the culture media. Protein expression was then induced with 0.5mM isopropyl-dithiogalactopyranoside (IPTG) at 25°C for 5–7 h. Cells were harvested by centrifugation, resuspended in lysis buffer (50mM Tris buffer, pH 8.0, 10% glycerol, and 10mM β -mercaptoethanol) at 10ml/g of wet cells, and lysed by

sonication. The sonication was carried out for 30 sec with 3 intermittent cooling steps of 10 sec each. The cell debris was separated from the supernatants by centrifugation for 30 min at 10,000 X g. The resulting supernatants were used for purification in an affinity chromatography.

2.2.3.2 Purification of recombinant 6x His-fusion proteins

For the purification of the 6x His fusion proteins, an affinity chromatography on NiNTA-matrix resin (Qiagen) was performed according to the manufacturer's instructions. Histidine has a high affinity for bivalent transition metals e.g. Ni²⁺. Protein fusions like oligo-histidine-peptide can be purified by affinity chromatography with immobilized metal ions. A soluble fraction of proteins produced from the above-mentioned expression and purification procedure was transferred to a fresh tube and incubated with 4 ml of the 50% Ni-NTA-matrix resin and the suspension was mixed gently for 1 hour at 4°C. This mix was loaded into a column and left for 30 min to settle down by gravity. The unbound supernatant was drained away and the column was washed with low salt concentration buffer (20mM imidazole). Similar amount of gel volume (1.5-2ml) of elution buffer containing gradually increasing imidazole concentration was used to elute the bound protein in different fractions. The eluates were stored at -20 °C till further use.

2.2.3.3 Protein concentrations determination

Protein concentration was estimated by two different methods. A colorimetric assay was used to determine the concentration from proteins extracted without detergent usage according to Bradford [119]. The assay was conducted by pipetting equal amounts of protein extract into a microtiter plate containing 200µl of 5-fold diluted Bradford reagent. The OD₅₉₅ was measured with a MRX plate reader (Dynex). Protein concentrations were calculated with the help of a standard curve derived from different BSA protein amounts (1, 4, 8 µg etc) on the same plate. Protein isolated using buffers containing detergent was loaded in a SDS-PAGE with variable amounts (1, 2, 4µl etc) together with known concentration of BSA standards. After staining and scanning, band intensity was used to estimate the concentration of protein in the sample.

2.2.3.4 Denaturing sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE)

In sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE), proteins are separated largely on the basis of polypeptide length. The electrophoresis of the protein was done using a discontinuous buffer system, in which a non-restrictive large pore gel, called a stacking gel, is layered on top of a separating gel called a resolving gel [120]. The recipe for the resolving gel was consisting of: 10-12% (w/v) acrylamide/bisacrylamide (19:1), 400 mM Tris-HCl pH 8.8, 0.1% (w/v) SDS, 0.1% (w/v) TEMED and 0.1% (w/v) ammonium persulfate. The stacking gel was consisting of: 5% (w/v) acrylamide/bisacrylamide (37.5:1), 125 mM Tris-HCl pH 6.8, 0.1% (w/v) SDS, 0.2% (w/v) TEMED and 0.1% (w/v) ammonium persulfate. The denatured protein extract samples were boiled at 95°C for 5 minutes, cooled on ice and loaded into the gel. The native extracted protein samples were mixed with 10µl of protein loading buffer and denatured at 95°C for 5 minutes, cooled on ice and then loaded on the gel. The electrophoresis was performed at 120V in 1x SDS-PAGE running buffer until the bromphenol blue band run out of the gel. Six-µl prestained protein ladder was loaded on each gel for the estimation of the size of the separated proteins.

2.2.3.5 Coomassie staining of proteins separated on SDS-PAGE

The Coomassie Brilliant Blue G-250 dye was used to detect proteins separated on SDS-PAGE. The gels were fixed in a Coomassie fixing solution for 30 minutes under continuous agitation then the fixing solution was removed. The fixed gel was incubated with Coomassie staining solution for at least 2 hours at room temperature. The gel was destained using Coomassie destain solution for 2 hours.

2.2.3.6 Western blot analysis

The protein ran in the SDS-PAGE were blotted into a PVDF membrane using semi-dry blotting method [121] in an electric field between two graphite plates. The PVDF membrane was activated before blotting using MeOH. For the transfer of proteins from the gel to the membrane, the gel and the membrane were sandwiched between two 3-layers of Whatman papers (pre-soaked with transfer buffer). The whole arrangement was

placed within a blot apparatus and transfer was performed using an electric flow of $2\text{mA}/\text{cm}^2$ for 1 hour. Ponceau S staining was done to observe whether the transfer was successful or not. Destaining was done using 1XPBS. Non-specific binding was avoided by blocking with non-fat dried milk powder. Since our protein constructs had 6XHis tag, we used a Universal His Western Blot kit (BD Bioscience) for the detection of specific protein in the membrane. The basis of the detection method was a unique detection reagent derived from BD TALON resin technology that combines specificity for His tags with high affinity. This reagent utilized metal ion affinity to specifically recognize and bound His tags. After washing, the detection reagent was bound with high affinity by streptavidin conjugated to horseradish peroxidase (HRP) and unbound reagent was again washed away. Addition of the enhanced chemiluminescent substrate, which reacted with the bound HRP conjugate, allowed visualization of the His-tagged protein. The chemiluminescent signal was detected using autoradiography where the exposure time was 30 sec to 3 min depending upon the strength of chemiluminiscent signal generated by the bound protein.

2.2.3.7 β -galactosidase activity (ONPG assay)

Determination of transcriptional activation of reporter gene *lacZ* was done from its coded product, β -galactosidase catalyzed breakdown of substrate *o*-nitrophenyl- β -D-Galactopyranoside (ONPG). The β -galactosidase catalyzed the colorless substrate ONPG into the fluorescent *o*-nitrophenyle substance, which could be quantified at 420nm. A modified protocol from Thurow [122] was followed where fresh clones were used to inoculate 2ml SD media with appropriate selection for the plasmid and supplemented with ligand and grown overnight at 30°C with 200-220rpm shaking. Cells were harvested from 200 μl of overnight culture and resuspended in 865 μl H buffer [100mM HEPES/KOH, pH 7.0; 150mM NaCl; 2mM MgCl_2 ; 1% (w/v) BSA]. Two hundred μl aliquot of this cell suspension was taken in a round-bottom 96-well plate for measuring the OD_{595} . The rest 665 μl cells were broken down by addition of 55 μl chloroform and 55 μl 0.1% (w/v) SDS along with vigorous vortexing. One hundred and twenty five μl of ONPG substrate (4mg/ml in H buffer) was added into the broken cells and incubated at room temperature. Necessary time required for the solution to become yellow was recorded. The reaction

was stopped upon appearance of yellow color with 400 μ l 1M Na₂CO₃, centrifuged at 13000 rpm (Biofuge Pico) for 5 min and the supernatant was taken into flat-bottom 96-well plate for quantifying at 420nm in a spectrophotometer. The calculation of β -galactosidase activity was done according to the following formula-

$$\beta\text{-galactosidase activity} = \frac{1000 \times \text{OD}_{420}}{V \times T \times \text{OD}_{600}}$$

Where, V was the volume (200 μ l) and T was reaction time. β -galactosidase activity was expressed in Miller Unit (MU).

2.2.3.8 *In vitro* ligand binding assay

A competitive binding assay was done to determine the affinity of WT and mutant hER α LBD towards E2 and synthetic analogs based on enzyme fragmentation complementation using Hithunter Estrogen Assay kit (DiscoverX, Birmingham, UK). Estrogen and estrogen analogs compete for binding to the estrogen receptor. Estrogen possesses the same property when conjugated with other chemicals or proteins. In the assay kit, the manufacturer has fragmented β -galactosidase protein into two fragments, namely enzyme donor coupled estrogen (ED-E2) and enzyme acceptor (EA). A functional β -galactosidase enzyme would be reconstituted when ED and EA interact together and catalyze substrate turn over where product formation could be monitored in a fluorescent intensity reader with excitation and emission at 530 \pm 25 and 620 \pm 20 nm, respectively. Addition of estrogen receptor would quench the ED-E2 conjugate and would reduce signal intensity whereas addition of synthetic ligand would bind to the receptor depending on their affinity and ligands with lower and higher affinity would generate decreased and increased signal, respectively. The assay was done with commercially available full length hER α LBD (Sigma) or purified WT and V400A mutant LBD (residue 302-553) expressed in heterologous *E. coli* BL21 (DE3) and purified through affinity chromatography. The concentrations of commercially available or purified receptor were adjusted to 5nM. The assay was performed according to the manufacturer's manual. Briefly, 1 μ l standard (known E2 concentration) or ligand was added to the 25 μ l ER/ED-

E2 conjugate (1.5 parts of receptor: 1 part of ED-E2 reagent) in a 384-well black microtitre plate and incubated for 90 min in dark. Ten μ l EA and 10 μ l substrate were added to the each wells and again incubated for 60 min at room temperature and in dark. Finally, the fluorescence was read and EC₅₀ values were calculated from a standard curve fitting the fluorescence value and ligand concentration in Y and X-axis, respectively. Each sample was included in triplicate in the assay.

2.2.3.9 Chemical complementation based library selection

2.2.3.9.1 One hybrid system

A chemical complementation based nutritional screening was done to select mutant from a random mutagenesis library with higher affinity towards synthetic ligand. Yeast strain PJ69 4A was transformed with randomized library in the pGBT-LBD vector backbone and the transformation mix was plated on transactivation plate lacking adenine (Ade), histidine (His) and tryptophan (Trp). The transactivation plate was supplemented with appropriate concentration (10 μ M) of selected ligand (for chemical complementation) and when necessary, 10-50mM 3-AT (to remove background growth due to the leakiness of *HIS3*). Colonies appeared on the transactivation plate after 3 days were then screened in β -galactosidase assay.

2.2.3.9.2 Two hybrid system

For library selection, a two-tiered strategy consisting of an agar plate-based selection (on-plate growth assay) followed by a 96-well plate-based screening was used. In the selection method, *S. cerevisiae* YRG2 cells harboring pGAD424 SRC-1 were co-transformed with mutagenized LBD fragments and BglIII-BsaI-digested pBD-GAL4 hER α vector. The transformed cells were plated on an agar plate containing minimum medium lacking tryptophan, leucine, and histidine and supplemented with an appropriate synthetic ligand concentration. The ligand concentration was chosen such that yeast cells bearing the parental hER α LBD in each round of directed evolution cannot form colonies, whereas yeast cells bearing a variant with 5–10-fold improvement in ligand binding affinity may form colonies. Colonies that formed on the selection plates after incubation

at 30°C for 3 days were picked and streaked onto two agar plates containing the same minimum medium as mentioned above, one with and one without the ligand. The colonies appearing on the agar plate with ligand but not on the agar plate without ligand were picked and assayed in the 96-well plate to determine their EC₅₀ values.

2.2.3.10 Chemical complementation based screening

2.2.3.10.1 One hybrid system

In the one hybrid system, clones selected for screening were included in the β -galactosidase assay (See 2.2.3.7) for determining the activity of ligand in the mutant.

2.2.3.10.2 Ligand dose-response/titration assay

The ligand dose-response/titration assay was done according to Chokalingam *et al.* (2005) in a 96-well plate format. Shortly, candidate mutants from the selection plates were grown to log phase (A_{600} : 2–4) in synthetic complete minimal medium lacking leucine and tryptophan (selection for plasmid only) at 30°C overnight (12–16 h) with shaking. The resulting cell culture was diluted to A_{600} : 0.002 in liquid minimal medium lacking leucine, tryptophan, and histidine (transactivation media). Each well in the 96-well plate contained 200 μ l of diluted yeast cells and the appropriate ligand or E2 at specific concentrations dissolved in EtOH. These ligand containing microtiter plates were incubated at 30°C for 24 h, after which they were visually inspected for identification of mutants with strengthened response toward the target ligand (higher cell density than parental mutant control) and weakened response toward E2 (lower cell density than parent). The cell density at 600nm was measured using a SpectraMax 340 PC plate reader (Amersham Biosciences).

2.2.4 Toxicity test

2.2.4.1 Germination of seed

A germination test was carried out according to the OECD in presence of synthetic ligand to find out general toxicity properties of the ligands. Tobacco seeds (SNN) were germinated in sterile condition and seeds were sterilized before in 6% (v/v) sodium

hypochlorite solution for 25 min and then washed with sterile ddH₂O for 5 times. Sterile tobacco seeds were used to inoculate 2XMS [123] solid media supplemented with 10 μ M concentration of E2, CV5407, CV6019 or polysorbate-ethanol (10:1) as solvent control in 5cm diameter Petri dish. Wheat seeds were soaked in ddH₂O and plated on 5cm diameter Petri dish containing 2-layers of filter paper. These filter papers were previously soaked with ddH₂O containing 10 μ M E2, CV5407, CV6019 and the solvent. The Petri dishes were sealed with Leucopore and kept in dark for 5 days at 25°C. When appearance of seedling observed in the plate, they were transferred to a growth chamber with 16 hours light and 8 hours dark rhythm where the temperature was adjusted to 24 and 22°C in day and night respectively. Growth of seedlings was observed for further 2 weeks.

2.2.4.2 Adventitious shoot induction test

In vitro grown sterile tobacco (SNN) leaf was cut into approximately 1 square cm and placed on Petri plates containing shoot induction media supplemented with 10 μ M concentration of E2, CV5407, CV6019 or polysorbate: ethanol (10:1). The plates were sealed with Leucopore and placed in a growth chamber where the condition was similar to one described above. Adventitious shoot formation and development were observed up to 4 weeks.

2.2.5 Molecular modeling

The hER α LBD crystal structure (Protein Data Bank code 3ERD) was imported into MOE (Molecular Operating Environment, Chemical Computing Group Inc., Montreal, Quebec, Canada). A model of hER α LBD complexed with the synthetic ligand CV3320 was built from the hER α -diethylstilbestrol (DES) structure using the following attributes: (i) the forcefield MMFF94s [124] was applied, (ii) hydrogen atoms were added, (iii) partial charges were assigned to all atoms, and (iv) the structure was subsequently energy-minimized by using a sequential combination of steepest descent, conjugate gradient, and truncated Newton algorithms. Subsequently, a docking box with a grid consisting of 52X38X30 points was drawn around the DES ligand to specify the boundaries for the movement of the ligand to be docked. In this orientation, the box includes the entire DES ligand and a few atoms of the interacting residues. The DES ligand was subsequently

deleted from the structure, and the CV3320 ligand (which had previously been assigned partial charges and minimized by using the MMFF94s force field) was docked into the docking box by using a simulated annealing algorithm [125] with the following parameters: initial temperature 12,000 K, 25 runs involving six cycles per run, and 20,000 iterations per cycle. The five structures with the best docking score (lowest overall energy) from these docking runs were compared and found to be within a root mean square deviation (RMSD) of 0.5 Å from each other. The lowest energy of these five was then subjected to energy minimization as described earlier, to determine the most favorable conformation and orientation of DHB in the ligand binding pocket. Residues within 4.6 Å of the docked DHB were considered to be in contact with the ligand for purposes of receptor engineering. For gauging the individual role played by the L384F, L387M and Y537S mutations, the appropriate amino acid substitutions were made to the docked DHB-hER α structure, and the resulting structure was energy-minimized.

3 Results

3.1 Validation of test system

To isolate human hER α variants with altered transcriptional activation activity (towards synthetic non-steroidal estrogen analogs), we utilized two efficient and sensitive high throughput screening methods based on macromolecular interaction in yeast. The first system was based on a modified yeast one hybrid screening which could be used for *lacZ* reporter gene assay and complemented phenotypic selection of library variants *in vivo*. The second system was a very sensitive two hybrid assay for chemical complementation in yeast and thus could be used for high throughput ligand titration assay as well as for screening of large libraries.

3.1.1 Yeast one hybrid system:

To allow efficient identification of mutants of hER α LBD having an altered response to E2 as well as synthetic ligands, we used a modified yeast one hybrid system where a chimeric protein was generated consisting of Gal4 DNA binding domain and VP16 activation domain placed in the amino- and carboxylic- terminal end of ER LBD (GLV). This chimeric construct was introduced into yeast strain PJ69 4A, where *HIS3* and *ADE2* genes are driven by Gal1 and Gal2 promoter, respectively. Upon binding of ligand, the chimeric protein would be released from hsp chaperon, dimerize, transport inside the nucleus and bind to the different Gal promoters, thus allow the yeast cell to biosynthesize histidine (His) and adenine (Ade) in a media lacking His and Ade. This chemical complementation allowed us first to validate E2-LBD binding and later to look for orthogonal combination of ligand-mutant receptors *in vivo* on transformation plates. The yeast strain PJ69 4A has also β -galactosidase gene (*lacZ*) integrated under the control of another Gal promoter, Gal7. In presence of a functional ligand-receptor interaction, expression of *lacZ* would take place and a quantitative β -galactosidase assay could be done to reveal the quantity of induction, which was correlated to the binding of ligand to

the receptor. The validation results from on plate chemically complemented nutritional selection and β -galactosidase assays are mentioned separately below:

3.1.1.1 On plate assay

The *in vivo* validation was done in yeast strain PJ69 4A by transforming with pGBT-LBD expressing the chimeric GLV or with the empty vector pGBT and plating the transformation mixture on minimal plates lacking adenine, histidine and tryptophan (SD-Ade-His-Trp). These plates were termed transactivation plate because of nutritional selection via inclusion of *HIS3* and *ADE2* genes as transcription reporter. These plates were supplemented with two different concentrations (0.1 and 1 μ M) of E2 or with EtOH as solvent control. Transformation mix was also plated on SD-Trp plates for the selection of plasmid, elucidation of transformation efficiency and production of transformants for β -galactosidase assay. As shown in the figure 3.1, in the transactivation plates containing 0.1 or 1 μ M E2, cells harbouring LBD could grow whereas there was no growth on plates where cells were only harbouring the empty plasmid. There was no growth in the absence of ligand on transactivation plates, showing the stringency of the plate assay.

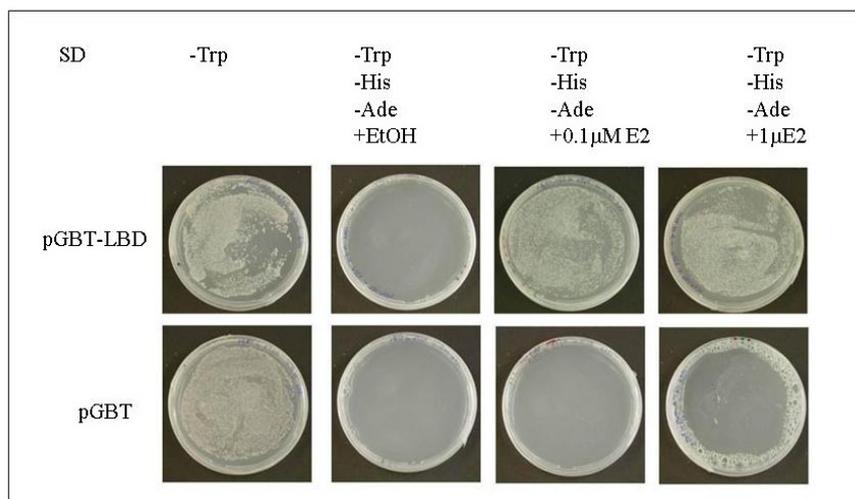


Figure 3.1: Ligand-dependent chemical complementation based on plate growth assay in yeast one hybrid system. Yeast strain PJ69 4A was transformed with plasmid pGBT (empty plasmid) or pGBT-LBD (chimeric construct Gal4 DBD- hER α LBD- VP16 AD) and the transformation mixture was plated on transactivation media (SD-Ade-His-Trp) with E2 or with solvent only. Growth was only observed with plates containing E2 where cells harbouring pGBT-LBD were plated.

3.1.1.2 β -galactosidase assay

As *Gal7-lacZ* was introduced in the strain PJ69 4A, the *lacZ* gene product β -galactosidase could be used also as quantitative transcription reporter. In presence of β -galactosidase, the substrate *o*-nitrophenyl- β -D-galactopyranoside (ONPG) was hydrolyzed into galactose and *o*-nitrophenol, which is yellow in color and measured in a spectrophotometer at 420nm. We tested E2 and ligands CV2807, CV5407 and CV6019 at 10 μ M concentration. β galactosidase activity (Miller Unit) for E2 was considered to be 100% and relative activities were calculated for the above-mentioned ligands from the β -galactosidase activity (Figure 3.2). When only the solvent was used instead of any ligand, the activity was negligible.

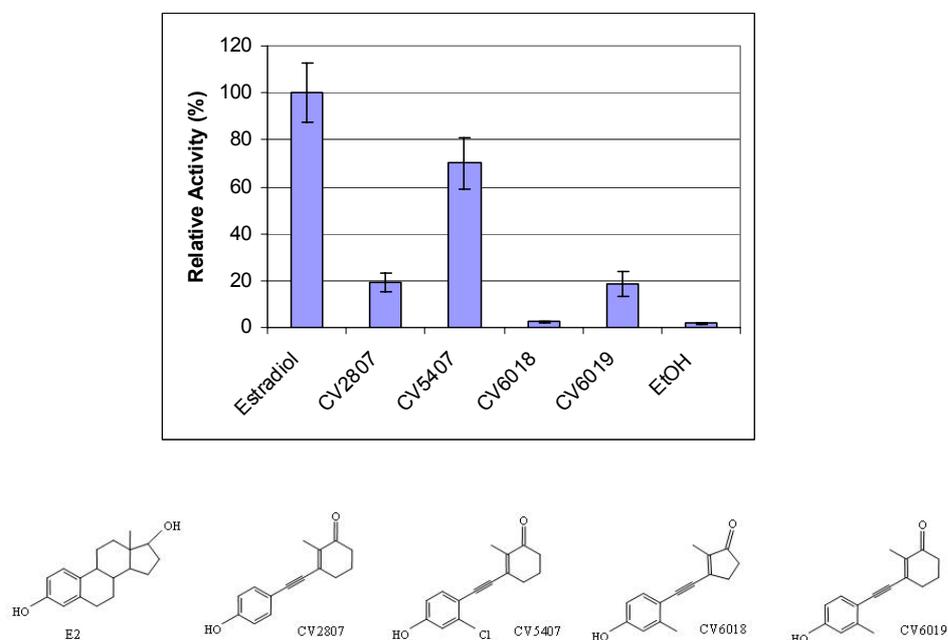
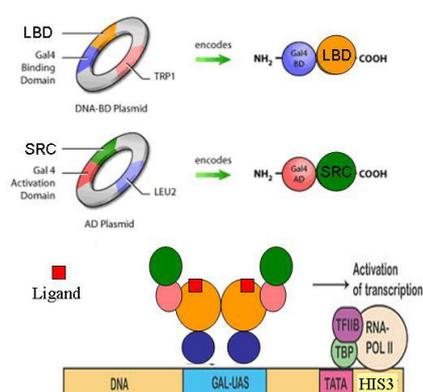


Figure 3.2: β -galactosidase reporter gene assay in yeast one hybrid system, where PJ69 4A containing pGBT-LBD was inoculated with different ligands of 10 μ M concentration. Ligand-induced transcriptional activation of *lacZ* was proportional to the β -galactosidase mediated breakdown of substrate ONPG into colored product *o*-nitrophenol, which is measured at OD420 nm and plotted against corresponding ligands in the graph to get the activity in Miller Unit. Taking E2 as 100%, the comparative activity was calculated for other ligands and plotted in the graph.

3.1.2 Yeast two hybrid system

In the two hybrid system, the cDNA encoding hER α amino acids 312-595 containing most of the LBD domain (hER α amino acids 302-553) and the F domain (hER α amino acids 554-595) was fused to the gene encoding the Gal4 DBD in plasmid pBD-Gal4-Cam to form the “bait plasmid” pBD-Gal-LBD-Cam. The gene encoding human steroid receptor coactivator-1 (SRC-1) was fused to the Gal4 AD in plasmid pGAD424 to form “prey plasmid” pGAD424SRC-1. Both plasmids were transformed and expressed in *S. cerevisiae* YRG-2, which contains a Gal4-regulated *HIS3* reporter construct integrated on its chromosome. In the presence of agonistic ligands, the LBD undergoes a conformational change and binds to SRC-1, which brings the Gal4 DBD and the Gal4 AD in proximity, thus activating the transcription of the reporter gene. A simple diagram is presented in Figure 3.3 (A) showing the two hybrid system in yeast. The cell growth rate in medium lacking histidine is proportional to the strength of ligand-receptor interaction. We have validated this system using the yeast cells with WT LBD in 96-well plates where cell density (A_{600}) was measured. We could show (Figure 3.3 B) that the cells bearing pBD-Gal4-LBD-Cam and pGAD424-SRC-1 responded to E2 at sub-nanomolar concentration and the EC_{50} value was 5×10^{-10} M. No growth was observed in 96-wells where only solvent (EtOH) was added instead of E2.

A.



B.

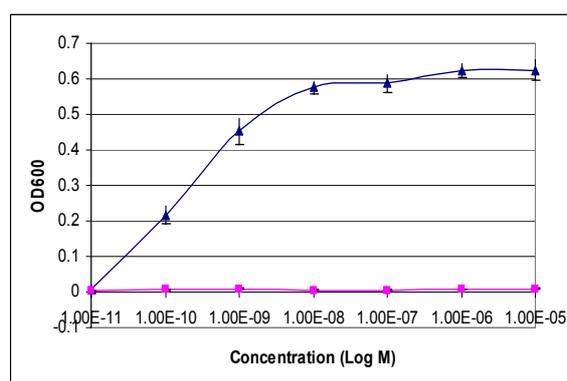


Figure 3.3: A. Diagram showing the classical two hybrid system that was used in YRG2 yeast strain. The LBD and SRC-1 were encoded by two different plasmids where they have in their amino terminal end Gal4 DBD and Gal4 AD, respectively. Ligand induced conformational change, dimerization and nuclear transport enable the chimera to bind to the Gal UAS of reporter gene *HIS3*, enabling nutritional complementation in selection plate. B. Transactivation profile in the yeast two hybrid cells for the WT LBD construct in presence of E2 (blue triangle) or solvent (pink

square). The transactivation potential of ligand correlates to the growth of cells in this assay in a concentration dependent manner.

3.2 Toxicity test

Because the long-term goal of our project is to transfer the established regulatory system in plant that could be used for commercial agriculture, we were interested in the general toxicity of the synthetic ligands we used. We used the following two short-term bioassays in plant system for toxicity test.

3.2.1 Germination Test

The germination test assessed the effects on seedling emergence and early growth of two higher plants i.e. tobacco and wheat, following exposure to the ligand. E2 and solvent (Polysorbate/EtOH- 1:10) were also included in the test as reference and negative control. All treatments were carried out at 10 μ M concentrations. Sterilized seeds were placed on solid media or paper towel containing appropriate concentrations of treatment in Petri dishes and evaluated for effects following 7-10 days after 50% or greater emergence of the seedling in solvent control plate. Endpoint measured were percentage of germination, visual assessment of seedling emergence and detrimental effects of plant development. As shown in Figure 3.4 and 3.5, there were no significant differences between germination of tobacco and wheat seeds in presence of E2, CV5407, CV6019 or the solvent alone. Since the ligands used had no negative effect on germination at 10 μ M concentration, we did not proceed further to find out range-finding tests to generate a dose-response curve. The growth of seedlings was homogenous and there were no visual evidence of altered biomass production like shoot size and growth stunting as well as visual detrimental effects like chlorosis and seedling mortality.

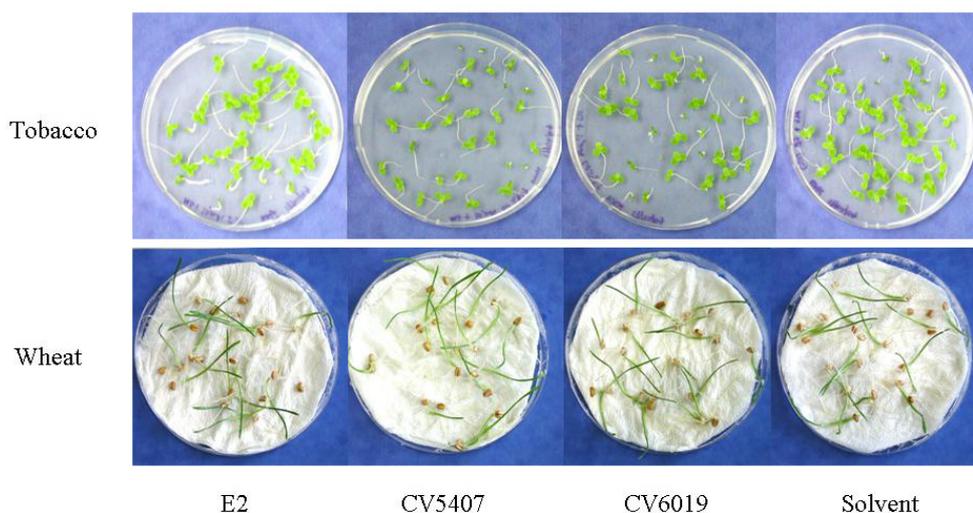


Figure 3.4: Germination of tobacco and wheat seeds in presence of ligands. Endpoints measured in this germination test were percentage of seedling appearance and visual assessment of adverse effect on seedling development.

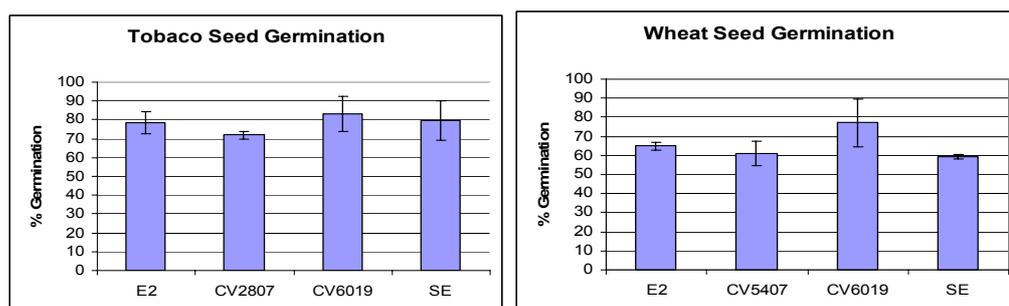


Figure 3.5: Percentage of tobacco and wheat seed germination in the presence of E2, CV5407, CV6019 or the solvent only. Three individual plates were used as replication for each of the treatment mentioned.

3.2.2 Adventitious shoot induction

Leaf sections of tobacco plants were placed on solid media containing 10 μ M concentration of E2, CV5407, CV6019 or the solvent alone and observed for adventitious shoot formation. Comparable shoot inductions were observed in all treatments (Figure 3.6). No growth abnormalities like bleaching and stunting were observed with any of the ligands in use. These observations demonstrated that the ligands in use did not have any adverse effect on shoot induction and development.



Figure 3.6: Adventitious shoot induction from tobacco leaf disc in presence of E2, CV5407, CV6019 or the solvent alone. Visual assessment of shoot induction and development was made to find out any adverse role of ligands used in this assay.

3.3 Rational design: Site-directed mutagenesis at M388 and L428

Brozowski *et al.* demonstrated 26 residues involved in estradiol contact and binding in agonist conformation whereas a slightly different set of 26 residues involved in raloxifene contact and binding in antagonist conformation [73]. Interestingly, in agonist and antagonistic conformation, residues vary significantly to come into contact and bind the ligand. Since compounds with halogen group at position 5' showed significant activity in our yeast one hybrid system (Table 3.1), we were interested in corresponding residues within a reasonable distance (6\AA) which may make any contact with that particular position of the ligand. Through molecular modeling, we have identified two residues (shown in Figure 3.7)- M388 and L428 which are situated in close proximity (3.95 and 5.98\AA , respectively) and both these residues were mentioned in previous literature to have interaction with E2 as well as different ligands [73, 107]. Once we identified these two residues, we exchanged them with smaller hydrophobic residue alanine (A) and added bulky aliphatic or aromatic hydrocarbon groups in the 5' position of ligand so that we can rematch the ligand-receptor pair. We have decided to exchange these two residues into alanine because of the general hydrophobic nature of the ligand binding pocket as well as the aromatic or aliphatic group substitution in the ligand. We tested M388A, L428A and in combination (ML-AA) in β -galactosidase assay in presence of natural ligand E2, compounds showing significant activity in WT LBD like CV5407 and compounds having bulky substitution in position 5'- IK308, IK377 and IK519 (Figure 3.8). These compounds displayed very poor activity in WT LBD construct. Unfortunately

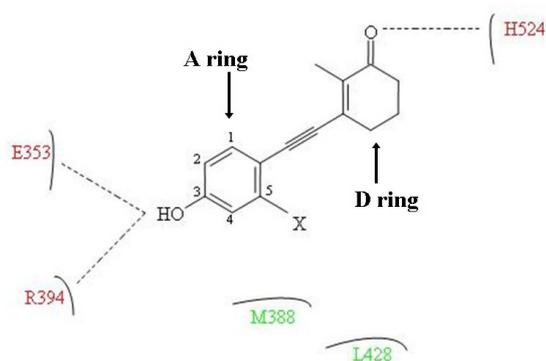
none of these compounds was active in any of the mutants created. Interestingly, when L428 is exchanged to A, the protein cannot even bind to the natural ligand E2, where M388A mutant retains the capability to bind E2. When combined together (M388A L428A), the double mutant can bind E2, although to reduced affinity and the relative activity went down to 60% compared to the WT.

Ligand	Substitution at C 5' position	Relative Activity (%) *
E2	-	100
CV5407	-Cl	68
CV6104	-Br	61
CV6407	-F	56

* Relative activity (%) was calculated from the β -galactosidase assay using PJ69 4A strain containing WT-LBD. The Miller Unit of E2 was fixed to 100% activity and the relative values for other ligands were calculated comparing their corresponding Miller Unit values with E2.

Table 3.1: Relative activity of E2 and ligands containing halogen group in 5' position. Relative activity was calculated from the β -galactosidase assay in yeast one hybrid system where E2 activity was fixed to 100% and the activity of other ligands was calculated from their corresponding Miller Unit values.

A.



B.

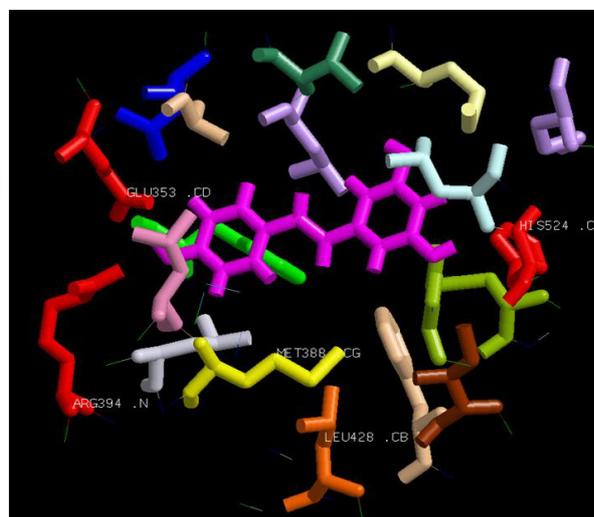
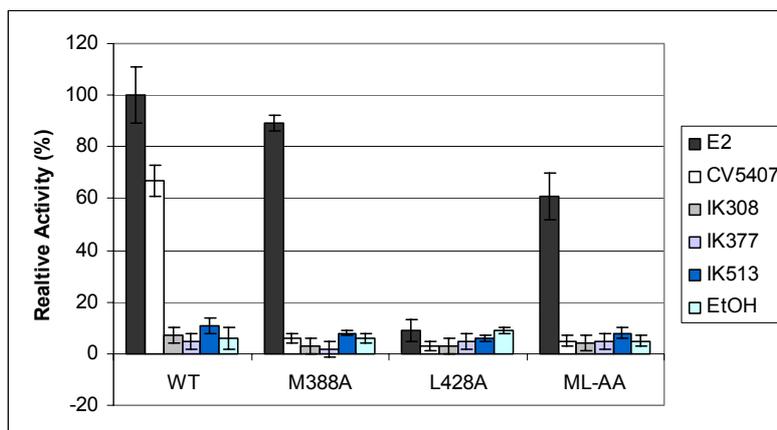


Figure 3.7: A. Selection of residues for manipulation. Ligand with halogen substitution at C5 is shown in the ligand binding pocket in a simple drawing. The hydroxyl and keto groups in the A and D ring interact with the E353, R394 and H524. These three residues are very important for ligand interaction and thus shown in red. Residues within 6Å distance from the C5 halogen group were shown in green. B. Residues forming the ligand binding pocket for docked DES in hER α LBD (PDB code 3ERD). M388 and L428 are shown in yellow and orange respectively.

A.



B.

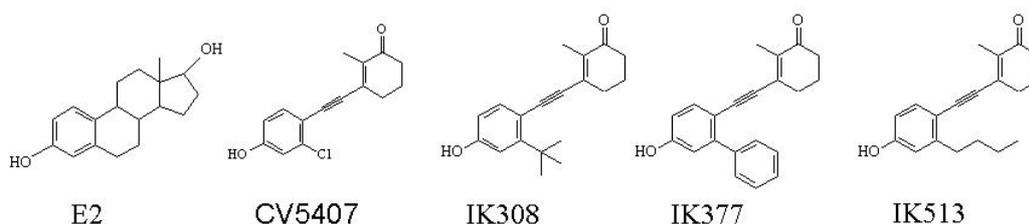


Figure 3.8: A. Relative activity of ligands with bulky substitution at C5 position compared to E2 in the WT as well as mutant M388A, L428A and double mutant ML-AA in yeast one hybrid system. B. The structure of ligands use in the assay.

3.4 Irrational design using random mutagenesis

While rational design failed to deliver a suitable ligand-receptor combination, we focused on the so-called “irrational design” to create uniquely interacting ligand-receptor pairs. To that end, we have developed an on-plate selection method where only mutants with altered positive affinity towards the ligand in question. We have chosen CV6019 for the selection of mutant in the random mutagenesis screen at concentration 10 μ M. Criteria for ligand selection was based on minimal activity of the ligand in the WT receptor- in this case, the ligand had around 20% relative activity in WT receptor compared to E2 which enabled us to start with and provided the opportunity for further improvement of ligand-receptor binding affinity. Besides, we had included this ligand in tobacco and wheat seed germination assay for possible toxicity and CV6019 showed no adverse effect in germination and seedling growth in our test system (See the 3.2 Toxicity Test). For the

random mutagenesis purpose, we chose error-prone PCR on the entire LBD keeping the exchange rate of amino acid residue ~ 1 per kb of gene [126]. A relatively small library with 4×10^4 variants was created and transformed into competent yeast strain PJ69 4A. A ligand-induced nutritional complementation selection was applied so that PJ69 4A cells carrying mutated LBD with improved affinity towards ligand in use, CV6019, would only grow on the transformation plate (SD-Ade-His-Trp) supplemented with ligand at $10 \mu\text{M}$ concentration, which is a synthetic dropout medium lacking tryptophan for selection for plasmid and adenine and histidine for transactivation. Figure 3.9 briefly outlines the steps involved in creating random mutagenesis library followed by selection, identification and characterization of mutant.

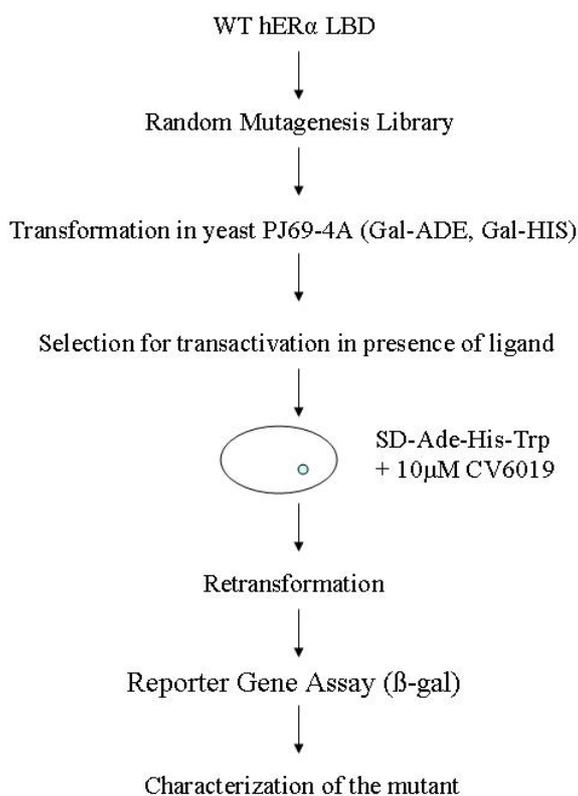


Figure 3.9: Scheme showing the steps in random mutagenesis library creation, screening, identification and characterization of mutant

In the random mutagenesis approach we found one mutant, which was ~ 3.5 fold more active in presence of $10 \mu\text{M}$ CV6019 compared to the wild type (Figure 3.10.A). Sequencing revealed that the mutant has an exchange of V to A at position 400. When we made a dose-dependent β -gal reporter gene assay in presence of CV6019, the mutant was

found to be active at concentrations starting from 1 μ M, which was also the case for WT (data not shown). Interestingly, the mutant was ~15 fold more active than the WT in presence of 1 μ M CV6019 (Figure 3.10.B).

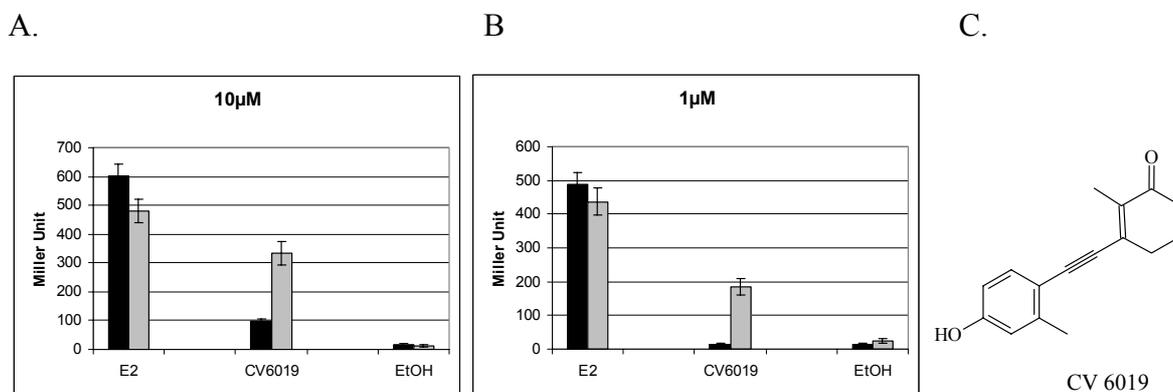


Figure 3.10: Activity of mutant V400A compared to the WT in presence of CV6019 in reporter gene assay. β -galactosidase assay was performed in yeast one hybrid system using A. 10 μ M and B. 1 μ M concentration of CV6019 and β -galactosidase activity is plotted as Miller Unit. WT and V400A LBD are depicted in black and grey, respectively. C. Structure of CV6019.

3.4.1 Characterization of V400A mutant

3.4.1.1 3-D Model of LBD with V400

After sequencing, we found that the increased activity towards CV6019 was due to the exchange of residue 400 from valine to alanine. We wanted to look where this residue lies and how it interact with our ligands using molecular modeling program PyMOL because there exists literature indicating a role of residue 400 in hsp interaction [127-129]. Since this interaction takes place on the surface of the receptor protein, there was no wonder when it was found that the position 400 is situated around 13 \AA away from ligand, as shown in Figure 3.11.

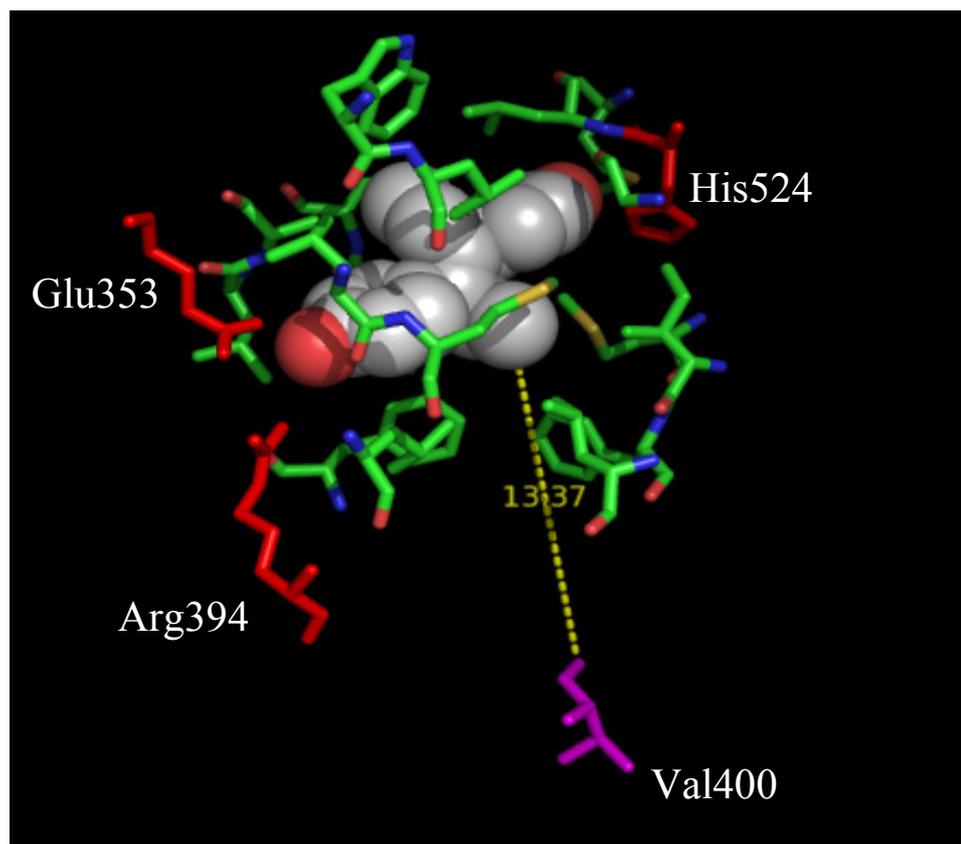
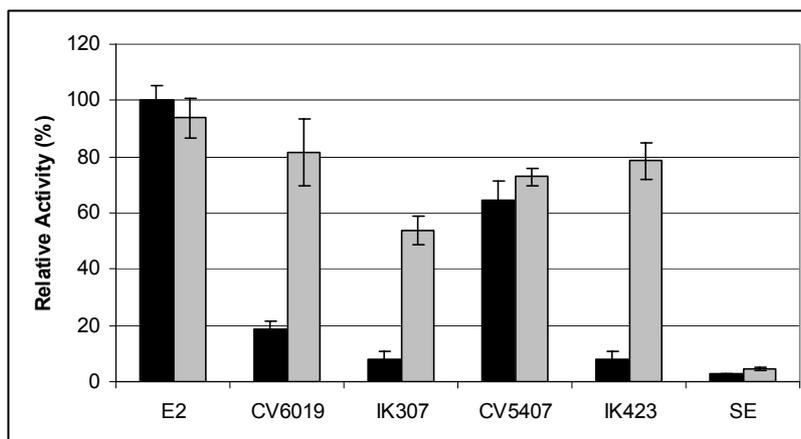


Figure 3.11: Residues forming the ligand binding pocket (LBP) for DES are shown in crystal structure of bound DES in hER LBD α (PDB code 3ERD). The ligand itself is shown in grey (sphere) while residue E353, R394 and H524 interacting with the OH-groups from A and D rings are shown in red. The V400 is shown in pink and is positioned 13Å away from the ligand and does not constitute the LBP.

3.4.1.2 Ligand-dependent activation of V400A

After observing increased activity of CV6019 in mutant V400A, we were interested in other ligands to know whether there was a general trend of increased activity towards this mutant or this ligand was more active specifically in the mutant. For this purpose, we tested several ligands with variable C 5' substitution and variable D ring structure. The result from the β -galactosidase assay showed all ligands included in the assay at 10 μ M concentration were active in the mutant and in combination with IK423, the highest activity compared to WT was observed, i.e. ~10 fold better activity compared to the WT (Figure 3.12). It was also clear from the figure that the solvent alone (as negative control) had no activity in the mutant, thus the mutant was exclusively dependent on ligand for transactivation.

A.



B.

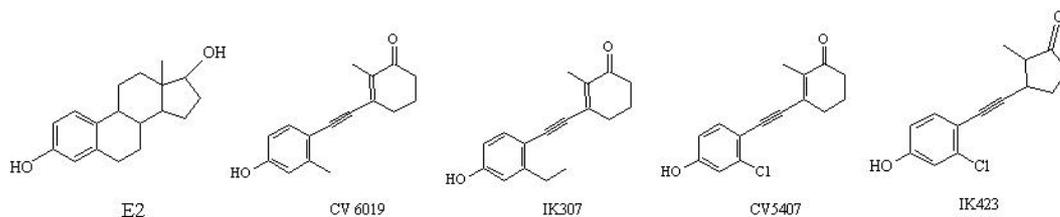


Figure 3.12: A. Relative activity of WT (black) and mutant V400A (grey) in β -galactosidase assay tested in yeast one hybrid system where ligands varying in C5 substitution as well as in the D ring structure were included at $10\mu\text{M}$ concentration. Although all ligands were active in the mutant, these activity were of reporter gene was ligand-dependent, as there was no activity found in the absence of ligand (only solvent). B. Structure of ligands included in the above-mentioned assay.

We also wanted to compare the result found from the above experiment using chemical complementation in our yeast two hybrid system. While comparing the EC_{50} values from WT and V400A mutant, we could see that there was no improvement in the V400A mutant for the affinity towards the ligands (Figure 3.13). The general tendency was that V400A mutant showed weakend affinity towards all ligands used (including E2) compared to WT.

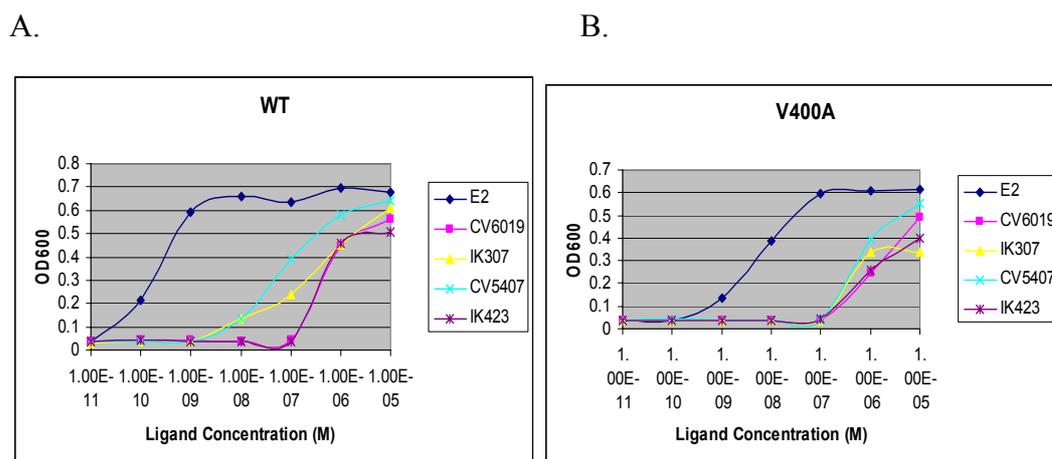


Figure 3.13: Comparison of EC_{50} values of A. WT and B. mutant V400A in the yeast two hybrid system in presence of compounds used in the previous experiments (see Figure 3.12). Compared to the WT, the mutant V400A showed reduced affinity towards E2 as well as ligands used in the assay.

3.4.1.3 *In vitro* binding assay

To establish whether the improved transactivation potency of the V400A mutant was the result of a corresponding change in the ligand binding affinity, we used an *in vitro* reconstitutive enzyme assay for competitive ligand binding study, which is a commercial estrogen assay kit from DiscoverRx (Hithunter Estrogen Assay kit). The assay is based on enzyme fragmentation complementation of β -galactosidase where the enzyme is fragmented into two distinct parts- an enzyme donor (ED) and an enzyme acceptor (EA). When the two fragments are separated, the enzyme is inactive. A simplified scheme is presented in Figure 3.14 showing the enzyme fragmentation complementation.

- A. ED is chemically coupled with E2, which does not interfere with the complementation of ED and EA to form functional enzyme and hydrolyse substrates like β -galactosides. This substrate degradation can be monitored and a standard curve can be generated to find the optimal concentration of ED and EA.
- B. When purified estrogen receptor is added to the reaction, the receptor binds E2 from ED-E2 and complementation of the functional β -galactosidase enzyme is not possible-thus no signal is generated.

- C. In presence of competing ligand and purified receptor, there would be a ligand-receptor pair as well as a functional enzyme and signal would be generated depending on the affinity of the ligand to the receptor.

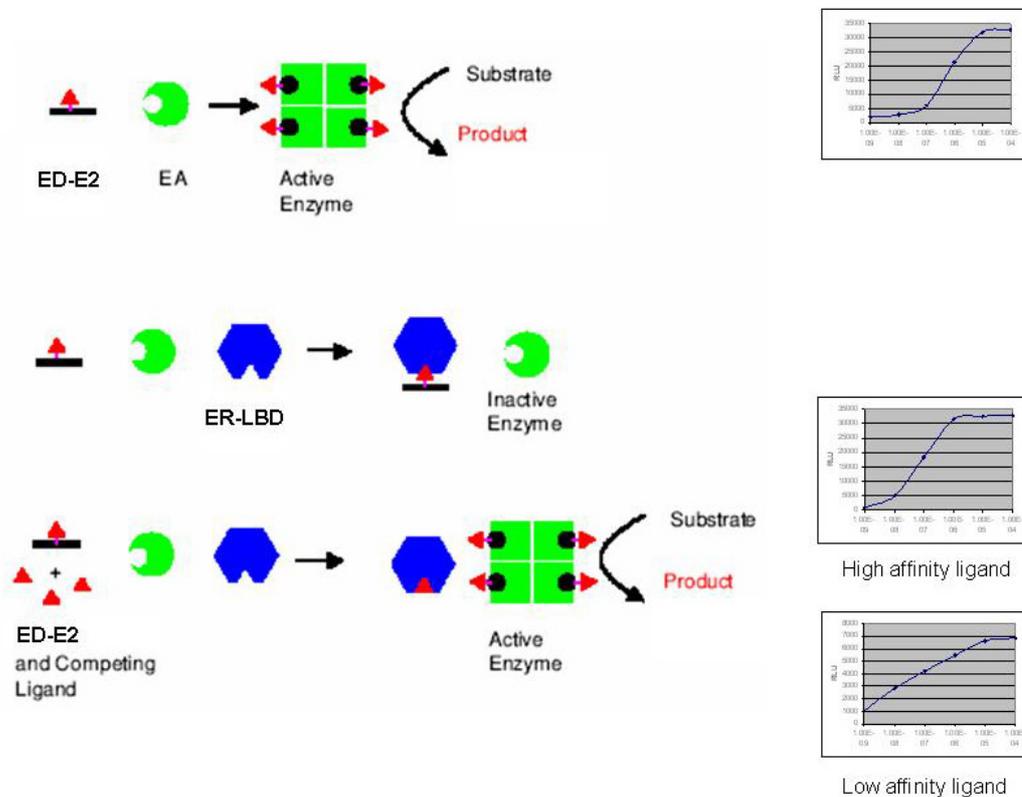


Figure 3.14: Schematic representation of the enzyme fragmentation complementation based *in vitro* assay for competitive binding. β -galactosidase enzyme is fragmented into two components—enzyme donor (ED, which is conjugated with E2) and enzyme acceptor (EA). Upon interaction, the two fragments of β -galactosidase (ED and EA) reconstitute functional enzyme and hydrolyze substrate, which could be quantified. In presence of ER, the ED-E2 binds to the receptor and a functional enzyme cannot be reconstituted. In presence of a competing ligand and ER, some of the ligand will bind to the receptor and ED-E2 could couple with EA to form functional enzyme and catalyze reaction. When receptor concentration is kept static, EC_{50} values could be obtained from the plotted dose vs enzyme activity (fluorescence) graph.

Using on the above-mentioned assay, we have compared first the EC_{50} of our ligand CV6019 and E2 in the commercially available full length WT LBD (Figure 3.15). The commercially available full-length receptor was used as a proof of principle that the assay could be used to characterize ligand binding *in vitro*. The EC_{50} values of E2 and CV6019 are 200nM and 30 μ M, respectively obtained from this *in vitro* assay towards the WT receptor.

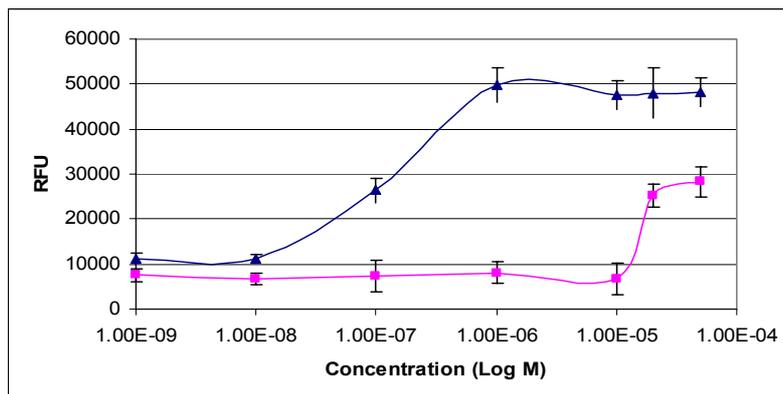


Figure 3.15: *In vitro* binding assay of E2 and CV6019 in the WT receptor. The enzyme activity was measured in presence of variable concentrations of E2 and CV6019 and the half-saturation was taken for calculation of the affinity of WT receptor towards E2 and CV6019.

Once we had the values of E2 and CV6019 from WT LBD, we wanted to know whether there was an improvement of the affinity of CV6019 in V400A mutant by comparing the EC_{50} values between the WT and the mutant. For this purpose, we had to purify the WT as well as V400A mutant receptor in native and ligand-free form. *E. coli* BL21 (DE3) cells transformed with the pET15b plasmid containing the WT or the V400A mutant were used for IPTG based over-expression followed by sonication and centrifugation to recover the receptor in soluble supernatant fraction. In the competition assay, the EC_{50} values for CV6019 were very much similar in the WT and V400A mutant (Figure 3.16 and Table 3.2). The WT and V400A mutant also showed similar binding affinity towards E2.

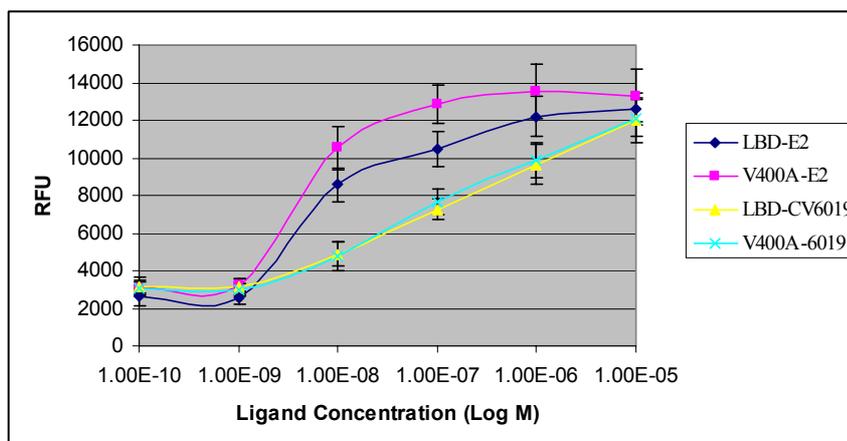


Figure 3.16: *In vitro* binding assay using purified WT and mutant V400A LBD to determine their affinity towards E2 and CV6019.

Estrogen Receptor	EC ₅₀ (M)	
	E2	CV6019
LBD	7X10 ⁻⁹	5X10 ⁻⁷
LBD V400A	5X10 ⁻⁹	6X10 ⁻⁷

Table 3.2: Half-saturation values obtained for E2 and CV6019 in WT and V400A LBD from the *in vitro* binding assay.

3.5 Directed Evolution

3.5.1 Ligand classification in yeast two hybrid system

We had relatively a large number (56) of ligands with diverse structure and activation properties in our test systems, we decided to build ligand titration curve/yeast transcription response curve in the yeast two hybrid system. This ligand titration curve would provide us with the valuable information like the affinity of our ligand towards the WT receptor and thus help us decide which ligand we would use for developing orthogonal ligand-receptor pair through directed evolution. The half-saturation (EC₅₀) values found in the ligand titration curve is the basis for ligand classification. E2 has the EC₅₀ value of 5X10⁻¹⁰ M in our test system whereas we have ligands with EC₅₀ values ranging from 5X10⁻⁸ to 5X10⁻⁶ M. We also had some ligands which did not show any transcription response and thus no ligand titration curve and EC₅₀ values could be obtained. Based on these results, we classified the ligands into the following four groups (Figur 3.17 and Table 3.3). Ligands with halogen substitution at position 5' of A ring (i.e. CV5407) showed the lowest EC₅₀ value (greatest affinity) in the range of 5X10⁻⁸ and they are termed as group D. Compounds with no transcription response (group A) have two distinct properties: the 3'-MeO or other substitution instead of -OH and 5-carbon D ring. There were no distinct relevance of structural similarities and EC₅₀ values in the rest of the groups (B and C)

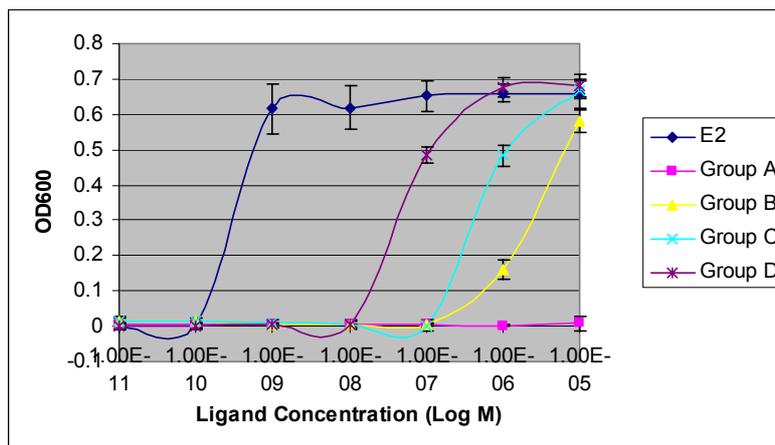


Figure 3.17: Ligand classification based on their transactivation properties in yeast two hybrid system. The half-saturation (EC_{50}) of natural ligand E2 was 0.5nM, whereas group B, C and D had EC_{50} values of 5 μ M, 0.5 μ M and 50nM, respectively.

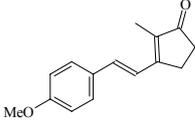
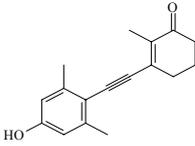
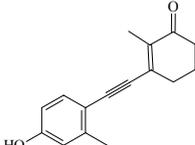
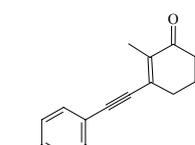
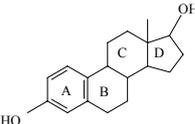
Ligand category	Example	EC_{50}
Group A	 CV3519	Could not be detected
Group B	 CV3320	$\sim 5 \times 10^{-6}$
Group C	 CV6019	$\sim 5 \times 10^{-7}$
Group D	 CV5407	$\sim 5 \times 10^{-8}$
Natural ligand	 E2	$\sim 5 \times 10^{-10}$

Table 3.3: Based on the transactivation properties of ligands in yeast two hybrid system, we classified them into 4 different group. A representative from each of these groups was included in the table with its EC_{50} value. The four different groups of ligands were compared with natural ligand E2.

3.5.2 Identification of ligand for the saturation mutagenesis

Directed evolution attempts to recapitulate the natural evolution of proteins with new structures and functions, but on an experimentally accessible timescale. To create orthogonal ligand-receptor pair, we considered starting with ligands, which have minimum transcriptional activation properties. This strategy provides the possibility of increased fold improvement (increased affinity of ligand towards the mutated receptor as well as reduced affinity towards WT receptor). Ligand selection criteria were also determined in such a way that the ligand would not bind to the WT hER α - so that the orthogonal ligand-receptor pair could be generated. From the ligand classification mentioned in the previous section, we identified five ligands from group B, which have the lowest affinity (next to zero, group A) towards the WT hER α LBD. We confirmed this result from the reporter gene assay (β -galactosidase, ONPG assay) using yeast one hybrid system (Figure 3.18). In that β -galactosidase assay, all those five compounds showed background-level activity. We have used CV6019 in our previous rational design approach. This ligand belongs to group C with EC₅₀ value around 5×10^{-7} , which is relatively high compared to the other ligands from group B. Group B ligands were ideal for creating orthogonal ligand-receptor pair because of their poor affinity towards WT LBD and thus there were greater potential to create orthogonal ligand-receptor pair with significant fold improvement.

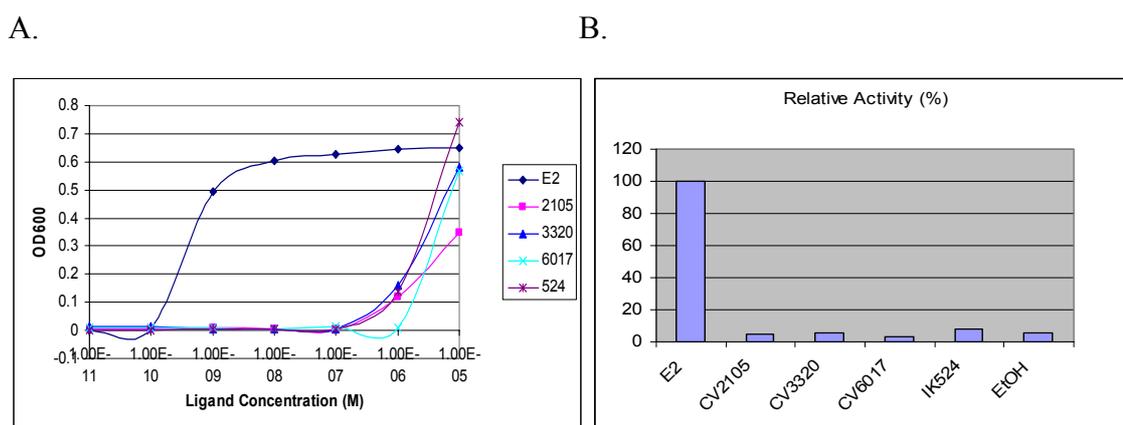
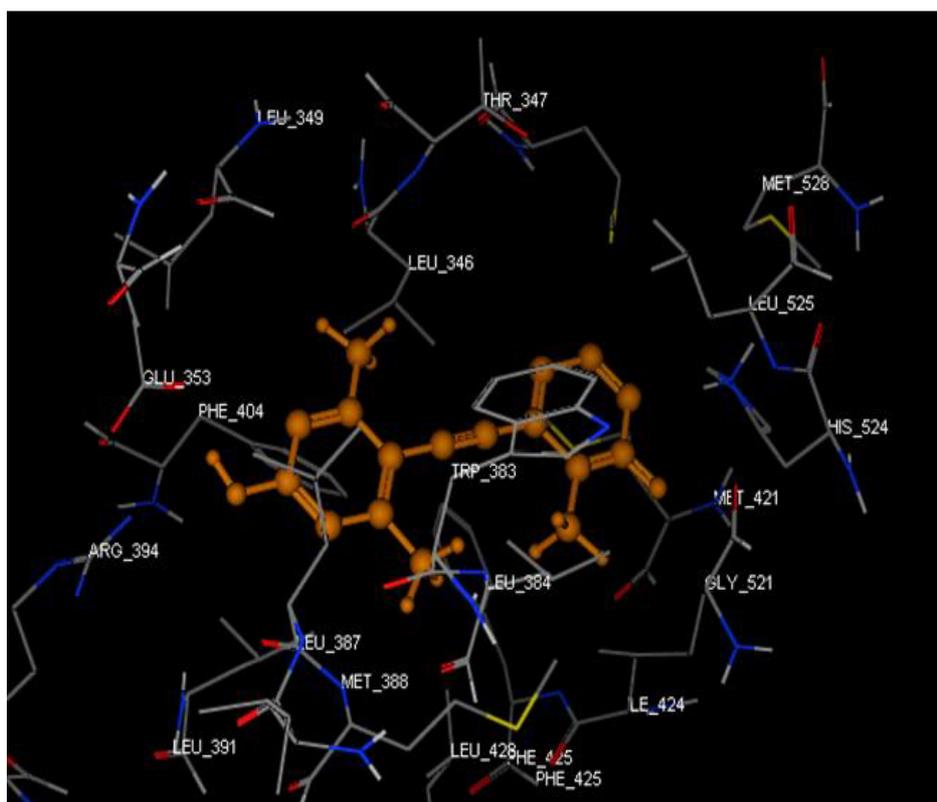


Figure 3.18: Affinity of ligands belonging to group B compared to E2 in WT LBD. Ligand affinity was calculated by (A) Ligand-dependent transactivation and growth of yeast cells containing WT LBD in yeast two hybrid system. The group B ligands had EC₅₀ values of $\sim 5 \mu\text{M}$ compared to the 0.5 nM of that of E2 in WT LBD. (B) Relative activity of group B ligands compared to E2 in yeast one hybrid system calculated from ligand-dependent β -galactosidase reporter gene assay. Group B ligands showed poor activity which was comparable to that of solvent only.

3.5.3 Identifying ligand-contacting residues in hER LBD

Molecular simulation studies were made by loading the known crystal structure of hER α LBD crystallized with E2 or DES and replacement of the ligand with group B ligand CV3320. Using Molecular Operating Environment (MOE), twenty-one residues were identified which were thought to be in direct contact (within 4.6 Å) with the docked CV3320 in the hER LBD, as shown in Figure 3.19. It was reasoned that these 21 residues could be considered lining the ligand binding pocket for CV3320. For our receptor re-engineering work towards CV3320, 3 of the 21 residues were not subjected for manipulation: E353, R394 and H524- because of their known role in hydrogen bonding with the terminal hydroxyl groups of the ligand [73]. The rest 18 residues were selected for stepwise individual site saturation mutagenesis.

A.



B.

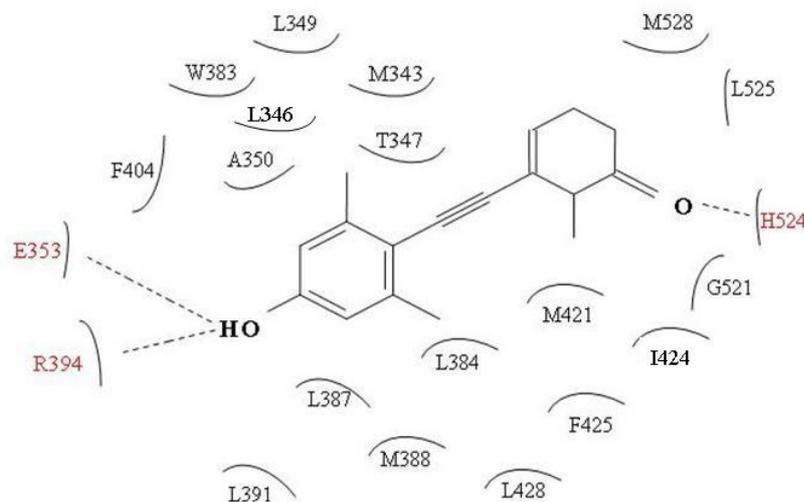


Figure 3.19: Computational simulation-based selection of residues for saturation mutagenesis. (A) hER α LBD crystal structure with DES was loaded in MOE and the ligand was replaced with CV3320. The ligand was allowed to rotate freely in energy-minimized state and residues within a reasonable distance ($\sim 5\text{\AA}$) were selected for saturation mutagenesis approach. (B) Two dimensional representation of CV3320 in the LBP of hER α LBD where 21 residues were identified within $\sim 5\text{\AA}$ distance. Among these 21 residues, E353, R394 and H524 are reported to be important for receptor activity and could not be replaced; these three residues were depicted in red. The rest 18 residues (depicted in black) were considered for manipulation to create orthogonal ligand-receptor pair.

3.5.4 Saturation mutagenesis

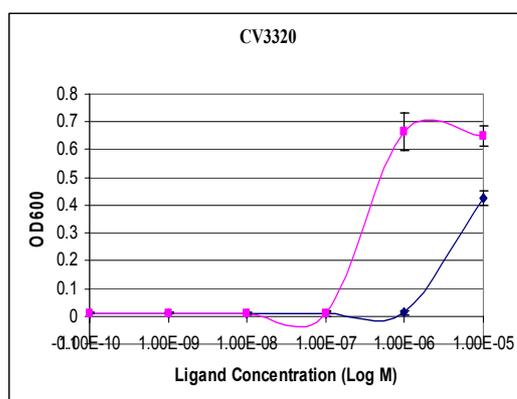
Saturation mutagenesis works for creating CV3320-specific receptor were performed by stepwise individual site saturation mutagenesis on 18 functionally important ligand-contacting residues. The saturation mutagenesis library with the linearized backbone pBDGal4Cam was co-transformed with pAD SRC-1 plasmid. Each saturation mutagenesis library was plated onto non-selective minimal media agar plates (selective only for plasmid containing cells, not for mutant with strengthened target ligand response). One hundred and ninety yeast colonies from each saturation mutagenesis library were picked and subjected to yeast two hybrid screening approach to achieve chemical complementation. Chemical complementation was achieved using CV3320-induced transcriptional activation of *HIS3* and biosynthesis of histidine in a minimal media lacking histidine (selection for transactivation). The concentration of synthetic ligand used for the

identification of mutants with enhanced response was chosen to be relatively lower (5-10 fold) than the concentration at which the WT receptor started to respond, while the concentration of E2 used for identification of mutants with weakend response was chosen such that the parent shows moderate response at this concentration. In addition, each saturation mutagenesis library was also accompanied by two controls consisting of yeast cells expressing WT construct. Mutants were considered candidates for enhancement in specificity for target ligand (CV3320) if two criteria were met:

- i. The growth of mutant yeast cells appeared more confluent than the WT in the media containing CV3320
- ii. The growth of mutant yeast appeared less confluent than the parent in media containing E2

In the first round of saturation mutagenesis, six putative candidate clones were identified, and when retransformed after plasmid rescue from yeast, only one showed improved affinity towards CV3320 and weakend response towards E2. The overall fold improvement in this clone, where the luciferase residue at 384 is exchanged to phenylalanine (L384F), was 1.6×10^3 . The transactivation profile is shown in Figure 3.20 and the EC_{50} values in comparison with WT are summarized in Table 3.4.

A.



B.

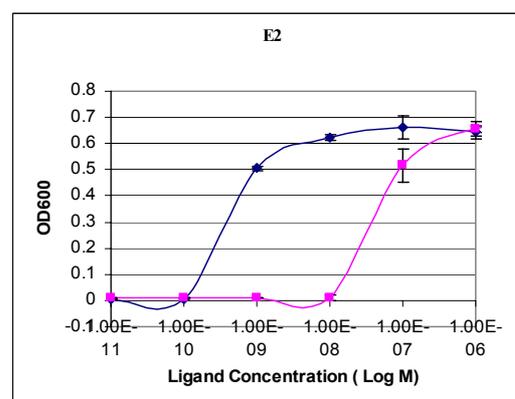


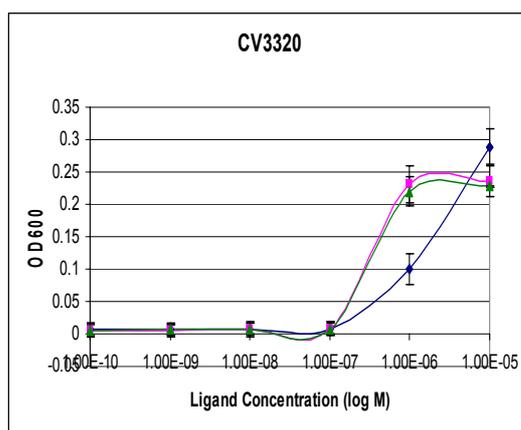
Figure 3.20: Transactivation profile of yeast two hybrid cells for WT (round, blue) and mutant (square, pink) showing the best transactivation from the first round of saturation mutagenesis. (A) Transactivation profile in presence of CV3320. (B) Transactivation profile in presence of E2. See Table 3.5.1 for mutant identity and EC_{50} values.

Construct	EC ₅₀ , CV3320 (M)	EC ₅₀ , E2 (M)	Selectivity	Fold improvement
WT	3X10 ⁻⁶	5X10 ⁻¹⁰	1.6X10 ⁻⁴	1
L384F	5X10 ⁻⁷	5X10 ⁻⁸	10 ⁻¹	1.6X10 ³

Table 3.4: Summary of the first round of saturation mutagenesis based on yeast two hybrid system. Selectivity refers to the preference of CV3320 relative to E2, and is the ratio of EC₅₀ values (EC₅₀, CV3320/ EC₅₀, E2). Fold improvement is the selectivity enhancement for CV3320 relative to the WT construct.

The mutant from the first round of saturation mutagenesis (L384F) was taken as template for creating further saturation mutagenesis library in rest 18 positions and screened for further improved affinity towards CV3320 and weakend affinity towards E2. As shown in the Figure 3.21, there was no improvement in the affinity towards CV3320 but the affinity towards E2 was 10-fold reduced, so that the overall fold improvement was 10-fold increased. The result from second round of saturation mutagenesis screen is summarized in Table 3.5. Sequence information revealed exchange of leucine into methionine at position 387 in the mutant clone from second round of saturation mutagenesis.

A.



B.

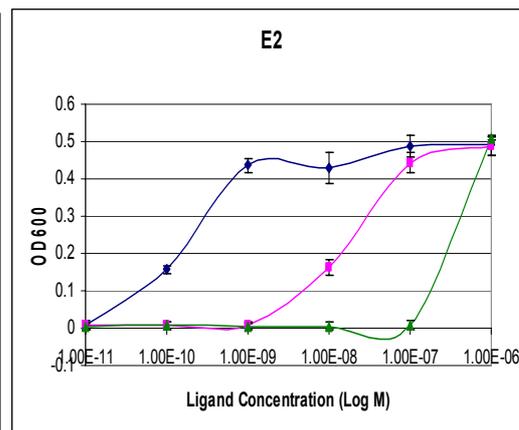


Figure 3.21: Transactivation profile of yeast two hybrid cells for WT (round, blue) and mutants showing the best transactivation from the first (square, pink) and second (triangle, yellow) round of saturation mutagenesis. (A) Transactivation profile in presence of CV3320. (B) Transactivation profile in presence of E2.

Construct	EC50, CV3320 (M)	EC50, E2 (M)	Selectivity	Fold improvement
WT	3×10^{-6}	5×10^{-10}	1.6×10^{-4}	1
L384F	5×10^{-7}	5×10^{-8}	10^{-1}	1.6×10^3
L384F L387M	5×10^{-7}	5×10^{-7}	1	1.6×10^4

Table 3.5: Summary of the second round of saturation mutagenesis based on yeast two hybrid system where the mutant from the first round was also included.

We have also created third round of saturation mutagenesis library and selected for further fold improvement but there was no improvement achieved. Several candidate mutants were identified with reduced estrogen affinity but those mutants also had dramatic reduced response towards CV3320. The inability to improve further after repeated attempt indicates that the level of saturation one could achieve using saturation mutagenesis was reached.

3.5.5 Random Mutagenesis

Once we arrived in such a position that no further improvement could be made to improve the affinity of the protein to CV3320 through saturation mutagenesis, we approached the random mutagenesis approach since it is well documented that residues outside the ligand binding pocket could also influence the binding and affinity towards a ligand through secondary structure rearrangements [107, 130]. In our experiments, we have chosen to use error-prone PCR to generate random mutations mainly because the mutation rate of error prone PCR can be easily and precisely controlled by the MnCl_2 concentration in the reaction mixture [131]. We used 0.15 mM MnCl_2 in the error prone PCR reaction at which concentration it was estimated one to two amino acid substitutions per gene on average was generated [131]. CamL ER and CamR ER, the two primers flanking the L384F L387M LBD (a.a. 312-595) were used for amplification. The amplified PCR product was co-transformed with linearized pBDGal4Cam into yeast competent cells YRG2-SRC-1. The two cotransformed linear DNAs share 40 bp of homology at their ends, allowing the yeast cell to recombine the linear DNA *in vivo*, giving rise to circular plasmid pBDGal4CamLBD(m). In the first round of random

mutagenesis, 378 clones were first picked from the transactivation plate containing appropriate concentration of ligand CV3320, where the library size was 2×10^6 , into four 96-well master plates (minimal media selecting for plasmid) and incubated overnight at 30°C. These candidate clones were used for inoculation of 96-well transactivation plates containing water, minimal media (selecting for plasmid as well as for transactivation) with CV3320 or E2 into the corresponding well compared to the masterplate. After 12-16 hours of incubation, clones were only picked from the corresponding master plate according to the following criteria observed in transactivation plate-

- No growth in water (discarding constitutive activation), and
- more confluent growth compared to control (here the mutant from the second round of saturation mutagenesis, L384F L387M LBD) in presence of CV3320 at the concentration of 50nM

Although we included weakend response of the candidate clone towards E2 compared to the positive control, no such clone were found which fulfill this particular criteria. So, we focused on increased affinity towards CV3320 and selected 4 clones from the corresponding master plates. After plasmid rescue and retransformation, only one clone showed increased affinity towards CV3320 with further 10-fold increase compared to the mutant found from second round of saturation mutagenesis (this clone also showed 100-fold increased affinity towards E2 compared to L384F/L387M). Sequencing revealed an additional exchange in Y537 to S in this mutant. The result from the first round of random mutagenesi is shown in Figure 3.22 and summarized in Table 3.6.

We took this mutant (L384F L387M Y537S) as a template for second round of random mutagenesis for further engineering the ligand-receptor combination in a library of 3.2×10^6 variants but we could not improve the affinity towards CV3320 nor decrease the affinity towards E2 separately. The candidates showing improved affinity towards CV3320 also showed improved affinity towards E2 and mutant showing decreased affinity towards E2 also showed significantly decreased affinity towards CV3320. One candidate clone behaved slightly according our

target but after retransformation and sequencing, it proved out to be the same clone from the first random mutagenesis (L384F L387M Y537S).

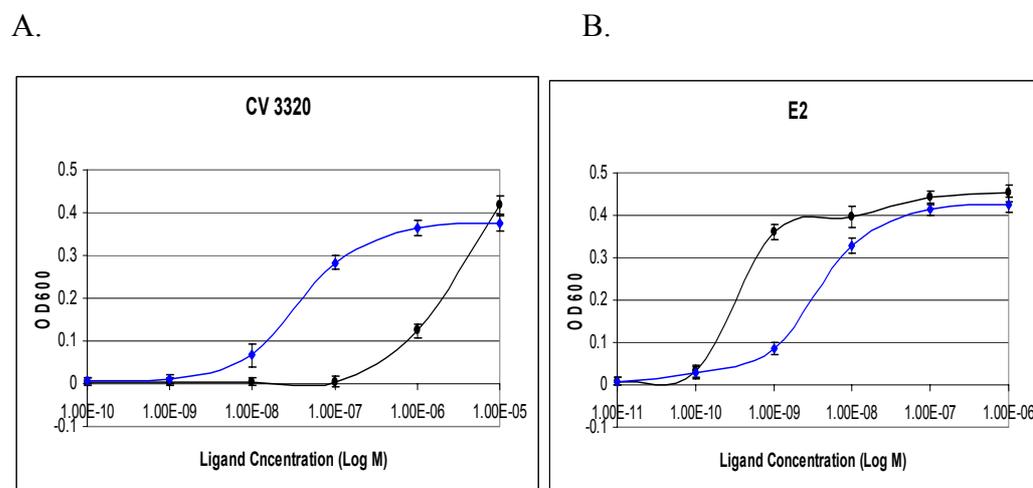


Figure 3.22: Transactivation profile of mutant generated from two rounds of saturation and the first round of random mutagenesis (royal blue, quadrate) compared to WT (round, black). Transactivation profile was obtained from the yeast two hybrid system using (A) CV3320, and (B) E2.

Round	Construct	EC ₅₀ , 3320 (M)	EC ₅₀ , E2 (M)	Selectivity	Fold improvement
0	WT	3X10 ⁻⁶	5X10 ⁻¹⁰	1.6X10 ⁻⁴	1
1-S	L384F	5X10 ⁻⁷	5X10 ⁻⁸	10 ⁻¹	1.6X10 ³
2-S	L384F L387M	5X10 ⁻⁷	5X10 ⁻⁷	1	1.6X10 ⁴
3-R	L384 F L387M Y537S	5X10 ⁻⁸	5X10 ⁻⁹	10 ⁻¹	1.6X10 ³
4-R	L384 F L387M Y537S	5X10 ⁻⁸	5X10 ⁻⁹	10 ⁻¹	1.6X10 ³

Table 3.6: Summary of results from two rounds of saturation mutagenesis and two rounds of random mutagenesis based on two hybrid system in yeast in presence of CV3320 and compared with E2. S and R are saturation and random mutagenesis, respectively. The numbers in the column round are depicting the numeric of rounds.

3.5.6 Additive effect of mutations for improved affinity and better transactivation

Since recombining the two rounds of saturation mutagenesis and one round of random mutagenesis improved the affinity of the mutant to CV3320 in fold improvement of 1.6X10³, we were interested to see if any of these residual

exchanges had dramatic effect alone or this improvement was additivity of the functional mutations. Particularly, several mutants at position 537 including serine were reported to have constitutive activation in yeast and mammalian system [132-134]. To that end, we separated the exchanges from saturation mutagenesis (L384F L387M) and random mutagenesis (Y537) and tested those mutants in our one or two hybrid system to decipher their role. In the yeast one hybrid system (shown in Figure 3.23), the ligand CV3320-L384F L387M Y537S combination was able to transactivate ~70% to that of E2-WT combination at 10 μ M concentration. When we compared these two combinations at 1 μ M concentration, the transactivation property increased upto 84%. Mutant L384F L387M had only 14% activity in presence of E2 and completely lost its activity in presence of CV3320. The Y537S mutant had retained 35% activity with E2 and had relatively increased activity with CV3320 as well as with the solvent alone. This could be explained by the increased background activity of this mutant but in general, we could not see any constitutive activation in our one hybrid system.

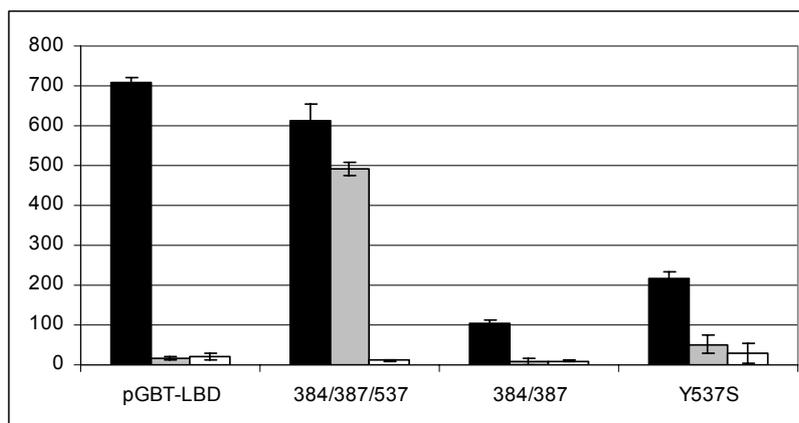


Figure 3.23: Transactivation profile based on the β -galactosidase reporter gene assay (Miller Unit) in yeast one hybrid system. Mutations achieved through two rounds of saturation and one round of random mutagenesis were separated according to the manipulation approach (i.e. saturation mutagenesis- L384F L387M, random mutagenesis Y537S) and compared in presence of E2 (black), CV3320 (grey) or the solvent (white) at 10 μ M concentration.

Because the yeast two hybrid system that we employed was more sensitive than that of one hybrid, we tested the Y537S mutant with the WT in our two hybrid system in presence of E2 and CV3320 (Figure 3.24). The Y537S mutant has EC_{50} value 5×10^{-9} M, which explained the reduced activity of E2 (35%) in this mutant in one hybrid system. There were no differences between WT and Y537S in their

affinity towards CV3320 from their EC_{50} values in yeast two hybrid system. Thus the Y537S alone could not increase the activity of CV3320 in the yeast one hybrid system and the affinity towards CV3320 was also not improved in this mutant.

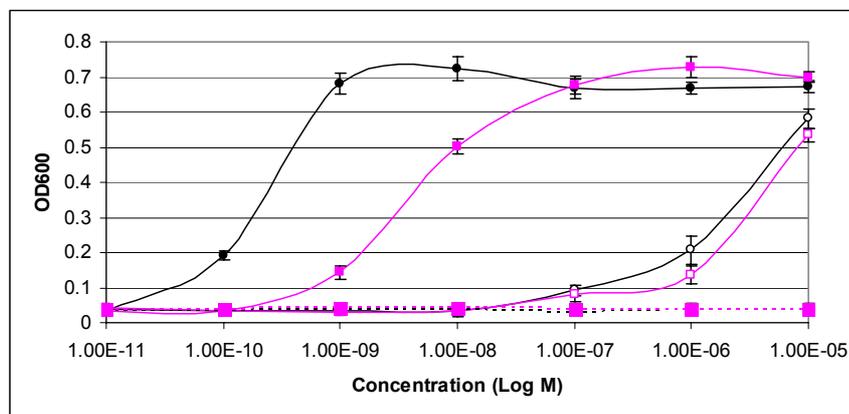


Figure 3.24: Transactivation profile of WT LBD and Y537S mutant in presence of E2, CV3320 or with solvent only in the yeast two hybrid system. Legends for receptor-treatment are as follows: WT-E2: closed circle, black, Y537S-E2: closed square, pink; WT-CV3320: open circle, black; Y537S-CV3320: open square, pink; WT-EtOH: closed circle, broken black; Y537S-EtOH: closed square, broken pink.

3.6 Combinatorial improvement of transactivation: combining V400A with L384F L387M Y537S

Since the mutant V400A did not have any improved affinity towards CV6019 in the *in vitro* binding assay (see Figure 3.16), we hypothesized that the hsp interaction is weakened in the V400A mutant. This led to the better release from hsp even in presence of a weak ligand interaction, which at the end contributed to the increased activity found in our test system. On the other hand, we could clearly show the improvement of affinity towards CV3320 in the mutant L384F L387M Y537S. We wanted to know whether weakening chaperon interaction (mediated by V400A) and increased affinity (engineered mutant L384F L387M Y537S) could be combined together. If their combination led to the cooperative influence, we would get dramatic improvement in reporter gene activation in presence of very low concentration of ligand. To validate our hypothesis, we brought the two mutants in pGBT-LBD background and tested in our one hybrid system for β -

galactosidase reporter gene assay. In the assay, we included CV3320 and CV6019, ligands which were used for two mutants separately in previous experiments.

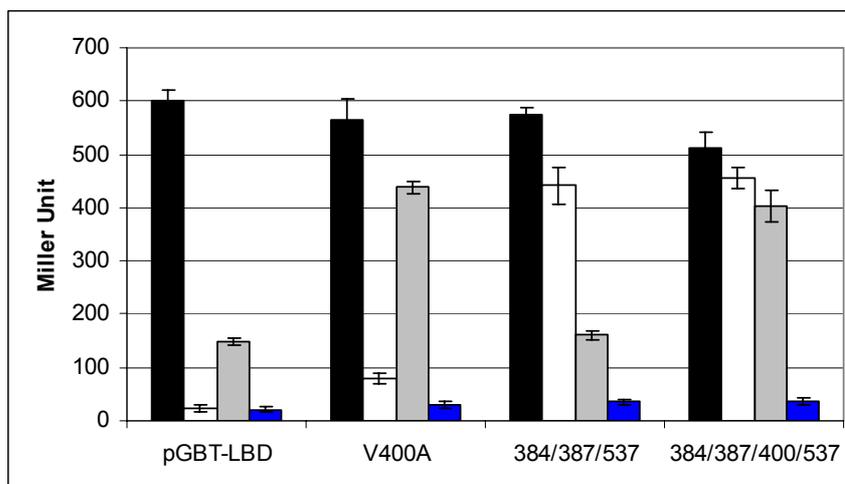


Figure 3.25: Comparison of β -galactosidase reporter gene activity (Miller Unit) among WT, V400A, L384F L387M Y537S and L384F L387M V400A Y537S mutant LBDs in presence of E2 (black), CV3320 (white), CV6019 (grey) and EtOH (royal blue) in yeast one hybrid assay.

As shown in Figure 3.25, the reporter gene activity found in combined mutant (L384F L387M V400A Y537S) for CV3320 and CV6019 were similar to their corresponding single mutant and there were no additive effect observed. This was also evident for E2- if improved affinity and weakened chaperon interaction would have worked cooperatively, we could expect dramatic increase of E2 action as well as for those ligands.

4 Discussion

4.1 Choice of ligand

In order to develop an orthogonal ligand-receptor pair for transcriptional regulation, the most important criterion to be met was that the receptor must be manipulated in such a way that it responded to the ligand only after the manipulation, i.e. the ligand was not binding in the WT receptor (which makes the system safe to be used in the field). Presumably, such a unique ligand should be estrogen-like, in order to be recognized at all by the hER α LBD, while at the same time being sufficiently different from the native ligand so that it does not trigger a strong transactivation response with WT hER α .

Although the generation of functionally orthogonal ligand-ER pairs have been reported [40, 41, 85, 107, 135], a major distinguishing feature of our work relies on the use of an entirely new class of compounds structurally unrelated to existing natural ligands. We wanted a general method for discovery of orthogonal ligand-receptor pair, so we imposed the additional criterion that there should be similar structural analogs of potential ligand so that once a ligand proved to be unsuitable due to its solubility, charge or substitution, there would be alternatives to be used for the same receptor.

Typically, ER ligands comprise two hydroxyl groups separated by a rigid hydrophobic linker region. In addition, effective ligands possess a phenolic hydroxyl group at the C3 position of 'A' ring in a steroid structure [136]. The underlying determinants of ER's promiscuous ligand binding preferences were revealed by the initial hER α LBD complexes [73]. Ligand recognition is achieved through a combination of specific hydrogen bonds and the complementarity of the hydrophobic residues that line the cavity to the non-polar nature of ER ligands. A triumvirate of hydrogen-bonding residues (E353, R394 and a structurally conserved water molecule) forms a polar pocket. Access to this polar pocket is somehow restricted by a 'pincer-like' arrangement of the side chains of L387 and F404. Consequently, the planar moieties of ER ligands, such as the 'A' ring of E2, bind at this end of cavity. In contrast to the extensive network of hydrogen-bonding

residues at the 'A' ring region of the binding cavity, interaction at its distal end, where 'D' ring of E2 lies, typically involves a single hydrogen bond to a histidine residue (H524). Although the 'pincer-like' arrangement around the 'A' ring imposes a general requirement for ligands to possess an aromatic ring with -OH group, the remainder of the binding pocket can accept a variety of hydrophobic groups [73, 74, 137].

While designing ligands for chemical synthesis, we focused on two particular characteristics such as:

- a) The presence of both 3-OH and 17-O group spatially separated by 10-12Å to ensure that the ER can recognize the ligand
- b) The absence of 'B' and 'C' ring of steroid receptor so that the interaction provided by this region is abolished and the ligands are only weakly interacting or not interacting at all with the WT receptor

We designed two different classes of ligands having above-mentioned characteristics where the 'D' ring was consist of either 6 (class 1) or 5 (class 2) carbon ring, as shown in Figure 4.1.

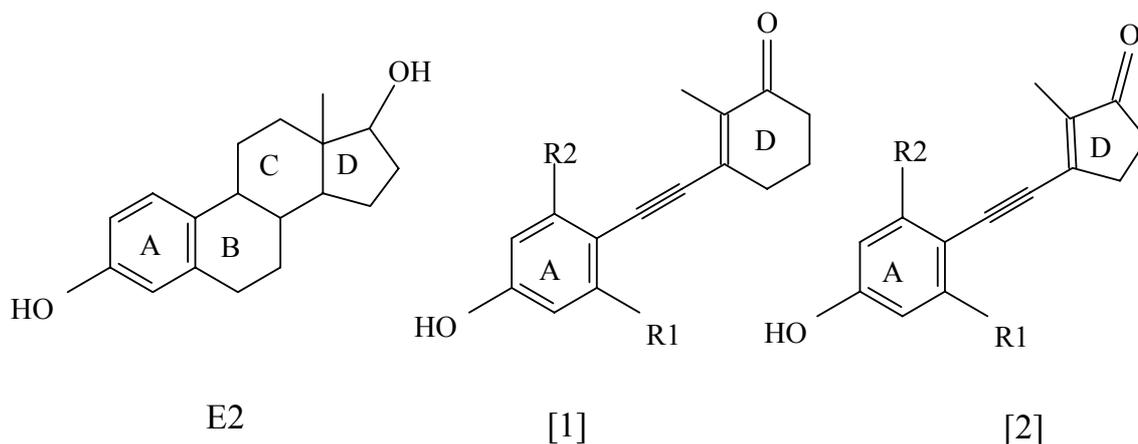


Figure 4.1: General ligand structure showing the difference in the 'D' ring in comparison with E2.

The initial focus was concentrated on ligands with R1 substitutions with halogen, OH and aliphatic hydrocarbon groups because of their activity in the WT receptor. Compounds which showed 10-20% activity in the WT LBD in yeast one hybrid system were primarily

selected for engineering LBD through rational design and later on using random mutagenesis. We have also tested these compounds for their possible role in general toxicity in planta according to the OECD guideline [138] where compounds with C5 –OH (CV6019) or Cl (CV5407) substitution did not show any adverse effects on germination and seedling development as well as adventitious shoot formation and general development. Although compound CV5407 did not show any general toxicity in those tests, we did not choose this compound for two reasons i.e. (i) it showed ~ 70% activity with the WT LBD (Table 3.1), and (ii) being an organochloric compound, it may have adverse effect on the environment. Compound CV6019, on the other hand, showed a minimum activity with the WT LBD (15-20% relative activity, Figure 3.2) in one hybrid system and was therefore chosen for creating orthogonal ligand-receptor pair based on rational design.

Because the yeast two hybrid system that we used was more sensitive than the one hybrid system and due to the high throughput of the selection of mutants in our directed evolution approach, we picked ligands from group B showing a very low affinity in WT LBD (see Figure 3.17 and 3.18). We have chosen ligands with minimum transcriptional activity in yeast two hybrid system (group B, EC_{50} value: $\sim 5 \times 10^{-6}$ M). Deciding to start with a minimum activity ligand minimizes the possibility of activating other estrogen receptors found in mammals, thus making the system suitable to use in other eukaryotic organisms. Additionally, there remains the greatest potential for improvement of ligand-receptor interaction, which is calculated as “fold improvement”. Because our test system in two hybrid screen was of high throughput, we had the choice to start with a ligand having very low affinity towards the WT receptor and improve the affinity in multiple rounds of selection in a relatively short period of time.

4.2 Yeast-based test system

The budding yeast *S. cerevisiae* has indeed proved to be a powerful model for studying the functions of more complex eukaryotes. The advantages of yeast arise in part from the relatively short doubling time in culture, the ability to use defined culture conditions and the availability of well established genetic tools for analyzing yeast mutants. In addition, a

number of fundamental processes, such as transcription initiation and regulation, appear to have been conserved from yeast to higher eukaryotes [139].

The first steroid receptors expressed in yeast were reported by two independent groups at same time [140, 141]. To date, the activity of virtually all the classical steroid receptors together with a number of non-steroid members of the nuclear receptor family has been reconstituted in the yeast *S. cerevisiae*. These pioneering studies illustrated that the different steps in steroid receptor signaling could be reconstituted in yeast cells by co-expression of the receptor protein and a reporter gene driven by the appropriate hormone response element [140, 141].

The ability to faithfully reconstitute ligand-dependent activity of steroid receptors in yeast has allowed yeast-based transactivation assays to be used to investigate the role of specific amino acids in hormone binding. Thus, yeast has been used to screen for altered ligand binding by the estrogen [142] and glucocorticoid [143] receptors after random mutagenesis of the receptor hormone binding domain. Significantly, the phenotypes observed for the mutant estrogen receptors were comparable in both yeast and Chinese hamster ovary cells [142]. Furthermore, the residues identified map to the recently described ligand binding pocket observed in the crystal structures of the estrogen and progesterone receptor ligand binding domains [144].

Yeast cells transformed with the estrogen receptor have thus proved a readily accessible and powerful screening tool to help identify potential “synthetic estrogens” and to compare the potency of such chemicals with the natural ligand 17 β -estradiol [145, 146]. Therefore, the yeast system can be seen as complementary and offers the advantage of increased sensitivity, adaptability and analysis of single receptor populations [146].

4.3 Chemical complementation based one and two hybrid test system

Genetic selection is a powerful tool and widely used in multiple aspects of molecular biology for evaluating macromolecular interactions. The general idea behind genetic selection is that the host cell will only survive if the desired entity, such as a desired function of the macromolecule, is present inside the cell [147, 148]. Chemical complementation is an attractive tool for genetic selection, which allows testing for activation of protein variants by specific ligands. However, when large libraries (>100,000 variants) of receptor variants are expressed in yeast, the number of hits (receptors activated by the desired ligand) may be too numerous to practically analyze. Many of the variants may only slightly activate transcription or have high EC₅₀s (in the μ M range). It is desirable to have a versatile selection method with which the receptors that have high efficacy and low EC₅₀s can be identified. To this end, we utilized two different chemical complementation selection methods with variable stringency. With variable stringency genetic selection, chemical complementation can identify not only receptors that are activated by the targeted small molecule, but also those that have the desired property of high activation levels. The concentration of ligand can be varied to identify receptors with low EC₅₀s. Hits with desired properties originally identified through chemical complementation on growth medium can be characterized using variable stringency chemical complementation screens.

The focus has been the development of a genetic selection system for detecting the interaction of a small molecule with a protein; in this case synthetic estrogen analogs were used to test the viability of chemical complementation for discovery of ligand-receptor pairs from large libraries. Combining chemical complementation with a large library of protein variants decreases the time, effort, and resources necessary to find new ligand-receptor pairs. Detection of ligand binding to nuclear receptors and transcriptional activation has been performed in yeast. Binding of the ligand to the nuclear receptor initiates transcription of the reporter gene. We had utilized a chemical complementation based coupled selection and screening in yeast one hybrid and a chemical complementation based selection in two hybrid system.

4.3.1 One hybrid system

We have constructed an estrogen inducible transcription factor Gal4 DBD-hER α LBD- VP16 AD, which consists of the DNA-binding domain of the yeast Gal4 protein (the first 74 amino acids) linked to the hormone-binding domain of the human estrogen receptor (amino acids 285-595) and the transactivation domain of herpes virus. The human estrogen receptor is a modular enhancer factor with two independent non-acidic transactivation functions (TAF-1 and TAF-2) each of which exhibits some cell type specificity [60, 149]. To render the chimeric transcription factor less cell type-dependent, we have modified this transcription factor by fusing onto it the transactivation domain of the VP16 [150]. This VP16 domain encodes one of the strongest acidic activation functions [151], does not transcriptionally interfere with TAF-2 activity [149] and functions in species from yeast to man [152]. Coupling the VP16 transactivation domain onto Gal4 DBD-hER α LBD did not lead to elevated basal activity in the absence of estrogen but only increased its transactivation strength in the presence of estrogen [153]. By replacing the DNA-binding domain of the ER with that of GAL4, we have eliminated the possibility of simultaneous activation of any endogenous E2-responsive genes if this system would be used in mammalian cells.

The reporter plasmid consists of the *ADE2*, *HIS3* and *lacZ* gene stably integrated in the yeast strain PJ69 4A [106]. The chimeric transcription factor was binding to the Gal4 response elements in the promoter regions and controlling the expression of the *ADE2*, *HIS3* and *lacZ* gene only upon ligand binding. The specific feature of this regulatory system was that it was only activating the target gene containing the GAL4-binding sites in the presence of E2 or matching ligand.

The three reporter genes present in PJ69 4A provided us two different types of assay: the on plate growth assay or selection assay quantifies yeast growth occurring through transcriptional activation of the *ADE2* and *HIS3* gene, and the screen quantified β -galactosidase activity occurring through transcriptional activation of the *LacZ* gene. The selection assay was \sim 10-fold more sensitive than

the screen as growth was observed at 10-fold lower ligand concentrations than β -galactosidase activity. However, the selection assay did not quantify activation level (efficacy) as well as the screen. In the selection assay, there was either growth or no growth, whereas the screen more accurately quantified different activation levels at various concentrations of ligands. We used this one hybrid system initially to validate the test system in growth based on plate assay and characterized the ligand binding in the WT LBD. We then generated variants through rational design and random mutagenesis, selected for positive variants and finally characterized the ligand binding for the mutant.

The sensitivity of yeast one hybrid based system was in nano- to micromolar range depending upon the affinity of the ligand. Although the system was not high throughput for screening of large libraries, the coupling of on plate growth assay eliminated the negative constructs (WT or mutants with no improvement in ligand binding) and thus reduced the load for screening quantitatively in β -galactosidase assay. The system was easy to use and produced highly reproducible data from the screening.

4.3.2 Two hybrid system

In an attempt to make the test system more sensitive where we could determine the ligand-receptor interaction in nanomolar range, we used a yeast two hybrid based interaction between hER α LBD and human steroid receptor coactivator (SRC-1). In this system, the cDNA encoding hER α amino acids 312–595 containing most of the LBD domain (hER α amino acids 303–553) and the F domain (hER α amino acids 554–595) was fused to the gene encoding the GAL4 DNA binding domain in plasmid pBD-GAL4-Cam (Stratagene) to form the “bait plasmid” pBD-GAL4 hER α , and the gene encoding human SRC-1 was fused to the gene encoding the GAL4 activation domain in plasmid pGAD424 (Clontech) to form the “prey plasmid” pGAD424 SRC-1. Both plasmids were transformed and expressed in *S. cerevisiae* YRG-2, which contains a *GAL4*-regulated *HIS3* reporter construct on its chromosome. The *HIS3* reporter provides strong

nutritional selection. In the presence of agonistic ligands, the LBD undergoes a conformational change and binds to SRC-1, which brings the GAL4 DNA binding domain and the GAL4 AD in proximity, thus activating the transcription of the reporter gene. In general, the cell growth rate is proportional to the strength of the ligand-receptor interaction. In the absence of agonistic ligands, no transcription of reporter genes would occur. While the survival of the yeast could indeed be linked to a small molecule ligand in the yeast one hybrid system, ligand-activated growth was only observed at 10^{-5} to 10^{-6} M of ligand concentrations. In the two hybrid system, introduction of the adapter fusion protein Gal4 AD-SRC-1 in chemical complementation increased the sensitivity 10-1000-fold over the one hybrid system based on the ligand used, which was also observed by several other authors [92, 148]. The functional interaction of hER α LBD with the coactivator SRC-1 is critical for effecting transcription in mammalian cells [67] and our chimeric construct could overcome these differences in the transcription machinery between human and yeast which we observed in one hybrid system. Thus, the Gal4 AD-SRC-1 fusion protein also humanizes the yeast response, allowing a complete correlation of nuclear receptor function in yeast and in mammalian cell culture.

This chemical complementation system was highly sensitive (producing growth with EC_{50} value of E2 5×10^{-10} M), with zero background (no growth without ligand), and high dynamic range (permitted variability in ligand concentration, incubation time and growth). Since we utilized a 96-well based assay format, the system provided high throughput analysis of multiple ligand receptor combinations. The assay provided an initial analysis of the ligands of interest through growth. The ligands that were able to cause growth of the yeast cells could then be subjected to a dose-response growth assay. Based on this growth assay, we initially classified and selected ligands for directed evolution, and characterized the mutations in terms of improved and weakened affinity towards CV3320 and E2, respectively.

4.4 Rational design

Rational protein design by site-directed mutagenesis is a very effective strategy to improve a protein function or define a new function for the existing protein. Rational design approach usually requires both the availability of the structure of the protein and knowledge about the relationships between sequence, structure and mechanism/function, and is therefore very information-intensive. Rational designs have been successfully applied to improve the thermal stability [154, 155] and oxidative stability [156, 157] of various proteins and to produce proteins with altered substrate specificities [158-160] and creating orthogonal ligand-receptor pairs [161]. Rational design of orthogonal specificity is especially significant because it can provide a general means for engineering biological macromolecules for useful chemical and biochemical reactions *in vitro* and *in vivo*. However, specificity is a challenging property to engineer predictably because it is determined by a complex and poorly understood mix of chemical binding forces that includes electrostatic interactions, steric and hydrophobic effects. So far, only a few successful orthogonal combinations have been developed based on principles of molecular recognition, as opposed to random screening. Multiple libraries were designed by various groups based on the hypothesis that binding affinity arises from hydrophobic contacts and that specificity arises from binding pocket size, shape, hydrogen bonding, and electrostatics. The targeted residues to be exchanged were chosen based on their proximity to the bound ligand as observed in the crystal structure [161-163].

The rational design approach that we followed was based on so called “bump-hole” approach used previously for creating orthogonal ligand receptor pairs [135]. We have identified two amino acids (M388 and L428) within 6Å, which might have interaction with the ‘A/B’ ring interface in E2 and then generated additional space in the LBP by exchanging the M388 and L428 into alanine separately, or simultaneously. We used a relatively inactive ligand (CV6019) carrying a C5 -methyl substitution in the ‘A’ ring where the concept was this additional CH₃ group could fill in and promote interaction. However, the mutants that we created through this approach were completely inactive with our ligand CV6019. M388A mutant bound E2 similar to that of WT LBD, whereas L428A completely lost interaction and was silent. When we combined these two mutations, the double mutant could bind E2 and had up to 60% relative activity. From

these results, we postulate that the exchange of M388 into alanine has no (or little) effect on E2 binding, whereas exchange of L428 into A made the protein completely inactive. Thus, the L at residue 428 provides a structural scaffold which is essential for correct tertiary structure and functional interaction with ligands. Residue 388 showed relatively higher tolerance and when these two residues were exchanged together into A, residue 388 could provide structural support for the entire structure and the double mutant could bind the natural ligand E2, although with weakened affinity.

4.5 Irrational design

Once we failed to generate suitable ligand-receptor combination through rational design approach, we contemplated manipulating the entire LBD through the so called “irrational design” or random mutagenesis approach. Because we had the one hybrid based test system providing chemical complementation through selection of growth on plate in presence of ligand and eventually screening of positive variants from the plate through β -galactosidase assay, we created a relatively small library through error-prone PCR. Error-prone PCR enabled us to control the amino acid exchange rate to ~ 1 residue/ kilo base pair of template DNA. Although the strain PJ69 4A had two reporter gene for growth based selection in the transactivation plate (*ADE2* and *HIS3*), this selection was somehow leaky as several clones appeared on the transactivation plates which later on characterized proved to be false-positive. While we used varying concentrations of 3-AT upto 50mM, we could still detect false positives in our candidate clones. One explanation could be the transactivation property of our chimeric construct was relatively strong in this strain PJ69 4A due to presence of VP16 AD coupled with Gal4 DBD and hER α LBD into one single construct. An alternative explanation might be the emergence of constitutive mutants through random mutagenesis which outgrew the true mutant in the transactivation plate where the true mutant might not grew in similar or better rate like the constitutive ones.

The V400A mutant had ~ 4 and 15-fold higher activity with CV6019 at 1 and 10 μ M concentrations, respectively compared to the WT LBD in our yeast one hybrid system (see Figure 3.10). Molecular modeling demonstrated that the residue 400 was situated in the surface of the LBD, around 13Å away from the docked C5 DES in crystal structure of

hER α LBD (Figure 3.11). Thus, this residue is not part of the LBP and not directly involved in the ligand interaction. Interestingly, this residue lies in such a position which was previously mentioned to be interacting with the chaperon hsp90 [73]. Hsp chaperons were implicated in the steroid receptor function where ligand-free (apo) receptors are anchored in the cytoplasm with the chaperon molecules and released from it after the ligand binding and subsequent conformational change [79, 164, 165].

A potential role of hsp90 *in vivo* in controlling ligand-inducible transactivation by the glucocorticoid receptor (GR) has been well supported by genetic studies in *S. cerevisiae* [166]. Reduced expression of the hsp90 gene strongly inhibited GR-dependent transactivation, suggesting that the hsp90 stabilized the ligand-free GR. However, the ER was less affected in similar experiments [166], suggesting that the hsp90 may not be necessary for regulating ligand-inducible transcription by ER. Moreover, several immunocytochemical studies have suggested that the hormone-free ER is at least partially nuclear [56, 167-170]. Lee *et al.* suggested that the mechanism of action of ER is intermediate between that of GR subfamily of steroid receptor and those of the thyroid hormone/retinoid/vitamin D₃ nuclear receptors [171]. The apo-GR is cytoplasmic, and ligand binding leads to its translocation to the nucleus [172]. In contrast, the thyroid hormone and related receptors do not interact with hsp90 *in vitro* [173] and are nuclear in the absence of hormone [174].

The first human ER α cDNA isolated [55] contains, within its ligand binding domain, a point mutation relative to ER that results in a G to V substitution at residue 400 (G400V). The two variants with G and V at residue 400 are termed HEGO and HEO, respectively. It has been [175] revealed that the HEO is hsp90-bound in absence of ligand and upon ligand binding, is dimerized, transported into the nucleus and transactivates the target gene, which is very much similar to that of glucocorticoid receptor. On the other hand, HEGO is relatively stable without any hsp interaction and without ligand binding, could be found at least partially nuclear located in dimer form but depends on ligand binding for transactivation.

While V400A showed ~4-fold improved activity with CV6019 at 10 μ M concentration compared to WT LBD, this improvement was not specific to a particular ligand. As shown in Figure 3.12, several other ligands exhibited improved activity with this mutant, suggesting that the mutant has elaborate specificity for ligand responsiveness. The increased activity found in the mutant was strictly ligand-dependent, although V400A mutant showed relatively increased background with the solvent. Aumais *et al.* [175] observed a similar phenomenon with HEO (V400) and HEGO (G400) variants where HEGO showed increased transactivation of reporter gene in their chimeric ER (VP16-GAL-ER) only in presence of E2.

When we made a competitive *in vitro* binding assay with WT LBD (V400) and the mutant V400A to determine the affinity towards E2, we could not find any significant difference among these two variants. The half-saturation value (EC_{50}) of WT LBD obtained from our binding assay was 10-fold higher (or 10-fold weakened affinity) than that of the reported values for WT hER α . The half-saturation value for full length hER α is in the range of 0.5nM which is obtained by *in vivo* experiments [176]. The difference might have been arisen due to the difference in test system as we have used the purified receptor and made an *in vitro* binding assay (competitive *in vitro* binding assay based on enzyme fragmentation complementation). Another reason for the difference might be due to the length of receptor construct as we used the construct 6XHis-LBD (aa residue 302-553) instead of the full length hER α . Nevertheless, we could include the purified V400A mutant along with the WT LBD in the competitive binding assay to find out if there was any improvement in the binding of CV6019 in V400A compared to WT LBD. The EC_{50} values for V400A and WT LBD were similar and thus there were no improvement in terms of ligand binding for CV6019 (Table 3.2). Thus the increased activity found in the mutant V400A with CV6019 was not due to the better ligand-receptor interaction and better affinity.

We postulate that the V400A mutant that we obtained is similar to the HEGO variant G400, because of the similar structure and hydrophobic charge of glycine and alanine. Thus, where as the V400 is relatively unstable and needs either hsp90 or ligand binding for stability, V400A is more stable and is more nuclear even in the absence of ligand. Due

to its partial nuclear presence, ligands with minimal activity in HEO are sufficient to mount increased response in this mutant.

4.6 Directed evolution

We have used a systematic directed evolutionary approach to develop an orthogonal ligand-receptor pair based on synthetic estrogen analog and engineered estrogen receptor LBD. Manipulation for engineering the LBD was achieved through saturation and random mutagenesis which were coupled with a sensitive, high throughput selection based on chemical complementation in yeast two hybrid system. The ligand induced interaction partner were the hER α LBD and the human SRC-1 as a coactivator. Since we followed the protein engineering approach to match an orthogonal combination of ligand-receptor pair rather than chemical synthesis, our starting point was to find a suitable ligand having minimum activity in the WT receptor. In that end, we have tested 65 ligands with varying structure and identified CV3320 as the suitable one to start with. It should be noted that the strategy shown in this work and by some other authors [107] can be used as a general approach to generate alternative combination of ligand-receptor pair based on a different structure with similar or different starting protein molecule.

Once the ligand has been chosen, a systematic approach was followed where molecular simulation studies were made to identify 21 residues involved in the interaction, and 18 residues were identified for manipulation among these 21. These residues were subsequently altered in each position into 20 possible amino acid variants using the saturation mutagenesis. So, the first round of saturation mutagenesis would result in 18 different libraries in yeast two hybrid system. Each library was screened in transactivation media in 96-well plates where ligand induced growth enhancement and E2 induced reduction of growth were the criteria for selection of mutant candidates. Once a candidate was found, it was characterized in dose response curve to find out the improvement in CV3320 affinity and decrease of E2 affinity. Such a mutant would be the template for the second round of saturation mutagenesis and the whole processes were repeated holding the most suitable exchange of residue and manipulating in the rest amino acid positions. Once such a point was reached where no improvement could be made anymore, random

mutagenesis strategy over the whole LBD was followed repeatedly until no further improvement could be achieved.

In this work, we have performed 3 rounds of saturation and 2 rounds of random mutagenesis to create our desired ligand-receptor pair. At the end of first two rounds of saturation mutagenesis, we could obtain a fold improvement of 1.6×10^4 in combination of CV3320-L384F L387M LBD. When we made another round of saturation mutagenesis (the third round), we could not shift the fold improvement in any positive direction meaning we had reached a point of saturation. Although it was reported that more manipulation through saturation mutagenesis and fold improvement could be achieved [107], it was also reported that a very restricted number of mutations could be tolerated in the hER α LBP to maintain a productive interaction [135]. Residues distant to the LBP are also thought to be important in providing suitable tertiary structure that can bind the ligand with improved affinity. The mutant found from saturation mutagenesis was thus used as template for random mutagenesis over the whole LBD and after two rounds of randomization, we could manipulate the fold change to 1.6×10^3 . Finally, when we tested the mutant in our yeast one hybrid system, the ligand-mutant receptor pair showed around 50-fold increased activity than the WT receptor. A second round mutagenesis produced similar result whereas a third round of random mutagenesis was performed but the fold change could not be improved anymore revealing the saturation. Although the fold change is 10-fold lower than the result obtained from two rounds of saturation mutagenesis, the affinity towards CV3320 could further be improved 10-fold through the random mutagenesis (random mutagenesis $EC_{50} 5 \times 10^{-8}$ compared to $EC_{50} 5 \times 10^{-7}$ of saturation mutagenesis, also see the table 3.5). We considered the improvement of affinity towards CV3320 to be more important than that of reducing affinity towards E2, and thus we focused on the mutant found and characterized it further. Because one of the major long term goals of this work is to develop an inducible system for plant, we were less concerned to increase the affinity of our ligand and the affinity of E2 could be less emphasized due to the absence of E2 or phytoestrogens in several crop species (like potato). Whereas the developed orthogonal ligand-receptor pair is tightly regulated and well-induced in absence and presence of the ligand CV3320 respectively, the system

needs more assessment before application in plants, specially plants with high phytoestrogen levels like soybean and grape.

4.7 Evolution of novel ligand-receptor combination through systematic protein engineering

A systematic directed evolution approach using saturation mutagenesis would provide advantages like universal applicability for creating ligand-receptor pair, a relatively small library to screen, very good sensitivity in two hybrid system and cumulative effect of exchanges in subsequent rounds of mutagenesis [107]. Additionally, and more importantly, such an approach provides a bi-directional selection in terms of improvement for desirable characteristic and weakened response in terms of undesirable attributes. Such tandem positive and negative selection led us to a combination of ligand-receptor pair where we could improve the affinity of the mutant towards CV3320 and weaken the response to E2, thus making the ligand-receptor combination true orthogonal. Another advantage of this approach was that it could be coupled to other engineering approaches to achieve greater positive attributes desirable, thus leaving the further potential open for improvement. In our case, we had combined a random mutagenesis approach (error-prone PCR) once we arrived to a point when no further improvement could be made using saturation mutagenesis alone. We expect that other mutagenesis methods could be used to introduce amino acid substitutions otherwise inaccessible by error-prone PCR, so as to further enhance potency of response toward the ligand CV3320.

4.8 Functional mutations identified in the evolved hER α variants: L384F L387M Y537S

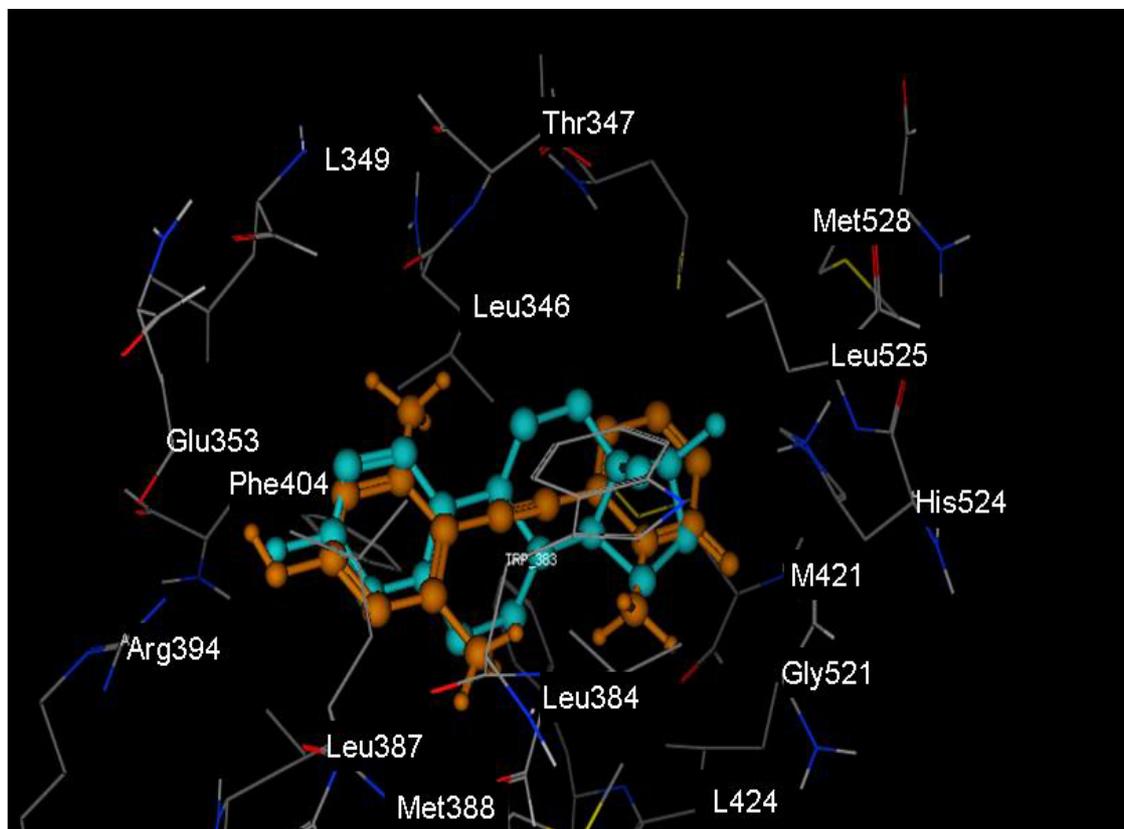


Figure 4.2: Superpositioning of ligand CV3320 (yellow) on E2 (magenta)-docked hER α LBD, showing the C5 methyl substitution in the 'A' ring of CV3320.

When we mapped the mutant receptor found after saturation and random mutagenesis screen, the roles played by the individual mutations in determining selectivity towards the target ligand were not immediately clear. Figure 4.2 and 4.3 show the LBP of WT and mutant receptor docked with CV3320 using DES-bound hER α LBD crystal structure. The energy minimized molecular simulation and cursory analysis revealed the potential role played by L384F, L387M and Y537S in the CV3320-specific receptor modeling. The individual exchanges are described below:

4.8.1 L384F

An analysis of the LBPs from ER α and ER β [each composed of 23 amino acids in direct vicinity (4Å) of E2] revealed two amino acid differences: ER α L384/ ER β M336 and ER α M421/ER β I373. There are ligands which are recognized differentially by two ER subtypes called selective estrogen receptor modulators (SERMs). While designing ER β -selective SERM, Hillisch *et al.* [177] demonstrated the importance of L384 in forming LBP and recognition of the ligand. Site directed mutagenesis and transactivation studies suggested that difference conformation of LBPs of ER α and ER β are due to the difference in ER α L384/ ER β M336. This difference plays important role for the high isotype selectivity of diarylpropionitrile (DPN) [178]. However, it has been shown that single point mutations of those amino acids differing between ER α and ER β alone could not account for the ER α selectivity of propyl pyrazole triol (PPT). Stauffer *et al.* [179] suggested additional parameters might be needed for selective binding of a ligand. Whereas leucine has a large aliphatic side chain, phenylalanine possesses even larger aromatic ring providing greater hydrophobicity. The aromatic ring of F contains delocalized pi (π) electron clouds that enable it to interact with other π -systems and to transfer the electrons. As shown in Figure 4.3, exchange of L384 to F led to the positioning of benzyl ring of F in close vicinity to 5'-methyl substitution of the 'A' ring of the ligand CV3320. This could explain the ten-fold improvement of EC₅₀ value of CV3320 in the mutant compared to the WT LBD. Additionally, the specific interaction that we observed in the mutant with CV3320 but not with E2 might be a contribution of van der Waals bonding. The relative proximity of F384 and the methyl group in CV3320 at C5 position (around 2Å) created a transient asymmetry in the electronic charge, which in turn created a steric complementarity to played role in specific interaction. On the other hand, due to the absence of any methyl group in 'A' ring, E2 could not form any interaction with F384.

4.8.2 L387M

The long sulfur-containing chain in methionine residue at 387 forms a parallel structure towards F384 only if the CV3320 is occupying the LBP. When E2 is occupying the LBP, it might produce congestion in the LBP due to the space occupation of B and C ring of

4.8.3 Y537S

It has been demonstrated previously that, rather remarkably, some single point mutations in the nuclear hormone receptors can mimic the effect of the binding of agonist ligands, resulting in receptors that have partial or full constitutive transcriptional activity [132, 133, 181, 182]. These constitutively active mutant receptors appear to have conformations that even when unoccupied by ligands, resemble that of the liganded, transcriptionally active receptor in terms of sensitivity to proteases and the binding of coactivator proteins [132, 183]. Nevertheless, the constitutively active receptors still bind ligands with high affinity, with estradiol maintaining the high transcriptional activity seen in their ligand-unoccupied state and antiestrogens blocking their constitutive and estradiol-occupied transcriptional activity [132, 133, 183]. Amino acid 537 is such a residue identified in the hER α LBD leading to constitutive activation when mutated to N, A and S [134, 181]. Tyrosine 537 of the hER is located in a flexible loop region at the base of the amphipathic R-helix (H12) of the ligand binding domain which undergoes extreme conformational changes in response to hormone binding [72, 73]. Y537S ER was shown to be fully constitutive i.e., the complete transcriptional activity in the absence or presence of estrogen [181, 184]. Observation with A, N, S, F and L mutants at residue 537 lead to proposal of a model (Figure 4.4) where the unoccupied WT receptor is staying as a collapsed pocket form (state I) and no transactivation is possible in this conformation. The close pocket conformation (state III) is transcriptionally active whether or not a ligand is occupying to the receptor. A receptor has to go through the open pocket conformation (state II) to bind and later on dissociate from the ligand. Y537S and other constitutive mutants can take the close pocket conformation in absence of ligand and thus are constitutively active [184, 185]. When tested in our one hybrid system, the Y537S mutant separately in presence of ligand or solvent alone, we could see a decrease of activity with E2 (30% relative activity) as shown in Figure 3.22. The Y537S mutant showed an increase in the activity with CV3320 and the solvent, which was almost at the background level. When we tested this mutant in the two hybrid system, we could not find any significant increase in the activity with solvent alone. The transactivation in presence of E2 was significantly decreased (Figure 3.23). We postulate thus the difference found with the published literature might have arisen from the difference in the test system as well as from the construct.

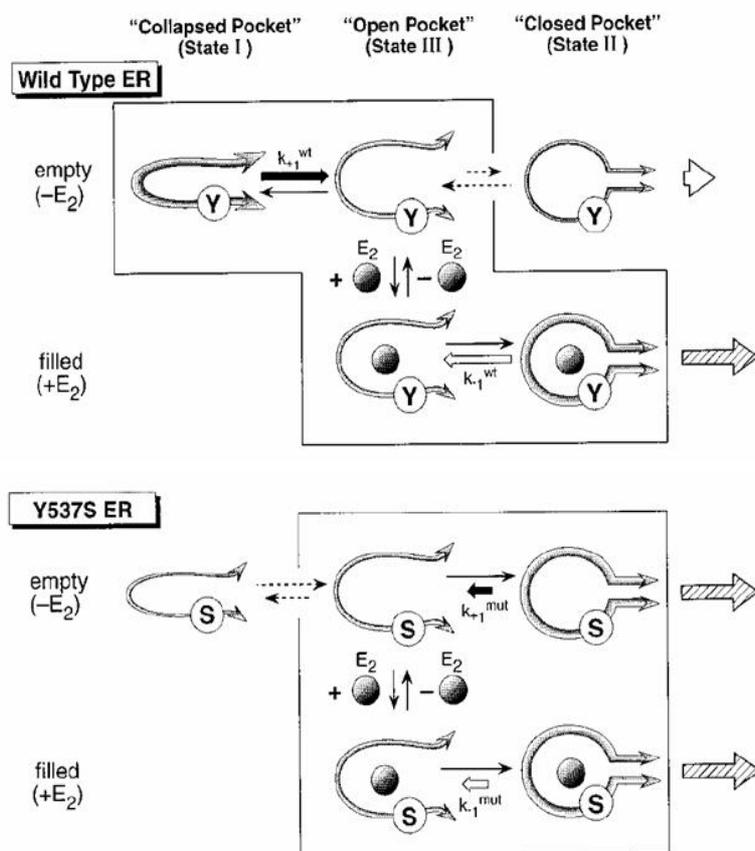


Figure 4.4: Proposed model of conformational states, ligand binding and transcriptional activation in WT and Y537S mutant ER LBD. Conformational reorganization involved in ligand association and dissociation in wild-type ER (top) and Y537S ER (bottom). The residue at position 537 (Y or S) is shown in the small circle. The magnitude of receptor transcriptional activity is illustrated by the large shaded arrow at the right side of the figures. The ligand (E2) is illustrated by a shaded circle [184].

4.8.4 Additivity of the functional mutations

Since recombining the two rounds of saturation mutagenesis and random mutagenesis variants yield a variant with further improved affinity to CV3320, we were interested to know whether any of these exchanges contributed to the beneficial attribute alone or this was a cumulative effect. Thus, we separated mutants from different mutagenesis approach and tested them in our one and two hybrid based test system. The double mutant generated from the saturation mutagenesis approach (L384F L387M) or the Y537S from random mutagenesis failed to show the increased activity shown by the combined triple mutant in β -galactosidase assay (Figure 3.22). When we compared the Y537S mutant in

the two hybrid system, the improved affinity shown by the triple mutant towards CV3320 could not be found (Figure 3.23). Thus, these three exchanges did not add independently the attribute of high affinity towards CV3320 but they were rather cumulative in function. This phenomenon was also observed in hER α LBD engineered for enhanced androgen specificity and affinity [92]. We postulate that the triple exchanges contribute to the suitable tertiary structure of the LBP where CV3320 could interact better with the residues involved in LBP formation through subtle modulation of the van der Waals interactions and probably hydrogen bonds between the receptor and ligand by many residues far away from the ligand binding pocket and thus showed improved binding. Recently, such long range interactions were also demonstrated to be important in determining the ligand selectivity of human estrogen receptor subtypes [92, 186]. Due to the subtlety and complexity of these long-range interactions, it is not surprising that some independently obtained functional mutations are not individually role-playing in their effect on ligand specificity and affinity.

We could show from our “irrational design” that the mutant identified (V400A) did not have any improvement in ligand binding; instead the increased activity that we have seen could be due to the weakened interaction with hsp90 which led to the partial dimerization and localization of receptor in the nucleus even in absence of ligand. We postulated that if the release from hsp90 is favoured due to the weakening of interaction, we could perhaps combine this phenomenon with the increased affinity that we found in the triple mutant L384F L387M Y537S. Theoretically, this combination would result in significant increase in transactivation due to the additive effect of chaperon release and better binding of the ligand with the receptor. However, when combined, we could not find any improvement in transactivation (Figure 3.25). This may be due to the non-additiveness of discrete properties the two independent mutant. Other reasons could be the point of saturation with transactivation, limitation in our test window, unfavourable conformation of the receptor which might lead to altered protein stability or turn over.

4.9 Orthogonal ligand-receptor pair with unique specificity is useful in controlling gene expression

Unique specificity regulatory systems are under active investigation not only as means to control the transcription of transgene [35, 187] but also to regulate enzyme activity [188-190] and to probe the function of certain enzymes [189]. There are a number of classical approaches to the regulation of gene expression with ligands [191, 192]. There are a number of well documented cases where nuclear receptor mutants were found that had useful shift in hormone specificity [193-195]. A potentially serious problem encountered with most of these systems is that the ligand being used to regulate the mutant receptor still has biological activity on endogenous receptors. Thus, the activating ligands in those systems have not been sufficiently mismatched with their endogenous receptor for the specificity of the system to be unique.

Rationally designed mutagenesis in retinoid X receptor (RXR) system led to significant changes in hormone-receptor specificity, such that the mutants were more responsive synthetic ligands than to the natural ligand *9-cis* retinoic acid [38]. In a number of studies on ER, random mutagenesis and phenotypic selection have been used to generate mutant ERs that are responsive to a ligand that has very low potency on WT ER [130, 196]. At least in these examples, one is starting with a ligand that is well mismatched with the WT receptor, and ER mutant that show significantly enhanced responsiveness towards the novel ligand have been found in this fashion.

4.10 Outlook

Here we demonstrated a combined mutagenesis approach where rational design guided directed evolution approach could successfully create an orthogonal ligand-receptor pair with significant improvement of affinity. This orthogonal system has the potentiality to be tested in eukaryotic systems, especially in plants. However, to use the system in crop plants by physical ligand application (for example, spraying), the inducing concentration of the ligand should be in nanomolar range. Further improvement could also be made using this mutant as a start point. To get further hints for manipulations on the structure of the inducing ligand CV3320 and generation of rational or irrational compensatory

mutagenesis in the L384F L387M Y537S-mutant, a crystal structure of this mutant in complex with CV3320 might give valuable insights. This might lead to a further increase in the affinity towards the ligand or a decrease in the affinity towards E2. In parallel the existing mutant could be tested in transient plant cell-culture reporter gene systems to find the transactivation potential in planta. In addition the uptake of the ligand into the plant cell should be tested. While we have tested some of the ligands in general toxicity tests in plant system, further studies are needed in mammalian system to find out whether the ligand in question has any adverse effect or not and whether it is safe to use externally. Additional studies regarding chemical and pharmacological properties and possibly modification may be required to apply this system for ligand inducible transcriptional regulation.

5 Summary

To develop an inducible gene-switch based on a modified estrogen receptor and non-steroidal estradiol analogs, we developed an *in vivo* test system in yeast that allowed us to qualitatively and quantitatively assess the transactivation properties of the artificial ligands in WT and mutated receptors. Here compounds with halogen groups at position 5' showed significant transactivation activity in the WT receptor. In a **rational design approach** via site-directed mutagenesis, two putative interaction sites (M388 and L428) were exchanged against alanine to generate a cavity which we tried to fill/compensate with ligands with a bulky side-chain at position 5' like IK308, IK377 and IK519. We tested M388A, L428A and a combination (ML-AA). Unfortunately none of these compounds showed activity in any of the mutants screened. Interestingly the L428A-mutant lost the interaction even with E2 completely, whereas M388A was still able to interact with E2. The double mutant can bind E2, but with reduced affinity (60 % compared to WT).

Because the rational design strategy was not successful in delivering a suitable ligand-receptor combination we started a **random mutagenesis approach** over the entire ER-LBD with CV6019 as the screening ligand in a one-hybrid based assay. This approach led to the identification of a mutant (V400A), which showed approximately 4- and 15-fold enhancement in activity at concentrations of 10 and 1 μ M CV6019, respectively, compared to WT. Another ligand, IK423, even showed 10-fold better activity compared to WT at 10 μ M concentration. Because V400 is not constituting the LBP and known to be involved in hsp interaction, we performed an *in vitro* binding test to analyze if the enhanced activity of the V400A mutant towards the ligands compared to WT-receptor is really due to an enhanced binding, or to a weakened interaction with hsp, which might lead to increased transactivation even upon binding of weak ligands. The EC₅₀ values for binding of E2 in WT and V400A as well as for binding of CV6019 in WT and V400A were similar, which led to the conclusion that ligand-binding was not improved in this mutant but rather the interaction with hsp weakened.

Because random mutagenesis over the entire LBD did not lead to mutants with increased binding activity, we continued with a systematic **directed evolution approach**. Here saturation mutagenesis libraries for 18 functionally important ligand-contacting residues were screened for altered transcriptional activation in a sensitive two-hybrid based *in vivo* assay with CV3320 as the target ligand. In the first round of saturation mutagenesis one mutant (L384F) showed improved affinity towards CV3320 and weakened response towards E2 ($1,6 \times 10^3$ -fold improvement). The 2nd round of saturation mutagenesis with L384F as a template led to the identification of the mutant L384F L387M. This mutant did not show an improvement in the affinity towards CV3320 but the affinity towards E2 was 10-fold reduced, so that the overall fold improvement was 10-fold increased ($1,6 \times 10^4$ -fold improvement). The 3rd round of saturation mutagenesis with L384F L387M as template did not lead to a new mutant with improved affinity, which indicated that the level of saturation was achieved.

Because it is documented that residues outside the LBP could also influence the binding and affinity through secondary structure rearrangements, we superimposed a random mutagenesis over the entire LBD to the L384F L387M mutant. One mutant (L384F L387M Y537S) showed a further 10-fold increase in affinity towards CV3320, but also a 100-fold increased affinity towards E2 compared to the L384F/L387M mutant, leading to an overall fold improvement of $1,6 \times 10^3$. When the Y537S mutation was analyzed separately, it showed no increase in activity or affinity towards CV3320.

Analysis of a **combination** of the V400A-mutation found in the first random mutagenesis approach, which led to a weakened chaperon interaction, with the L384F L387M Y537S-mutant did not lead to a further improvement in transactivation compared to the corresponding single mutants. We hypothesized that a combination of reduced anchoring (due to the weakened hsp interaction) and increased affinity (found in the L384F L387M Y537S mutant) would lead to a cooperative effect and significant improvement in ligand-induced transactivation. We conclude that these two different mechanisms were non-cooperative in our chimeric constructs and test system.

6 Bibliography

1. Gatz, C.a.L., I (1998). Promoters that respond to chemical inducers. *Trends Plant Sci* 3, 352.
2. Aoyama, T. (1999). Glucocorticoid-inducible gene expression in plants. In *Inducible gene expression in plants*, P. Reynolds, ed. (Wallingford: CAB International), pp. 44-59.
3. Zuo, J., and Chua, N.H. (2000). Chemical-inducible systems for regulated expression of plant genes. *Curr Opin Biotechnol* 11, 146-151.
4. Guntas, G., Mitchell, S.F., and Ostermeier, M. (2004). A molecular switch created by in vitro recombination of nonhomologous genes. *Chem Biol* 11, 1483-1487.
5. Winkler, W.C.a.B., R.R. (2003). Gene control by metabolite-binding riboswitches. *ChemBioChem* 4, 1024.
6. Winkler, W.N., A. and Breaker, R.R. (2002). Thiamine derivatives bind messenger RNAs directly to regulate bacterial gene expression. *Nature* 419, 952.
7. Suess, B., Fink, B., Berens, C., Stentz, R., and Hillen, W. (2004). A theophylline responsive riboswitch based on helix slipping controls gene expression in vivo. *Nucleic Acids Res* 32, 1610-1614.
8. Kozak, M. (2006). Rethinking some mechanisms invoked to explain translational regulation in eukaryotes. *Gene* 382, 1-11.
9. Buskirk, A.R., and Liu, D.R. (2005). Creating small-molecule-dependent switches to modulate biological functions. *Chem Biol* 12, 151-161.
10. Johnson, P.F.a.M., S.L. (1989). Eukaryotic transcriptional regulatory proteins. *Ann. Rev. Biochem.* 41, 799-839.
11. Brent, R., and Ptashne, M. (1985). A eukaryotic transcriptional activator bearing the DNA specificity of a prokaryotic repressor. *Cell* 43, 729-736.
12. Hope, I.A., and Struhl, K. (1986). Functional dissection of a eukaryotic transcriptional activator protein, GCN4 of yeast. *Cell* 46, 885-894.
13. Green, S., and Chambon, P. (1987). Oestradiol induction of a glucocorticoid-responsive gene by a chimaeric receptor. *Nature* 325, 75-78.
14. Ransone, L.J., Wamsley, P., Morley, K.L., and Verma, I.M. (1990). Domain swapping reveals the modular nature of Fos, Jun, and CREB proteins. *Mol Cell Biol* 10, 4565-4573.
15. Guarente, L. (1995). Transcriptional coactivators in yeast and beyond. *Trends Biochem Sci* 20, 517-521.
16. Kuo, C.J., Conley, P.B., Chen, L., Sladek, F.M., Darnell, J.E., Jr., and Crabtree, G.R. (1992). A transcriptional hierarchy involved in mammalian cell-type specification. *Nature* 355, 457-461.
17. Tronche, F., and Yaniv, M. (1992). HNF1, a homeoprotein member of the hepatic transcription regulatory network. *Bioessays* 14, 579-587.

18. Karin, M., and Hunter, T. (1995). Transcriptional control by protein phosphorylation: signal transmission from the cell surface to the nucleus. *Curr Biol* 5, 747-757.
19. Hunter, T., and Karin, M. (1992). The regulation of transcription by phosphorylation. *Cell* 70, 375-387.
20. Blank, V., Kourilsky, P., and Israel, A. (1992). NF-kappa B and related proteins: Rel/dorsal homologies meet ankyrin-like repeats. *Trends Biochem Sci* 17, 135-140.
21. Beg, A.A., and Baldwin, A.S., Jr. (1993). The I kappa B proteins: multifunctional regulators of Rel/NF-kappa B transcription factors. *Genes Dev* 7, 2064-2070.
22. Liou, H.C., and Baltimore, D. (1993). Regulation of the NF-kappa B/rel transcription factor and I kappa B inhibitor system. *Curr Opin Cell Biol* 5, 477-487.
23. Gronemeyer, H. (1992). Control of transcription activation by steroid hormone receptors. *Faseb J* 6, 2524-2529.
24. Gatz, C. (1997). Chemical control of gene expression. *Ann. Rev. Plant Phys. Plant Mol. Biol.* 48, 89.
25. Prandl, R., and Schoffl, F. (1996). Heat shock elements are involved in heat shock promoter activation during tobacco seed maturation. *Plant Mol Biol* 31, 157-162.
26. Kuhlemeier, C.G., P.J. and Chua, N.H. (1987). Regulation of gene expression in higher plant. *Ann. Rev. Plant Phys. Plant Mol. Biol.* 38, 221.
27. Wright, C.F., Hamer, D.H., and McKenney, K. (1988). Autoregulation of the yeast copper metallothionein gene depends on metal binding. *J Biol Chem* 263, 1570-1574.
28. Mett, V.L., Lochhead, L.P., and Reynolds, P.H. (1993). Copper-controllable gene expression system for whole plants. *Proc Natl Acad Sci U S A* 90, 4567-4571.
29. Gatz, C., Frohberg, C., and Wendenburg, R. (1992). Stringent repression and homogeneous de-repression by tetracycline of a modified CaMV 35S promoter in intact transgenic tobacco plants. *Plant J* 2, 397-404.
30. Gossen, M., and Bujard, H. (1992). Tight control of gene expression in mammalian cells by tetracycline-responsive promoters. *Proc Natl Acad Sci U S A* 89, 5547-5551.
31. Zuo, J., Niu, Q.W., and Chua, N.H. (2000). Technical advance: An estrogen receptor-based transactivator XVE mediates highly inducible gene expression in transgenic plants. *Plant J* 24, 265-273.
32. Bohner, S., Lenk, I.I., Rieping, M., Herold, M., and Gatz, C. (1999). Technical advance: transcriptional activator TGV mediates dexamethasone-inducible and tetracycline-inactivatable gene expression. *Plant J* 19, 87-95.
33. Padidam, M. (2003). Chemically regulated gene expression in plants. *Curr Opin Plant Biol* 6, 169-177.
34. Bishop, A., Buzko, O., Heyeck-Dumas, S., Jung, I., Kraybill, B., Liu, Y., Shah, K., Ulrich, S., Witucki, L., Yang, F., Zhang, C., and Shokat, K.M. (2000). Unnatural ligands for engineered proteins: new tools for chemical genetics. *Annu Rev Biophys Biomol Struct* 29, 577-606.

35. Belshaw, P.J., Schoepfer, J., Liu, K., Morrison, K and Schreiber, S.L. (1995). Rational design of orthogonal receptor-ligand combination. *Angew. Chem. Int. Ed. Eng.* *34*, 2129-2132.
36. Shah, K., Liu, Y., Deirmengian, C., and Shokat, K.M. (1997). Engineering unnatural nucleotide specificity for Rous sarcoma virus tyrosine kinase to uniquely label its direct substrates. *Proc Natl Acad Sci U S A* *94*, 3565-3570.
37. Nguyen, J.T., Porter, M., Amoui, M., Miller, W.T., Zuckermann, R.N., and Lim, W.A. (2000). Improving SH3 domain ligand selectivity using a non-natural scaffold. *Chem Biol* *7*, 463-473.
38. Peet, D.J., Doyle, D.F., Corey, D.R., and Mangelsdorf, D.J. (1998). Engineering novel specificities for ligand-activated transcription in the nuclear hormone receptor RXR. *Chem Biol* *5*, 13-21.
39. Koh, J.T., Putnam, M., Tomic-Canic, M. and McDaniel, C.M. (1999). Selective regulation of gene expression using rationally-modified retinoic acid receptor. *J Am Chem Soc* *121*, 1984-1985.
40. Shi, Y., and Koh, J.T. (2001). Selective regulation of gene expression by an orthogonal estrogen receptor-ligand pair created by polar-group exchange. *Chem Biol* *8*, 501-510.
41. Tedesco, R., Thomas, J.A., Katzenellenbogen, B.S., and Katzenellenbogen, J.A. (2001). The estrogen receptor: a structure-based approach to the design of new specific hormone-receptor combinations. *Chem Biol* *8*, 277-287.
42. Baniahmad, A., Leng, X., Burris, T.P., Tsai, S.Y., Tsai, M.J., and O'Malley, B.W. (1995). The tau 4 activation domain of the thyroid hormone receptor is required for release of a putative corepressor(s) necessary for transcriptional silencing. *Mol Cell Biol* *15*, 76-86.
43. Beato, M., Herrlich, P., and Schutz, G. (1995). Steroid hormone receptors: many actors in search of a plot. *Cell* *83*, 851-857.
44. Damm, K., Thompson, C.C., and Evans, R.M. (1989). Protein encoded by v-erbA functions as a thyroid-hormone receptor antagonist. *Nature* *339*, 593-597.
45. Kastner, P., Mark, M., and Chambon, P. (1995). Nonsteroid nuclear receptors: what are genetic studies telling us about their role in real life? *Cell* *83*, 859-869.
46. Mangelsdorf, D.J., Thummel, C., Beato, M., Herrlich, P., Schutz, G., Umesono, K., Blumberg, B., Kastner, P., Mark, M., Chambon, P., and Evans, R.M. (1995). The nuclear receptor superfamily: the second decade. *Cell* *83*, 835-839.
47. Sap, J., Munoz, A., Schmitt, J., Stunnenberg, H., and Vennstrom, B. (1989). Repression of transcription mediated at a thyroid hormone response element by the v-erb-A oncogene product. *Nature* *340*, 242-244.
48. Steinmetz, A.C., Renaud, J.P., and Moras, D. (2001). Binding of ligands and activation of transcription by nuclear receptors. *Annu Rev Biophys Biomol Struct* *30*, 329-359.
49. Rich, R.L., Hoth, L.R., Geoghegan, K.F., Brown, T.A., LeMotte, P.K., Simons, S.P., Hensley, P., and Myszka, D.G. (2002). Kinetic analysis of estrogen receptor/ligand interactions. *Proc Natl Acad Sci U S A* *99*, 8562-8567.
50. Yaghmai, R.a.C., G.R. (2002). Optimized regulation of gene expression using artificial transcription factor. *Molecular Therapy* *5*, 685-694.

51. Ansari, A.Z., and Mapp, A.K. (2002). Modular design of artificial transcription factors. *Curr Opin Chem Biol* 6, 765-772.
52. Jensen, E.V.a.J., H.I. (1962). Basic guides to the mechanism of estrogen action. *Recent Prog Horm res* 18, 387-414.
53. Jensen, E.V.a.D., E.R. (1973). Estrogen receptor interaction. *Science (Wash. DC)* 182, 126-134.
54. Walter, P., Green, S., Greene, G., Krust, A., Bornert, J.M., Jeltsch, J.M., Staub, A., Jensen, E., Scrace, G., Waterfield, M., and et al. (1985). Cloning of the human estrogen receptor cDNA. *Proc Natl Acad Sci U S A* 82, 7889-7893.
55. Green, S., Walter, P., Kumar, V., Krust, A., Bornert, J., Argos, P. and Chambon, P. (1986). Human oestrogen receptor cDNA: sequence, expression and homology to *v-erb-A*. *Nature* 320, 134-139.
56. Greene, G.L., Gilna, P., Waterfield, M., Baker, A., Hort, Y. and Shine, J. (1986). Sequence and expression of human estrogen receptor complementary DNA. *Science (Wash. DC)* 231, 1150-1154.
57. Krust, A., Green, S., Argos, P., Kumar, V., Walter, P., Bornet, J.M. and Chambon, P. (1986). The chicken oestrogen receptor sequence: Homology with *v-erbA* and the human oestrogen and glucocorticoid receptor. *EMBO J* 5, 891-897.
58. Macgregor, J.I.a.J., V.C. (1998). Basic guide to mechanism of antiestrogen action. *Pharmacological Reviews* 50, 151-196.
59. Nilsson, S., Mäkelä, S., Treuter, E., Tujague, M., Thomson, J., Andersson, G., Enmark, E., Petersson, K., Warner, M. and Gustafsson, J. (2001). Mechanism of estrogen action. *Physiol. Rev.* 81, 1535-1565.
60. Tora, L., White, J., Brou, C., Tasset, D., Webster, N., Scheer, E., and Chambon, P. (1989). The human estrogen receptor has two independent nonacidic transcriptional activation functions. *Cell* 59, 477-487.
61. Berry, M., Metzger, D., and Chambon, P. (1990). Role of the two activating domains of the oestrogen receptor in the cell-type and promoter-context dependent agonistic activity of the anti-oestrogen 4-hydroxytamoxifen. *Embo J* 9, 2811-2818.
62. Kraus, W.L., McInerney, E.M., and Katzenellenbogen, B.S. (1995). Ligand-dependent, transcriptionally productive association of the amino- and carboxyl-terminal regions of a steroid hormone nuclear receptor. *Proc Natl Acad Sci U S A* 92, 12314-12318.
63. McInerney, E.M., and Katzenellenbogen, B.S. (1996). Different regions in activation function-1 of the human estrogen receptor required for antiestrogen- and estradiol-dependent transcription activation. *J Biol Chem* 271, 24172-24178.
64. Schwabe, J.W., Neuhaus, D., and Rhodes, D. (1990). Solution structure of the DNA-binding domain of the oestrogen receptor. *Nature* 348, 458-461.
65. Schwabe, J.W., Chapman, L., Finch, J.T., and Rhodes, D. (1993). The crystal structure of the estrogen receptor DNA-binding domain bound to DNA: how receptors discriminate between their response elements. *Cell* 75, 567-578.
66. Schwabe, J.W., Chapman, L., and Rhodes, D. (1995). The oestrogen receptor recognizes an imperfectly palindromic response element through an alternative side-chain conformation. *Structure* 3, 201-213.

67. Aranda, A.a.P., A. (2001). Nuclear hormone receptors and gene expression. *Physiol. Rev.* *81*, 1269-1304.
68. Wurtz, J.M., Bourguet, W., Renaud, J.P., Vivat, V., Chambon, P., Moras, D., and Gronemeyer, H. (1996). A canonical structure for the ligand-binding domain of nuclear receptors. *Nat Struct Biol* *3*, 87-94.
69. Pike, A.C., Brzozowski, A.M., Hubbard, R.E., Bonn, T., Thorsell, A.G., Engstrom, O., Ljunggren, J., Gustafsson, J.A., and Carlquist, M. (1999). Structure of the ligand-binding domain of oestrogen receptor beta in the presence of a partial agonist and a full antagonist. *Embo J* *18*, 4608-4618.
70. Henttu, P.M., Kalkhoven, E., and Parker, M.G. (1997). AF-2 activity and recruitment of steroid receptor coactivator 1 to the estrogen receptor depend on a lysine residue conserved in nuclear receptors. *Mol Cell Biol* *17*, 1832-1839.
71. Feng, W., Ribeiro, R.C., Wagner, R.L., Nguyen, H., Apriletti, J.W., Fletterick, R.J., Baxter, J.D., Kushner, P.J., and West, B.L. (1998). Hormone-dependent coactivator binding to a hydrophobic cleft on nuclear receptors. *Science* *280*, 1747-1749.
72. Shiau, A.K., Barstad, D., Loria, P.M., Cheng, L., Kushner, P.J., Agard, D.A., and Greene, G.L. (1998). The structural basis of estrogen receptor/coactivator recognition and the antagonism of this interaction by tamoxifen. *Cell* *95*, 927-937.
73. Brzozowski, A.M., Pike, A.C., Dauter, Z., Hubbard, R.E., Bonn, T., Engstrom, O., Ohman, L., Greene, G.L., Gustafsson, J.A., and Carlquist, M. (1997). Molecular basis of agonism and antagonism in the oestrogen receptor. *Nature* *389*, 753-758.
74. Anstead, G.M., Carlson, K.E., and Katzenellenbogen, J.A. (1997). The estradiol pharmacophore: ligand structure-estrogen receptor binding affinity relationships and a model for the receptor binding site. *Steroids* *62*, 268-303.
75. Ruff, M., Gangloff, M., Wurtz, J.M., and Moras, D. (2000). Estrogen receptor transcription and transactivation: Structure-function relationship in DNA- and ligand-binding domains of estrogen receptors. *Breast Cancer Res* *2*, 353-359.
76. Bourguet, W., Germain, P. and Gronemeyer, H. (2000). Nuclear receptor ligand-binding domains: three-dimensional structures, molecular interactions and pharmacological implications. *Trends in Pharmacological Science* *21*, 381-388.
77. Mckenna, N.J., Lanz, R.B. and O'Malley B. W. (1999). Nuclear receptor coregulators: cellular and molecular biology. *Endocrine Reviews* *20*, 321.
78. Heery, D.M., Kalkhoven, E., Hoare, S., and Parker, M.G. (1997). A signature motif in transcriptional co-activators mediates binding to nuclear receptors. *Nature* *387*, 733-736.
79. Smith, D.F.a.T., D.O. (1993). Steroid receptors and their associated proteins. *Molecular Endocrinology* *7*, 4.
80. Petersson, K., Grandien, K., Kuiper, G.J.M. and Gustafsson, J. (1997). Mouse estrogen receptor β forms estrogen response element-binding heterodimers with estrogen receptor alpha. *Molecular Endocrinology* *11*, 1486-1496.
81. Fujimoto, N.a.K., B.S. (1994). Alteration in the agonist/antagonist balance of antiestrogens by activation of protein kinase A signaling pathway in breast cancer cells: antiestrogen selectivity and promoter dependence. *Molecular Endocrinology* *8*, 296-304.

82. Tsai, M.J., and O'Malley, B.W. (1994). Molecular mechanisms of action of steroid/thyroid receptor superfamily members. *Annu Rev Biochem* 63, 451-486.
83. Harvey, D.M., and Caskey, C.T. (1998). Inducible control of gene expression: prospects for gene therapy. *Curr Opin Chem Biol* 2, 512-518.
84. Fussenegger, M. (2001). The impact of mammalian gene regulation concepts on functional genomic research, metabolic engineering and advanced gene therapies. *Biotechnology Progress* 17, 1.
85. Shi, Y.a.K., J.T. (2002). Functionally orthogonal ligand-receptor pairs for the selective regulation of gene expression generated by manipulation of charged residues at the ligand-receptor interface of ER -alpha and -beta. *Journal of American Chemical Society* 124, 6921-6928.
86. Arnold, F.H. (1998). Design by directed evolution. *Accounts of Chemical Research* 31, 125.
87. Arnold, F.H. (2001). Combinatorial and computational challenges for biocatalyst design. *Nature* 409, 253-257.
88. Penning, T.M., and Jez, J.M. (2001). Enzyme redesign. *Chem Rev* 101, 3027-3046.
89. Schmidt-Dannert, C. (2001). Directed evolution of single proteins, metabolic pathways and viruses. *Biochemistry* 40, 12125-13136.
90. Brannigan, J.A., and Wilkinson, A.J. (2002). Protein engineering 20 years on. *Nat Rev Mol Cell Biol* 3, 964-970.
91. Yano, T., Oue, S., and Kagamiyama, H. (1998). Directed evolution of an aspartate aminotransferase with new substrate specificities. *Proc Natl Acad Sci U S A* 95, 5511-5515.
92. Chen, Z., Katzenellenbogen, B.S., Katzenellenbogen, J.A., and Zhao, H. (2004). Directed evolution of human estrogen receptor variants with significantly enhanced androgen specificity and affinity. *J Biol Chem* 279, 33855-33864.
93. Chen, R., Greer, A., and Dean, A.M. (1995). A highly active decarboxylating dehydrogenase with rationally inverted coenzyme specificity. *Proc Natl Acad Sci U S A* 92, 11666-11670.
94. Mace, J.E., and Agard, D.A. (1995). Kinetic and structural characterization of mutations of glycine 216 in alpha-lytic protease: a new target for engineering substrate specificity. *J Mol Biol* 254, 720-736.
95. Liu, D.R., Magliery, T.J., Pastrnak, M., and Schultz, P.G. (1997). Engineering a tRNA and aminoacyl-tRNA synthetase for the site-specific incorporation of unnatural amino acids into proteins in vivo. *Proc Natl Acad Sci U S A* 94, 10092-10097.
96. Lin-Goerke, J.L., Robbins, D.J., and Burczak, J.D. (1997). PCR-based random mutagenesis using manganese and reduced dNTP concentration. *Biotechniques* 23, 409-412.
97. Dower, W.J., Miller, J.F., and Ragsdale, C.W. (1988). High efficiency transformation of *E. coli* by high voltage electroporation. *Nucleic Acids Res* 16, 6127-6145.

98. Gietz, R.D., and Woods, R.A. (2002). Transformation of yeast by lithium acetate/single-stranded carrier DNA/polyethylene glycol method. *Methods Enzymol* *350*, 87-96.
99. Moore, J.C., and Arnold, F.H. (1996). Directed evolution of a para-nitrobenzyl esterase for aqueous-organic solvents. *Nat Biotechnol* *14*, 458-467.
100. Zacco, M., and Gherardi, E. (1999). The effect of high-frequency random mutagenesis on in vitro protein evolution: a study on TEM-1 beta-lactamase. *J Mol Biol* *285*, 775-783.
101. Drummond, D.A., Iverson, B.L., Georgiou, G., and Arnold, F.H. (2005). Why high-error-rate random mutagenesis libraries are enriched in functional and improved proteins. *J Mol Biol* *350*, 806-816.
102. Nixon, A.E., and Firestone, S.M. (2000). Rational and "irrational" design of proteins and their use in biotechnology. *IUBMB Life* *49*, 181-187.
103. Chica, R.A., Doucet, N., and Pelletier, J.N. (2005). Semi-rational approaches to engineering enzyme activity: combining the benefits of directed evolution and rational design. *Curr Opin Biotechnol* *16*, 378-384.
104. Studier, F.W., and Moffatt, B.A. (1986). Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *J Mol Biol* *189*, 113-130.
105. Hanahan, D. (1983). Studies on transformation of *Escherichia coli* with plasmids. *J Mol Biol* *166*, 557-580.
106. James, P., Halladay, J., and Craig, E.A. (1996). Genomic libraries and a host strain designed for highly efficient two-hybrid selection in yeast. *Genetics* *144*, 1425-1436.
107. Chockalingam, K., Chen, Z., Katzenellenbogen, J.A., and Zhao, H. (2005). Directed evolution of specific receptor-ligand pairs for use in the creation of gene switches. *Proc Natl Acad Sci U S A* *102*, 5691-5696.
108. Ding, X.F., Anderson, C.M., Ma, H., Hong, H., Uht, R.M., Kushner, P.J., and Stallcup, M.R. (1998). Nuclear receptor-binding sites of coactivators glucocorticoid receptor interacting protein 1 (GRIP1) and steroid receptor coactivator 1 (SRC-1): multiple motifs with different binding specificities. *Mol Endocrinol* *12*, 302-313.
109. Vock, C. (2005). *Synthese neuartiger Estron- und Estradiol-Analoga als Liganden für die Aktivierung von Maxi-K+-Kanälen und zur Induktion von Pflanzenpromotoren*, 1 Edition (Göttingen: Cuvillier Verlag).
110. Le Gouill, C., Parent, J.L., Rola-Pleszczynski, M., and Stankova, J. (1994). Analysis of recombinant plasmids by a modified alkaline lysis method. *Anal Biochem* *219*, 164.
111. Mullis, K.B., and Faloona, F.A. (1987). Specific synthesis of DNA in vitro via a polymerase-catalyzed chain reaction. *Methods Enzymol* *155*, 335-350.
112. Ho, S.N., Hunt, H.D., Horton, R.M., Pullen, J.K., and Pease, L.R. (1989). Site-directed mutagenesis by overlap extension using the polymerase chain reaction. *Gene* *77*, 51-59.
113. Sanger, F., Nicklen, S., and Coulson, A.R. (1977). DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci U S A* *74*, 5463-5467.

114. Dohmen, R.J., Strasser, A.W., Honer, C.B., and Hollenberg, C.P. (1991). An efficient transformation procedure enabling long-term storage of competent cells of various yeast genera. *Yeast* 7, 691-692.
115. Louvion, J.F., Havaux-Copf, B., and Picard, D. (1993). Fusion of GAL4-VP16 to a steroid-binding domain provides a tool for gratuitous induction of galactose-responsive genes in yeast. *Gene* 131, 129-134.
116. Schiermeyer, A., Thurow, C., and Gatz, C. (2003). Tobacco bZIP factor TGA10 is a novel member of the TGA family of transcription factors. *Plant Mol Biol* 51, 817-829.
117. Wrenn, C.K., and Katzenellenbogen, B.S. (1993). Structure-function analysis of the hormone binding domain of the human estrogen receptor by region-specific mutagenesis and phenotypic screening in yeast. *J Biol Chem* 268, 24089-24098.
118. Eiler, S., Gangloff, M., Duclaud, S., Moras, D., and Ruff, M. (2001). Overexpression, purification, and crystal structure of native ER alpha LBD. *Protein Expr Purif* 22, 165-173.
119. Bradford, M.M. (1976). A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal Biochem* 72, 248-254.
120. Laemmli, U.K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* 227, 680-685.
121. Kyhse-Andersen, J. (1984). Electrophoretic transfer of multiple gels: a simple apparatus without buffer tank for rapid transfer of proteins from polyacrylamide to nitrocellulose. *J Biochem Biophys Methods* 10, 203-209.
122. Thurow, C. (2001). Biochemische Aufreinigung, funktionelle Analyse und Identifikation eines Interaktionspartners des an der Salicylsäure-induzierbaren Genexpression beteiligten SARP-Komplexes aus Tabak, 1st Edition (Göttingen: Cuvillier Verlag).
123. Murashige, T.a., F (1962). A revised medium for rapid growth and bioassays with tobacco tissue culture. *Plant Physiol* 15, 473-497.
124. Halgren, E., Dale, A.M., Sereno, M.I., Tootell, R.B., Marinkovic, K., and Rosen, B.R. (1999). Location of human face-selective cortex with respect to retinotopic areas. *Hum Brain Mapp* 7, 29-37.
125. Hart, T.N., and Read, R.J. (1992). A multiple-start Monte Carlo docking method. *Proteins* 13, 206-222.
126. Cadwell, R.C.a.J., G.F. (1994). Mutagenic PCR. *PCR Methods and Applications* 3, S136-S140.
127. Bohlen, S.P., and Yamamoto, K.R. (1993). Isolation of Hsp90 mutants by screening for decreased steroid receptor function. *Proc Natl Acad Sci U S A* 90, 11424-11428.
128. Pratt, W.B., and Welsh, M.J. (1994). Chaperone functions of the heat shock proteins associated with steroid receptors. *Semin Cell Biol* 5, 83-93.
129. Pratt, W.B., and Toft, D.O. (1997). Steroid receptor interactions with heat shock protein and immunophilin chaperones. *Endocr Rev* 18, 306-360.
130. Miller, N.a.W., J (1998). Random mutagenesis of human estrogen receptor ligand binding domain identifies mutations that decrease sensitivity to estradiol and

- increase sensitivity to a diphenol indene-ol compound: basis for regulatable expression system. *J. Steroid Biochem. Mol. Biol.* *64*, 129-135.
131. Zhao, H.M., J.C.; Volkov, A.A. and Arnold, F.H. ed. (1999). *Methods for optimizing industrial enzymes by directed evolution*, 2nd Edition (Washington, DC: ASM Press).
 132. Weis, K., Dingwall, C., and Lamond, A.I. (1996). Characterization of the nuclear protein import mechanism using Ran mutants with altered nucleotide binding specificities. *Embo J* *15*, 7120-7128.
 133. White, R., Sjoberg, M., Kalkhoven, E., and Parker, M.G. (1997). Ligand-independent activation of the oestrogen receptor by mutation of a conserved tyrosine. *Embo J* *16*, 1427-1435.
 134. Yudt, M.R., Vorobjikina, D., Zhong, L., Skafar, D.F., Sasson, S., Gasiewicz, T.A., and Notides, A.C. (1999). Function of estrogen receptor tyrosine 537 in hormone binding, DNA binding, and transactivation. *Biochemistry* *38*, 14146-14156.
 135. Gallinari, P., Lahm, A., Koch, U., Paolini, C., Nardi, M.C., Roscilli, G., Kinzel, O., Fattori, D., Muraglia, E., Toniatti, C., Cortese, R., De Francesco, R., and Ciliberto, G. (2005). A functionally orthogonal estrogen receptor-based transcription switch specifically induced by a nonsteroid synthetic ligand. *Chem Biol* *12*, 883-893.
 136. Pike, A.C., Brzozowski, A.M., and Hubbard, R.E. (2000). A structural biologist's view of the oestrogen receptor. *J Steroid Biochem Mol Biol* *74*, 261-268.
 137. Sippl, W. (2000). Receptor-based 3D QSAR analysis of estrogen receptor ligands-merging the accuracy of receptor-based alignments with the computational efficiency of ligand-based methods. *J Comput Aided Mol Des* *14*, 559-572.
 138. (2003). OECD Guideline for Testing of Chemicals: Seedling Emergence and Seeding Growth Test. 1-19.
 139. Guarente, L. (1992). Messenger RNA transcription and its control in *Saccharomyces cerevisiae*. In *The molecular and cellular biology of yeast Saccharomyces: Gene expression*, E.W. Jones, Pringle, J.R. and Broach, J.R., ed. (New York: Cold Spring Harbor Laboratory Press), pp. 49-98.
 140. Metzger, D., White, J.H., and Chambon, P. (1988). The human oestrogen receptor functions in yeast. *Nature* *334*, 31-36.
 141. Schena, M., and Yamamoto, K.R. (1988). Mammalian glucocorticoid receptor derivatives enhance transcription in yeast. *Science* *241*, 965-967.
 142. Katzenellenbogen, B.S., Bhardwaj, B., Fang, H., Ince, B.A., Pakdel, F., Reese, J.C., Schodin, D., and Wrenn, C.K. (1993). Hormone binding and transcription activation by estrogen receptors: analyses using mammalian and yeast systems. *J Steroid Biochem Mol Biol* *47*, 39-48.
 143. Garabedian, M.J., and Yamamoto, K.R. (1992). Genetic dissection of the signaling domain of a mammalian steroid receptor in yeast. *Mol Biol Cell* *3*, 1245-1257.
 144. Tanenbaum, D.M., Wang, Y., Williams, S.P., and Sigler, P.B. (1998). Crystallographic comparison of the estrogen and progesterone receptor's ligand binding domains. *Proc Natl Acad Sci U S A* *95*, 5998-6003.

145. Arnold, S.F., Collins, B.M., Robinson, M.K., Guillette, L.J., Jr., and McLachlan, J.A. (1996). Differential interaction of natural and synthetic estrogens with extracellular binding proteins in a yeast estrogen screen. *Steroids* 61, 642-646.
146. Coldham, N.G., Dave, M., Sivapathasundaram, S., McDonnell, D.P., Connor, C., and Sauer, M.J. (1997). Evaluation of a recombinant yeast cell estrogen screening assay. *Environ Health Perspect* 105, 734-742.
147. Lin, H.N.a.C., V.W. (2002). Screening and selection methods for large-scale analysis of protein function. *Angewandte Chemie-International Edition* 41, 4403-4425.
148. Azizi, B., Chang, E.I., and Doyle, D.F. (2003). Chemical complementation: small-molecule-based genetic selection in yeast. *Biochem Biophys Res Commun* 306, 774-780.
149. Tasset, D., Tora, L., Fromental, C., Scheer, E., and Chambon, P. (1990). Distinct classes of transcriptional activating domains function by different mechanisms. *Cell* 62, 1177-1187.
150. Triezenberg, S.J., Kingsbury, R.C., and McKnight, S.L. (1988). Functional dissection of VP16, the trans-activator of herpes simplex virus immediate early gene expression. *Genes Dev* 2, 718-729.
151. Sadowski, I., Ma, J., Triezenberg, S., and Ptashne, M. (1988). GAL4-VP16 is an unusually potent transcriptional activator. *Nature* 335, 563-564.
152. Cousens, D.J., Greaves, R., Goding, C.R., and O'Hare, P. (1989). The C-terminal 79 amino acids of the herpes simplex virus regulatory protein, Vmw65, efficiently activate transcription in yeast and mammalian cells in chimeric DNA-binding proteins. *Embo J* 8, 2337-2342.
153. Braselmann, S., Graninger, P., and Busslinger, M. (1993). A selective transcriptional induction system for mammalian cells based on Gal4-estrogen receptor fusion proteins. *Proc Natl Acad Sci U S A* 90, 1657-1661.
154. Villafranca, J.E., Howell, E.E., Voet, D.H., Strobel, M.S., Ogden, R.C., Abelson, J.N., and Kraut, J. (1983). Directed mutagenesis of dihydrofolate reductase. *Science* 222, 782-788.
155. Perry, L.J., and Wetzel, R. (1984). Disulfide bond engineered into T4 lysozyme: stabilization of the protein toward thermal inactivation. *Science* 226, 555-557.
156. Courtney, B.C., Williams, K.C., and Schlager, J.J. (1995). A phage display vector with improved stability, applicability and ease of manipulation. *Gene* 165, 139-140.
157. Rosenberg, S., Barr, P.J., Najarian, R.C., and Hallewell, R.A. (1984). Synthesis in yeast of a functional oxidation-resistant mutant of human alpha-antitrypsin. *Nature* 312, 77-80.
158. Almo, S.C., Smith, D.L., Danishefsky, A.T., and Ringe, D. (1994). The structural basis for the altered substrate specificity of the R292D active site mutant of aspartate aminotransferase from *E. coli*. *Protein Eng* 7, 405-412.
159. diSioudi, B., Grimsley, J.K., Lai, K., and Wild, J.R. (1999). Modification of near active site residues in organophosphorus hydrolase reduces metal stoichiometry and alters substrate specificity. *Biochemistry* 38, 2866-2872.

160. Stuible, H.P., and Kombrink, E. (2001). Identification of the substrate specificity-conferring amino acid residues of 4-coumarate:coenzyme A ligase allows the rational design of mutant enzymes with new catalytic properties. *J Biol Chem* *276*, 26893-26897.
161. Schwimmer, L.J., Rohatgi, P., Azizi, B., Seley, K.L., and Doyle, D.F. (2004). Creation and discovery of ligand-receptor pairs for transcriptional control with small molecules. *Proc Natl Acad Sci U S A* *101*, 14707-14712.
162. Allegretto, E.A., McClurg, M.R., Lazarchik, S.B., Clemm, D.L., Kerner, S.A., Elgort, M.G., Boehm, M.F., White, S.K., Pike, J.W., and Heyman, R.A. (1993). Transactivation properties of retinoic acid and retinoid X receptors in mammalian cells and yeast. Correlation with hormone binding and effects of metabolism. *J Biol Chem* *268*, 26625-26633.
163. Zhao, Q., Chasse, S.A., Devarakonda, S., Sierk, M.L., Ahvazi, B., and Rastinejad, F. (2000). Structural basis of RXR-DNA interactions. *J Mol Biol* *296*, 509-520.
164. Chambrud, B., Berry, M., Redeuilh, G., Chambon, P., and Baulieu, E.E. (1990). Several regions of human estrogen receptor are involved in the formation of receptor-heat shock protein 90 complexes. *J Biol Chem* *265*, 20686-20691.
165. Landel, C.C., Kushner, P.J., and Greene, G.L. (1994). The interaction of human estrogen receptor with DNA is modulated by receptor-associated proteins. *Mol Endocrinol* *8*, 1407-1419.
166. Picard, D., Kumar, V., Chambon, P., and Yamamoto, K.R. (1990). Signal transduction by steroid hormones: nuclear localization is differentially regulated in estrogen and glucocorticoid receptors. *Cell Regul* *1*, 291-299.
167. Zhuang, Y., Katzenellenbogen, B.S., and Shapiro, D.J. (1995). Estrogen receptor mutants which do not bind 17 beta-estradiol dimerize and bind to the estrogen response element in vivo. *Mol Endocrinol* *9*, 457-466.
168. Tzukerman, M., Zhang, X.K., Hermann, T., Wills, K.N., Graupner, G., and Pfahl, M. (1990). The human estrogen receptor has transcriptional activator and repressor functions in the absence of ligand. *New Biol* *2*, 613-620.
169. Fuxe, K., Cintra, A., Agnati, L.F., Harfstrand, A., Wikstrom, A.C., Okret, S., Zoli, M., Miller, L.S., Greene, J.L., and Gustafsson, J.A. (1987). Studies on the cellular localization and distribution of glucocorticoid receptor and estrogen receptor immunoreactivity in the central nervous system of the rat and their relationship to the monoaminergic and peptidergic neurons of the brain. *J Steroid Biochem* *27*, 159-170.
170. Press, M.F., Holt, J.A., Herbst, A.L., and Greene, G.L. (1985). Immunocytochemical identification of estrogen receptor in ovarian carcinomas. Localization with monoclonal estrophilin antibodies compared with biochemical assays. *Lab Invest* *53*, 349-361.
171. Lee, H.S., Aumais, J., and White, J.H. (1996). Hormone-dependent transactivation by estrogen receptor chimeras that do not interact with hsp90. Evidence for transcriptional repressors. *J Biol Chem* *271*, 25727-25730.
172. Pratt, W.B. (1992). Control of steroid receptor function and cytoplasmic-nuclear transport by heat shock proteins. *Bioessays* *14*, 841-848.

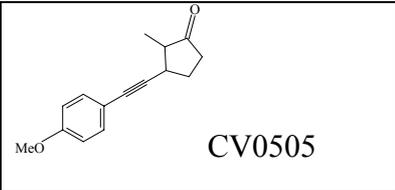
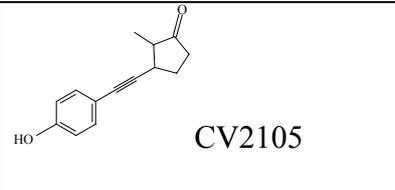
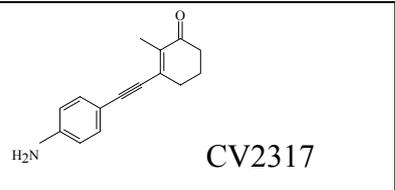
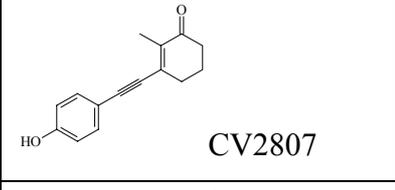
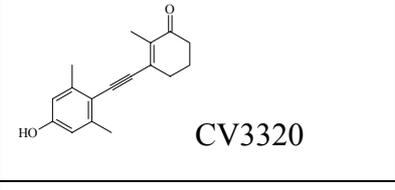
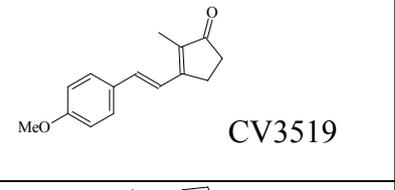
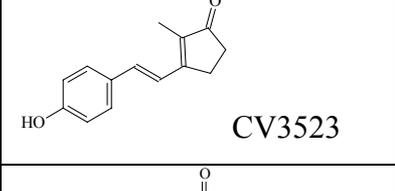
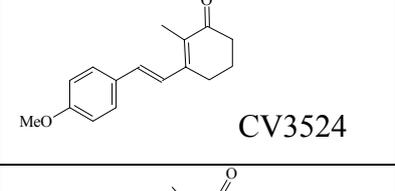
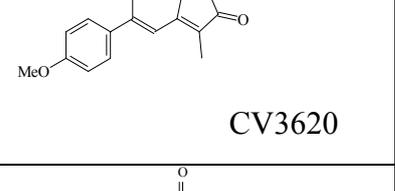
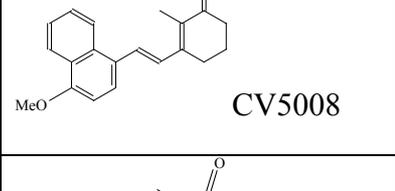
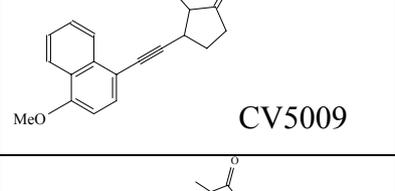
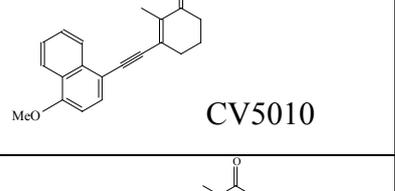
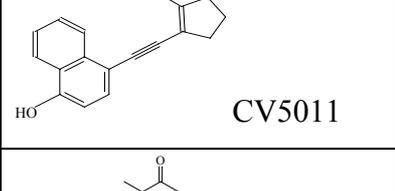
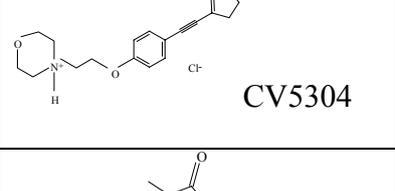
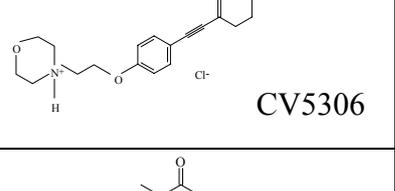
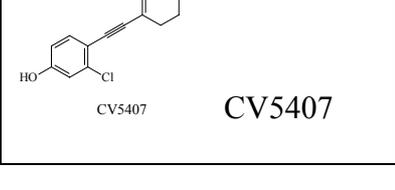
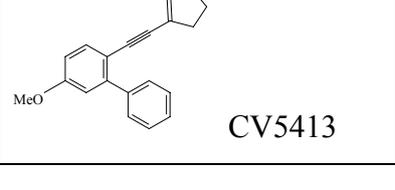
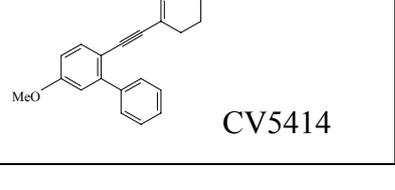
173. Dalman, F.C., Sturzenbecker, L.J., Levin, A.A., Lucas, D.A., Perdew, G.H., Petkovitch, M., Chambon, P., Grippo, J.F., and Pratt, W.B. (1991). Retinoic acid receptor belongs to a subclass of nuclear receptors that do not form "docking" complexes with hsp90. *Biochemistry* *30*, 5605-5608.
174. Yu, V.C., Delsert, C., Andersen, B., Holloway, J.M., Devary, O.V., Naar, A.M., Kim, S.Y., Boutin, J.M., Glass, C.K., and Rosenfeld, M.G. (1991). RXR beta: a coregulator that enhances binding of retinoic acid, thyroid hormone, and vitamin D receptors to their cognate response elements. *Cell* *67*, 1251-1266.
175. Aumais, J.P., Lee, H.S., Lin, R., and White, J.H. (1997). Selective interaction of hsp90 with an estrogen receptor ligand-binding domain containing a point mutation. *J Biol Chem* *272*, 12229-12235.
176. Palkowitz, A.D., Glasebrook, A.L., Thrasher, K.J., Hauser, K.L., Short, L.L., Phillips, D.L., Muehl, B.S., Sato, M., Shetler, P.K., Cullinan, G.J., Pell, T.R., and Bryant, H.U. (1997). Discovery and synthesis of [6-hydroxy-3-[4-[2-(1-piperidinyl)ethoxy]phenoxy]-2-(4-hydroxyphenyl)]b enzo[b]thiophene: a novel, highly potent, selective estrogen receptor modulator. *J Med Chem* *40*, 1407-1416.
177. Hillisch, A., Peters, O., Kosemund, D., Muller, G., Walter, A., Schneider, B., Reddersen, G., Elger, W., and Fritzemeier, K.H. (2004). Dissecting physiological roles of estrogen receptor alpha and beta with potent selective ligands from structure-based design. *Mol Endocrinol* *18*, 1599-1609.
178. Sun, J., Baudry, J., Katzenellenbogen, J.A., and Katzenellenbogen, B.S. (2003). Molecular basis for the subtype discrimination of the estrogen receptor-beta-selective ligand, diarylpropionitrile. *Mol Endocrinol* *17*, 247-258.
179. Stauffer, S.R., Coletta, C.J., Tedesco, R., Nishiguchi, G., Carlson, K., Sun, J., Katzenellenbogen, B.S., and Katzenellenbogen, J.A. (2000). Pyrazole ligands: structure-affinity/activity relationships and estrogen receptor-alpha-selective agonists. *J Med Chem* *43*, 4934-4947.
180. Zheng, Q.C., Li, Z.S., Xiao, J.F., Sun, M., Zhang, Y., and Sun, C.C. (2005). Homology modeling and PAPS ligand (cofactor) binding study of bovine phenol sulfotransferase. *J Mol Model (Online)* *11*, 97-104.
181. Zhang, Q.X., Borg, A., Wolf, D.M., Oesterreich, S., and Fuqua, S.A. (1997). An estrogen receptor mutant with strong hormone-independent activity from a metastatic breast cancer. *Cancer Res* *57*, 1244-1249.
182. Pakdel, F., Reese, J.C., and Katzenellenbogen, B.S. (1993). Identification of charged residues in an N-terminal portion of the hormone-binding domain of the human estrogen receptor important in transcriptional activity of the receptor. *Mol Endocrinol* *7*, 1408-1417.
183. Lazennec, G., Ediger, T.R., Petz, L.N., Nardulli, A.M., and Katzenellenbogen, B.S. (1997). Mechanistic aspects of estrogen receptor activation probed with constitutively active estrogen receptors: correlations with DNA and coregulator interactions and receptor conformational changes. *Mol Endocrinol* *11*, 1375-1386.
184. Carlson, K.E., Choi, I., Gee, A., Katzenellenbogen, B.S., and Katzenellenbogen, J.A. (1997). Altered ligand binding properties and enhanced stability of a constitutively active estrogen receptor: evidence that an open pocket conformation is required for ligand interaction. *Biochemistry* *36*, 14897-14905.

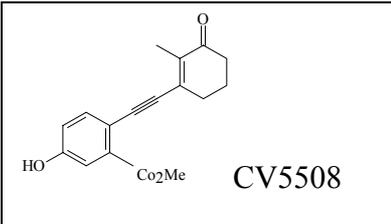
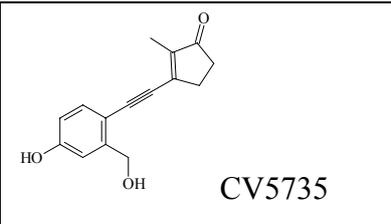
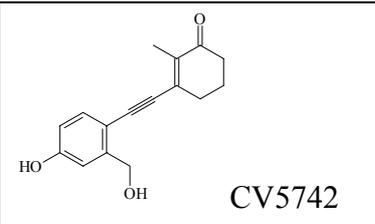
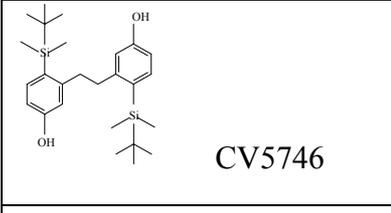
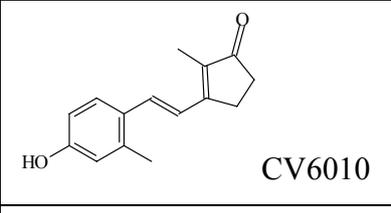
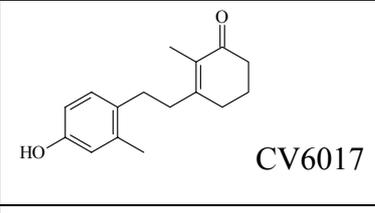
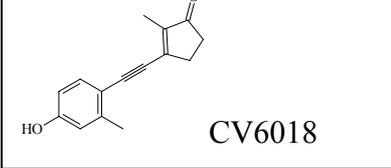
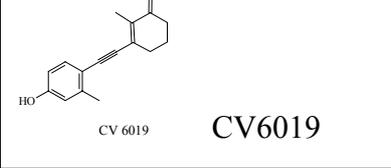
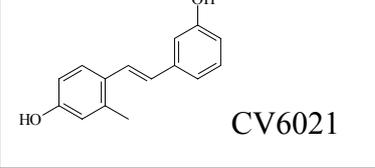
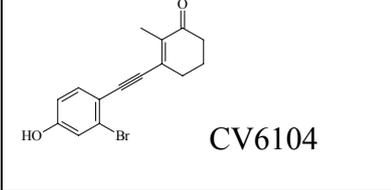
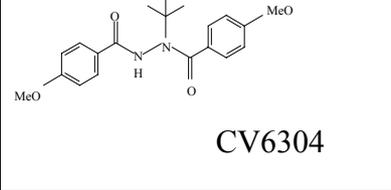
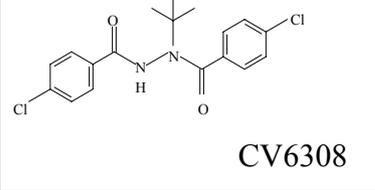
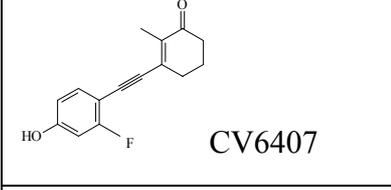
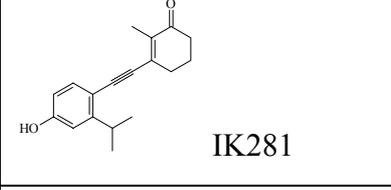
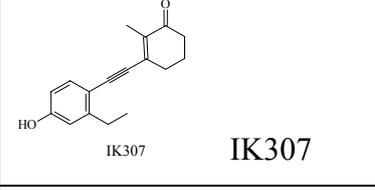
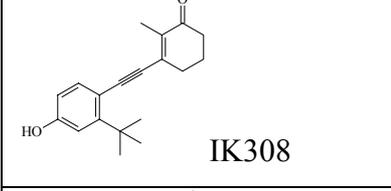
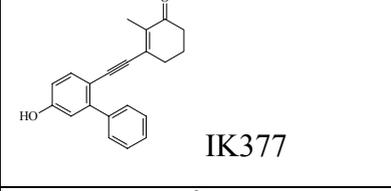
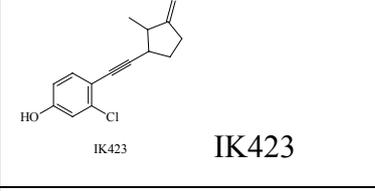
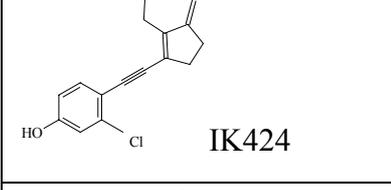
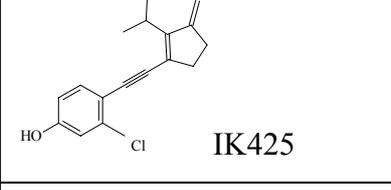
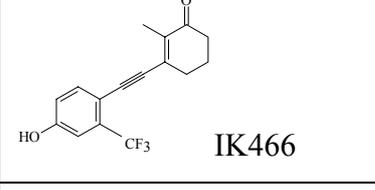
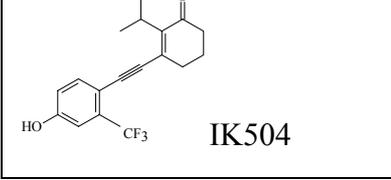
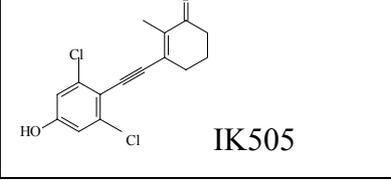
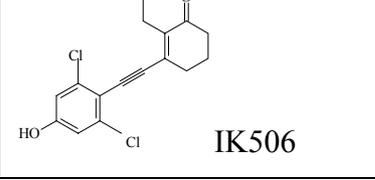
185. Zhong, L., and Skafar, D.F. (2002). Mutations of tyrosine 537 in the human estrogen receptor- α selectively alter the receptor's affinity for estradiol and the kinetics of the interaction. *Biochemistry* 41, 4209-4217.
186. Nettles, K.W., Sun, J., Radek, J.T., Sheng, S., Rodriguez, A.L., Katzenellenbogen, J.A., Katzenellenbogen, B.S., and Greene, G.L. (2004). Allosteric control of ligand selectivity between estrogen receptors α and β : implications for other nuclear receptors. *Mol Cell* 13, 317-327.
187. Doyle, J.L., DeSilva, U., Miller, W., and Green, E.D. (2000). Divergent human and mouse orthologs of a novel gene (WBSCR15/Wbscr15) reside within the genomic interval commonly deleted in Williams syndrome. *Cytogenet Cell Genet* 90, 285-290.
188. Feil, I.K., Hendle, J., and Schomburg, D. (1997). Modified substrate specificity of L-hydroxyisocaproate dehydrogenase derived from structure-based protein engineering. *Protein Eng* 10, 255-262.
189. Liu, Y., Shah, K., Yang, F., Witucki, L., and Shokat, K.M. (1998). Engineering Src family protein kinases with unnatural nucleotide specificity. *Chem Biol* 5, 91-101.
190. Bishop, S.M., Ross, J.B., and Kohanski, R.A. (1999). Autophosphorylation dependent destabilization of the insulin receptor kinase domain: tryptophan-1175 reports changes in the catalytic cleft. *Biochemistry* 38, 3079-3089.
191. No, D., Yao, T.P., and Evans, R.M. (1996). Ecdysone-inducible gene expression in mammalian cells and transgenic mice. *Proc Natl Acad Sci U S A* 93, 3346-3351.
192. Shockett, P.E., and Schatz, D.G. (1996). Diverse strategies for tetracycline-regulated inducible gene expression. *Proc Natl Acad Sci U S A* 93, 5173-5176.
193. Veldscholte, J., Berrevoets, C.A., Brinkmann, A.O., Grootegoed, J.A., and Mulder, E. (1992). Anti-androgens and the mutated androgen receptor of LNCaP cells: differential effects on binding affinity, heat-shock protein interaction, and transcription activation. *Biochemistry* 31, 2393-2399.
194. Mahfoudi, A., Roulet, E., Dauvois, S., Parker, M.G., and Wahli, W. (1995). Specific mutations in the estrogen receptor change the properties of antiestrogens to full agonists. *Proc Natl Acad Sci U S A* 92, 4206-4210.
195. Feil, R., Wagner, J., Metzger, D., and Chambon, P. (1997). Regulation of Cre recombinase activity by mutated estrogen receptor ligand-binding domains. *Biochem Biophys Res Commun* 237, 752-757.
196. Whelan, J., and Miller, N. (1996). Generation of estrogen receptor mutants with altered ligand specificity for use in establishing a regulatable gene expression system. *J Steroid Biochem Mol Biol* 58, 3-12.

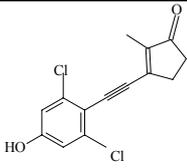
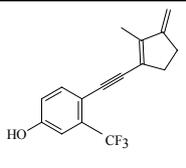
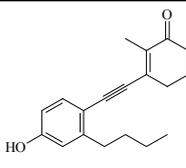
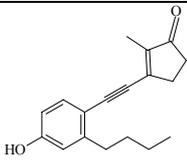
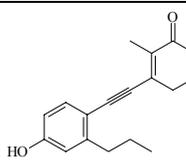
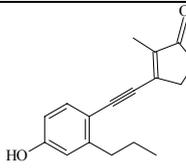
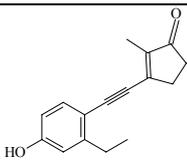
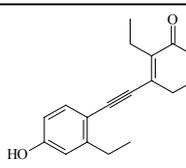
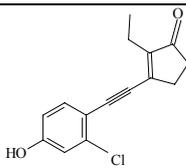
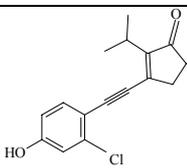
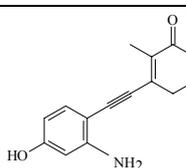
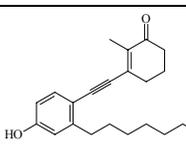
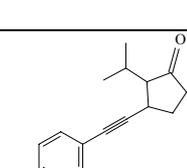
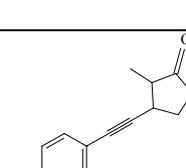
7 Appendix

7.1 List of ligands

Structure of ligands synthesized in the Institute of Organic and Biomolecular Chemistry, Georg August University Göttingen.

 CV0505	 CV2105	 CV2317
 CV2807	 CV3320	 CV3519
 CV3523	 CV3524	 CV3620
 CV5008	 CV5009	 CV5010
 CV5011	 CV5304	 CV5306
 CV5407	 CV5413	 CV5414

 CV5508	 CV5735	 CV5742
 CV5746	 CV6010	 CV6017
 CV6018	 CV 6019 CV6019	 CV6021
 CV6104	 CV6304	 CV6308
 CV6407	 IK281	 IK307 IK307
 IK308	 IK377	 IK423 IK423
 IK424	 IK425	 IK466
 IK504	 IK505	 IK506

 IK507	 IK508	 IK513
 IK519	 IK524	 IK525
 IK531	 IK534	 IK551
 IK585	 IK602	 IK603
 IK608	 IK612	

7.2 The amino acid code

Single letter code	Abbreviation	Full name
A	Ala	Alanine
R	Arg	Arginine
N	Asn	Asparagine
D	Asp	Aspartic acid
C	Cys	Cystein
Q	Gln	Glutamine
E	Glu	Glutamic acid
G	Gly	Glycine
H	His	Histidine
I	Ise	Isoleucine
L	Leu	Leucine
K	Lys	Lysine
M	Met	Methionine
F	Phe	Phenylalanine
P	Pro	Proline
S	Ser	Serine
T	Thr	Threonine
W	Trp	Tryptophan
Y	Tyr	Tyrosine
V	Val	Valine

Acknowledgement

Sincere gratitude to Prof Dr. C. Gatz for providing me the opportunity to pursue the doctoral study in her innovative and exciting group. Her creative ideas, instructive discussions and helpful insights made this work possible. The interest, time and patience that she maintained throughout the entire period are astounding and needs special mentioning. I would also like to thank PD Dr. Wolfgang Dröge-Laser for his constant interest in this work from the very beginning and his kind assertion in being the Co-referent of this thesis.

Special thanks to Dr. Meik Dilcher for his versatile support and inspiring discussions, even in difficult periods. All the efforts and support that he provided in academic and non-academic areas were really instrumental. I would like to also mention the constant support that Dr. Corinna Thurow provided- particularly the exceptionally creative ideas in dead-end situations. Heike Runne, while working as a Master student, started creating some of the vector and I would like to appreciate her contribution. I would also like to thank Dr. Guido Kriete and Ms. Heike Freundt for helping in numerous administrative matters. I appreciate the contribution from Anna Herrman, Annette Gunkel and Ronald Scholz for their technical assistance and invaluable knowledge of special know-how which made the life easier for a researcher.

I would like to thank the collaboration partners at the University of Göttingen, namely, Dr. Carsten Vock and Dr. Ilga Krimmelbein (Organic Chemistry) for synthesis of ligands as well as Dr. Christina Schlicker (Structural Chemistry) and Simone Brauns (Molecular Structural Biology) for their assistance in protein biochemistry.

Prof. Dr. Humin Zhao (Dept. of Biomolecular Engineering, University of Illinois at Urbana Champaign) deserves special gratitude for providing me the opportunity to perform some of the protein engineering approaches (directed evolution) in such a great setup. His critical thinking, dynamic drive and unorthodox supervision enriched the work significantly.

Finally, many warm thanks to my colleagues at the Dept. of General and Developmental Physiology for creating a comfortable work environment.

Curriculum Vitae

Name: Kazi Mohammed Didarul Islam

Date and Place of Birth: 18 January 1977, Gopalganj

1982-1986: Primary School Study (Class I-V)
BAF Shaheen School, Dhaka, Bangladesh

1987-1991: Secondary School Study (Class VI-X)
BAF Shaheen School, Dhaka, Bangladesh

1993-1994: Higher Secondary Certificate
Notre Dame College, Dhaka, Bangladesh

1995-2000: BSc in Biotechnology
Khulna University, Bangladesh

2001-2003: MSc in Biology (Plant Cell Biotechnology)
Transport of catharanthine and vindoline into the isolated vacuoles of *Catharanthus roseus*
Under the supervision of Prof. Robert Verpoorte
Department of Pharmacognosy
Leiden University, The Netherlands

2003- 2007: Completed the thesis for PhD in Biology (Botany)
Development of an orthogonal ligand-receptor pair based on synthetic estrogen analogs and engineered estrogen receptor for transcriptional regulation
Under the supervision of Prof Christiane Gatz
Dept. of General and Developmental Physiology
Georg August Universität Göttingen, Germany