ENGINEERING PROTEIN MOLECULAR SWITCHES TO REGULATE GENE EXPRESSION WITH SMALL MOLECULES

A Thesis
Presented to
The Academic Faculty

by

Priyanka Rohatgi

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in the
School of Chemistry and Biochemistry

Georgia Institute of Technology
December 2006

ENGINEERING PROTEIN MOLECULAR SWITCHES TO REGULATE GENE EXPRESSION WITH SMALL MOLECULES

Approved by:

Dr. Donald F. Doyle, Advisor
School of Chemistry and Biochemistry
Georgia Institute of Technology

Dr. Joseph M. LeDoux
School of Biomedical Engineering
Georgia Institute of Technology

Dr. Nicholas V. Hud
School of Chemistry and Biochemistry
Georgia Institute of Technology

Dr. H.Trent Spencer
School of Medicine
Emory University

Dr. C. David Sherrill
School of Chemistry and Biochemistry
Georgia Institute of Technology

Date Approved: September 11, 2006

I dedicate this work to my parents Col. Pradeep Rohatgi and Ms. Vinita Rohatgi.

My sister Ms. Prapti Rohatgi and my husband Mr. Nalin Rohatgi.

ACKNOWLEDGEMENTS

Doing a Ph.D. is not something that one can do all by oneself. There are many people who help you along the way and I would like to thank everyone who has helped in my quest. The first person I want to thank is my advisor Dr. Donald Doyle. He has been a great teacher and mentor, training me with the skills I would need, inculcating in me with the thirst for learning and for giving me the opportunity to go for conferences. He guided me while always encouraging me to think on my own and to come up with my own ideas. He also has been a very good friend, always understanding of what was going on in my personal life.

I would also thank my committee members Dr. Joseph M. LeDoux, Dr. Nicholas V. Hud, Dr. C. David Sherrill and Dr. H. Trent Spencer for developing my critical thinking skills. I would like to give an additional thank you to Dr. Trent Spencer for training me in viral transduction, Dr. Christoph Fahrni for helping me with the in vivo labeling of protein experiments and Dr. Dickson for the emission spectra experiments.

Next I would like to thank my laboratory mates, who have been more like my family in last five years. A big thank you to Dr. Bahareh Azizi to whom I am very grateful for all the help she has given me over all the years of my Ph.D., for her help with training me with cloning and presentation skills, for raising my interest in scientific literature, troubleshooting with experiments and analyzing them, for helping with my original proposal and for being a very good sounding board. She was always there to raise my spirits and remind me to be patient when experiments didn't go exactly the way I wanted. She has been a very good friend in good as well as bad times.

Thank you to Kenyetta A. Johnson for helping with the writing and editing of my thesis and for reading and re-reading it multiple times. She has helped me with many experiments, burning the midnight oil with me when necessary. I am grateful to have such a wonderful friend who has always been there for me and given me good advice. She always made me laugh, making all the hard times easier.

Thank you to Eileen I. Chang for helping with all the experiments and her meticulous handling of them. Thank you to Dr. Lauren J. Schwimmer for teaching me molecular modeling, Christopher Skipwith and Mary E. Johnson, my undergraduate scholars, for helping with my experiments. Dr. Eva L.-H. Heintz for collaborating with the superparamagnetic nanoparticles experiments. Terry J. Watt for all his constructive ideas while discussing materials with him and Jennifer L. Taylor for helping with experiments. I would also like to thank all the group members of Dr. Doyle, Dr. Bommarius, and Dr. LeDoux labs for sharing their resources and for being helpful.

On a personal basis I would like to thank Vikram Subramanian for his help with seminars, editing skills and with classes. Thank you to Ashish Labroo for all his help with applying for scholarships. Thank you to my friends Kunal Seth, Mugdha Wagle and Sonali Tare for always being there, encouraging me when I was down, listening and giving me positive outlook, believing in me.

In the end, I would like to thank all my family members, especially my parents Col. Pradeep Rohatgi and Vinita Rohatgi for always being my strength and for the early lessons they taught me about the importance of education, and that is one of the biggest reasons I am here today. They never held me back, making sure that I knew that resources would never be lacking where my education was concerned. They have given

me opportunities that many can only dream of. They pushed me to excel and have always been there to give advice and guide me in my career goals. I wouldn't have been here without their love and support. My sister Prapti Rohatgi, though being married, having a family and career of her own always made time for me and helped me out when I was in a tough spot. I can't thank her enough for her love and inspiration. My brother-in-law Dr. Shashank Rohatagi, for giving me career counseling and introducing me with new avenues and ideas. My grandparents Hari Krishan Rohatgi and Urmila Rohatgi for their goodwishes.

Lastly, I want to thank my husband, Nalin Rohatgi, for his love and his understanding. For being patient while I lived in Atlanta and he in Seattle as I finished my Ph.D. He has been a wonderful husband, a great shoulder to lean on. He has helped me tremendously with my self-confidence, always giving me a boost. Nalin has been a great motivator and always pointed out the positives in every situation. I can't thank him enough for his love and support.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

DBD	DNA binding domain
LBD	Ligand binding domain
TetR	Tetracycline repressor
HSV	Herpes simplex virus
Tet	Tetracycline
tTA	Tetracycline transactivator
rtTA	Reverse tetracycline transactivator
ER	Estrogen receptor
PR	Progesterone receptor
EcR	Ecdysone receptor
CID	Chemical inducers of dimerization
AD	Activation domain
RXR	Retinoid X receptor
RAR	Retinoid acid receptor
NR	Nuclear receptor
RE	Response element
OLRP	Orthogonal ligand receptor pair
9cRA	9 *cis* retinoic acid
5'- FOA	5'- Fluoroorotic acid
5'- FU	5'- Fluorouracil
GFP	Green fluorescence protein

Neo	Neomycin resistant gene
Hrl	*Renilla* luciferase gene
Ttk	Thymidine kinase gene

SUMMARY

Small molecule dependent molecular switches that control gene expression are important tool in understanding biological cellular processes and for regulating gene therapy. Nuclear receptors are ligand activated transcription factors that have been engineered to selectively respond to synthetic ligands and used as regulators of gene expression. In this work the retinoid X receptor (RXR), has been used to develop an inducible molecular switch with a near drug like compound LG335. Three RXR variants (Q275C; I310M; F313I), (I268A; I310A; F313A; L436F), (I268V; A272V; I310M; F313S; L436M) were created via site-directed mutagenesis and a structure based approach, such that they preferentially bind to the synthetic ligand LG335 and not its natural ligand, 9-*cis* retinoic acid. These variants show reverse ligand specificity as designed and have an EC_{50} for LG335 of 80 nM, 30 nM, 180 nM, respectively. The ligand binding domains of the RXR variants were fused to a yeast transcription factor Gal4 DNA binding domain. This modified chimeric fusion protein showed reverse response element specificity as designed and recognized the Gal4 response element instead of the RXR response element. The modified RXR protein did not heterodimerize with wild type RXR or with other nuclear receptor such as retinoic acid receptor. These RXR-based molecular switches were tested in retroviral vectors using firefly luciferase and green fluorescence protein and they maintain their inducible behavior with LG335. These experiments demonstrate the orthogonality of RXR variants and their possible use in regulating gene therapy.

CHAPTER 1

INTRODUCTION

Molecular switches

Gene therapy has developed into a promising therapeutic to treat a diverse array of diseases such as cancer, AIDS, cystic fibrosis, Parkinson's, Alzheimer's, cardiovascular disease and arthritis. This technology of gene therapy uses functional genes to provide a desired treatment. Gene therapy has become effective due to the advances in gene delivery systems and gene regulatory systems. The gene regulatory system is an important aspect of gene therapy because over or under-production of the therapeutic protein can cause side effects. Some regulatory systems have been developed to control gene expression in *vitro* and in *vivo*. Initial attempts to regulate gene expression have used endogenous cellular elements such as promoter and enhancer that respond to physiological changes such as heat[1], metal ions[2], interferons[3] and hypoxia[4]. Other regulatory systems such as the lac operator-IPTG-based system[5-8], the FKB12-rapamycin-associated protein/FK106 binding protein[9] and mRNA aptamer based system[10] have also been developed. Many of these systems are not suitable for clinical gene therapy for reasons such as toxicity, lack of specificity and background transcriptional activity in the uninduced state.

Inducible gene expression systems have been developed that are regulated by administration of specific small molecules or ligands. These ligand-dependent inducible systems are usually based on two components: the first component is a chimeric

transcription factor containing a DNA-binding domain (DBD) fused to a ligand binding domain (LBD), where the DBD recognizes a DNA sequence; the second component is an artificial promoter consisting of binding sites for the DBD followed by a minimal promoter and the therapeutic gene (Figure 1.1). The genetically engineered transcription factors used in these systems are generally derived from bacterial repressor proteins or eukaryotic receptors. The engineered transcription factors function as molecular switches, i.e. they are either turned on or off in the presence of a suitable ligand. To utilize these switches for gene therapy, genes encoding for the molecular switch and the therapeutic gene will be delivered to the patient. In the presence of the ligand the molecular switch will be recruited to the artificial promoter and express the therapeutic gene downstream of this promoter (Figure 1.2). This thesis focuses on developing new molecular switches to control gene expression using small molecules.

The molecular switches that are currently being developed can control three processes: transcriptional activation, translational initiation and posttranslational protein activity. Among the transcriptional switches, several researchers have tried creating DNA binding protein to bind arbitrarily chosen DNA sequence[11, 12]. One of the earlier works was done by Barbas, Schultz and coworkers[13, 14]. They engineered the natural zinc finger protein Zif268, to bind to DNA in the presence of a small molecule ligand. They evolved zinc fingers that bind to individual GNN codons[15]. Two residues, histidine 125 and phenylalanine 116, which participate in zinc binding, were mutated to an alanine and a glycine resulting in a cavity around the zinc ion and abolishing transcriptional activation. They used the bump-hole approach[16-18] to alter ligand

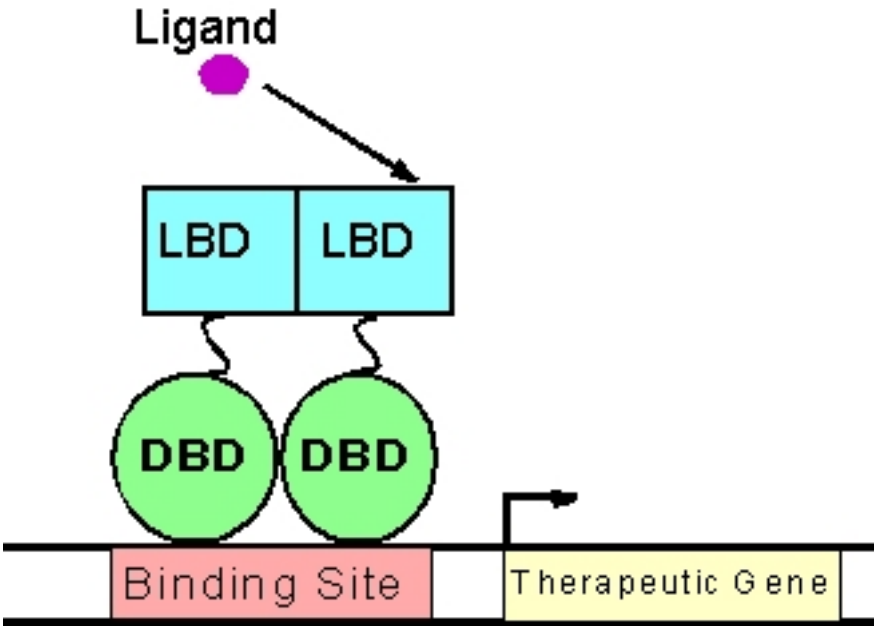


Figure 1.1 Ligand dependent inducible systems.

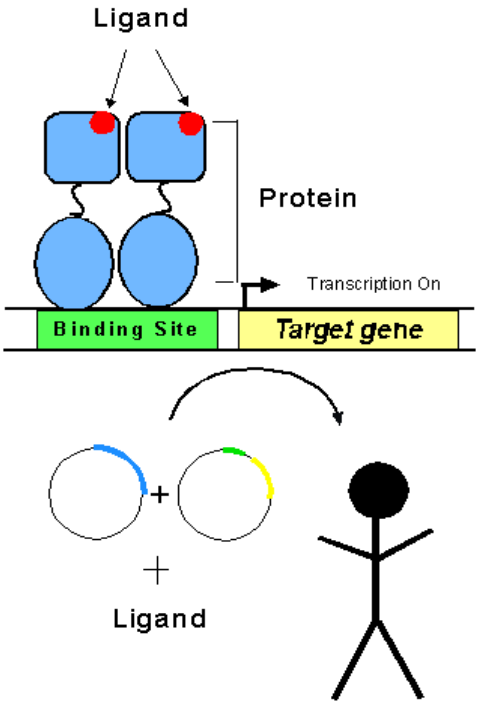


Figure 1.2 Schematic depiction of molecular switch.

binding specificity and screened 250 heterocyclic molecules. They found a small molecule 2-(4-quinoline)-benzimidazole, that increased gene activity by 18 fold. The small molecule bound mutant protein recognizes the DNA site with increased affinity. The fusion of three Zif268 proteins to the VP16 activation domain and the estrogen receptor LBD created transcriptional switches[14]. Another example of transcription switch based on the DNA binding protein is tetracycline repressor (TetR). Hillen and coworkers created an orthogonal regulatory system in eukaryotics cells by incorporating elements of the tetracycline-resistance operon[19, 20]. TetR was fused to a herpes simplex virus (HSV) VP16 transactivation domain to form a tet-controlled transactivator (tTA). This chimeric protein could control transcription in mammalian cells from a tet-responsive control element consisting of tet operators fused to a minimal promoter. In the presence of tet, tTA is unable to bind operator sequences to activate transcription, making this a tet-repressible system. One of the drawbacks of this tTA system is that tetracycline must be present to keep gene expression repressed. To overcome this problem directed evolution approach was used to create TetR mutants that bind the operator in the presence of tetracycline or a derivative doxycycline. Correspondingly, the reverse tet transactivator (rtTA) activates gene expression in the presence of either drug, rendering the system more suitable for therapeutic applications[20].

In contrast to altering the DNA binding domain, researchers have made efforts to create molecular switches by altering the ligand specificity of the activation domain. Nuclear receptors are extensively used to generate orthogonal ligand receptor pairs to control gene expression. Parker and coworkers created orthogonal mutants of estrogen receptor (ER) LBD. They created a mutant glycine 521 arginine that is induced by 4-

hydroxytamoxifen by 10 fold and does not respond to its natural ligand β-estradiol[21, 22]. Crystal structure of ER LBD with β-estradiol and 4-OHT reveals that the side chain of arginine 521 reduces the cavity size hindering β-estradiol to bind, while 4-OHT remains unaffected. In another systems developed by O'Malley and colleagues, progesterone receptor LBD was modified. It was based on the truncated progesterone receptor ligand-binding domains (PR-LBD). This modified PR-LBDs lost the ability to respond to its natural ligand progesterone, but have gained the ability to respond to antiprogestins as agonists, rather than as antagonists[23]. This modified PR-LBD is fused to yeast Gal4 DNA binding domain and an activator domain such as VP16 or hNF-$_{\kappa}$B p65. The inducible target gene has a minimal promoter that consists of multiple binding sites for the Gal4 DBD. Binding of the antiprogestin inducer mifepristone triggers a conformational change that causes the regulator protein to become an activated homodimer, which binds to Gal4 sites in the inducible promoter stimulating transcription of the target gene[24, 25].

Schreiber and Clackson pioneered the strategy of creating molecular switches using chemical inducers of dimerization (CID)[26]. Here the transcription is controlled by using two separate protein components, a DNA binding domain (DBD) and an activation domain (AD), each fused to a small molecule binding domain[27, 28]. In the presence of a small molecule, both fusion proteins bind the small molecule, recruiting the activation domain to the promoter of interest. The most widely used chemical dimerizer system is FK506-binding protein (FKBP) and FRAP which binds to a ligand rapamycin. In one application two fusion proteins was constructed, one being FRB (FKBP and rapamycin binding domain of FRAP) fused to the transcriptional activator from NF-κB

and the other was a zinc finger DBD, ZFHD1 fused three repeats of FKBP as the second protein. In the presence of both rapamycin and the target gene containing the binding site for ZFHD1, the transcription increased by more then 10000 fold[9].

Translational switches are another way of regulating gene expression. Breaker and coworkers are the pioneers in discovering translational switches. They discovered RNA-mediated small molecules termed riboswitches that regulate translation[29]. These riboswitches are found in the 5'-untranslated region of mRNA. They contain two domains, an aptamer region that binds to a metabolite and an expression platform that interacts with regulatory elements in the mRNA, such as the Shine-Dalgarno sequence. Riboswitches can be turned on and off through allosteric regulation by ligands. The mechanism by which gene expression is regulated involves the formation of alternative structures that, in the repressing conformation, cause premature termination of transcription or inhibition of translation initiation. These riboswitches regulate several metabolic pathways including the biosynthesis of vitamins, metabolism of methionine, lysine and purines[30]. Hillen and coworkers designed an artificial riboswitch using RNA aptamer for theophylline and structural bridge evolved by Breaker[31] such that upon addition of theophylline, a conformational change occurs and translation is increased by eight fold[32]. This was the first riboswitch that increased translation rather then repressing it.

Some ligand dependent switches that act posttranslationally are based on protein splicing. Trans-splicing is a phenomenon in which an intein is split into inactive N- and C- terminal halve, when combined these halves reconstitute an active intein that is capable of splicing. Muir developed a trans-splicing system in which two intein

fragments are brought together by addition of rapamycin[33]. Three previously developed molecular switches that are commonly used to date and are similar to the one developed in this thesis are now reviewed in detail.

Progesterone receptor regulatory system

Recently, a progesterone receptor (PR) mutant with a C- terminal truncation was isolated that prevents binding with the natural ligand progesterone[24, 25]. The truncated receptor retains the ability to bind the antagonist mifepristrone (RU 486). Mifepristrone (known for abortive function in humans) acts as an inducer on the mutant PR and promotes transcription of target genes containing progesterone responsive elements[34]. Wang and coworkers made modifications to the PR to increase its sensitivity to mifepristrone and to create a ligand inducible gene expression system[25]. These modifications allowed PR to respond to mifepristrone at concentrations one order of magnitude lower than those needed to cause an abortive activity[24]. The truncated PR with 10 to 34 glutamine residues was fused with the Gal4 DNA binding domain (a yeast transcription factor) and a eukaryotic transcriptional activation domain from viral protein (VP), VP16 or the p65 subunit of human NF-κB. This chimeric transcription factor and its inducer, mifepristrone could now regulate the expression of target gene with a Gal4 binding site[35]. To reverse the mifepristrone responsiveness behavior the VP16 domain was replaced by the KRAB motif (krüppel-associated box protein), a potent repressor domain of the kidney-specific transcription factor Kid-1[23]. Using this KRAB-

containing transrepressor resulted in mifepristrone dependent repression of target genes[25].

The advantage of this system is that the majority of the system comprises modified human proteins and does not provoke an immune response. The disadvantage is that despite various improvements, the regulation performance of mifepristrone-based systems in mammalian cells remains relatively poor, due to the high basal activity in an uninduced state. The high basal activity results in a low induction ratio, generally about 20 folds. Also, mifepristrone may affect the ovarian cycle and exert a contraceptive activity.

Ecdysone receptor regulatory systems

Ecdysone receptor (EcR) is an insect steroid hormone receptor, that functions as a heterodimer of the ecdysone receptor (EcR) and the ultraspiracle protein (USP)[36]. EcR triggers metamorphosis in Drosophila melanogaster and other insects. The class of ligands that bind the EcR are called ecdysteroids[37]. Modifications were made to the EcR to create a molecular switch. The N-terminal activation domain of EcR was replaced by the corresponding domain of glucocorticoid receptor (GR), its natural heterodimeric partner USP, was replaced by its mammalian homologue Retinoid X receptor (RXR). These modification resulted in a 34-fold induction of the target gene[38]. However the N-terminal truncated EcR fused to the VP16 (VpEcR) heterodimerizing with RXR increased the induction ratio to 212-fold[38]. To minimize the potential interference with endogenous factors, the specificity of VpEcR and its binding site was further improved.

VP16-EcR- GR (VgEcR)/RXR heterodimer and a synthetic binding site consisting of a hybrid between the GR, EcR and RXR for (VgEcR)/RXR complexes were constructed[38].

An advantage of EcR based regulatory system is the low level of basal activity as compared to other inducible systems. The low basal activity results in high induction level of the target gene. Despite of many advantages over other systems, it has several shortcomings that may prevent its use in gene therapy applications. Firstly, insect hormones and their agonists are not orally bioavailable and may not get approved for human therapeutic use. Secondly, this system requires simultaneous expression of two proteins VgEcR and RXR, which may complicates its use in certain viral delivery systems. Thirdly, over expression of mammalian RXR, may have pleiotropic effects in the mammalian cells. Finally, RXR is a reluctant dimer partner of EcR, and very high endogenous levels of RXR are necessary for stimulation[36, 39].

Tetracycline dependent regulatory system

The tetracycline (Tet) dependent gene regulation system comprises of two complementary ideas known as the Tet-OFF and the Tet-ON system. This system is of prokaryotic origin and its core components are the Tet repressor (TetR), its cognate binding site, the tet operator, and an antibiotic tetracycline[19]. In bacteria, the TetR hinders transcription by docking on the Tet operator in the absence of tetracycline. In Tet- OFF system, a fusion of TetR and VP16 was made to convert tetR from a repressor to an activator termed tTA[40]. tTA interacts with its responsive promoter i.e. tandem

repeats of tet operator sequence and minimal promoter (TRE) to drive the expression of a target gene. In the presence of tetracycline, tTA is hindered to dock on its binding site TRE, thereby impeding transcription of the target gene. In Tet-ON system, random mutations were made in the tTA. Four amino acid changes yielded a protein that exhibits opposite function. This mutant, rtTA triggers activation of TRE controlled target gene only in the presence of tetracycline[20].

The Tet system offers many advantages such as tetracycline and its analogue doxycycline is well characterized, safe, rapidly metabolized with a half-life of 14 to 22 hours and nontoxic at doses required for gene activation in preclinical and clinical studies. Doxycycline is orally bioavailable, it has a good tissue penetration and does not interfere with native proteins[41]. Tet transcription factors are target specific i.e. tTA or rtTA docks only on TRE controlled target genes, hence reducing the risk of serious side effects. However, these proteins are prokaryotic in origin and may be immunogenic. Further study is required to determine whether the immune system can recognize components of the Tet system over the long time periods for the treatment of chronic illnesses such as Parkinson disease and multiple sclerosis.

In the last decade, technologies for regulating gene expression in mammalian cells have advanced from cell lines to animal models. These technologies have impacted genomic research in an attempt to unravel molecular pathways. It has also contributed to the creation of animal models for currently untreatable human diseases, such as Parkinson's and Alzheimer's. Next generation gene regulation will have major challenges

including long-term applications, precise regulation of target genes over the lifetime of the patient, and human compatibility of the molecular switch.

Nuclear receptors

Nuclear receptors are a superfamily of eukaryotic ligand activated transcription factors that regulate development, cell proliferation, endocrine signaling and metabolism. Nuclear receptors comprise of a diverse super-family in terms of physiological roles ranging from receptors for regulating metabolic pathway such as uptake, oxidation, and processing of extracellular lipids to regulation of developmental pathways[42-44]. This super-family is composed of steroid receptors such as the ER, androgen receptor, the non-steroidal receptors such as the thyroid hormone receptor, retinoic acid receptor (RAR) and orphan receptors. Nuclear receptors are associated with numerous human diseases, for example, RARs with types of leukemia[45], ER with breast cancer, and peroxisome proliferator activated receptor with diabetes[43]. Hence, they have become a primary target of drug discovery and the pharmaceutical industry has a great interest in discovering agonist and antagonist for these receptors[46]. Some drugs that are currently used are tamoxifen against breast cancer, dexamethasone for inflammatory diseases and thiazolidinediones as drugs for type II diabetes[44].

Nuclear receptors are ligand activated transcription factors that bind hydrophobic, fat-soluble small molecules. The role of the ligands can be to activate receptors or to deactivate constitutively active receptors. Structurally, nuclear receptors exhibit a modular structure with different domains corresponding to functional domains that can be

interchanged between related receptors without loss of function. Typically nuclear receptor consists of domains A-F, which consists of a variable NH_2 – terminal domain (A/B), a conserved DNA binding domain known as C domain, a linker domain D, a conserved E domain that contains C-terminal ligand binding domain[43, 47], and a F domain (Figure 1.3). The A/B domain is variable in size and sequence and contains an AF-1 region which is a ligand-independent activation domain[48-50]. The DBD is highly conserved and has the ability to recognize specific DNA sequences known as the response element (RE)[44, 51]. Nuclear receptors regulate transcription by binding to a RE that is normally present within 1 kb to the promoter or in the enhancer region of the transcription. Analysis of a large number of RE's revealed that a sequence of six base pairs constitutes the recognition motif. Most receptors bind as homo- or heterodimers to REs to two core hexameric motifs. To form the dimer's RE, the half-sites can be configured as palindromes, inverted palindromes, or direct repeats[52, 53]. This DBD or C domain comprises of two "zinc fingers"; each finger has four cysteines that coordinate one zinc ion[54]. Amino acids that are required for discrimination of DNA recognition motifs are present at the base of the first finger and the residues of the second zinc finger are involved in dimerization. The D domain serves as a hinge between the DBD and the LBD, allowing flexibility between the LBD and DBD[43]. Nuclear receptors are modular proteins such that the DBD and LBD can function independent of each other. The E region or LBD is a multifunctional domain that in addition to the binding of the ligand mediates homo and heterodimerization [55]. The crystal structure of the LBDs of many nuclear receptors have been solved. Generally, the LBDs are formed by 12 conserved α-helical regions and a conserved β turns between helices 5 and 6. The 12 α-helices in the

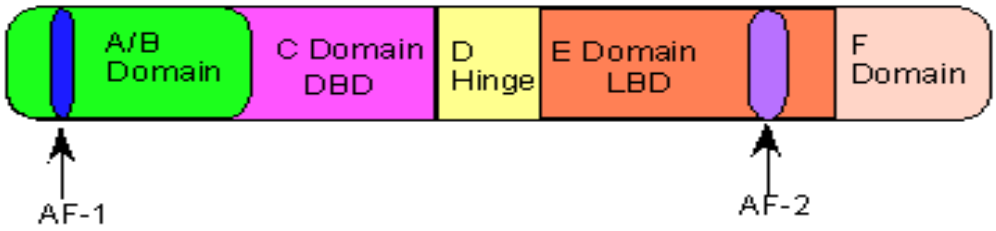


Figure 1.3 Nuclear receptor domains.

LBDs are folded into a three layered antiparallel helical sandwich[56]. Within the LBD a central core is formed called the ligand binding pocket which primarily consists of hydrophobic residues[56]. The size of the binding pocket varies among the different receptors. Several differences are evident in the LBD structures of a liganded and unliganded LBD. The ligand bound LBD is more compact, indicating there is a conformational change on ligand binding.

In the absence of ligand the nuclear receptors recruit corepressor complexes. Co-repressors do not bind to the DNA themselves but mediate gene silencing via DNA bound transcriptional factors[57]. Malfunctioning of co-repressor action can cause a variety of diseases such as human syndrome of resistance to thyroid hormone, familial type II diabetes based on reduced dissociation of the corepressor from peroxisome proliferator activated receptor γ. The binding motif for two major co-repressors, nuclear receptor corepressor and the silencing mediator of retinoid and thyroid hormone receptor (SMRT) consists of the LXXI/LXXXI/L motif[57, 58]. It involves the recruitment of histone deacetylase (HDAC) which leads to a more compact chromatin structure which inhibits the accessibility of transcription factors and represses transcription[59-61].

Upon ligand binding, the nuclear receptor LBD undergoes a conformational change that involves several helices present in the LBD, especially helix 12. In the apo form (unliganded) of the receptor, helix 12 is freely floating in the solution. In the holo form of the receptor helix 12 moves towards the protein and tightly packs against helix 3 or 4 making direct contact with the ligand[62, 63]. The proper positioning of helix-12 leads to formation of a hydrophobic cleft to create a proper dimerization interface for "LXXLL" NR-box sequences of transcriptional co-activators. Ligand binding induces

recruitment of co activators complexes. Some of these proteins have histone acetyltransferase activity and interact with the transcription machinery[64]. This result is chromatin decompactation allowing RNA polymerase to initiate the transcription of target gene.

Retinoid X receptor

The human retinoid X receptor (RXR) is a ligand activated transcription factor of the nuclear receptor superfamily, under the subclass of retinoid receptors. RXR plays an important role in differentiation, homeostasis and morphogenesis. RXR has three isotypes RXRα, RXRβ, and RXRγ. It serves as a homodimer or as a dimerization partner for other nuclear receptors such thyroid receptor, vitamin D receptor [65, 66]. It is a key binding partner with retinoid acid receptor. Structurally, both RXR and RAR are similar to other nuclear receptors, containing a variable A/B region, the DBD, hinge region and LBD (Figure 1.3). The natural ligands known for RXR include phytanic acid[67], docosahexaenoic acid[68] and 9-cis retinoic acid (9cRA)[69]. The crystal structure of RXR bound to 9cRA is one of the first nuclear receptor to be solved and the key interactions of the binding pocket with the ligand were determined (Figure 1.4)[70]. When RXR homodimerizes it binds to a DNA sequence called RXR response element (RE). In the presence of 9cRA, RXR undergoes a conformational change, binds to RXR RE and initiates transcription of the gene downstream of RXR RE (Figure 1.5). This RE is a direct repeat of six bases with one base spacer (DR1) such as CRBPII (TAGGTCA A AGGTCA GTAGGTCA A AGGTCA G)[71, 72]. . RXR is a modular protein, DBD

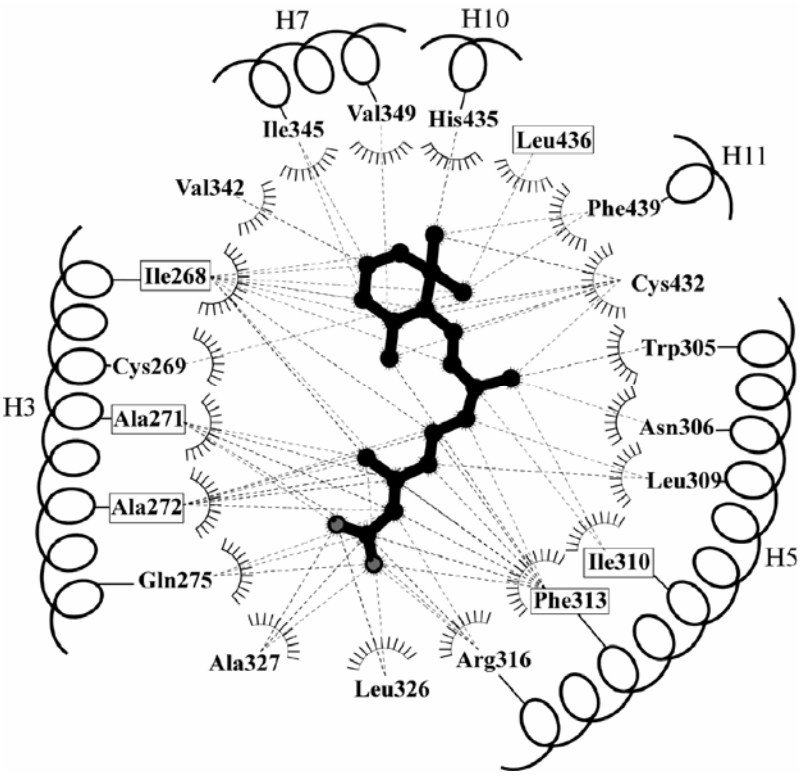


Figure 1.4 Binding pocket of RXR with the residues that interact with the ligand 9cRA.

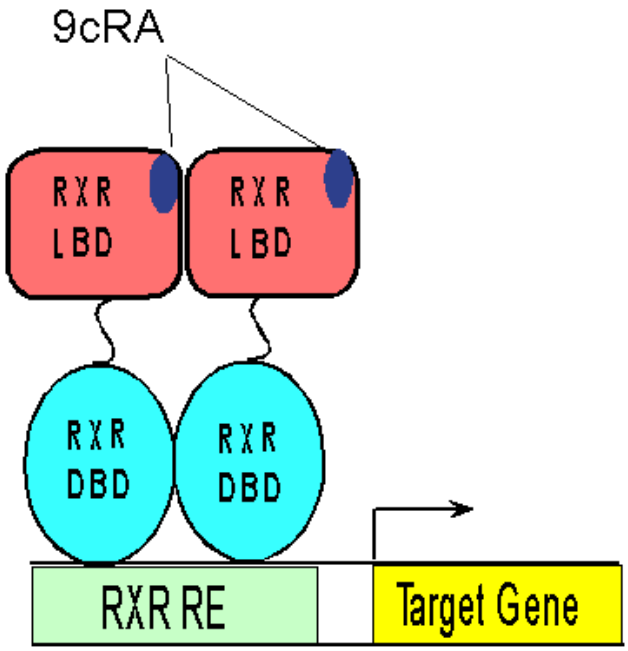


Figure 1.5 Retinoid X receptor ligand binding domain binds to 9cRA and RXR DNA binding domain recognizes RXR response element.

and LBD can function independently. Therefore, the LBD can be fused to other DBD's and retain the function. This feature can be utilized to engineer the LBD such that it binds to synthetic ligand and fusing them to a foreign DBD would make an orthogonal ligand receptor pair.

An orthogonal ligand receptor pair

The ability to manipulate protein to bind and respond to synthetic ligands in an orthogonal or independent manner is a challenge in protein engineering but an important tool for various biotechnology applications (Figure 1.6). The ability to control transcription using a ligand would be useful to study cellular processes such as apoptosis, signal transduction and facilitate the analysis of gene function during embryonic development and differentiation. It would reveal the function for thousand of genes that were discovered in various genome projects with unknown function. Creating such a matched ligand-receptor pair that is orthogonal can be utilized as molecular switches for inducible gene expression system. These molecular switches can be used for practical application such as gene therapy by regulating the expression of target gene. The molecular switches or orthogonal ligand receptor pair can be used as rheostats or as on/off switch depending on their response to the ligand. To utilize it for gene therapy, along with the therapeutic gene, the gene for orthogonal receptor would be delivered.

The orthogonal receptor would be expressed in the cell and would bind to its specific response element that is present only in the promoter region of the therapeutic gene. In the presence of ligand the receptor will get turned on and express the therapeutic

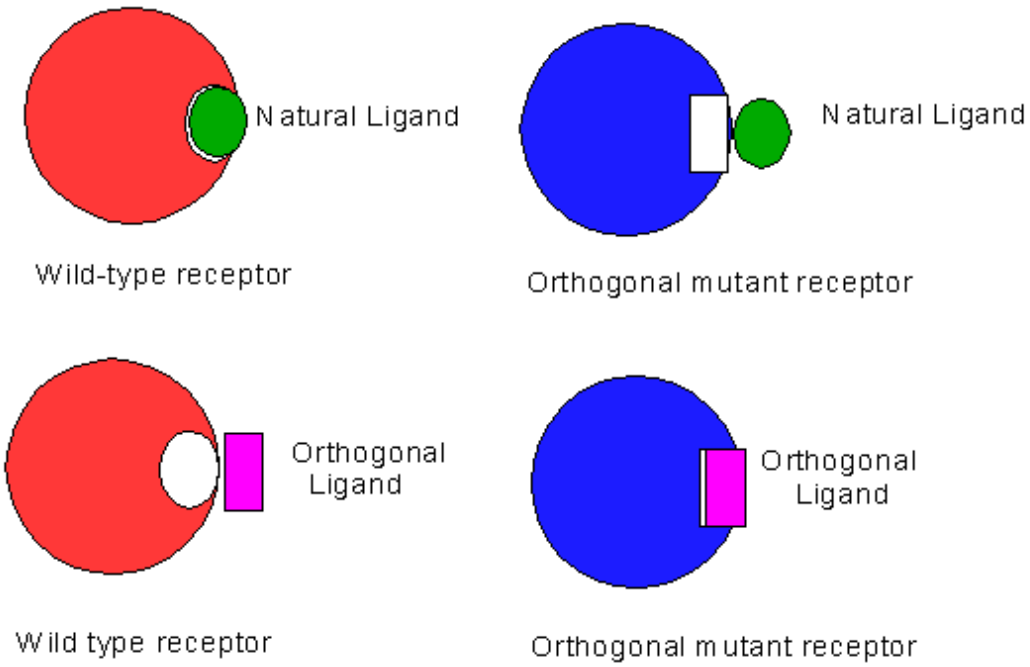


Figure 1.6 Schematic depiction of the interaction between wild type and orthogonal receptor with its natural or synthetic ligand.

gene to the desired level depending on the dose of the ligand and activation level of the receptor. Orthogonal ligand receptor pairs could also be utilized for enzyme engineering, where the receptor could be engineered to bind the desired product of an enzymatic reaction. A library of engineered enzymes would be screened and when it produces the desired product, the product would behave as the ligand for the receptor and activate transcription of the reporter gene, thereby identifying the cell containing the desired enzyme. The gene expressing the desired enzyme would then be sequenced and recovered.

Chemical complementation

Chemical complementation is a genetic selection system in which the survival of yeast or mammalian cells is linked to the ability of ligand binding and activation of the nuclear receptor RXR. It is a useful tool for engineering and selecting functional variants from libraries of receptors or to screen a library of compounds for a particular receptor[73]. Chemical complementation system is developed in yeast, where the PJ694A strain is engineered to contain Gal4 response element controlling the expression of *ADE2* gene (selection marker)[74]. Expression of the *ADE2* gene allows the yeast to produce adenine and to survive in media lacking adenine. Gal4 RE is a short DNA sequence that is recognized by the DBD of a yeast ligand independent transcriptional activator Gal4 protein. A fusion protein of RXR LBD and Gal4 DBD was constructed that would bind Gal4RE. Another fusion protein comprising of ACTR (nuclear receptor coactivator) and Gal4 activation domain is also expressed. In the presence of an

agonist this fusion protein undergoes a conformational change and recruits the coactivator ACTR, Gal4DBD recognizes Gal4 RE which allows the transcription of ADE2 gene, resulting to the survival of yeast in adenine lacking media (Figure 1.7)[74].

Chemical complementation is a high throughput method for selecting novel ligand receptor pairs in a single experiment. This method is being developed in mammalian cells by using a positive and negative selection (details in chapter 5). Analogous to the yeast, mammalian cell line is engineered to contain Gal4 response element controlling the expression of triple fusion gene: neomycin resistant gene, reporter gene and thymidine kinase gene. A library of RXR variants fused to Gal4 DBD is expressed in the cells. In the presence of an agonist RXR undergoes a conformation change and expresses the triple fusion gene. In the presence of selective media such as geneticin and gancyclovir the cell containing the functional variants will survive and form colonies. These colonies can then be easily separated and evaluated by additional selection and/or screening assays. The variants that have high activation of the reporter gene in the presence of ligand would qualify as orthogonal ligand receptor pairs.

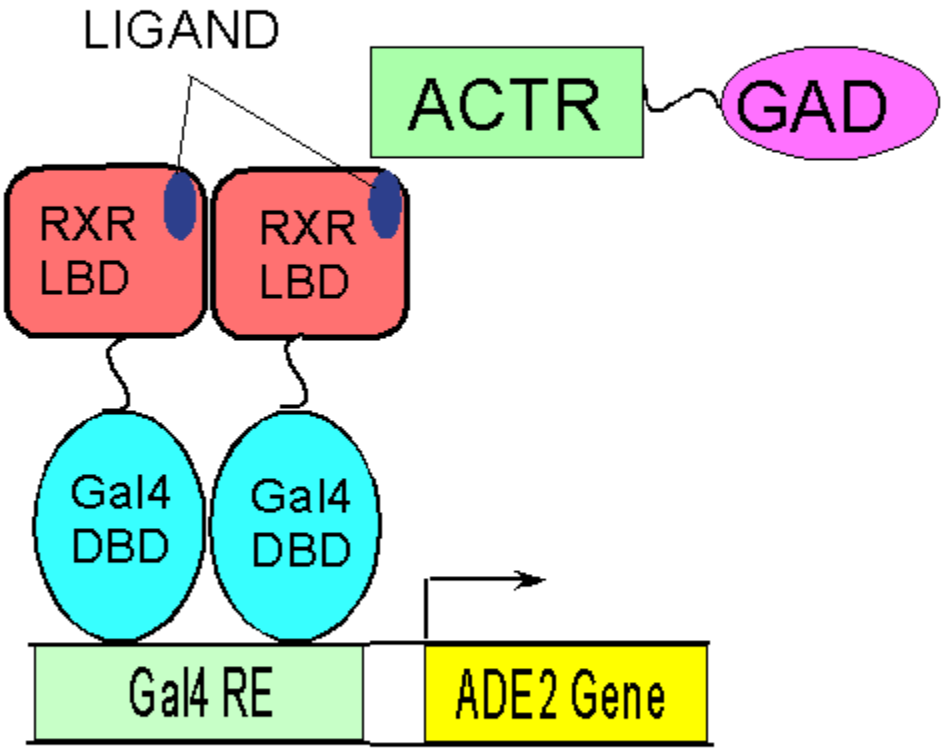


Figure 1.7 Schematic depiction of chemical complementation in yeast.

Literature Cited

1. Wurm, F.M., K.A. Gwinn, and R.E. Kingston, *Inducible overproduction of the mouse c-myc protein in mammalian cells.* Proc Natl Acad Sci U S A, 1986. **83**(15): p. 5414-8.

2. Mayo, K.E., R. Warren, and R.D. Palmiter, *The mouse metallothionein-I gene is transcriptionally regulated by cadmium following transfection into human or mouse cells.* Cell, 1982. **29**(1): p. 99-108.

3. Ryals, J., P. Dierks, H. Ragg, and C. Weissmann, *A 46-nucleotide promoter segment from an IFN-alpha gene renders an unrelated promoter inducible by virus.* Cell, 1985. **41**(2): p. 497-507.

4. Rinsch, C., E. Regulier, N. Deglon, B. Dalle, Y. Beuzard, and P. Aebischer, *A gene therapy approach to regulated delivery of erythropoietin as a function of oxygen tension.* Hum Gene Ther, 1997. **8**(16): p. 1881-9.

5. Jacob, F. and J. Monod, *Genetic regulatory mechanisms in the synthesis of proteins.* J Mol Biol, 1961. **3**: p. 318-56.

6. Hu, M.C. and N. Davidson, *The inducible lac operator-repressor system is functional in mammalian cells.* Cell, 1987. **48**(4): p. 555-66.

7. Hu, M.C. and N. Davidson, *A combination of derepression of the lac operator-repressor system with positive induction by glucocorticoid and metal ions provides a high-level-inducible gene expression system based on the human metallothionein-IIA promoter.* Mol Cell Biol, 1990. **10**(12): p. 6141-51.

8. Labow, M.A., S.B. Baim, T. Shenk, and A.J. Levine, *Conversion of the lac repressor into an allosterically regulated transcriptional activator for mammalian cells.* Mol Cell Biol, 1990. **10**(7): p. 3343-56.

9. Rivera, V.M., T. Clackson, S. Natesan, R. Pollock, J.F. Amara, T. Keenan, S.R. Magari, T. Phillips, N.L. Courage, F. Cerasoli, Jr., D.A. Holt, and M. Gilman, *A humanized system for pharmacologic control of gene expression.* Nat Med, 1996. **2**(9): p. 1028-32.

10. Tang, J. and R.R. Breaker, *Rational design of allosteric ribozymes.* Chem Biol, 1997. **4**(6): p. 453-9.

11. Klug, A., *Zinc finger peptides for the regulation of gene expression.* J Mol Biol, 1999. **293**(2): p. 215-8.

12. Tan, S., D. Guschin, A. Davalos, Y.L. Lee, A.W. Snowden, Y. Jouvenot, H.S. Zhang, K. Howes, A.R. McNamara, A. Lai, C. Ullman, L. Reynolds, M. Moore, M. Isalan, L.P. Berg, B. Campos, H. Qi, S.K. Spratt, C.C. Case, C.O. Pabo, J. Campisi, and P.D. Gregory, *Zinc-finger protein-targeted gene regulation: genomewide single-gene specificity.* Proc Natl Acad Sci U S A, 2003. **100**(21): p. 11997-2002.

13. Beerli, R.R., U. Schopfer, B. Dreier, and C.F. Barbas, 3rd, *Chemically regulated zinc finger transcription factors.* J Biol Chem, 2000. **275**(42): p. 32617-27.

14. Lin, Q., C.F. Barbas, 3rd, and P.G. Schultz, *Small-molecule switches for zinc finger transcription factors.* J Am Chem Soc, 2003. **125**(3): p. 612-3.

15. Segal, D.J., B. Dreier, R.R. Beerli, and C.F. Barbas, 3rd, *Toward controlling gene expression at will: selection and design of zinc finger domains recognizing each of the 5'-GNN-3' DNA target sequences.* Proc Natl Acad Sci U S A, 1999. **96**(6): p. 2758-63.

16. Clackson, T., *Redesigning small molecule-protein interfaces.* Curr Opin Struct Biol, 1998. **8**(4): p. 451-8.

17. Bishop, A.C., O. Buzko, and K.M. Shokat, *Magic bullets for protein kinases.* Trends Cell Biol, 2001. **11**(4): p. 167-72.

18. Bishop, A.C., K. Shah, Y. Liu, L. Witucki, C. Kung, and K.M. Shokat, *Design of allele-specific inhibitors to probe protein kinase signaling.* Curr Biol, 1998. **8**(5): p. 257-66.

19. Hillen, W. and C. Berens, *Mechanisms underlying expression of Tn10 encoded tetracycline resistance.* Annu Rev Microbiol, 1994. **48**: p. 345-69.

20. Gossen, M., S. Freundlieb, G. Bender, G. Muller, W. Hillen, and H. Bujard, *Transcriptional activation by tetracyclines in mammalian cells.* Science, 1995. **268**(5218): p. 1766-9.

21. Fawell, S.E., J.A. Lees, R. White, and M.G. Parker, *Characterization and colocalization of steroid binding and dimerization activities in the mouse estrogen receptor.* Cell, 1990. **60**(6): p. 953-62.

22. Feil, R., J. Brocard, B. Mascrez, M. LeMeur, D. Metzger, and P. Chambon, *Ligand-activated site-specific recombination in mice.* Proc Natl Acad Sci U S A, 1996. **93**(20): p. 10887-90.

23. Vegeto, E., G.F. Allan, W.T. Schrader, M.J. Tsai, D.P. McDonnell, and B.W. O'Malley, *The mechanism of RU486 antagonism is dependent on the*

conformation of the carboxy-terminal tail of the human progesterone receptor. Cell, 1992. **69**(4): p. 703-13.

24. Wang, Y., B.W. O'Malley, Jr., S.Y. Tsai, and B.W. O'Malley, *A regulatory system for use in gene transfer.* Proc Natl Acad Sci U S A, 1994. **91**(17): p. 8180-4.

25. Wang, Y., J. Xu, T. Pierson, B.W. O'Malley, and S.Y. Tsai, *Positive and negative regulation of gene expression in eukaryotic cells with an inducible transcriptional regulator.* Gene Ther, 1997. **4**(5): p. 432-41.

26. Spencer, D.M., T.J. Wandless, S.L. Schreiber, and G.R. Crabtree, *Controlling signal transduction with synthetic ligands.* Science, 1993. **262**(5136): p. 1019-24.

27. Clackson, T., *Controlling mammalian gene expression with small molecules.* Curr Opin Chem Biol, 1997. **1**(2): p. 210-8.

28. Pollock, R. and T. Clackson, *Dimerizer-regulated gene expression.* Curr Opin Biotechnol, 2002. **13**(5): p. 459-67.

29. Mandal, M. and R.R. Breaker, *Gene regulation by riboswitches.* Nature review molecular and cellular biology, 2004. **5**(6): p. 451.

30. Vitreschak, A.G., D.A. Rodionov, A.A. Mironov, and M.S. Gelfand, *Riboswitches: the oldest mechanism for the regulation of gene expression?* Trends Genet, 2004. **20**(1): p. 44-50.

31. Soukup, G.A. and R.R. Breaker, *Engineering precision RNA molecular switches.* Proc Natl Acad Sci U S A, 1999. **96**(7): p. 3584-9.

32. Suess, B., B. Fink, C. Berens, R. Stentz, and W. Hillen, *A theophylline responsive riboswitch based on helix slipping controls gene expression in vivo.* Nucleic Acids Res, 2004. **32**(4): p. 1610-4.

33. Mootz, H.D. and T.W. Muir, *Protein splicing triggered by a small molecule.* J Am Chem Soc, 2002. **124**(31): p. 9044-5.

34. Baulieu, E.E., *Contragestion and other clinical applications of RU 486, an antiprogesterone at the receptor.* Science, 1989. **245**(4924): p. 1351-7.

35. Burcin, M.M., G. Schiedner, S. Kochanek, S.Y. Tsai, and B.W. O'Malley, *Adenovirus-mediated regulable target gene expression in vivo.* Proc Natl Acad Sci U S A, 1999. **96**(2): p. 355-60.

36. Thomas, H.E., H.G. Stunnenberg, and A.F. Stewart, *Heterodimerization of the Drosophila ecdysone receptor with retinoid X receptor and ultraspiracle.* Nature, 1993. **362**(6419): p. 471-5.

37. Yao, T.P., B.M. Forman, Z. Jiang, L. Cherbas, J.D. Chen, M. McKeown, P. Cherbas, and R.M. Evans, *Functional ecdysone receptor is the product of EcR and Ultraspiracle genes.* Nature, 1993. **366**(6454): p. 476-9.

38. No, D., T.P. Yao, and R.M. Evans, *Ecdysone-inducible gene expression in mammalian cells and transgenic mice.* Proc Natl Acad Sci U S A, 1996. **93**(8): p. 3346-51.

39. Yao, T.P., W.A. Segraves, A.E. Oro, M. McKeown, and R.M. Evans, *Drosophila ultraspiracle modulates ecdysone receptor function via heterodimer formation.* Cell, 1992. **71**(1): p. 63-72.

40. Gossen, M. and H. Bujard, *Tight control of gene expression in mammalian cells by tetracycline-responsive promoters.* Proc Natl Acad Sci U S A, 1992. **89**(12): p. 5547-51.

41. Klein, N.C. and B.A. Cunha, *New uses of older antibiotics.* Med Clin North Am, 2001. **85**(1): p. 125-32.

42. Mangelsdorf, D.J., C. Thummel, M. Beato, P. Herrlich, G. Schutz, K. Umesono, B. Blumberg, P. Kastner, M. Mark, P. Chambon, and R.M. Evans, *The nuclear receptor superfamily: the second decade.* Cell, 1995. **83**(6): p. 835-9.

43. Gronemeyer, H., J.A. Gustafsson, and V. Laudet, *Principles for modulation of the nuclear receptor superfamily.* Nat Rev Drug Discov, 2004. **3**(11): p. 950-64.

44. Laudet, V. and H. Gronemeyer, The Nuclear Receptor FactBook, 2002.

45. Rego, E.M., D. Ruggero, C. Tribioli, G. Cattoretti, S. Kogan, R.L. Redner, and P.P. Pandolfi, *Leukemia with distinct phenotypes in transgenic mice expressing PML/RAR alpha, PLZF/RAR alpha or NPM/RAR alpha.* Oncogene, 2006. **25**(13): p. 1974-9.

46. McDonnell, D.P., E. Vegeto, and M.A. Gleeson, *Nuclear hormone receptors as targets for new drug discovery.* Biotechnology (N Y), 1993. **11**(11): p. 1256-61.

47. Krust, A., S. Green, P. Argos, V. Kumar, P. Walter, J.M. Bornert, and P. Chambon, *The chicken oestrogen receptor sequence: homology with v-erbA and the human oestrogen and glucocorticoid receptors.* Embo J, 1986. **5**(5): p. 891-7.

48. Warnmark, A., E. Treuter, A.P. Wright, and J.A. Gustafsson, *Activation functions 1 and 2 of nuclear receptors: molecular strategies for transcriptional activation.* Mol Endocrinol, 2003. **17**(10): p. 1901-9.

49. Folkertsma, S., P. van Noort, J. Van Durme, H.J. Joosten, E. Bettler, W. Fleuren, L. Oliveira, F. Horn, J. de Vlieg, and G. Vriend, *A family-based approach reveals*

the function of residues in the nuclear receptor ligand-binding domain. J Mol Biol, 2004. **341**(2): p. 321-35.

50. Lees, J.A., S.E. Fawell, and M.G. Parker, *Identification of two transactivation domains in the mouse oestrogen receptor.* Nucleic Acids Res, 1989. **17**(14): p. 5477-88.

51. Zilliacus, J., A.P. Wright, J. Carlstedt-Duke, and J.A. Gustafsson, *Structural determinants of DNA-binding specificity by steroid receptors.* Mol Endocrinol, 1995. **9**(4): p. 389-400.

52. Kumar, V. and P. Chambon, *The estrogen receptor binds tightly to its responsive element as a ligand-induced homodimer.* Cell, 1988. **55**(1): p. 145-56.

53. Luisi, B.F., W.X. Xu, Z. Otwinowski, L.P. Freedman, K.R. Yamamoto, and P.B. Sigler, *Crystallographic analysis of the interaction of the glucocorticoid receptor with DNA.* Nature, 1991. **352**(6335): p. 497-505.

54. Schwabe, J.W., L. Chapman, J.T. Finch, and D. Rhodes, *The crystal structure of the estrogen receptor DNA-binding domain bound to DNA: how receptors discriminate between their response elements.* Cell, 1993. **75**(3): p. 567-78.

55. Wurtz, J.M., W. Bourguet, J.P. Renaud, V. Vivat, P. Chambon, D. Moras, and H. Gronemeyer, *A canonical structure for the ligand-binding domain of nuclear receptors.* Nat Struct Biol, 1996. **3**(1): p. 87-94.

56. Moras, D. and H. Gronemeyer, *The nuclear receptor ligand-binding domain: structure and function.* Curr Opin Cell Biol, 1998. **10**(3): p. 384-91.

57. Chen, J.D. and R.M. Evans, *A transcriptional co-repressor that interacts with nuclear hormone receptors.* Nature, 1995. **377**(6548): p. 454-7.

58. Perissi, V., L.M. Staszewski, E.M. McInerney, R. Kurokawa, A. Krones, D.W. Rose, M.H. Lambert, M.V. Milburn, C.K. Glass, and M.G. Rosenfeld, *Molecular determinants of nuclear receptor-corepressor interaction.* Genes Dev, 1999. **13**(24): p. 3198-208.

59. Guenther, M.G., O. Barak, and M.A. Lazar, *The SMRT and N-CoR corepressors are activating cofactors for histone deacetylase 3.* Mol Cell Biol, 2001. **21**(18): p. 6091-101.

60. Nagy, L., H.Y. Kao, D. Chakravarti, R.J. Lin, C.A. Hassig, D.E. Ayer, S.L. Schreiber, and R.M. Evans, *Nuclear receptor repression mediated by a complex containing SMRT, mSin3A, and histone deacetylase.* Cell, 1997. **89**(3): p. 373-80.

61. Nagy, L., H.Y. Kao, J.D. Love, C. Li, E. Banayo, J.T. Gooch, V. Krishna, K. Chatterjee, R.M. Evans, and J.W. Schwabe, *Mechanism of corepressor binding and release from nuclear hormone receptors.* Genes Dev, 1999. **13**(24): p. 3209-16.

62. Onate, S.A., S.Y. Tsai, M.J. Tsai, and B.W. O'Malley, *Sequence and characterization of a coactivator for the steroid hormone receptor superfamily.* Science, 1995. **270**(5240): p. 1354-7.

63. Leo, C. and J.D. Chen, *The SRC family of nuclear receptor coactivators.* Gene, 2000. **245**(1): p. 1-11.

64. Spencer, T.E., G. Jenster, M.M. Burcin, C.D. Allis, J. Zhou, C.A. Mizzen, N.J. McKenna, S.A. Onate, S.Y. Tsai, M.J. Tsai, and B.W. O'Malley, *Steroid receptor coactivator-1 is a histone acetyltransferase.* Nature, 1997. **389**(6647): p. 194-8.

65. Mangelsdorf, D.J., E.S. Ong, J.A. Dyck, and R.M. Evans, *Nuclear receptor that identifies a novel retinoic acid response pathway.* Nature, 1990. **345**(6272): p. 224-9.

66. Bugge, T.H., J. Pohl, O. Lonnoy, and H.G. Stunnenberg, *RXR alpha, a promiscuous partner of retinoic acid and thyroid hormone receptors.* Embo J, 1992. **11**(4): p. 1409-18.

67. Lemotte, P.K., S. Keidel, and C.M. Apfel, *Phytanic acid is a retinoid X receptor ligand.* Eur J Biochem, 1996. **236**(1): p. 328-33.

68. de Urquiza, A.M., S. Liu, M. Sjoberg, R.H. Zetterstrom, W. Griffiths, J. Sjovall, and T. Perlmann, *Docosahexaenoic acid, a ligand for the retinoid X receptor in mouse brain.* Science, 2000. **290**(5499): p. 2140-4.

69. Levin, A.A., L.J. Sturzenbecker, S. Kazmer, T. Bosakowski, C. Huselton, G. Allenby, J. Speck, C. Kratzeisen, M. Rosenberger, A. Lovey, and et al., *9-cis retinoic acid stereoisomer binds and activates the nuclear receptor RXR alpha.* Nature, 1992. **355**(6358): p. 359-61.

70. Egea, P.F., A. Mitschler, N. Rochel, M. Ruff, P. Chambon, and D. Moras, *Crystal structure of the human RXRalpha ligand-binding domain bound to its natural ligand: 9-cis retinoic acid.* Embo J, 2000. **19**(11): p. 2592-601.

71. Mangelsdorf, D.J., K. Umesono, S.A. Kliewer, U. Borgmeyer, E.S. Ong, and R.M. Evans, *A direct repeat in the cellular retinol-binding protein type II gene confers differential regulation by RXR and RAR.* Cell, 1991. **66**(3): p. 555-61.

72. Zhao, Q., S.A. Chasse, S. Devarakonda, M.L. Sierk, B. Ahvazi, and F. Rastinejad, *Structural basis of RXR-DNA interactions.* J Mol Biol, 2000. **296**(2): p. 509-20.

73. James, P., J. Halladay, and E.A. Craig, *Genomic libraries and a host strain designed for highly efficient two-hybrid selection in yeast.* Genetics, 1996. **144**(4): p. 1425-36.

74. Azizi, B., E.I. Chang, and D.F. Doyle, *Chemical complementation: small-molecule-based genetic selection in yeast.* Biochem Biophys Res Commun, 2003. **306**(3): p. 774-80.

CHAPTER 2

INVESTIGATING ADDITIVITY OF MUTATIONS IN RETINOID X RECEPTOR

Introduction

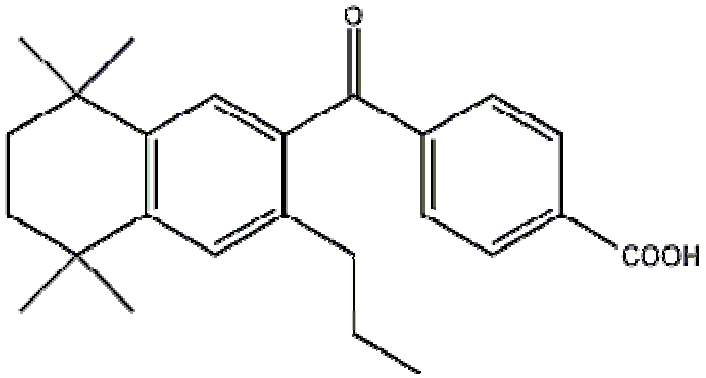
The ability to manipulate proteins to bind and respond to synthetic cell permeable small molecules is a significant challenge in protein engineering and an important tool for many biotechnology applications. The ability to control transcription using a ligand would be useful to study cellular processes such as apoptosis, signal transduction and facilitate the analysis of gene function during embryonic development and differentiation. Manipulating naturally occurring ligand-protein interactions can lead to the creation of orthogonal ligand receptor pairs (OLRP)[1]. OLRP must be orthogonal to the contents of the cell, meaning the receptor should not be activated by any endogenous small molecules and the ligand should not activate or interact with any endogenous proteins. Developing OLRPs can provide insights into the function of thousands of genes discovered in genome projects. It can be utilized to activate or repress transcription of specific gene by selective ligands, providing a useful tool to probe metabolic pathways.

The retinoid X receptor (RXR) is a good candidate for creating variants that bind different ligands through site directed mutagenesis, because side chain atoms, not main chain atoms, contribute the majority of the ligand contacts[2]. RXR is a member of the nuclear receptor superfamily, under the class of retinoid receptors[3, 4]. RXR plays a crucial role in cell differentiation, development and maintenance of homeostasis. This

group of receptors when bound to a selective ligand can single handedly provide a transcriptional signal to specific target genes[5]. The known natural ligand of RXR is 9-*cis*-retinoic acid (9cRA)[6]. RXR forms homodimers or heterodimers with other members of the nuclear receptor superfamily such as retinoid acid receptor (RAR), thyroid receptor, and vitamin D receptor[7]. The structure of RXR contains a variable N-terminal region, DBD, hinge, and LBD (Figure 1.1)[4]. The crystal structure for 9cRA-bound RXR has been solved and the residues that interact with the ligand have been identified (Figure 1.4)[8]. There are 20 hydrophobic and polar amino acids within 4.4 Å of the bound 9cRA[8].

Peet and Doyle constructed a variety of RXR variants to alter its activation in response to various synthetic ligands to create an OLRP[2, 9]. An orthogonal ligand is a compound that is inactive as a wild type RXR agonist but active as a variant RXR agonist. One such orthogonal ligand that was discovered during the drug development of RXR agonists was LG335 (Figure 2.1a). It is the inactive analogue of a synthetic RXR selective activator LG69, an approved drug known as Targretin® (Figure 2.1b)[10, 11]. Targretin® is used to treat AIDS related Kaposi sarcoma and cutaneous T cell lymphoma. LG335 did not bind or activate any RXR subtype[10, 11]. The Doyle group constructed RXR variants to reverse the ligand selectivity from its natural ligand 9cRA to LG335[9]. Site-directed RXR variants gave a variety of activation profiles with 9cRA and LG335. Q275C; I310M; F313I and I310M; F313I; F439L met the criteria for an orthogonal receptor; it is activated by LG335, not activated by all-*trans* retinoic acid, and slightly activated by 9cRA.

A



B

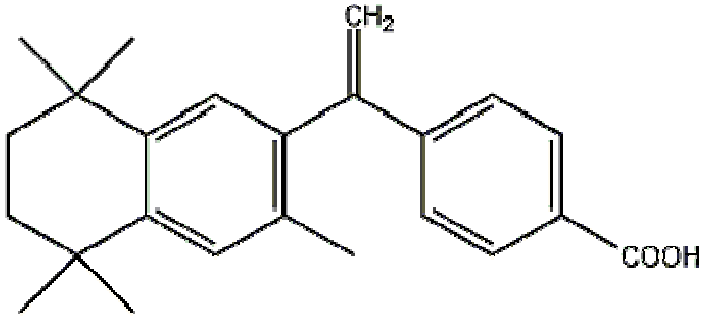
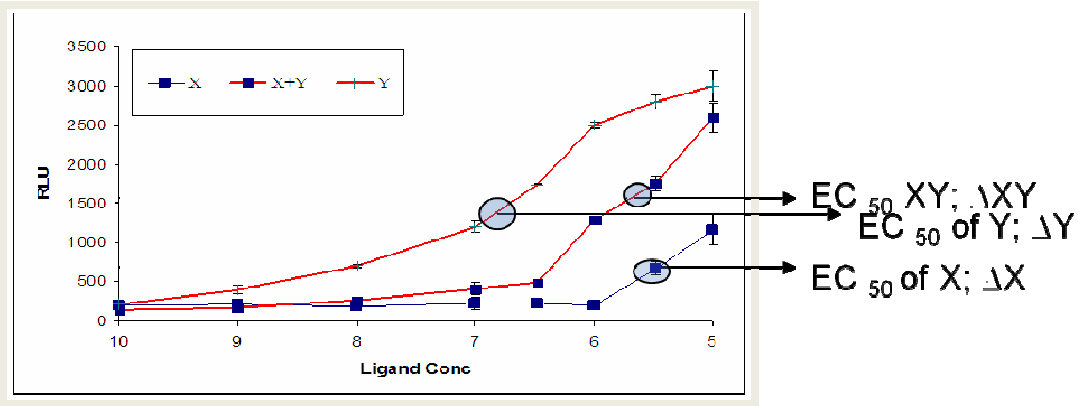


Figure 2.1 A) LG335
B) Targretin® (Bexarotene)

To create an orthogonal receptor such that it gives the desired activation profile with a ligand, it would helpful if the mutations were designed. In this chapter, it is investigated if the mutations in the ligand binding domain have an additive effect on function. If the mutations are additive, receptors may be designed for a specific ligand by making the appropriate mutations in the ligand binding pocket. Additivity means that if one mutation "X" in the LBD has a ΔXEC_{50} (where EC_{50} is the concentration of ligand producing half maximal activity, and ΔXEC_{50} is the EC_{50} of the RXR variant "X" with ligand minus the EC_{50} of the wild-type RXR with ligand), and another mutation in the LBD "Y" has a $\Delta YEC_{50,}$ then the two mutations together RXR "X+Y" have an additive effect, where the experimental ΔXY EC_{50} will be the same as $\Delta XEC_{50} + \Delta YEC_{50}$ (Figure 2.2). If the mutations are additive, the plot between the predicted ΔEC_{50} i.e. $\Delta XEC_{50} + \Delta YEC_{50}$ versus the experimental ΔEC_{50} i.e. ΔXY EC_{50} will have data points falling on the straight line, which has a slope of one and zero intercept. Assuming that a straight line provides a useful mathematical model of this relationship, regression analysis can be used to determine whether this particular straight line model is the best fit for the data.

Results and Discussion

To test if the mutations in the RXR LBD are additive, the RXR variant I310M; F313I; F439L was used. Single and double mutants were made, namely I310M, F313I, F439L, I310M; F313I, F313I; F439L and I310M; F439L (Figure 2.3, and 2.4). These RXR variants were tested in mammalian cell culture with varying



Δ X EC_{50} = Difference of EC_{50} between "X" and wt RXR
Δ Y EC_{50} = Difference of EC_{50} between "Y" and wt RXR
Δ XY EC_{50} = Difference of EC_{50} between "XY" and wt RXR
Additive if: Experimental ΔXY EC_{50} = Δ X EC_{50} + Δ Y EC_{50}

Figure 2.2 Concept of additivity.

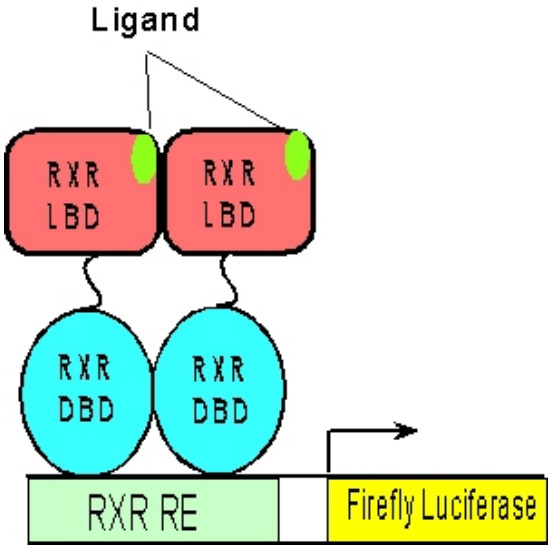


Figure 2.3 Schematic diagram showing ligand bound RXR initiating transcription of firefly luciferase.

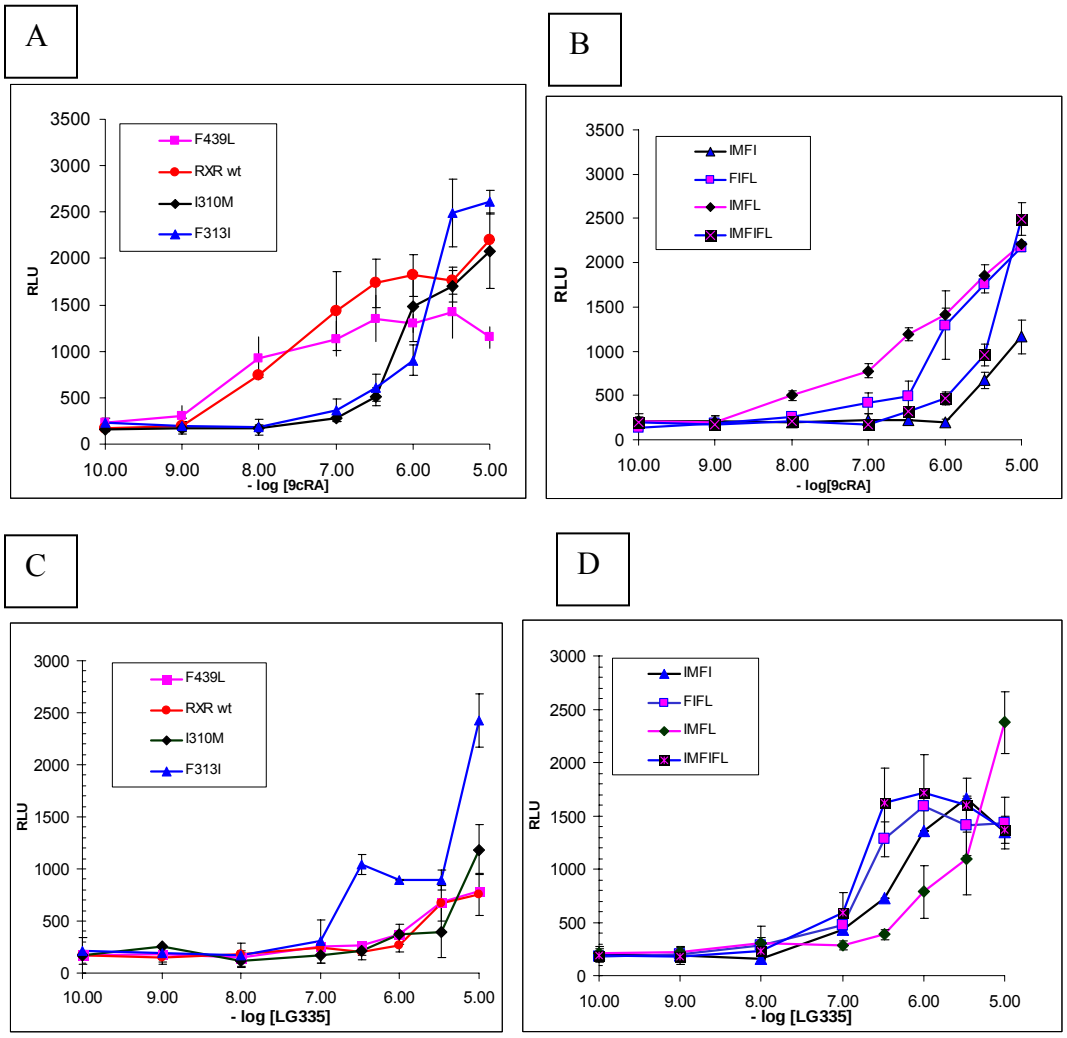


Figure 2.4 Dose response curves:

A) Wild type RXR and single mutants in response to 9cRA
B) RXR double and triple mutants in response to 9cRA
C) Wild type RXR and single mutants in response to LG335
D) RXR double and triple mutants in response to LG335

concentrations of 9cRA and LG335. The dose response curve of I310M; F313I; F439L, single and double mutants are shown in figure 2.4. Based on the response of each variant with LG335 and 9cRA, EC_{50}s, and ΔEC_{50} (variation of EC_{50} of each variant from wild-type RXR), the error of ΔEC_{50} were calculated for each mutant and shown in Table 2.1. To test the additivity, a plot was generated between the experimental ΔEC_{50} of double and triple mutant versus their respective predicted ΔEC_{50} (i.e. experimental ΔEC_{50} I310M; F313I versus predicted ΔEC_{50} i.e. (ΔEC_{50} I310M + ΔEC_{50} F313I)). The predicted ΔEC_{50} of the double and triple mutants are the sums of the ΔEC_{50} of the respective single mutants. If the mutations have an additive effect on the function, the data points will fall on the straight line with an intercept of zero.

The data points between experimental ΔEC_{50} and predicted ΔEC_{50} for the double and triple mutants with 9cRA do not fall on the straight line (Figure 2.5). The best fit line between the data points has a slope of 0.3000, intercept of $5.0 * 10^{-7}$, and R^2 value of 0.6857 (Table 2.2). The slope of 0.3000 means that with an increase of one unit of experimental ΔEC_{50} the value of predicted ΔEC_{50} is estimated to increase by an average of 0.3000 units. The standard error of the slope is 0.1467 (Table 2.2). R^2 value of the line is the coefficient of determination that measures the proportion of variation in the dependent variable i.e. predicted ΔEC_{50} that is explained by the independent variable i.e. experimental ΔEC_{50}, in the regression model. R^2 value is 0.6867; meaning 68% of the variation in the predicted ΔEC_{50} of the mutants can be explained by experimental ΔEC_{50}. The standard error of estimate or standard deviation of the line of regression is $3.9*10^{-7}$ (Table 2.2).

Table 2.1 Experimental EC_{50} and Predicted EC_{50} of RXR variants in HEK 293T cells EC_{50} is the concentration of ligand producing half maximal activity. Error EC_{50} is the standard deviation of EC_{50} of each variant from two experiments. Δ EC_{50} is the difference in EC_{50} between the variant and wild type RXR. Error Δ EC_{50} is the propagation of error in Δ EC_{50}. Values represent averages of experiment repeated twice with each variant in triplicate.

	9cRA					
RXR	Experimental EC50	error EC50	Δ EC50	error ΔEC50	Predicted Δ EC50	Predicted EC50
Wild type	1.0E-07	4.0E-08	0.0E+00			
F439L	4.0E-09	1.6E-09	-9.6E-08	4.0E-09		
I310M	5.6E-07	2.2E-07	4.6E-07	5.6E-07		
F313I	1.3E-06	5.0E-07	1.2E-06	1.3E-06		
I310M;F313I	2.6E-06	1.1E-06	**2.5E-06**	2.6E-06	**1.6E-06**	1.7E-06
I310M;F439L	6.3E-07	2.5E-07	**5.3E-07**	6.3E-07	**3.7E-07**	4.7E-07
F313I;F439L	1.0E-06	4.0E-07	**9.0E-07**	1.0E-06	**1.1E-06**	1.2E-06
I310M;F313I;F439L	4.0E-06	1.6E-06	**3.9E-06**	4.0E-06	**1.5E-06**	1.6E-06

RXR	LG335 Experimental EC50	error EC50	Δ EC50	error ΔEC50	Predicted Δ EC50	Predicted EC50
Wild type	1.6E-06	6.3E-07	0.0E+00			
F439L	2.0E-06	8.0E-07	4.1E-07	2.0E-06		
I310M	4.0E-06	1.6E-06	2.4E-06	4.0E-06		
F313I	7.9E-07	3.2E-07	-7.9E-07	7.9E-07		
I310M;F313I	3.2E-07	1.3E-07	**-1.3E-06**	3.2E-07	**1.6E-06**	3.2E-06
I310M;F439L	2.0E-06	8.0E-07	**4.1E-07**	2.0E-06	**2.8E-06**	4.4E-06
F313I;F439L	1.3E-07	5.0E-08	**-1.5E-06**	1.3E-07	**-3.8E-07**	1.2E-06
I310M;F313I;F439L	1.6E-07	6.3E-08	**-1.4E-06**	1.6E-07	**2.0E-06**	3.6E-06

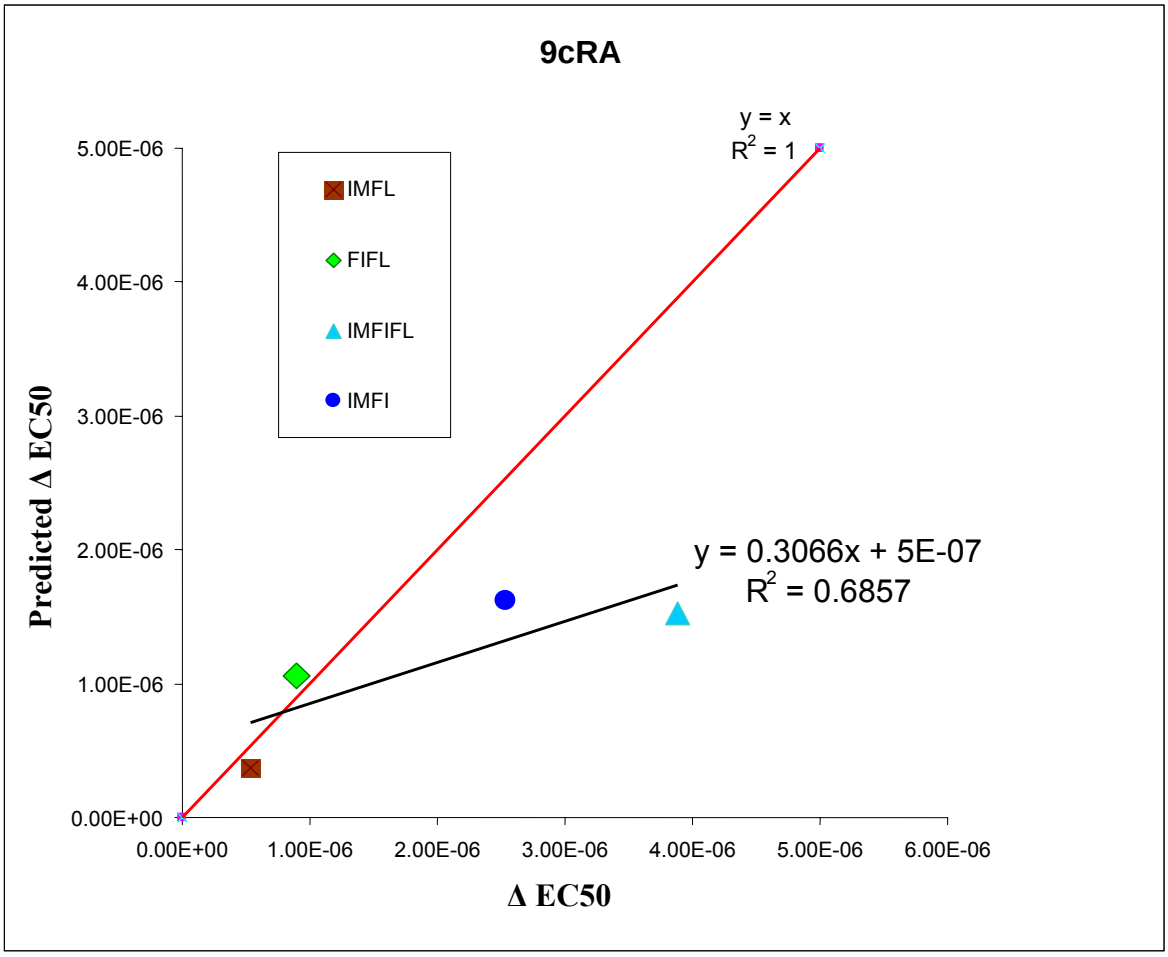


Figure 2.5 Graph of the predicted Δ EC_{50} and the Δ EC_{50} of RXR variants I310M; F439L, F313I; F439L, I310M; F313I; F439L and I310M; F313I in response to 9cRA. For comparison, the line with slope of one and intercept of zero is also shown in red.

Table 2.2 Regression analysis of the RXR variants with 9cRA.

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.828091
R Square	0.685735
Adjusted R Square	0.528602
Standard Error	3.93E-07
Observations	4

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	6.751E-13	6.75E-13	4.364052	0.171908928
Residual	2	3.0939E-13	1.55E-13		
Total	3	9.8449E-13			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	5.43E-07	3.4853E-07	1.557441	0.259672	-9.567E-07	2.042E-06
Slope	0.306604	0.14676826	2.089031	0.171909	-0.324889	0.938096

One can determine whether a linear relationship exists between the predicted ΔEC_{50} and experimental ΔEC_{50} by testing whether the slope (m) is equal to zero. If this hypothesis is rejected, one could conclude that there is evidence of linear relationship.

$H_{o:}$ $m = 0$ (There is no linear relationship),

H_1: $m \neq 0$ (There is a linear relationship).

To test the hypothesis a confidence interval is set up to determine whether the hypothesized value of slope is equal to zero ($m = 0$). The confidence interval estimate of the slope is calculated with the formula $m \pm t_{n-2}\ S_m = .3066 \pm 4.302\ (0.1467)$ (m is the slope of the line; S_m is the standard error of the slope). Hence with 95% confidence the slope lies between $-0.3245 \leq m \leq 0.9377$. The interval includes zero, therefore there is no significant linear relationship between the prediction ΔEC_{50} and experimental ΔEC_{50} of the double and triple mutants. But, when taking into account the propagation of error as seen in Figure 2.6 the uncertainty involved with each data point is relatively large. The propagation of error analysis suggests we cannot predict the additive behavior unless the error is reduced. There are two possible ways we can reduce propagation of error.

A. If the mutants have a large difference in EC_{50} from the wild type receptor, then the ratio of error would be small.
B. If the variation between the experiments reduces then the error involved with each reading will reduce and hence the propagation in error would reduce.

The data points between experimental ΔEC_{50} and predicted ΔEC_{50} for the double and triple mutants with LG335 also do not fall on the straight line (Figure 2.7). The best fit line between the data points has a slope of 1.007. The standard error of the slope is 0.7907, which means the error involved in calculating the slope is large (Table 2.3). The

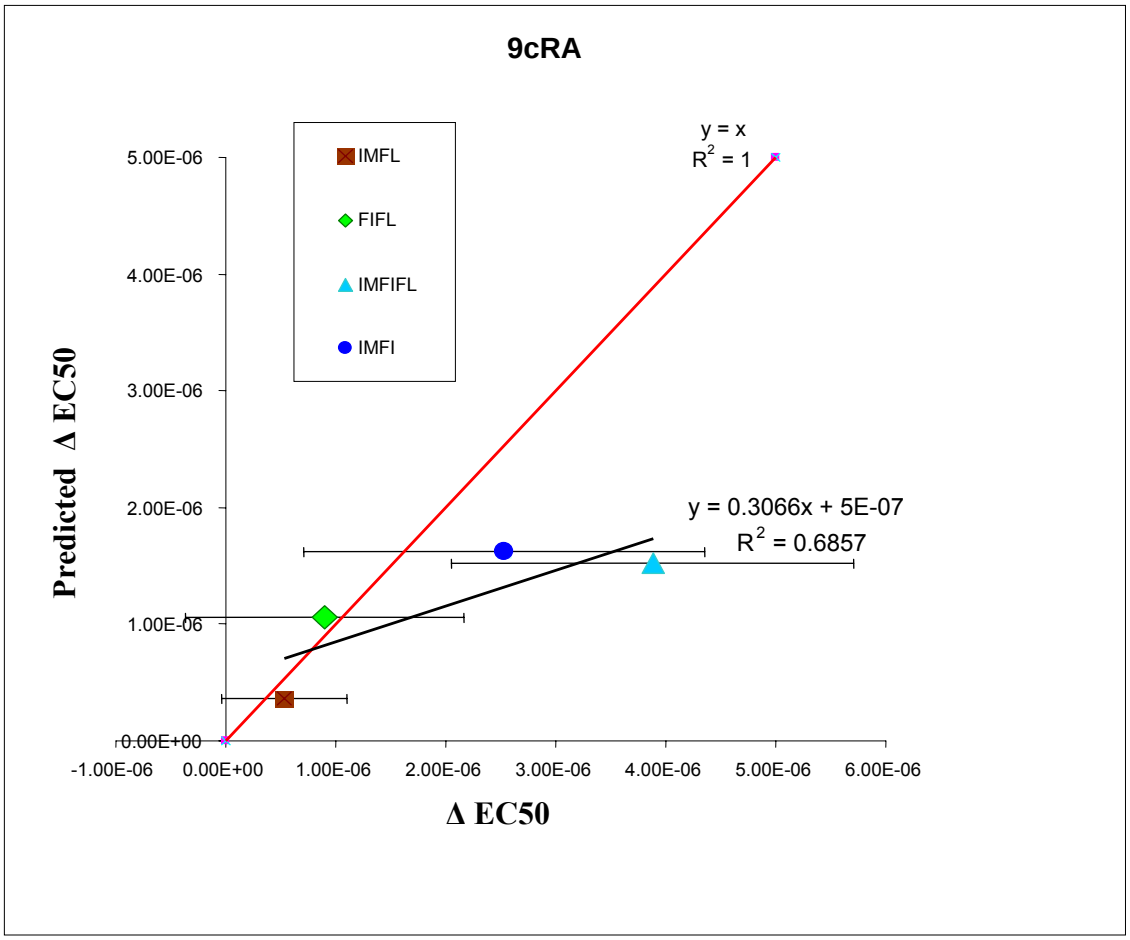


Figure 2.6 Graph of predicted Δ EC_{50} and the Δ EC_{50} of RXR variants I310M; F439L, F313I; F439L, I310M; F313I and I310M; F313I; F439L incorporated with the error of propagation, in response to 9cRA. For comparison, the line with slope of one and intercept of zero is also shown in red.

intercept of the best fit line is $2*10^{-6}$ and the R^2 value is 0.4478 (Table 2.3). R^2 value is the coefficient of determination that measures the proportion of variation in the dependent variable (predicted ΔEC_{50}) that is explained by the independent variable (experimental ΔEC_{50}). In this regression model R^2 value is 0.4478; meaning 44% of the variation in the predicted ΔEC_{50} of the mutants can be explained by experimental ΔEC_{50}. The standard error of estimate or standard deviation of the line of regression is $1.2*10^{-06}$(Table 2.3).

One can determine whether a linear relationship between the predicted ΔEC_{50} and experimental ΔEC_{50} exists by testing whether the slope (m) is equal to zero. If this hypothesis is rejected, one could conclude that there is evidence of linear relationship.

$H_{o:}$ $m = 0$ (There is no linear relationship),

H_1: $m \neq 0$ (There is a linear relationship).

To test the hypothesis a confidence interval is set up to determine whether the hypothesized value ($m = 0$). With 95% confidence the slope lies between $-2.395 \leq m \leq 0.409$. The interval includes zero, therefore these is no significant linear relationship between the prediction ΔEC_{50} and experimental ΔEC_{50} of the double and triple mutants. Also, when taking into account the propagation of error as seen in Figure 2.8 the uncertainty involved is relatively large. The propagation of error analysis suggests that the mutations do not have an additive function within the uncertainties of the EC_{50}'s.

Materials and Methods

Construction of plasmids

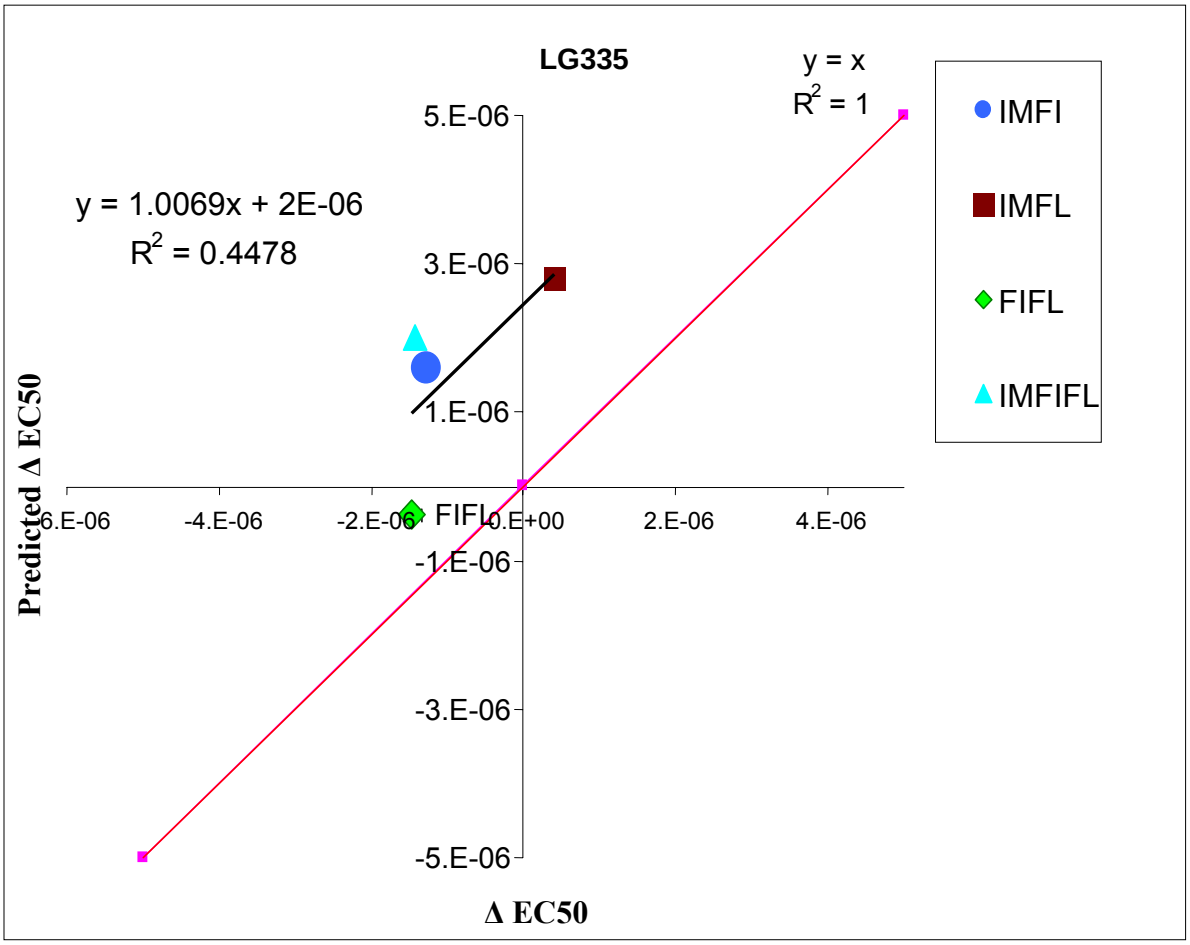


Figure 2.7 Graph of predicted Δ EC_{50} and the Δ EC_{50} of RXR variants I310M; F439L, F313I; F439L, I310M; F313I and I310M; F313I; F439L in response to LG335. For comparison, the line with slope of one and intercept of zero is also shown in red.

Table 2.3 Regression analysis of the RXR variants with LG335.

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.669154
R Square	0.447767
Adjusted R Square	0.171651
Standard Error	1.2345E-06
Observations	4

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	2.471E-12	2.47E-12	1.6216	0.33084
Residual	2	3.047E-12	1.52E-12		
Total	3	5.519E-12			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	2.4544E-06	9.6365E-07	2.546956	0.12573	-1.691E-06	6.600E-06
Slope	1.00691	0.790703	1.273445	0.33084	-2.39520	4.40903

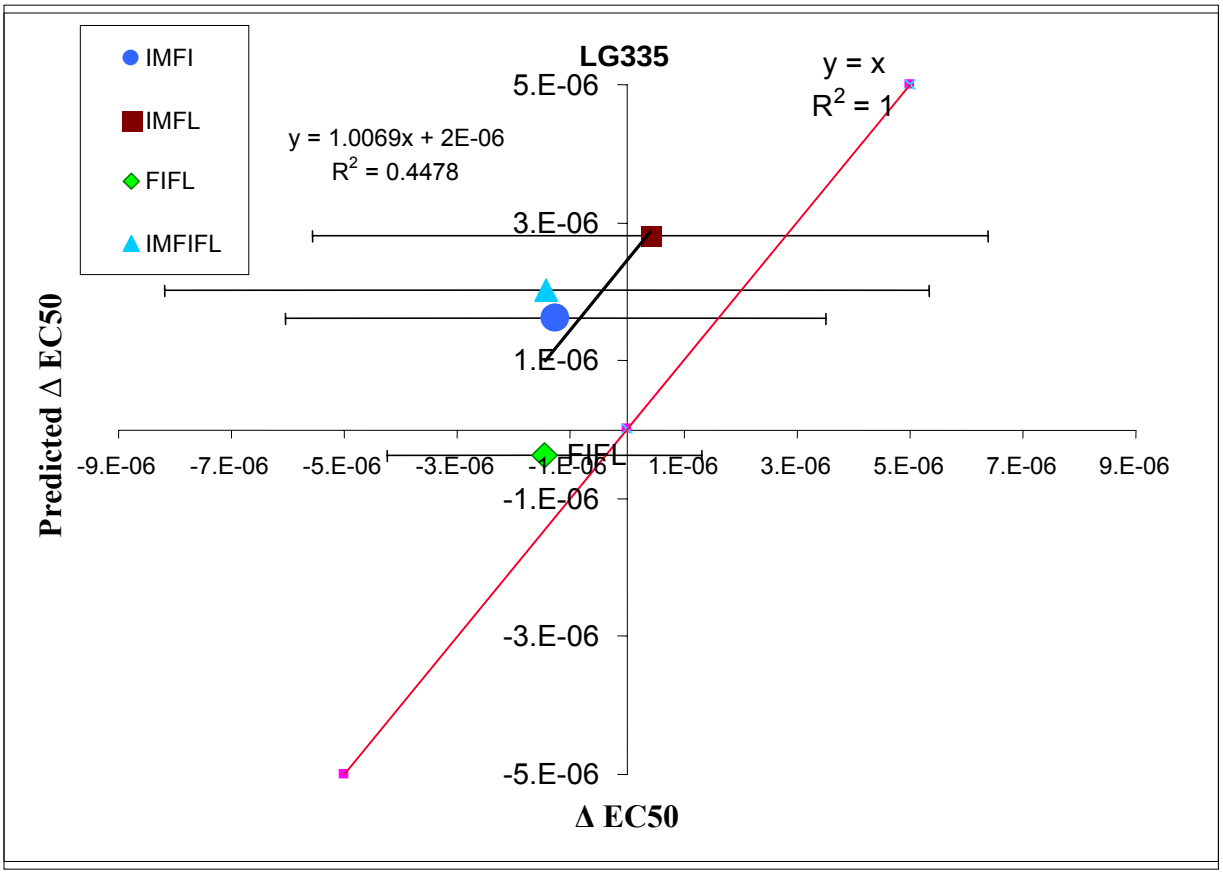


Figure 2.8 Graph of predicted Δ EC_{50} and the Δ EC_{50} of RXR variants I310M; F439L, F313I; F439L, I310M; F313I and I310M; F313I; F439L in response to LG335 with propagation of error incorporated. For comparison, the line with slope of one and intercept of zero is also shown in red.

Mutants RXR I310M, RXR F313I, RXR F439L, RXR I310M; F313I, RXR F313I; F439L and RXR I310M; F439L and RXR I310M; F313I; F439L were made via site directed PCR [9].

Cell culture and transfection

Day prior to the transfection, CV-1 cells were plated with the density of 25000 cells per well in 48-well culture plates. Transfection was performed using lipofectamine cationic lipid (Invitrogen). Each well contained 20 ng of the expression plasmid (RXR variant), 40 ng of pLucCRBPII and pCMXβgal. Eight hours after transfection, the media was removed and replaced with DMEM+10% charcoal resin stripped calf bovine serum containing appropriate concentration of ligand. The cells
were allowed to incubate with ligand for thirty six hours before harvesting. Cell extracts were assayed for luminescence using a luminometer. Activity is reporter in relative light units (RLU) determined as the ratio of the firefly luminescence divided by β gal reading (control).

Literature Cited

1. Belshaw, P.J., J. Schoepfer, K. Liu, K. Morrison, and S. Schreiber, *Rational Design of Orthogonal Receptor-Ligand Combinations.* Angew. Chem., Int. Ed. Eng, 1995. **34**: p. 2129-2132.

2. Peet, D.J., D.F. Doyle, D.R. Corey, and D.J. Mangelsdorf, *Engineering novel specificities for ligand-activated transcription in the nuclear hormone receptor RXR.* Chem Biol, 1998. **5**(1): p. 13-21.

3. Mangelsdorf, D.J., C. Thummel, M. Beato, P. Herrlich, G. Schutz, K. Umesono, B. Blumberg, P. Kastner, M. Mark, P. Chambon, and R.M. Evans, *The nuclear receptor superfamily: the second decade.* Cell, 1995. **83**(6): p. 835-9.

4. Laudet, V. and H. Gronemeyer, The Nuclear Receptor FactBook, 2002.

5. Weatherman, R.V., R.J. Fletterick, and T.S. Scanlan, *Nuclear-receptor ligands and ligand-binding domains.* Annu Rev Biochem, 1999. **68**: p. 559-81.

6. Heyman, R.A., D.J. Mangelsdorf, J.A. Dyck, R.B. Stein, G. Eichele, R.M. Evans, and C. Thaller, *9-cis retinoic acid is a high affinity ligand for the retinoid X receptor.* Cell, 1992. **68**(2): p. 397-406.

7. Mangelsdorf, D.J. and R.M. Evans, *The RXR heterodimers and orphan receptors.* Cell, 1995. **83**(6): p. 841-50.

8. Egea, P.F., A. Mitschler, N. Rochel, M. Ruff, P. Chambon, and D. Moras, *Crystal structure of the human RXRalpha ligand-binding domain bound to its natural ligand: 9-cis retinoic acid.* Embo J, 2000. **19**(11): p. 2592-601.

9. Doyle, D.F., D.A. Braasch, L.K. Jackson, H.E. Weiss, M.F. Boehm, D.J. Mangelsdorf, and D.R. Corey, *Engineering orthogonal ligand-receptor pairs from "near drugs".* J Am Chem Soc, 2001. **123**(46): p. 11367-71.

10. Boehm, M.F., L. Zhang, B.A. Badea, S.K. White, D.E. Mais, E. Berger, C.M. Suto, M.E. Goldman, and R.A. Heyman, *Synthesis and structure-activity relationships of novel retinoid X receptor-selective retinoids.* J Med Chem, 1994. **37**(18): p. 2930-41.

11. Boehm, M.F., M.R. McClurg, C. Pathirana, D. Mangelsdorf, S.K. White, J. Hebert, D. Winn, M.E. Goldman, and R.A. Heyman, *Synthesis of high specific activity [3H]-9-cis-retinoic acid and its application for identifying retinoids with unusual binding properties.* J Med Chem, 1994. **37**(3): p. 408-14.

CHAPTER 3

ENGINEERING AN ORTHOGONAL LIGAND RECEPTOR PAIR

Introduction

Small molecules that modulate the activity of biological signaling such as agonists and antagonists for receptors are a powerful tool for studying protein function. To date, many receptors have being engineered to respond to unnatural or synthetic ligand. One of the pioneer works in engineering proteins to accept an unnatural substrate by altering electrostatic interaction was done by Hwang and Miller, where specificity of elongation factor Tu, a GTPase, was altered to an XTPase by a single mutation of aspartate 138 to an aspargine[1]. This work started the era of protein engineering. Another earlier example of protein engineering was engineering κ-opioid, a seven transmembrane receptor to respond only to a synthetic ligand[2]. This change was done by swapping an extracellular loop from γ-opioid receptor Ro1 and by making the point mutation glutamic acid 297 to glutamine in the first external loop[2]. Shokat and coworkers used a structure-based design to engineer the ATP-binding site of v-Src, a protein kinase to uniquely accept N^6 ATP analogs by mutating the bulky residue in the active site, isoleucine 338, to an alanine or a glycine[3, 4]. Other notable work was done by Smith and coworkers on an orthogonal human carboxypeptidase A1 (hCPA1). They engineered hCPA1, such that the enzyme hydrolyzed an unnatural substrate methotrexate (MTX) prodrugs[5]. This change was achieved by synthesizing derivatives of MTX-Phe

with bulky substituents at the positions 2 and 3 of MTX and making mutants of hCPA1 at threonine 268 to have smaller amino acids such as glycine or alanine[5].

Creating a matched pair of a synthetic ligand and an engineered receptor that are orthogonal to the wild type interaction is called an orthogonal ligand receptor pair (OLRP)[6]. OLRP can provide insight in structural specificity between various ligand-receptor pairs, and to understand cellular and metabolic pathways. OLRPs offer the ability to control transcription by behaving as a small molecule dependent molecular switch. These molecular switches can be used for practical applications such as gene therapy by regulating the expression of target gene[7]. The molecular switches or OLRP may behave as a rheostat or as an on/off switch depending on their response to the ligand. To utilize these switches for gene therapy, genes for orthogonal receptor or molecular switch would be delivered along with the target gene (Figure 1.2). The molecular switch once expressed in the cells would bind to its unique binding site in the promoter region of the target gene. In the presence of ligand, the receptor will be activated and express the downstream target gene. Depending on the dosage of the ligand and the activation level of the receptor the target gene will be expressed. OLRP can also be utilized to reveal the function for thousands of gene that were discovered in the genome project with unknown function.

Nuclear receptors (NR) are good candidates for creating OLRPs. NR's are a super-family of ligand activated transcription factors that are composed of steroids and non steroidal receptors which control a broad range of physiological processes. These proteins contain an evolutionarily conserved DNA binding domain (DBD) and ligand-binding domain (LBD). The LBD binds to a small molecule and the DBD binds to a

DNA sequence called response element (RE). NRs are modular proteins, such that alterations within the ligand binding domain will not disrupt the DNA binding domain and vice versa. The proteins are attractive targets for constructing novel ligand receptor pairs and controlling transcription for two reasons: 1) Small molecule derivatives are available and can be synthesized to bind and activate these receptors. 2) NR response elements can be introduced into the promoter regions of any gene and hence enabling the regulation of gene expression.

Retinoid X receptor (RXR) is a nuclear receptor that belongs to the class of retinoid receptors[8]. Retinoid receptors contain of retinoic acid receptors (RAR) that bind and respond to all-*trans* retinoic acid, and the RXR is activated by the retinoic acid stereoisomer 9-*cis* retinoic acid (Figure 1.3)[9]. RXR forms a homodimer with itself (Figure 1.5) and heterodimers with other nuclear receptors such as thyroid hormone, vitamin D receptor, peroxisome proliferator-activated receptor and RAR[10, 11]. These heterodimers have distinct response element specificities to mediate the hormonal response to target gene transcription. Response elements are comprised of direct repeats (DRs) of a common half-site with the spacing between repeats which is a critical in mediating specificity. RARs activate preferentially through DRs spaced by two or five nucleotides, whereas RXR homodimers activate preferentially through DRs spaced by one nucleotide[12, 13]. Heterodimeric complexes of RXR-RAR have shown to mediate transcription via response elements composed of palindromes as well as inverted palindromes. The dimerization interface has been identified within the DBD of the receptors that promotes DNA binding[14, 15]. But there is additional dimerization interface within the LBD of RAR, and RXR[16-18]. A 40-amino acid region in RXR

(amino acid 389-429) and RAR constitute critical regions in their heterodimerization and efficient DNA binding[19, 20].

Koh and colleagues at the University of Delware have rationally designed retinoid acid receptor, a member of NR superfamily, to respond to novel synthetic ligands and selectively activate genes [7]. Crystal structure of all *trans* retinoic acid (atRA) with RAR was used to design a receptor ligand pair. Two residues serine 289 and arginine 278, were shown to interact with carboxylic acid moiety of atRA, and these residues were critical for electrostatics and for ligand binding. Mutation of one or both of these residues to glycine or negatively charged residues such as aspartic or glutamic acid results in RAR activation by neutral and basic charged analogs of RA, and not its natural ligand, negatively charged atRA. Koh's group also created an orthogonal estrogen receptor-ligand pair by polar group exchange[8]. They changed the covalent connectivity of polar groups involved in an intra-molecular protein salt-bridge. The crystal structure of the human estrogen receptor (ER) and estradiol shows that the 3-hydroxyl of estrogen is held via hydrogen bonds to an intra-molecular protein salt-bridge formed by glutamine (G) 353 and arginine (A) 394 [9]. 3-hydroxyl group of estradiol and G353 have been shown to be important for the selectivity [10, 11]. They substituted the carboxylate of LG353 with alanine, and the variant ER G353A, lacks a carboxyl group critical for high-affinity binding of estradiol. This variant is able to activate by a carboxylate-functionalized estrogen analog, ES8 at nano molar concentration, whereas wild type ER is not. Ligand ES8 activates ER G353A at concentrations that do not activate the wild type ER.

Previously, Doyle and coworkers in an attempt to create OLRP's, constructed a variety of RXR variants to respond to various synthetic ligands[21, 22]. One of the

orthogonal ligand LG335, an inactive analogue of a synthetic RXR selective activator LG69 also known as Targretin® did not bind or activate any wild type RXR subtype (Figure 2.1) [23, 24]. The Doyle group constructed RXR variants to reverse the ligand selectivity from its natural ligand 9cRA to LG335 via site directed mutagenesis. They converted the RXR residues 3.9 Å of any ligand carbon atom to their RAR counterparts, with the reasoning that the changes will alter the specificity for ligand without major structural modification of RXR. Three residues within the 3.9 Å of the oxygen atoms of the carboxylate were untouched as LG335 also contain the carboxylic group. RXR variants gave a variety of activation profiles with 9cRA and LG335. RXR Q275C; I310M; F313I (QCIMFI) was activated by LG335 and slightly activated by 9cRA (Figure 3.1).

Here the RXR variant QCIMFI was modified to create a fully orthogonal receptor and behave as a molecular switch to control transcription. For the engineered ligand receptor pair to behave like a molecular switch and aid in gene therapy it should meet the following criteria (Figure 3.2)

1. The DNA binding domain should only recognize a unique binding site
2. The ligand binding domain should preferentially bind to a synthetic small molecule and not any endogenous ligand.
3. The receptor should not dimerize with wild type RXR or similar protein such as RAR

To engineer a fully OLRP that fulfils all the above criteria's, the DBD of RXR variant QCIMFI was replaced with the yeast transcription factor DBD, Gal4 DNA

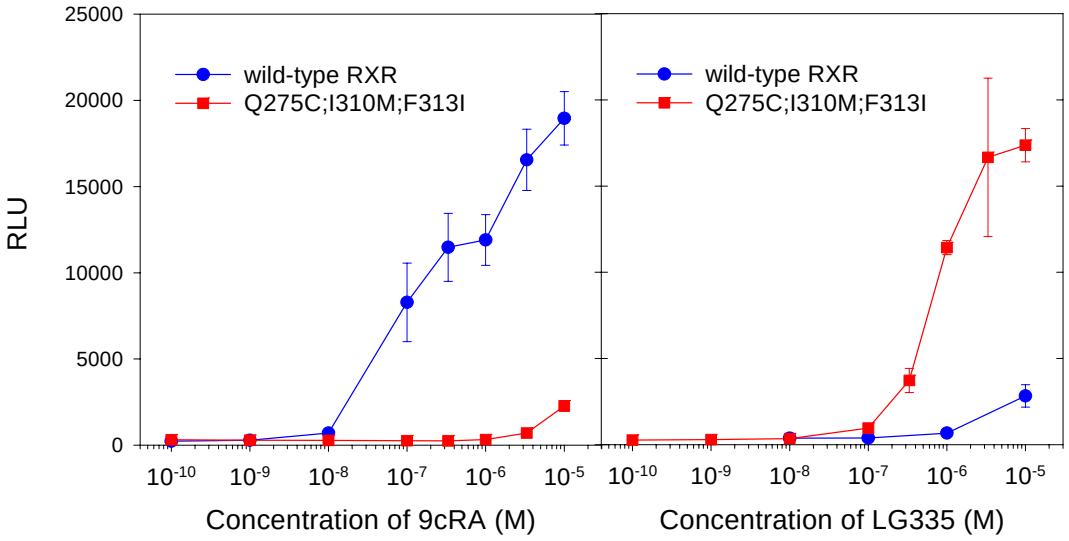


Figure 3.1 Dose response curves of wild type RXR and RXR variant Q275C; I310M; F313I with 9cRA and LG335.

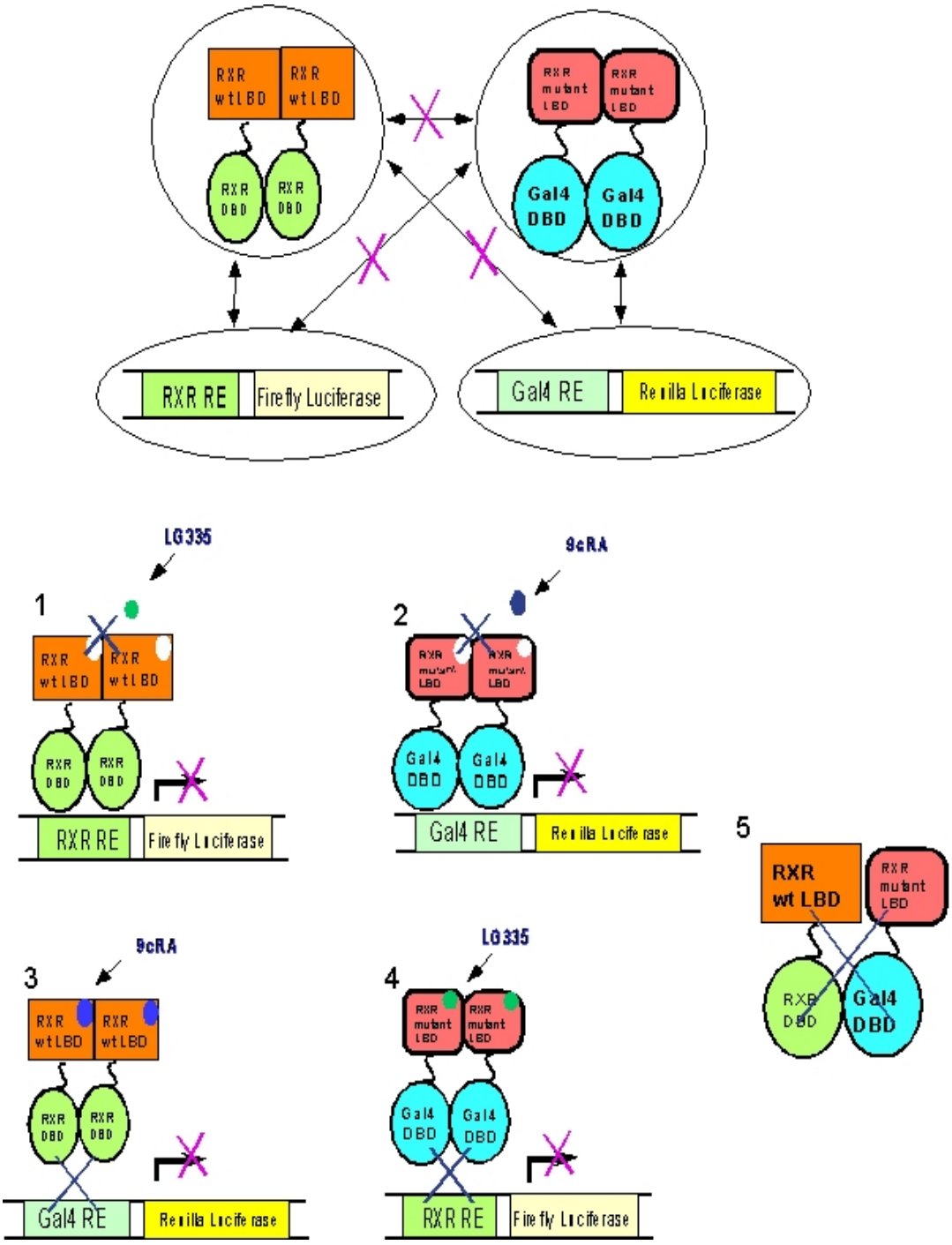


Figure 3.2 Experimental setup to show the GR Q275C; I310M; F313I and LG335 are orthogonal. Tested for interactions between 1) Wild type receptor and LG335. 2) GR variant and ligand (LG335, 9cRA or atRA) 3) Wild type RXR and RAR with Gal4 response element. 4) GR variant and Gal4 RE and RXR RE. 5) Dimerization between the wild type receptor and the GR variant.

binding domain (Gal4 DBD). The variant Gal4 DBD- RXR QCIMFI (GR QCIMFI) would solve two purposes; it would prevent activation of any RXR responsive genes i.e. genes controlled by RXR RE and the wild type RXR will not bind and activate Gal4 RE driven target gene.

Results and Discussion

Response element orthogonality

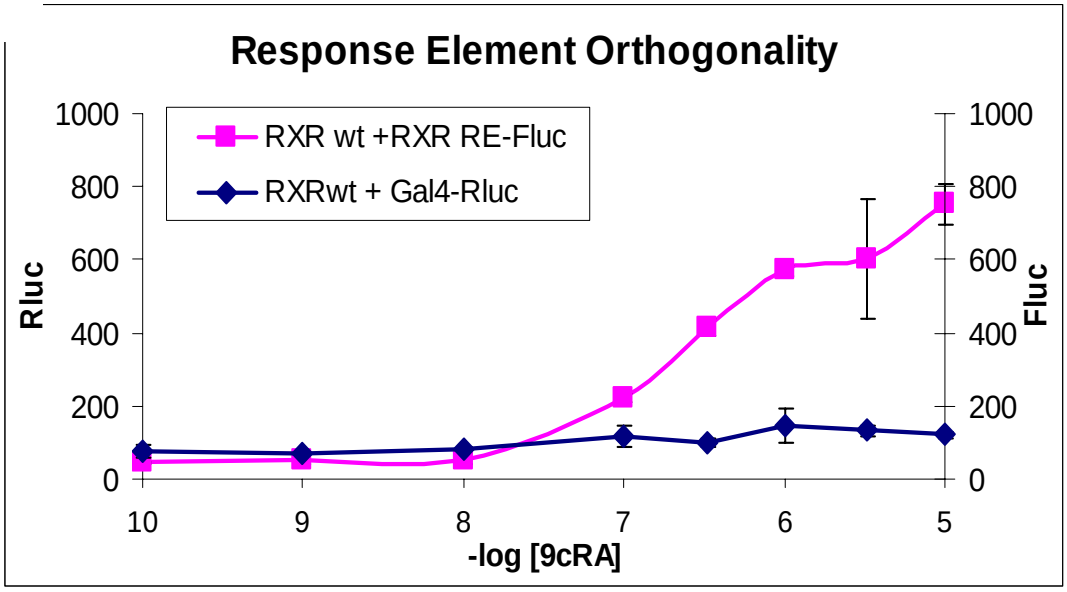
To achieve target specificity and to eliminate the possibility of simultaneous activation of any endogenous RXR responsive genes the DNA binding domain of the RXR variant was replaced by Gal4 DBD. Gal4 DBD is a yeast transcription factor that recognizes a unique DNA sequence of 17 base pairs called Gal4 RE. Gal4 activated genes are not known to exist in mammalian cells, hence this regulator should be specific for the target gene of interest. In theory, in the presence of LG335 the GR variant should only activate the target gene containing the Gal4 binding sites. HEK293T cells were co-transfected with plasmids expressing GR variant QCIMFI (pCMXGR QCIMFI), wild type RXR (pCMX RXR), four repeats of Gal4 RE driven *Renilla* luciferase (p17*4Tata Rluc) and RXR RE driven firefly luciferase (pLucCRBPII). In the presence of LG335, the GR variant was activated and expressed Gal4 RE driven *Renilla* luciferase but did not bind to RXR RE and hence did not activate RXR RE driven firefly luciferase (Figure 3.3). In the presence of 9cRA, wild type RXR was activated and expressed RXR RE controlled firefly luciferase but did not bind Gal4 RE or express *Renilla* luciferase

(Figure 3.3). This data shows that Gal4 DBD recognizes only the Gal4 RE not the RXR RE, and RXR does not recognize Gal4 RE. To further examine the specificity of Gal4 DBD and Gal4 RE, it was tested with retinoid acid receptor and RAR response element.

RAR belongs to the same family of retinoid receptors as the RXR. It is structurally similar to RXR and shares the common modular domains that can be interchanged without the loss of function [25, 26]. RAR and RXR form heterodimers that bind a specific DNA sequence called the RAR response element [13, 27]. HEK293T cells were co-transfected with plasmids expressing the GR variant QCIMFI, wild type RXR, Gal4 RE driven *Renilla* luciferase and RAR RE driven firefly luciferase. The result shows that in the presence of LG335, GR variant did not bind RARE and activate firefly luciferase (Figure 3.4). In the presence of atRA, wild type RAR did not bind and activate Gal4 RE controlled firefly luciferase (Figure 3.4). This result confirms the target specificity of the GR variant with Gal4 RE driven target gene and that RAR or RXR does not recognize Gal4 RE. This result confirms the target specificity of the GR variant with Gal4 RE driven target gene.

Testing for ligand orthogonality

Altering the ligand selectivity requires changing the binding pocket residues of the LBD such that it is activated only in the presence of the synthetic drug. Upon binding the ligand, the LBD of the nuclear receptor undergoes a conformational change that leads to the release of co-repressor proteins and the recruitment of co-activators. Helix 12 of



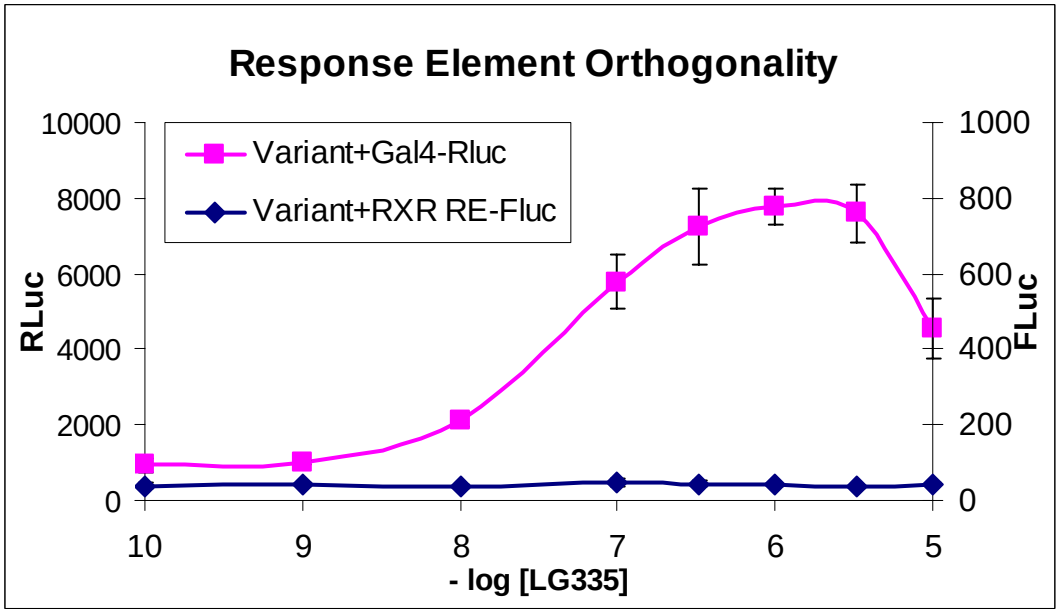
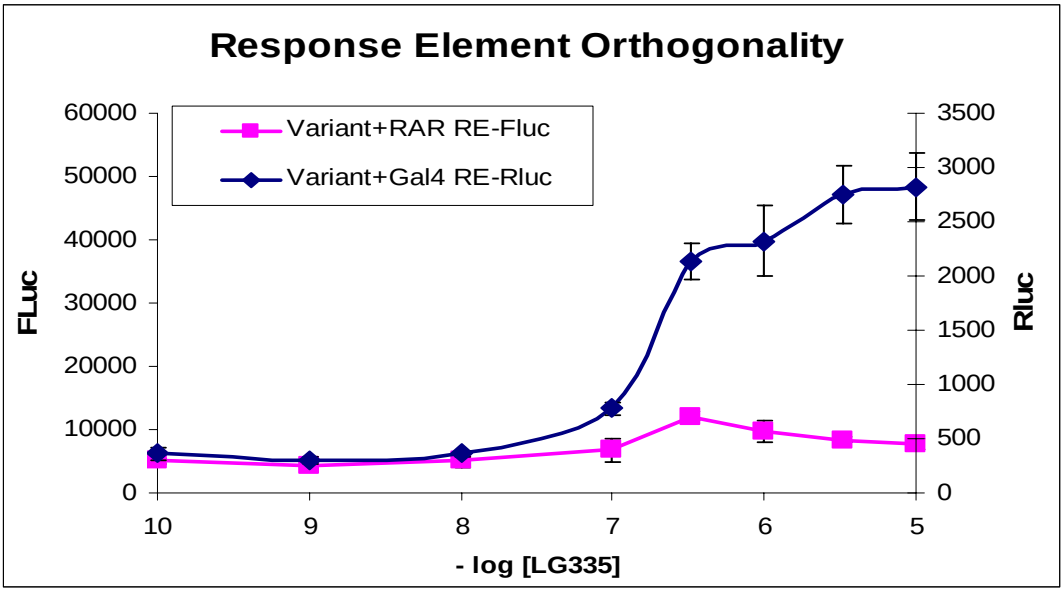


Figure 3.3 Activation profile of wild type RXR and GR Q275C; I310M; F313M in response to 9cRA and LG335, respectively. Two different reporter plasmids are employed containing the Gal4 response element driven *Renilla* luciferase or the RXR response element driven firefly luciferase.



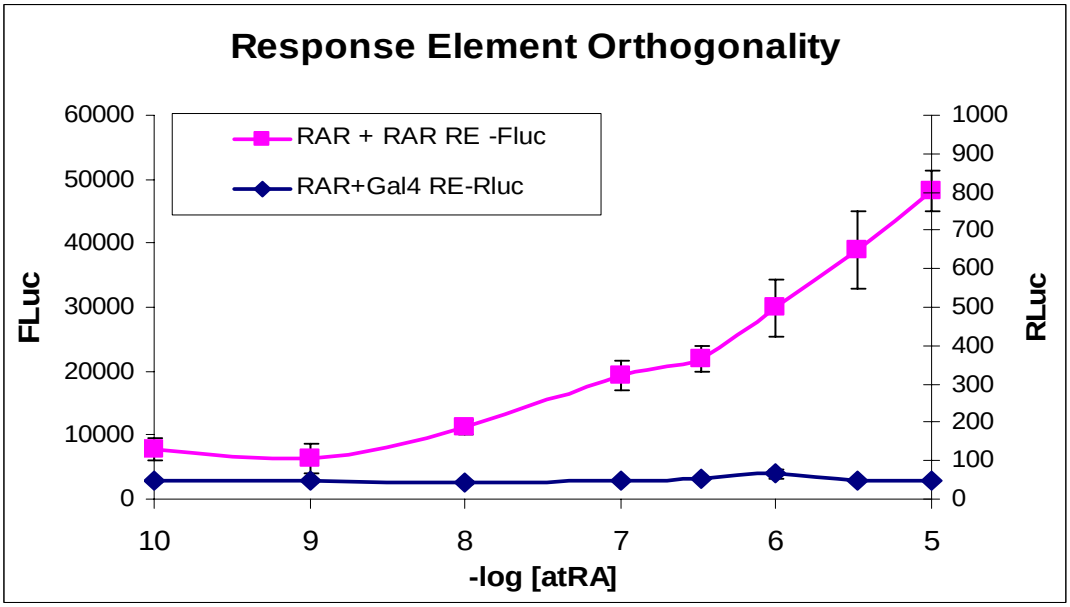
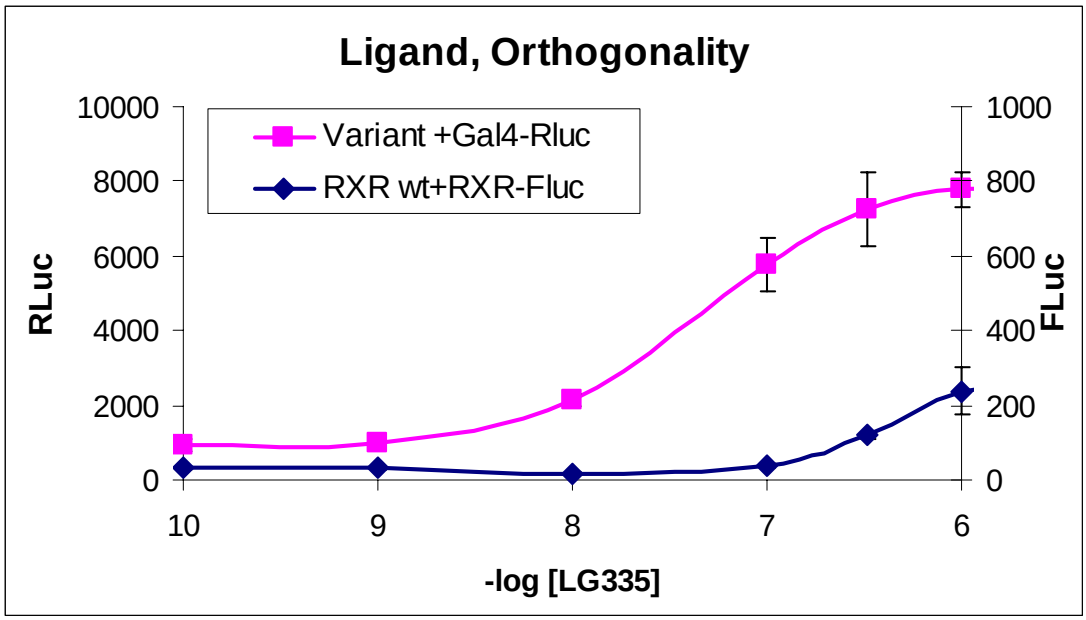


Figure 3.4 Activation profile of wild type RAR and GR Q275C; I310M; F313M in response to atRA and LG335, respectively. Two different reporter plasmids are employed containing the Gal4 response element driven *Renilla* luciferase or the RAR response element driven firefly luciferase.

the LBD plays a central role in controlling the ligand-induced conformational change and forms a part of the receptor/co-activator interface. This aspect of changing the RXR LBD

residues was previously performed by Doyle and coworkers in CV-1 cell line (green monkey kidney cells)[22]. Three residues Q275, I310 and F313 that are within the 4.4 Å of the bound 9cRA were mutated to cysteine, methionine and isoleucine respectively. These mutations were made so that the RXR variant would be preferentially activated by LG335 and not its natural ligand, 9cRA (Figure 3.1). Mutations at I310 and F313 would provide space for the propyl group of LG335. The RXR DNA binding domain of this variant was replaced with Gal4 DNA binding domain as explained in the previous section (GR variant). HEK293T cells were transfected with plasmids encoding for the wild-type RXR, GR QCIMFI, Gal4RE linked renilla luciferase gene and RXR RE driven firefly luciferase gene respectively. The GR QCIMFI had the same response with the ligands in HEK293T cells as in CV-1 cells (Figure 3.5). The mutation caused an altered specificity with little activation in response to 9cRA and dose response with LG335 (Figure 3.5). The EC_{50} of GR QCIMFI with LG335 was 50 nM and with 9cRA was 100 nM. On the other hand, the wild type RXR was not activated with LG335.

To further examine ligand orthogonality, the GR variant was tested with atRA and wild type RAR was tested with LG335. RAR shares 27% sequence identity with RXR in their LBD. If the GR variant is orthogonal to the ligand atRA, ligand orthogonality would be reinforced. In the presence of atRA, GR QCIMFI was not activated and did not express *Renilla* luciferase also wild type RAR did not express RAR RE driven firefly *Renilla* luciferase also wild type RAR did not express RAR RE driven firefly luciferase



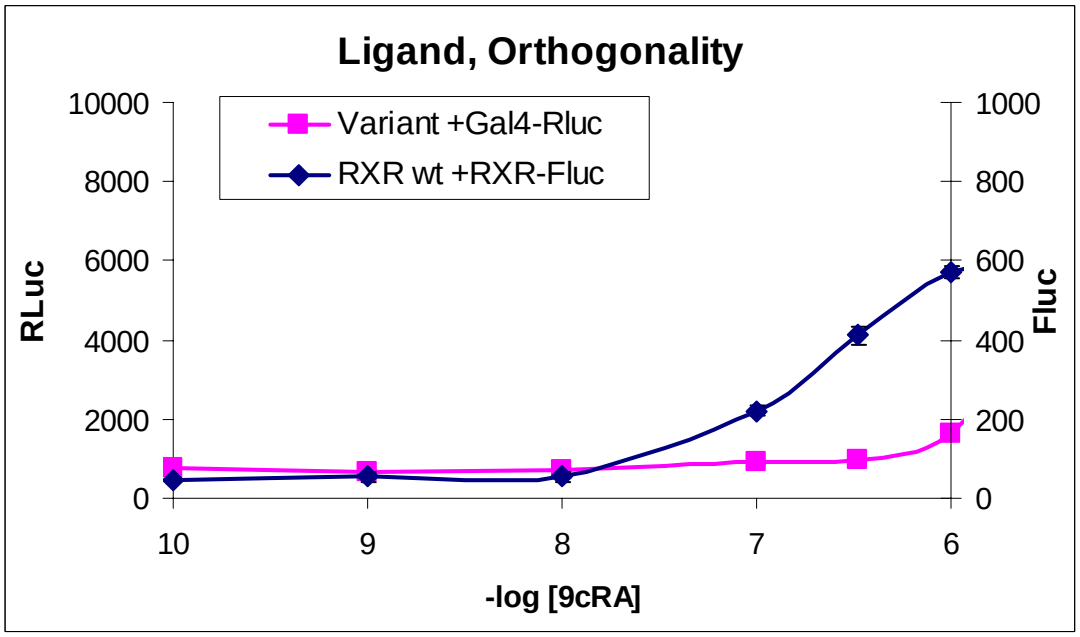


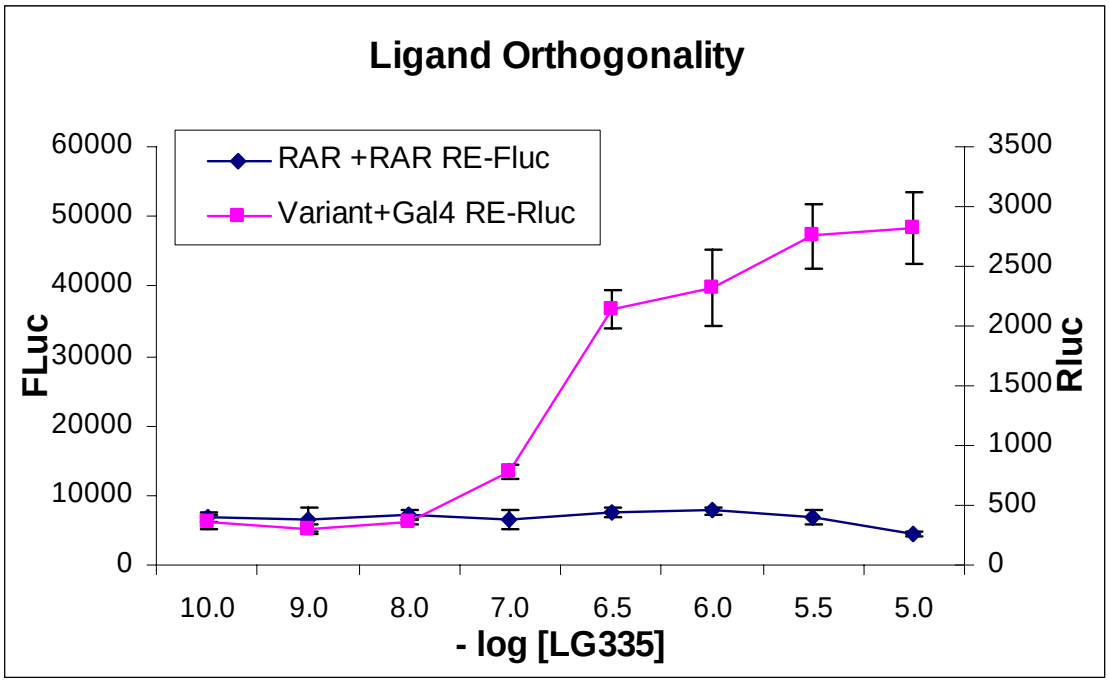
Figure 3.6 Activation profile of wild type RXR and GR Q275C; I310M; F313I in response to 9cRA and LG335.

in the presence of LG335 (Figure 3.6). This data shows that GR QCIMFI was activated by LG335, slightly activated by 9cRA and not activated by atRA. Also, RXR and RAR were not activated by the orthogonal ligand LG335.

To explore which mutation or combination of mutations in the LBD of RXR were critical for the ligand based orthogonal behavior, the variant RXR was characterized. Single mutants of RXR Q275C, I313M and F313I containing the RXR DBD were constructed and co-transfected in mammalian cell culture[22]. The result was analyzed using a reporter plasmid containing the RXR RE linked to firefly luciferase. The transfection result from the RXR variants Q275C, I310M, F313I individually and their combination gave a variety of EC_{50} with LG335 and 9cRA ranging from .1 μM to 1 μM (Figure 3.7). RXR variant Q275C had a low activation level with both 9cRA and LG335, F313I was activated with both 9cRA and LG335, and I310M had low activation with LG335 and slightly activated with 9cRA. Hence all the single mutations were not orthogonal. In combination of two single mutant's co-transfected, Q275C and I310M together was slightly activation with LG335, Q275C and F313I was activated by both 9cRA and LG335 to equal levels. RXR variant I310M and F313I together were activated by LG335 but the EC_{50} was 500 nM, which was higher then the triple mutant. In addition, with 9cRA this mutant was slightly activated. Hence all the three mutations are required for RXR variant to behave as an OLRP.

Dimerization Orthogonality

The last step towards making the variant GR QCIMFI orthogonal is changing the



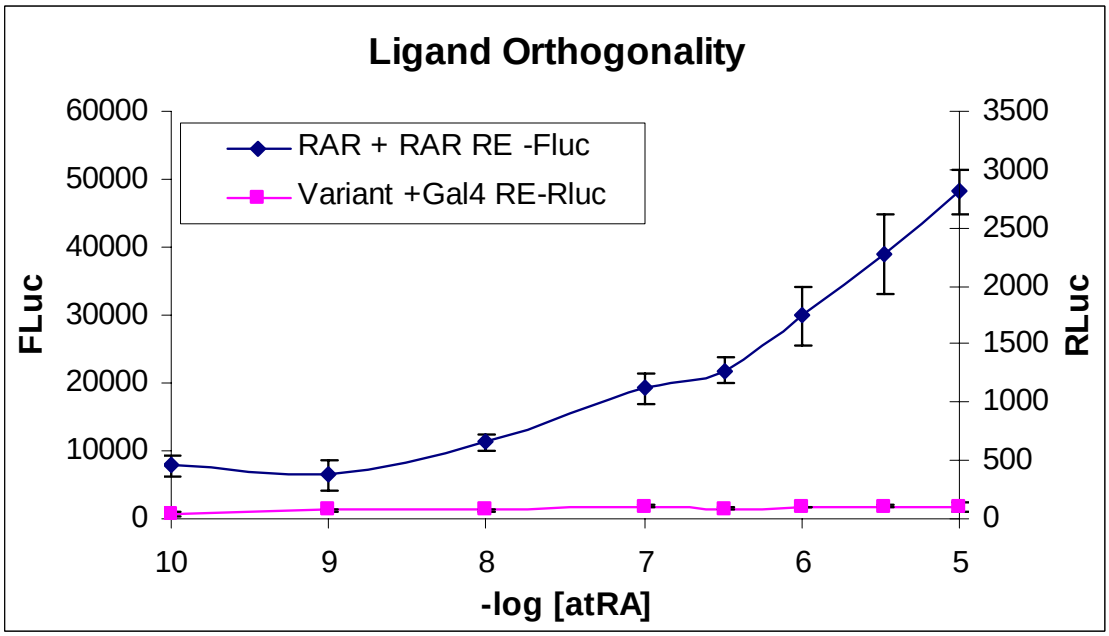
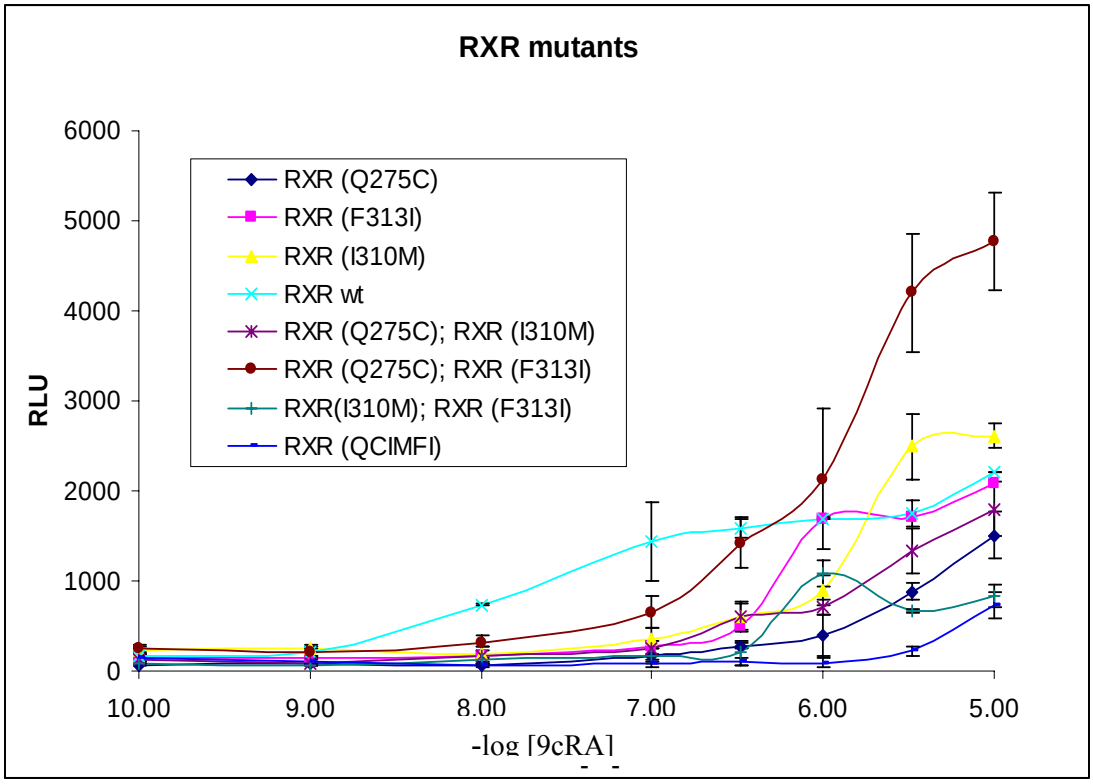


Figure 3.6 Activation profile of wild type RAR and GR Q275C; I310M; F313I in response to atRA and LG335.



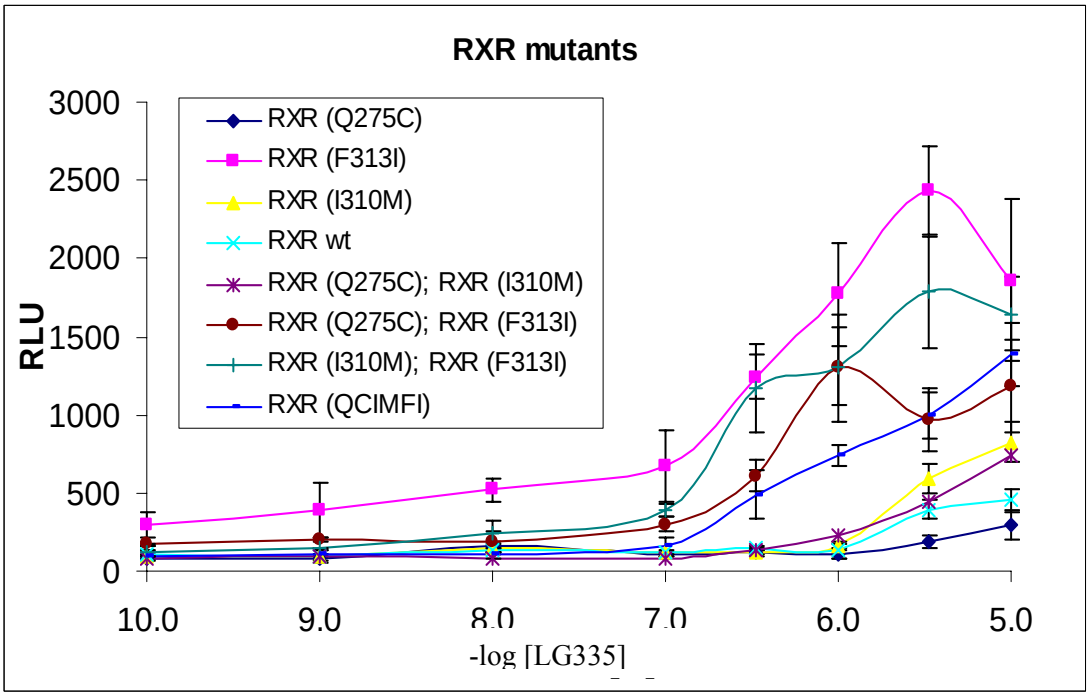
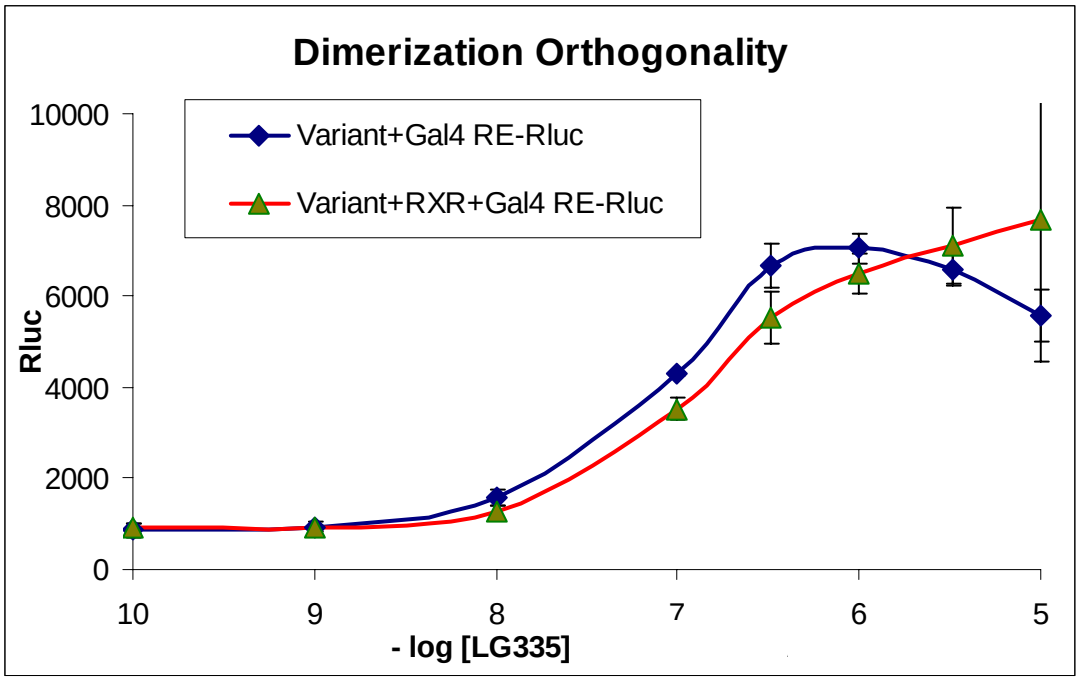


Figure 3.7 Dose response curves of RXR single, double and triple variant with 9cRA and LG335.

dimerization interface of RXR such that it does not dimerize with other proteins. RXR is a universal partner in the nuclear receptor family. It forms heterodimers with a variety of receptors including itself, RAR, vitamin D receptor, thyroid receptor and peroxisome proliferator activated receptor[28-30]. The DBD of RXR and its heterodimeric partners do not interact in the absence of response elements[14, 31]. Dimerization is mediated primarily by the interacting surface of the LBD on the receptor and is believed to stabilize the complex and promote the recognition of DNA[19].

Though the dimerization interface in the LBD was not modified, it was examined whether replacing the RXR DBD with Gal4 DBD would abolish GR variants dimerization with its partners. The GR variant is most similar to wild type RXR and RAR and was tested with these receptors for dimerization. It was observed that the HEK293T cells transfected with both GR variant and wild type RXR or RAR had a similar *Renilla* luciferase activity as observed from the cells containing the GR variant alone (Figure 3.8). This result implies that in the presence of LG335, the wild type RXR or RAR are not interacting with the GR variant to affect the expression of Gal4 RE controlled *Renilla* luciferase. If dimerization would have occurred, a decrease in *Renilla* luciferase activity would be observed as the RXR LBD-RXR DBD and RXR LBD-Gal4 DBD dimer would not be able to bind Gal4 RE with as much affinity as RXR LBD - Gal4 DBD homodimer. Also, in the presence of atRA, firefly luciferase activity from the cells transfected with both wild type RAR and GR variant was the same as firefly luciferase activity from the cells transfected with RAR alone (Figure 3.8). This data implies that the RAR and GR variant did not dimerize and bind with RAR RE.



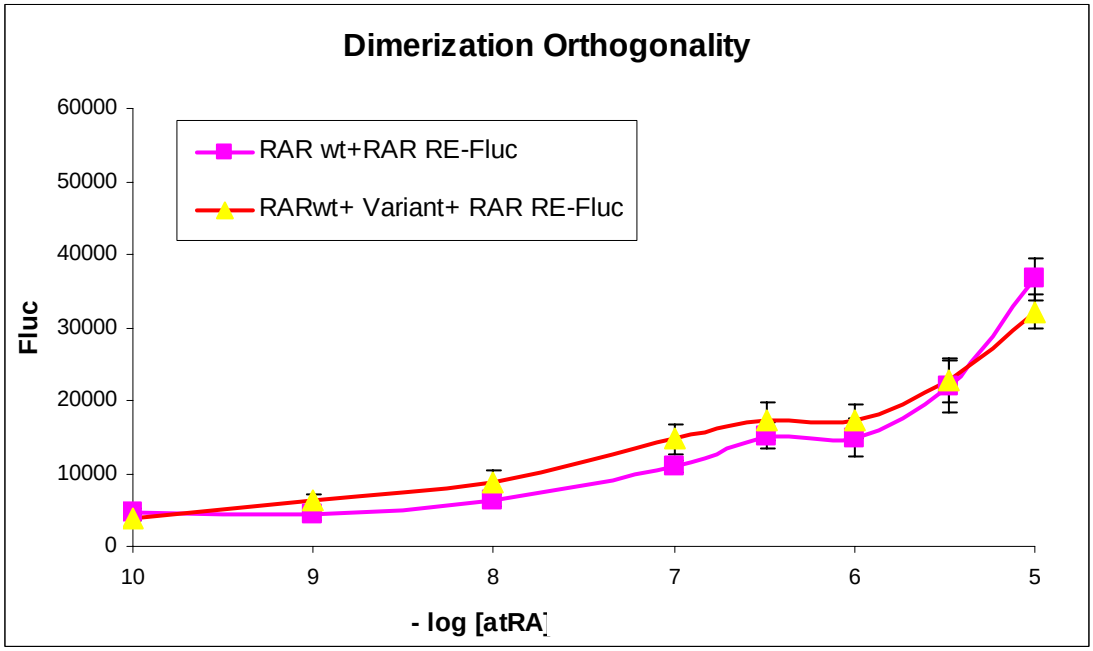


Figure 3.8 Graphs showing dimerization orthogonality between GR QCIMFI and wild type RXR and RAR

The data from response element, ligand and dimerization orthogonality imply that by generating three mutations in the LBD and replacing the RXR DBD with Gal4 DBD alters the response element, ligand specificity and abolish the dimerization between the variant and wild type RAR and RXR.

Summary

In conclusion a fully orthogonal ligand receptor pair was constructed using retinoid X receptor and synthetic ligand LG335. The three criteria for orthogonality were

1 The DNA binding domain of the orthogonal ligand receptor pair should only recognize a unique binding site.

2 The ligand binding domain should preferentially bind to a synthetic small molecule and not any endogenous ligand.

3 The receptor should not dimerize with other nuclear receptor such as wild type retinoid X receptor or retinoic acid receptor.

The receptor was made target specific by replacing the RXR DNA binding domain with a Gal4 DBD and introducing a Gal4 response element before the target gene. Three mutations Q275C; I310M; F313I (QCIMFI) were made in the ligand binding domain to alter the ligand specificity from 9cRA to LG335[22]. It was determined that not any of these mutations alone were enough to cause ligand orthogonality. Switching RXR DBD to Gal4 DBD also prevents the variant from dimerizing with other nuclear receptors such as wild type RAR and RXR. This variant GR (QCIMFI) qualifies to be an

orthogonal ligand receptor pair and can be used for practical applications such as a molecular switch to regulate gene expression.

Materials and Methods

Ligands

9-cis retinoic acid (MW=304.44g/mol) and all *trans* retinoic acid was purchased from ICN Biomedicals. LG335 was synthesized by Dr Lauren Schwimmer [32]

Expression and reporter plasmids

Plasmid were obtained from various sources, pCMX-hRXR was a gift from Dr Ronald Evans (Salk institute for biological studies, La Jolla, CA) [33], pCMX-hRXR (Q275C; I310M; F313I) was previously constructed by Dr Donald Doyle [22], To construct pCMXGR QCIMFI, PCR amplified fragment of Gal4 DBD fused to RXR LBD Q275C; I310M; F313I was cloned from pGBD RXR Q275C; I310M; F313I into pCMX-hRXR using KpnI and PstI. pCMXGR wt was constructed from pGBD RXR wt in the same way as pCMXGR QCIMFI. p17*4 Tata Luc expressing firefly luciferase under the control of four Gal4RE was a kind gift from Dr Sofia Tsai (Baylor college, Houston, TX)[34, 35], p17*4 Tata Hrl was made by replacing firefly luciferase from p17*4 TataLuc with NotI and SacII restriction site and inserting Renilla luciferase gene from pHRL (Clonetech), pCMXβGal [21], pBRE-Luc, pLucCRBPII expressing firefly

luciferase under the control of RXR RE was made by site-directed mutagenesis from pLucMCS (Stratagene, USA). Site-directed primers were designed to incorporate a CRBPII response element in the multiple cloning site. All the plasmids were sequencing confirmed.

Mammalian Cell Culture

HEK293T cells were cultured in DMEM supplemented with 10% calf bovine serum. Twenty four hours before transfection, HEK293T cells were plated with the density of 25000 cells per well in 48-well culture plates. After the cells were 75% confluent it was transfected with pCMX-hRXR (20-40ng), pLucCRBPII (40-80ng) and pCMXβGal (40-80ng) or pCMXGR QCIMFI(20-40ng), p17*4 Tata Hrl (40-80ng), and pCMXβGal (40-80ng) or all of them together. The transfection reagent used was either Lipofectamine® or Lipofectamine 2000® cationic lipid (Gibco BRL/Life technologies). Ten to twelve hrs after transfection, the media was removed and replaced with DMEM+10% charcoal resin stripped calf bovine serum containing appropriate concentration of ligand. The cells were allowed to incubate with ligand for twenty four-thirty six hours before harvesting. Cell extracts were assayed for firefly luciferase, Renilla luciferase and β-galactosidase activity using the luminometer and plate reader. Activity is reporter in relative light units (RLU) determined as the ratio of the firefly or Renilla luminescence divided by β-gal reading (control).

Preparation of Reagents

Mango lysis buffer For a 96 well plate, the mango lysis buffer consist of mango lysis base buffer (10 mL at 4 °C), 0.1 M EGTA (425 µL), 0.5 M $MgCl_2$ (170 µL), 1.0 M DTT (10.6 µL), and 250 mM PMSF (17 µL).

Firefly luciferin For a 96 well plate, the firefly luciferase assay buffer consist of 10X luciferin stock (1.35 mL) and 0.1 M KPO_4 at pH 7.8 (12.15 mL).

Firefly luciferase assay buffer For a 96 well plate, the firefly luciferase assay buffer consist of water (8.2 mL), 1.0 M KPO_4 (1 mL), 0.5 M $MgCl_2$ (430 µL), and 0.1 M ATP (370 µL).

Renilla luciferase assay buffer 7.5 mM Na_4PPi, 200 mM Na_2SO_4, 10mM CDTA and 1 µM coelentrazine

β-galactosidase assay buffer: For a 96 well plate, β-galactosidase buffer (10 mL), 2mg/mL ONPG (2.5 mL), and BME (35 µL).

Harvest Protocol

After thirty - forty eight hours of transfection the cells are assayed for firefly luciferase, Renilla luciferase and β-galactosidase activity using the luminometer and plate reader respectively. The media from the 48 well plate is aspirated and the cells are lysed using mango lysis buffer (100 µL/per well). The plate was gently shaken for five minutes and the cell extract was transferred to a 96 well plate, 100 µl cell extract per well. From this 96 well master plate, cell extract is transferred into three 96 well plates, 20 µl of cell lysate per well for firefly luminescence assay, 20 µl of cell lysate per well for *Renilla* luminescence and 40 µl of cell lysate per well for β-galactosidase activity (Figure 3.9).

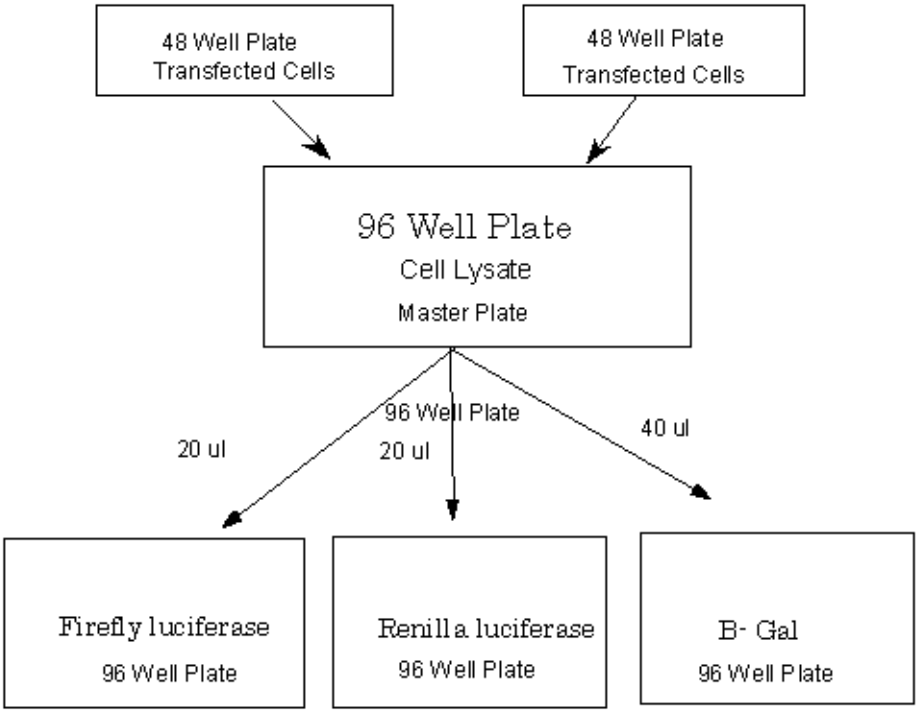


Figure 3.9 Schematic depiction of harvest protocol.

Literature Cited

1. Hwang, Y.W. and D.L. Miller, *A mutation that alters the nucleotide specificity of elongation factor Tu, a GTP regulatory protein.* J Biol Chem, 1987. **262**(27): p. 13081-5.

2. Coward, P., H.G. Wada, M.S. Falk, S.D. Chan, F. Meng, H. Akil, and B.R. Conklin, *Controlling signaling with a specifically designed Gi-coupled receptor.* Proc Natl Acad Sci U S A, 1998. **95**(1): p. 352-7.

3. Liu, Y., K. Shah, F. Yang, L. Witucki, and K.M. Shokat, *Engineering Src family protein kinases with unnatural nucleotide specificity.* Chem Biol, 1998. **5**(2): p. 91-101.

4. Shah, K., Y. Liu, C. Deirmengian, and K.M. Shokat, *Engineering unnatural nucleotide specificity for Rous sarcoma virus tyrosine kinase to uniquely label its direct substrates.* Proc Natl Acad Sci U S A, 1997. **94**(8): p. 3565-70.

5. Smith, G.K., S. Banks, T.A. Blumenkopf, M. Cory, J. Humphreys, R.M. Laethem, J. Miller, C.P. Moxham, R. Mullin, P.H. Ray, L.M. Walton, and L.A. Wolfe, 3rd, *Toward antibody-directed enzyme prodrug therapy with the T268G mutant of human carboxypeptidase A1 and novel in vivo stable prodrugs of methotrexate.* J Biol Chem, 1997. **272**(25): p. 15804-16.

6. Belshaw, P.J., J. Schoepfer, K. Liu, K. Morrison, and S. Schreiber, *Rational Design of Orthogonal Receptor-Ligand Combinations.* Angew. Chem., Int. Ed. Eng, 1995. **34**: p. 2129-2132.

7. Burcin, M.M., B.W. O'Malley, and S.Y. Tsai, *A regulatory system for target gene expression.* Front Biosci, 1998. **3**: p. c1-7.

8. Mangelsdorf, D.J., C. Thummel, M. Beato, P. Herrlich, G. Schutz, K. Umesono, B. Blumberg, P. Kastner, M. Mark, P. Chambon, and R.M. Evans, *The nuclear receptor superfamily: the second decade.* Cell, 1995. **83**(6): p. 835-9.

9. Levin, A.A., L.J. Sturzenbecker, S. Kazmer, T. Bosakowski, C. Huselton, G. Allenby, J. Speck, C. Kratzeisen, M. Rosenberger, A. Lovey, and et al., *9-cis retinoic acid stereoisomer binds and activates the nuclear receptor RXR alpha.* Nature, 1992. **355**(6358): p. 359-61.

10. Bugge, T.H., J. Pohl, O. Lonnoy, and H.G. Stunnenberg, *RXR alpha, a promiscuous partner of retinoic acid and thyroid hormone receptors.* Embo J, 1992. **11**(4): p. 1409-18.

11. Zhang, X.K., B. Hoffmann, P.B. Tran, G. Graupner, and M. Pfahl, *Retinoid X receptor is an auxiliary protein for thyroid hormone and retinoic acid receptors.* Nature, 1992. **355**(6359): p. 441-6.

12. Glass, C.K., *Differential recognition of target genes by nuclear receptor monomers, dimers, and heterodimers.* Endocr Rev, 1994. **15**(3): p. 391-407.

13. Mangelsdorf, D.J. and R.M. Evans, *The RXR heterodimers and orphan receptors.* Cell, 1995. **83**(6): p. 841-50.

14. Zechel, C., X.Q. Shen, J.Y. Chen, Z.P. Chen, P. Chambon, and H. Gronemeyer, *The dimerization interfaces formed between the DNA binding domains of RXR, RAR and TR determine the binding specificity and polarity of the full-length receptors to direct repeats.* Embo J, 1994. **13**(6): p. 1425-33.

15. Predki, P.F., D. Zamble, B. Sarkar, and V. Giguere, *Ordered binding of retinoic acid and retinoid-X receptors to asymmetric response elements involves determinants adjacent to the DNA-binding domain.* Mol Endocrinol, 1994. **8**(1): p. 31-9.

16. Forman, B.M., C.R. Yang, M. Au, J. Casanova, J. Ghysdael, and H.H. Samuels, *A domain containing leucine-zipper-like motifs mediate novel in vivo interactions between the thyroid hormone and retinoic acid receptors.* Mol Endocrinol, 1989. **3**(10): p. 1610-26.

17. Zhang, X.K., G. Salbert, M.O. Lee, and M. Pfahl, *Mutations that alter ligand-induced switches and dimerization activities in the retinoid X receptor.* Mol Cell Biol, 1994. **14**(6): p. 4311-23.

18. Au-Fliegner, M., E. Helmer, J. Casanova, B.M. Raaka, and H.H. Samuels, *The conserved ninth C-terminal heptad in thyroid hormone and retinoic acid receptors mediates diverse responses by affecting heterodimer but not homodimer formation.* Mol Cell Biol, 1993. **13**(9): p. 5725-37.

19. Perlmann, T., K. Umesono, P.N. Rangarajan, B.M. Forman, and R.M. Evans, *Two distinct dimerization interfaces differentially modulate target gene specificity of nuclear hormone receptors.* Mol Endocrinol, 1996. **10**(8): p. 958-66.

20. Perlmann, T., P.N. Rangarajan, K. Umesono, and R.M. Evans, *Determinants for selective RAR and TR recognition of direct repeat HREs.* Genes Dev, 1993. **7**(7B): p. 1411-22.

21. Peet, D.J., D.F. Doyle, D.R. Corey, and D.J. Mangelsdorf, *Engineering novel specificities for ligand-activated transcription in the nuclear hormone receptor RXR.* Chem Biol, 1998. **5**(1): p. 13-21.

22. Doyle, D.F., D.A. Braasch, L.K. Jackson, H.E. Weiss, M.F. Boehm, D.J. Mangelsdorf, and D.R. Corey, *Engineering orthogonal ligand-receptor pairs from "near drugs".* J Am Chem Soc, 2001. **123**(46): p. 11367-71.

23. Boehm, M.F., M.R. McClurg, C. Pathirana, D. Mangelsdorf, S.K. White, J. Hebert, D. Winn, M.E. Goldman, and R.A. Heyman, *Synthesis of high specific activity [3H]-9-cis-retinoic acid and its application for identifying retinoids with unusual binding properties.* J Med Chem, 1994. **37**(3): p. 408-14.

24. Boehm, M.F., L. Zhang, B.A. Badea, S.K. White, D.E. Mais, E. Berger, C.M. Suto, M.E. Goldman, and R.A. Heyman, *Synthesis and structure-activity relationships of novel retinoid X receptor-selective retinoids.* J Med Chem, 1994. **37**(18): p. 2930-41.

25. Moras, D. and H. Gronemeyer, *The nuclear receptor ligand-binding domain: structure and function.* Curr Opin Cell Biol, 1998. **10**(3): p. 384-91.

26. Tenbaum, S. and A. Baniahmad, *Nuclear receptors: structure, function and involvement in disease.* Int J Biochem Cell Biol, 1997. **29**(12): p. 1325-41.

27. Dilworth, F.J., C. Fromental-Ramain, E. Remboutsika, A. Benecke, and P. Chambon, *Ligand-dependent activation of transcription in vitro by retinoic acid receptor alpha/retinoid X receptor alpha heterodimers that mimics transactivation by retinoids in vivo.* Proc Natl Acad Sci U S A, 1999. **96**(5): p. 1995-2000.

28. Kliewer, S.A., K. Umesono, D.J. Mangelsdorf, and R.M. Evans, *Retinoid X receptor interacts with nuclear receptors in retinoic acid, thyroid hormone and vitamin D3 signalling.* Nature, 1992. **355**(6359): p. 446-9.

29. Kliewer, S.A., K. Umesono, D.J. Noonan, R.A. Heyman, and R.M. Evans, *Convergence of 9-cis retinoic acid and peroxisome proliferator signalling pathways through heterodimer formation of their receptors.* Nature, 1992. **358**(6389): p. 771-4.

30. Yu, V.C., C. Delsert, B. Andersen, J.M. Holloway, O.V. Devary, A.M. Naar, S.Y. Kim, J.M. Boutin, C.K. Glass, and M.G. Rosenfeld, *RXR beta: a coregulator that enhances binding of retinoic acid, thyroid hormone, and vitamin D receptors to their cognate response elements.* Cell, 1991. **67**(6): p. 1251-66.

31. Zechel, C., X.Q. Shen, P. Chambon, and H. Gronemeyer, *Dimerization interfaces formed between the DNA binding domains determine the cooperative binding of RXR/RAR and RXR/TR heterodimers to DR5 and DR4 elements.* Embo J, 1994. **13**(6): p. 1414-24.

32. Schwimmer, L.J., *Engineering ligand receptor pairs for small molecule control of transcription.* Thesis, Georgia Tech, 2005.

33. Heyman, R.A., D.J. Mangelsdorf, J.A. Dyck, R.B. Stein, G. Eichele, R.M. Evans, and C. Thaller, *9-cis retinoic acid is a high affinity ligand for the retinoid X receptor.* Cell, 1992. **68**(2): p. 397-406.

34. Wang, Y., B.W. O'Malley, Jr., S.Y. Tsai, and B.W. O'Malley, *A regulatory system for use in gene transfer.* Proc Natl Acad Sci U S A, 1994. **91**(17): p. 8180-4.

35. Wang, Y., J. Xu, T. Pierson, B.W. O'Malley, and S.Y. Tsai, *Positive and negative regulation of gene expression in eukaryotic cells with an inducible transcriptional regulator.* Gene Ther, 1997. **4**(5): p. 432-41.

CHAPTER 4

ENGINEERING A MOLECULAR SWITCH FOR GENE THERAPY

Introduction

Gene therapy involves introducing a functional gene that produces a therapeutic protein to treat or prevent diseases. One of the challenges of gene therapy is controlling the production of the therapeutic protein. Over or under-production of the therapeutic protein can cause side effects, hence it is important to regulate therapeutic gene expression. Initial attempts to control gene expression have used endogenous cellular elements such as promoter and enhancer that respond to exogenous signals. Some of these regulatory systems are temperature responsive system[1, 2], heavy metal ion based gene regulation[3], and oxygen tension driven system[4]. However, these approaches have had limitations due to a high basal expression in the uninduced state, because of interference due to inducing condition or low regulation performance i.e. low induction ratio. Currently the most widely used method to control gene expression and produce relatively high levels of the therapeutic protein is ligand inducible gene expression. Ligand-dependent inducible systems are usually based on two components: 1) a chimeric transcription factor containing a DNA-binding domain (DBD) fused to a ligand binding domain (LBD), where the DBD does not bind endogenous sequences, and the LBD interacts only with its unnatural ligand; 2) an artificial promoter consisting of multimeric-binding sites for the DBD followed by a minimal promoter and the target gene.

In the presence of the ligand the transcription factor is recruited to the promoter and

expresses the target gene downstream of this promoter. Neither the chimeric transcription factor nor the ligand should interact with any endogenous protein or endogenous small molecule, in other words they need to be orthogonal to cell function (as defined in chapter 3). A good inducible system is one that provides the ability to control protein levels quantitatively, in a timely manner, over an extended period of time without eliciting an immune response. To be used in human gene therapy, the molecular switch should meet the following criteria:

(1) Should be an ON switch i.e. the ligand should activate rather than repress transcription (OFF-switch). This is because an OFF-switch requires a prolonged exposure to the drug and induction kinetics are determined by the rate of drug clearance.

(2) The ligand and the molecular switch should be target specific and not interfere with endogenous metabolic pathways.

(3) The ligand should be orally bioavailable, be able to penetrate the target tissues and have a compatible metabolic profile over extended period of therapeutic use.

(4) Target gene expression should correlate with the dose of the ligand.

(5) The system should have low basal activity i.e. be inactive in the absence of the ligand but strongly stimulated by ligand administration, hence high fold activation.

(6) The molecular switch should not elicit an immune response in humans.

Here a transcription based molecular switch is developed using an orthogonal ligand receptor pair: LG335 and an RXR variant. Presently, there are three widely used inducible systems that are similar to the molecular switch proposed and each of them has advantages and disadvantages. 1) The progesterone receptor (PR) inducible system regulates gene expression using mifepristrone (RU486) and PR mutant, known as GeneSwitch® [5, 6]. The

advantage of this system is that it comprises of human protein and does not have an immune response, strongly induces transcription and responds to .1 nM concentrations of mifepristrone. But the performance of this system remains poor with a high level of basal activity in the absence of the inducer, resulting in a low induction level [7]. Mifepristrone is used in several countries as a treatment to terminate pregnancy. Though mifepristrone stimulates GeneSwitch® maximally by oral doses of at least one order of magnitude lower than those required to induce abortion (about 200–600 mg), it may also affect the ovarian cycle and exert a contraceptive activity. 2) The tetracycline (Tet) based inducible system is well characterized and widely used, however the proteins may be immunogenic because there are derived from bacteria. The Tet-system is apparently not immunogenic in several mouse strains, however recent studies indicate that intramuscularly delivered Tet-ON activators may elicit a cellular and humoral response in non-human primates[8, 9]. 3) The ecdysone receptor dependent gene regulatory system is based on a heterodimer between insect steroid hormone receptor (EcR) and RXR known as RheoSwitch®[10]. This system has very a low level of basal expression and hence high fold induction (>10000) [11, 12]. However, it requires the over expression of two genes (EcR and RXR) simultaneously which complicates its use in viral delivery systems. RXR is a reluctant dimer of EcR and efficient transactivation can only be achieved by overexpressing RXR in the target cells. This poses a safety concern as RXR is involved in many metabolic pathways. A recent finding has shown that RXR overexpression causes dilated cardiomyopathy in mice[13]. Also, the insect hormone and their derivates may not be approved for human therapeutic use.

Here, an attempt is made to create a molecular switch that requires over-expression of only one protein, produces a high induction level at low concentrations of a non-toxic

synthetic ligand. Previously a molecular switch was created using an orthogonal ligand receptor pair; a synthetic ligand LG335 and a RXR variant (Q275C; I310M; F313I) containing Gal4 DBD (described in chapter 3). This RXR variant is an ON switch, i.e. it turns on the transcription of the target gene in the presence of the ligand LG335. It is orthogonal to the endogenous proteins such as wild type RXR and RAR and is target specific i.e. the Gal4 DBD recognizes its binding site, Gal4 RE, and transcribes only Gal4 RE driven target genes. Also, the variant does not recognize the RXR RE and hence do not interfere with RXR responsive gene. The mutations in the variant Q275C; I310M; F313I altered ligand selectivity from the natural ligand 9cRA to LG335. The EC_{50} for LG335 is 50 nM and the induction ratio is 8 fold.

To design a more sensitive molecular switch that has a lower EC_{50} and high fold induction with LG335, structure-based approach was applied to generate libraries of RXR variants[14]. A general method was developed for docking ligands into the binding pocket of the RXRα LBD crystal structure which was obtained from the protein databank. LG335 was manually placed in the binding site of RXR and the atoms of LG335 were superimposed onto the atoms of the crystallographic 9cRA. The binding pocket of RXR is hydrophobic and the majority of substitutions were made on the basis of size, charge or hydrophobicity. The randomized amino acids were chosen on the basis of their proximity to the bound 9cRA as observed in the crystal structure. Mutations were made at six positions (I268, A271, A272, I310, F313, and L436). The first three positions had four possible mutations (leucine (L), valine (V), alanine(A) and proline (P)) and the other three positions had eight possible amino acids leucine (L), isoleucine (I), valine (V), phenylalanine (F), methionine (M) , serine (S),

alanine (A) and threonine (T)). The combination of six positions and number of possible amino acids kept the library size to 32,768 amino acid combinations.

To discover the functional variants within the libraries, a genetic selection assay method called chemical complementation. Chemical complementation is a method in which the survival of the yeast is dependent on a functional variant and a small molecule. The engineered receptor are fused to the Gal4 DBD and transformed into yeast with the appropriate coactivator ACTR-Gal4 activation domain (Figure 4.1). Yeast transformants were plated on adenine selective media with a certain concentration of LG335. The variants that are able to bind and activate the receptor will express ADE2 gene allowing the yeast cells to survive on the selective media. The variants that bind LG335 and not 9cRA were rescued from the yeast and sequenced. Table 4.1 summarizes the transcriptional activation profiles of all twelve variants in response to both 9cRA and LG335 compared to wild-type RXR.

The library produced ~380,000 transformants designed for the synthetic retinoid-like compound LG335. In yeast, wild-type RXR has an EC_{50} of 500 nM with 9cRA and an EC_{50} >10 μM LG335. The library produced ligand-receptor pairs with LG335 that have a variety of EC_{50}s (40nM to > 2 μM) and activation levels (10% to 80% of wild-type RXR with. 9cRA). Some of the variants behaved as switches and some as rheostats, they had wild-type levels of activation (> 50% efficacy) and some were low (<25%) as shown in figure 4.1

Results and Discussion

Evaluating RXR variants in mammalian cell culture

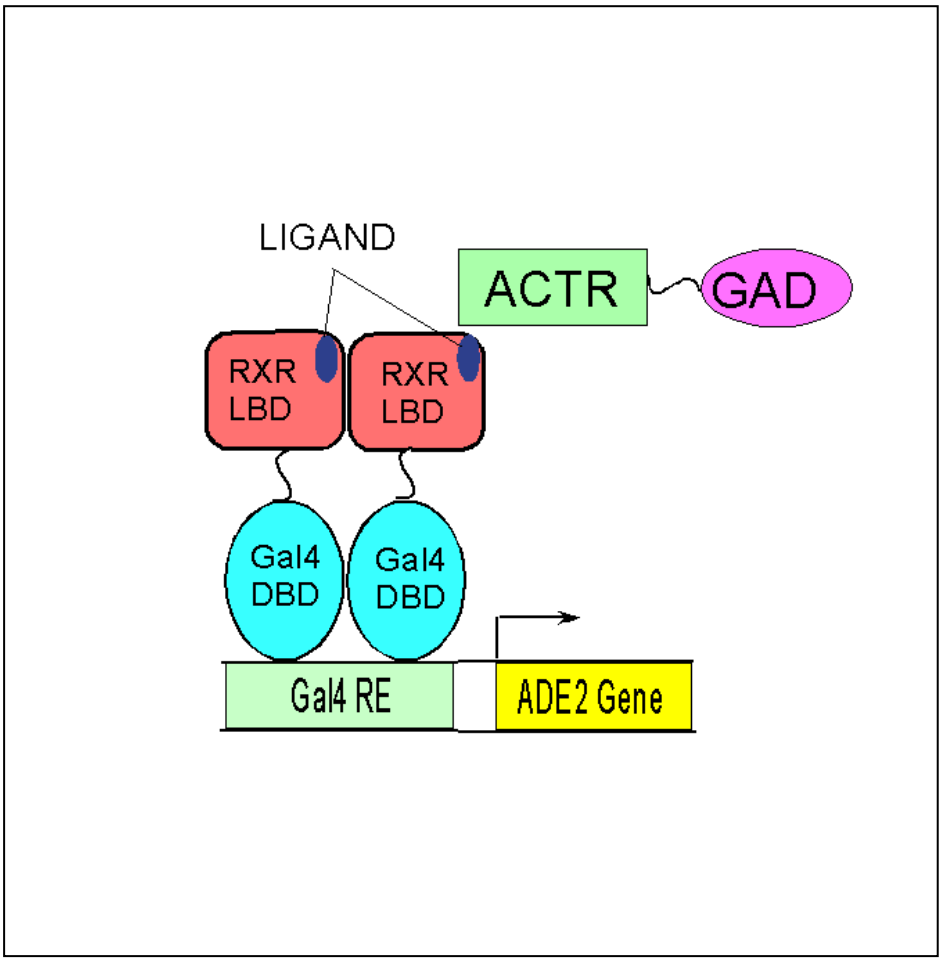


Figure 4.1 Chemical complementation in yeast: Yeast cells transformed with RXR variants fused to the Gal4 DBD and the appropriate coactivator ACTR-Gal4 activation domain. Yeast cells containing the RXR variants that are activated by the ligand express ADE2 gene and survive in media lacking adenine

Table 4.1 EC_{50} and Efficacy (Eff) of the twelve RXR variants in response to 9cRA and LG335 in yeast and mammalian (HEK293T) cells.
EC_{50}, is the concentration of ligand producing half maximal activity.
Eff, maximum increase in activation relative to the increase in activation of wild type with 10 µM 9cRA.

	9cRA				LG335			
	Yeast		Mammalian		Yeast		Mammalian	
Variant	EC50 (nM)	Eff (%)	EC50 (nM)	Eff (%)	EC50 (nM)	Eff (%)	EC50 (nM)	Eff (%)
WT	500	100	220	100	>10,000	10	300	10
I268A;I310A;F313A;L436F	>10,000	0	>10,000	0	220	70	30	50
I268V;A272V;I310L;F313M	>10,000	10	>1600	30	40	60	1	30
I268A;I310S;F313V;L436F	>10,000	10			470	60		
I268A;I310S;F313A;L436F	>10,000	0	>10,000	0	430	50	690	20
I268V;A272V;I310M;F313S;L436M	>10,000	10	>10,000	0	680	30	180	30
I268A;A272V;I310A;F313A;L436F	>10,000	0			530	30		
I268L;A271V;I310L;F313L	>10,000	0			530	20		
I268A;I310M;F313A;L436T	>10,000	0	>10,000	0	610	10	140	20
I268V;A271V;I310L;F313V	>10,000	0			650	10		
I268L;I310V;F313I	>10,000	0			>2000	10		
I268L;I310M;F313V	>10,000	20			610	20		
I268V;I310V;F313S	>10,000	0			440	10		

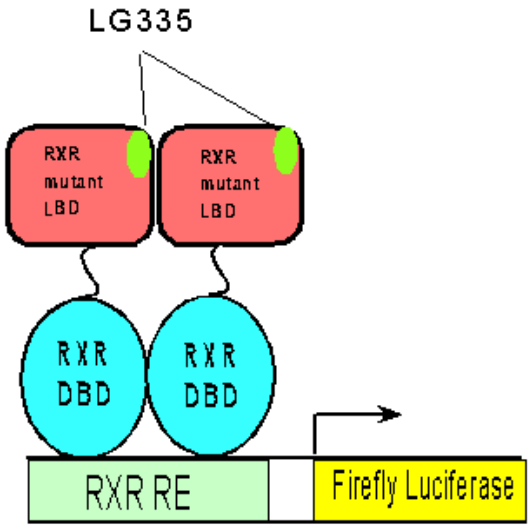


Figure 4.2 Schematic diagram showing LG335 bound RXR variant initiating transcription of firefly luciferase

With the data obtained from the RXR library, five RXR variants [I268A; I310A; F313A; L436F, I268V; A272V; I310L; F313M, I268A; I310S; F313A; L436F, I268V; A272V; I310M; F313S; L436M, and I268A; I310M; F313A; L436T] containing the RXR DBD was cloned in mammalian expression vectors. Plasmids encoding for the RXR variants and the RXR RE driven firefly luciferase were transfected in HEK293T cells (Figure 4.2). The activation profile of these variants in response to 9cRA and LG335 is shown in figure 4.3. Among the five RXR variants, two of them I268A; I310M; F313A; L436T, I268A; I310S; F313A; L436F have low activation level (similar to the activity of wild type RXR with LG335), relatively high EC_{50} (i.e. the concentration of ligand producing half maximal activity) of 140 nM and 690 nM respectively and efficacy of 20% (Efficacy is the maximum increase in activation relative to the increase in activation of wild type RXR with 10μM 9cRA) (Figure 4.1, 4.3). The other three RXR variants I268V; A272V; I310L; F313M, I268A; I310A; F313A; L436F and I268V; A272V; I310M; F313S; L436M were activated by LG335 to a higher level then wild type RXR, they have lower EC_{50} for LG335of 1 nM, 30 nM, and 180 nM respectively and their EC_{50} for 9cRA of > 1 μM. The efficacy of the variants was 30%, 50% and 30% respectively. The RXR variant I268V; A272V; I310L; F313M with 1 nM EC_{50} for LG335 was 25-fold lower than the previous best orthogonal receptor RXR Q275C; I310M; F313I [15]. It behaved more like a switch rather than a rheostat.

Converting the RXR variants into GR variants

To use the RXR variants as molecular switches it is necessary to minimize the

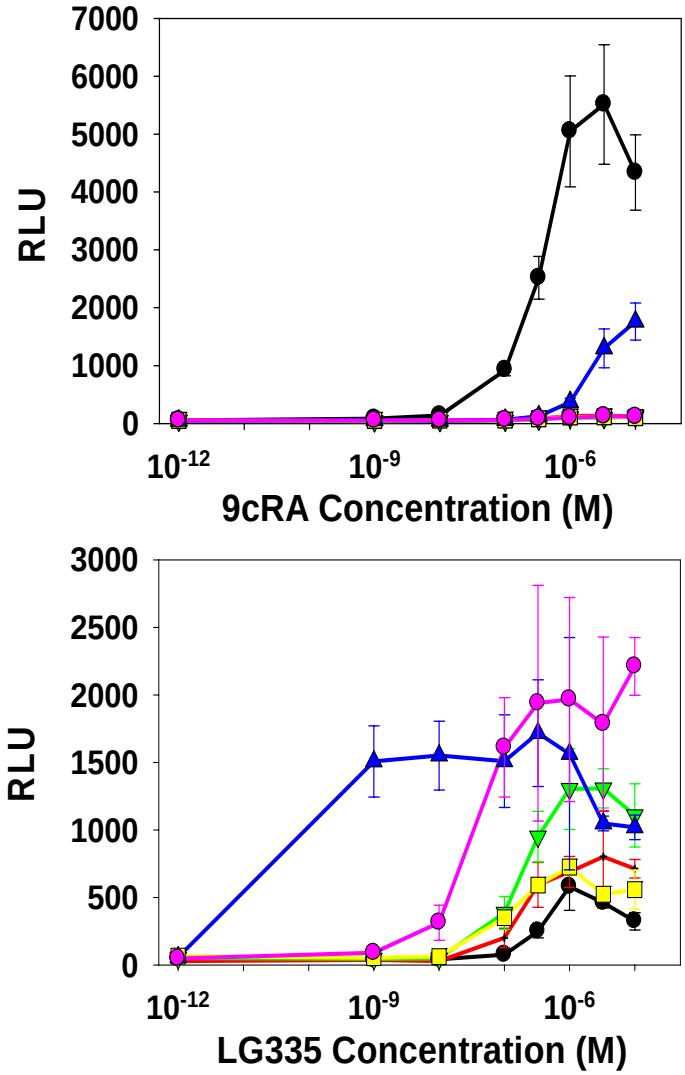


Figure 4.3. Activation profile of the RXR variants in response to 9cRA (a) and LG335 (b) in HEK293T cells. with wild-type RXR (wt RXR, filled circle), I268A;I310S;F313A;L436F (horizontal line), I268V;A272V;I310M;F313S;L436M (downward-pointing triangle), I268A;I310M;F313A;L436T (yellow square), I268V;A272V;I310L;F313M (upright pointing triangle), or I268A;I310A;F313A;L436F (pink circle).The 10^{-12} M point is 0 M ligand concentration. RLU, relative light units.

potential for interference with endogenous transcription factors. The specificity of the receptors and the response elements need to be addressed such that they have all the three properties of an orthogonal receptor i) ligand orthogonality, where the variant protein binds to LG335 and is unresponsive to its natural ligand, 9cRA ii) response element orthogonality, where the receptor's DBD recognizes an artificial promoter containing the multimeric-unique binding sites and iii) dimerization orthogonality, where the variant protein does not dimerize with wild type receptors. From the library, five RXR variants were shortlisted that fulfill the first criteria of ligand orthogonality and were selective for LG335. To achieve the response element and dimerization orthogonality the RXR DBD of the variants were replaced by Gal4 DBD to form a fusion protein, Gal4 DBD- RXR LBD (GR variant). Also, four copies of Gal4 RE were introduced in the promoter region of the target gene, firefly luciferase (Figure 4.4).

The GR variants had similar activation profile with ligands 9cRA and LG335 as observed with the RXR variants containing RXR DBD and LG335. GR variants I268V; A272V; I310L; F313M, I268A; I310A; F313A; L436F, and I268V; A272V; I310M; F313S; L436M have an EC_{50} for LG335 of 0.5 nM, 50 nM, 50 nM respectively (Figure 4.5). One major difference observed between GR variants and RXR variants was the high basal activity or leaky expression in the absence of LG335. Due to the high basal activity there was a low fold induction (the ratio of maximum activation observed in the presence of ligand versus the activation in the absence of ligand). The fold induction of GR variant I268V; A272V; I310L; F313M was 4 fold, and GR I268A; I310A; F313A; L436F, I268V; A272V; I310M; F313S; L436M was about 10 to 13 fold in response to LG335 (Figure 4.5). Although the fold

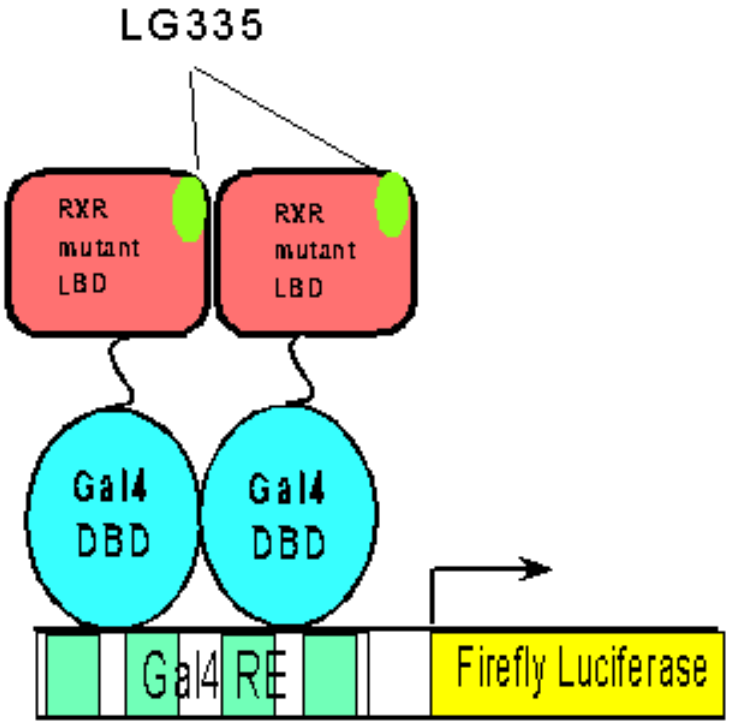
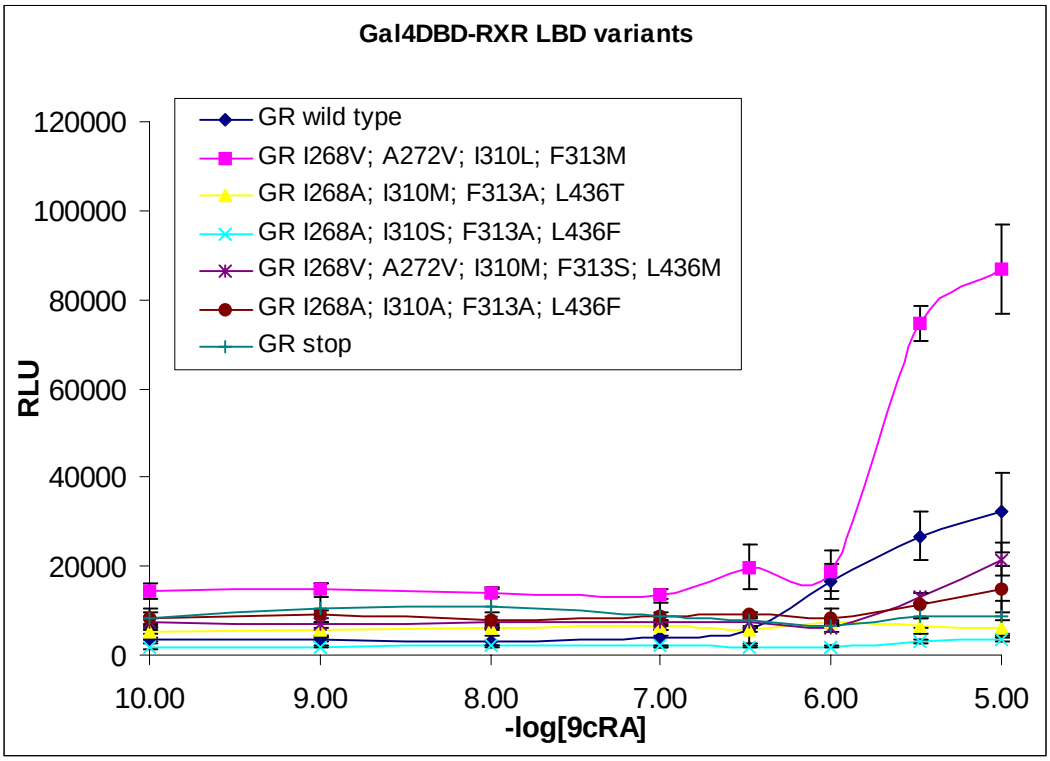


Figure 4.4 LG335 bound GR variants control the transcription of Gal4 RE driven firefly luciferase



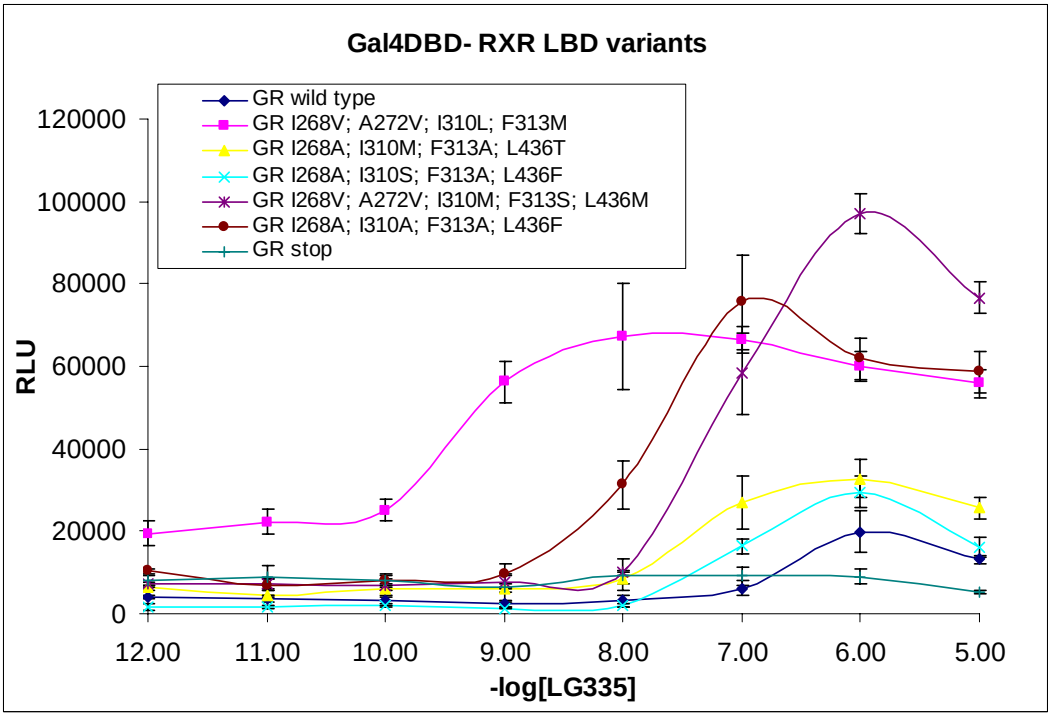


Figure 4.5 Activation profile of the GR variants containing the RXR LBD and Gal4 DBD and RXR LBD in response to 9cRA and LG335

induction is low, the variants were tested as potential molecular switches because of their remarkably low EC_{50}'s for LG335. However, the results indicate that to use these variants for gene therapy the high basal activity needs to be addressed.

Retroviral vectors containing the GR variant I268V; A272V; I310L; F313M and the reporter gene

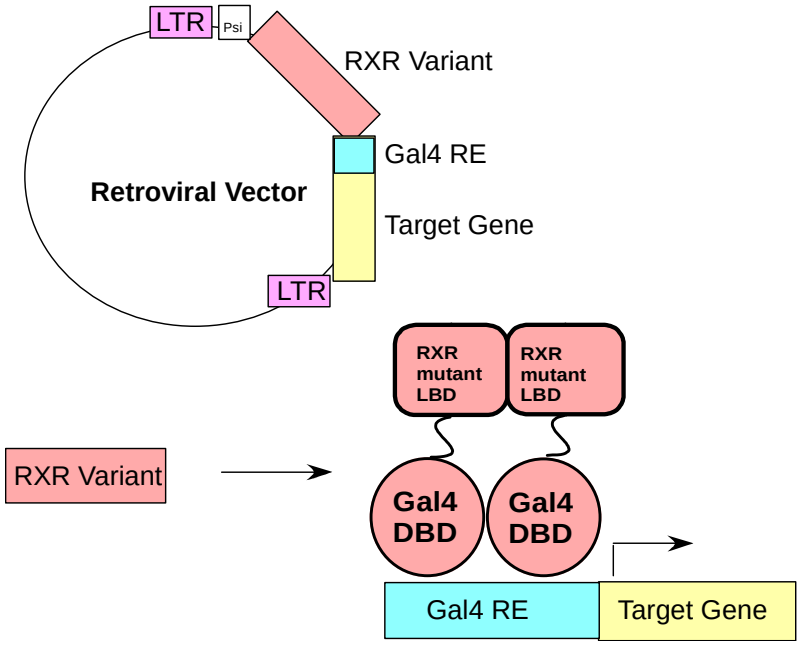
To use the GR variants as molecular switches in mice the variants were cloned into a retroviral vector along with the reporter gene, pMSCVGR (variant)_reporter (Figure 4.6). The GR variant was under the control of LTR promoter and the reporter gene was under the control of Gal4 RE and minimal promoter as shown in figure 4.6. The first variant cloned was GR I268V; A272V; I310L; F313M (GR 'I') with Gal4 RE driving the reporter gene, firefly luciferase, pMSCV GR'I'_Luc. The variant maintained a low EC_{50} of 5 nM for LG335 but had a high basal activity of 1000 RLU and low induction of 4 fold (Figure 4.7). To confirm the high basal activity the reporter gene was substituted from luciferase to green fluorescence protein (GFP). This construct contains the GR variant I268V; A272V; I310L; F313M with Gal4 RE driven GFP, pMSCV GR'I'_GFP. The assay changes from luminescence to fluorescence. Fluorescence was observed from GFP both with and without LG335 (Figure 4.8). This expression of GFP along with the previous data of luciferase by pMSCVGR 'I' confirmed the leaky expression of this variant in the absence of LG335. It was concluded that the basal expression of the reporter gene was not due to the read through of the stop codon of the Gal4 DBD-RXR LBD (GR) fusion gene in the plasmid. This is because the luciferase activity from the cells transfected with pMSCV GR 'I'_luc that

contains the GR variant and the Gal4 RE driven luciferase gene was the same as observed from the cells co-transfected with the two plasmids, one containing the GR variant (pMSCV GR 'I') and the other plasmid containing the luciferase gene (p17*4 Tata Luc) (Figure 4.5, 4.7).

Characterizing the GR variant I268V; A272V; I310L; F313M

To understand the protein ligand interaction between GR I268V; A272V; I310L; F313M and LG335 and to investigate the key residues in the binding pocket of RXR that have an effect on EC_{50}, activation level and high basal activity by LG335, the variant was characterized. To characterize the variant, single and triple mutants were made via site-directed mutagenesis (Stratagene, USA). GR I268V, A272V, I310L, F313M, I268V; A272V; I310L, I268V; A272V; F313M, I268V; I310L; F313M, A272V; I310L; F313M were tested in mammalian cell culture in the presence of ligand 9cRA or LG335 (Figure 4.9). Each of the single mutants except GR A272V maintained a high basal activity between 1000-5000 RLU and was activated by both 9cRA and LG335 (Figure 4.9). Hence, none of the single mutations alone contributed to the ligand orthogonality, it was a cooperative effect. Triple triple mutant was about 1 μM with LG335 (Figure 4.10). These results show that none of the single or triple mutants had an EC_{50} as low as the quadruple mutant GR I268V; A272V; I310L; F313M at 5 nM (Figure 4.7), nor were they as selective for LG335 over 9cRA (Figure 4.9, 4.10).

Attempt to decrease the basal activity of the GR variants



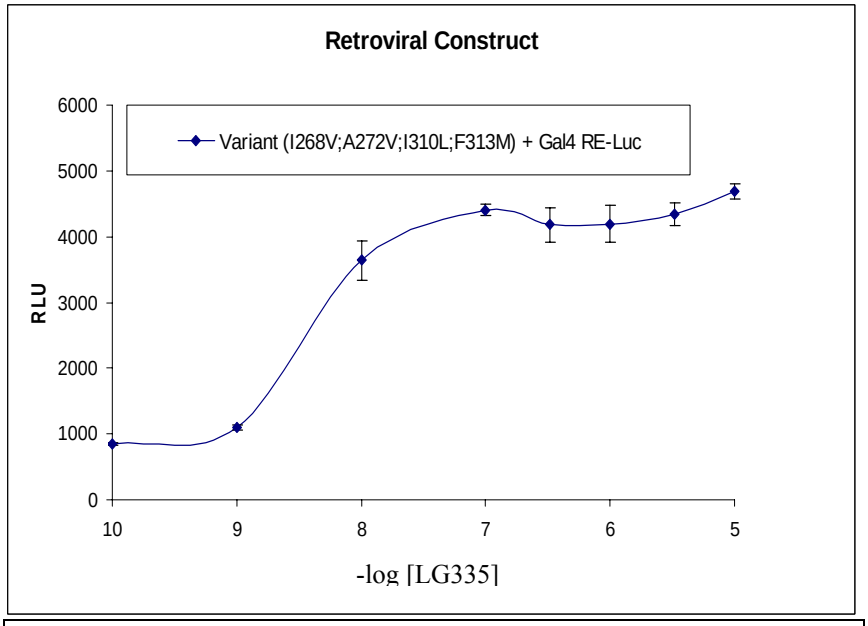


Figure 4.7 Dose response curve of the retroviral construct pMSCVGR I268V; A272V; I310L; F313M_luc at varying concentrations of LG335.

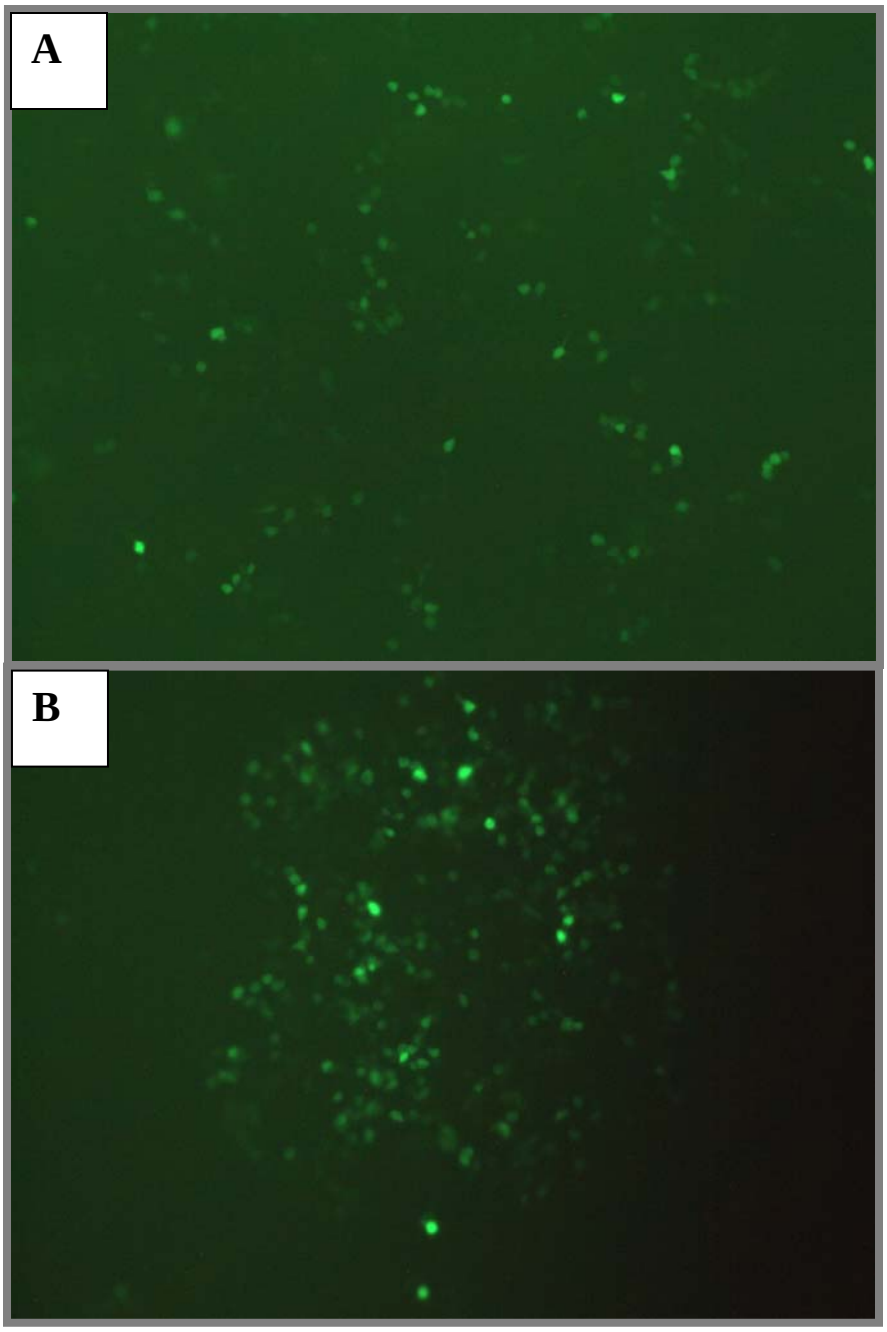
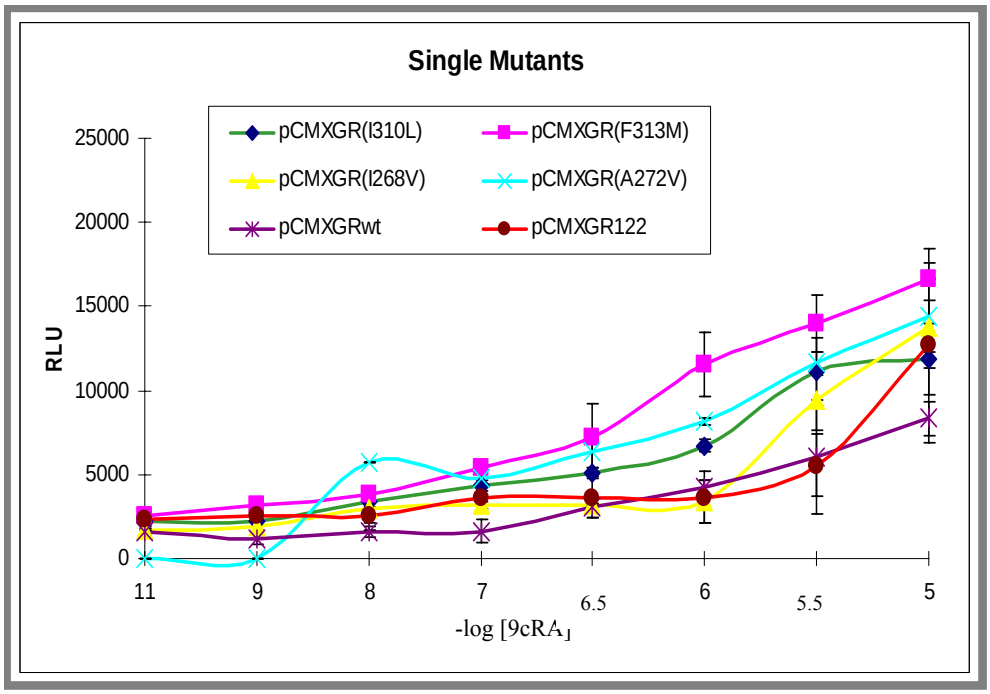


Figure 4.8 Detection of fluorescence from HEK293T cells expressing retroviral vector, pMSCVGR I268V; A272V; I310L; F313M _GFP (A) In the absence of ligand (B) In the presence of 1 μM LG335.



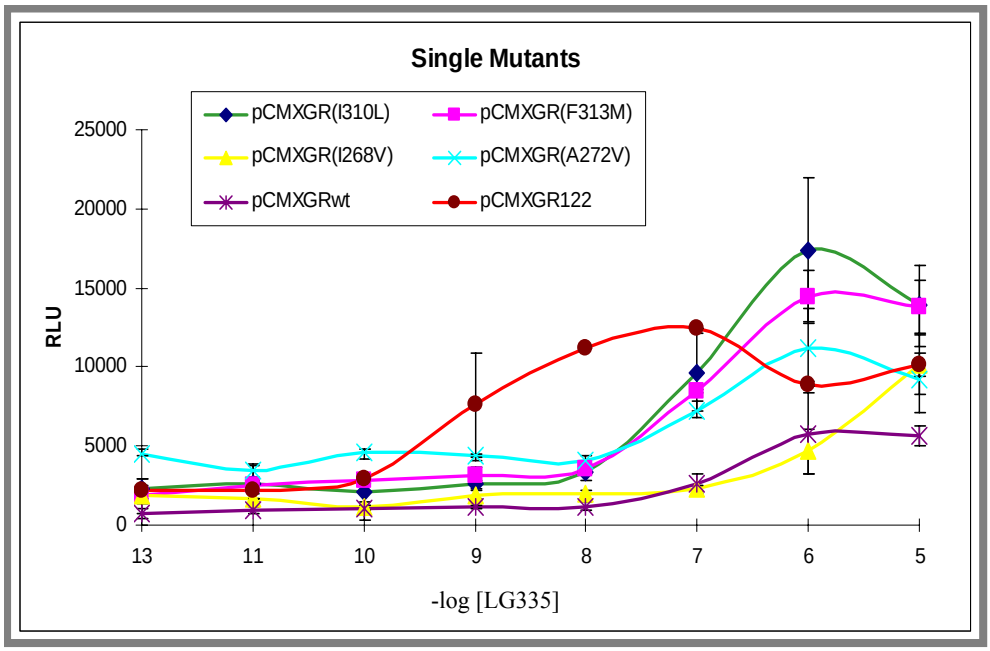
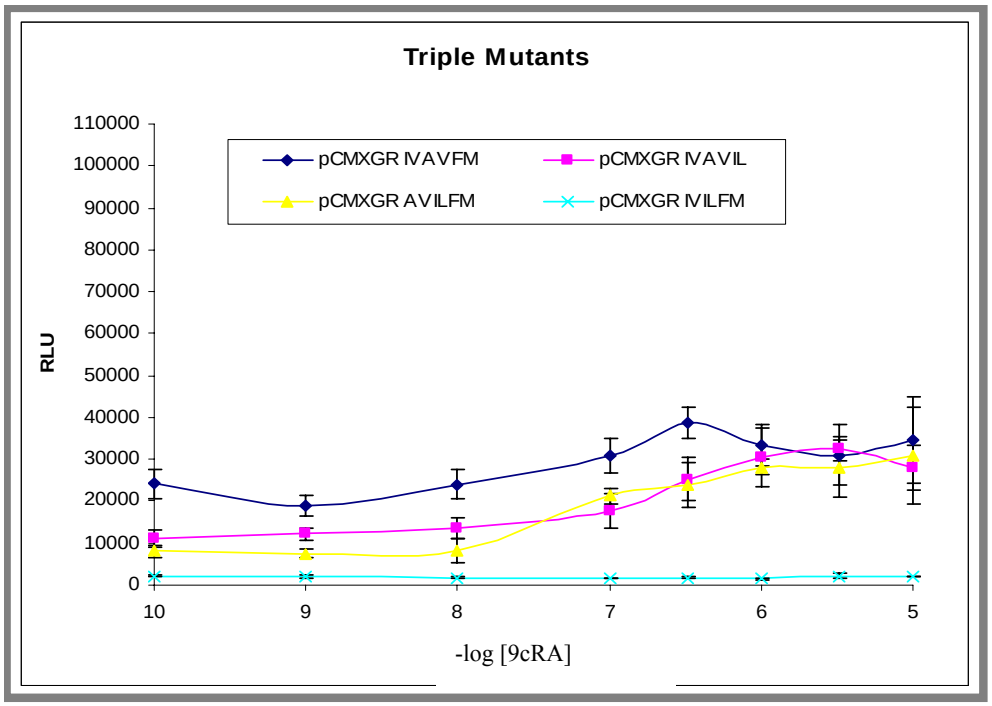


Figure 4.9 Activation profile of GR variant with single mutation.



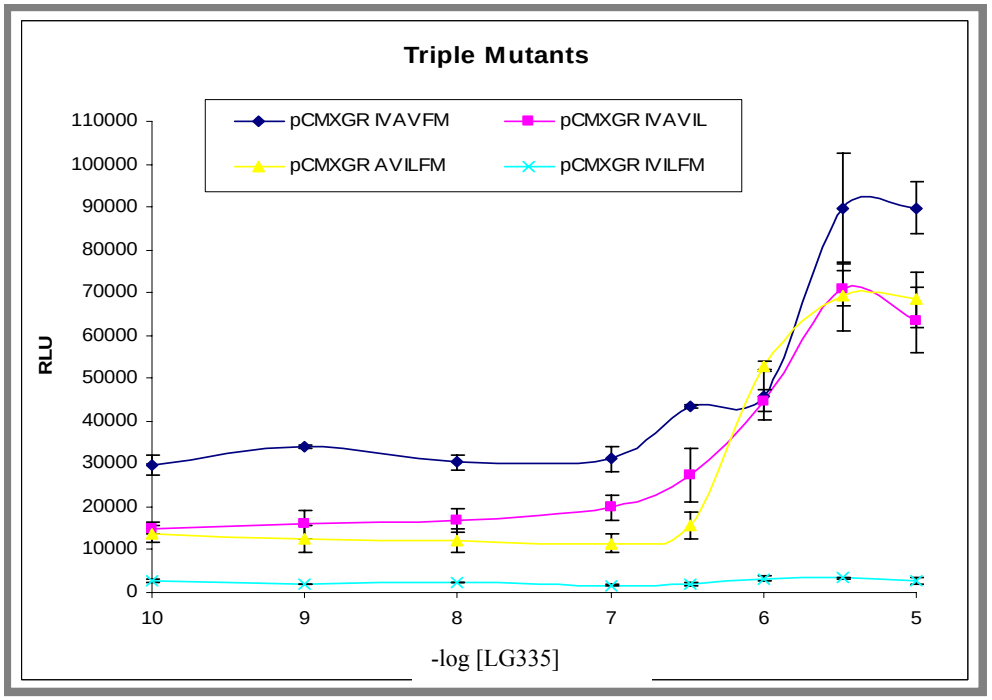


Figure 4.10 Activation profile of GR variants with three mutations.

As an effort to decrease the basal activity (leaky expression) the co-repressor silencing mediator of retinoid and thyroid hormone receptor (SMRT) was transfected with the GR variant in HEK293T cells. Co-repressors are proteins that are recruited by RXR in the absence in ligand. The co-repressor in turn recruits other proteins such as histone deactylase (HDAC), that modify histones resulting in tighter association of chromatin, preventing RNA polymerase access to transcribe the DNA, hence reducing the basal expression. Upon addition of the ligand, the co-repressor is displaced and the target gene is expressed.

To decrease the basal expression in the absence of ligand, equal concentration of the plasmid containing the co-repressor SMRT and GR I268V; A272V; I310L; F313M (40 ng/well) was transfected in HEK293T cells. The basal expression of GR I268V; A272V; I310L; F313M was not significantly decreased by adding the corepressor SMRT, as the standard deviation of the two groups (i.e. the GR variant with or without the corepressor) overlap (Figure 4.11 A). Similar result was observed on adding SMRT to another GR variant I268A; I310A; F313A; L436F (Figure 4.11 B). Hence, addition of the co-repressor SMRT had negligible effect on decreasing the basal activity of GR I268V; A272V; I310L; F313M and this GR variant cannot be used as a potential molecular switch.

Retroviral vector containing the GR variants

Two GR variants I268A; I310A; F313A; L436F (GR 'A') and I268V; A272V; I310M; F313S; L436M (GR 'B") from the "RXR library"[14] and GR Q275C; I310M; F313I (GR' C'), previously described as the best OLRP (Chapter 3), were cloned in the retroviral

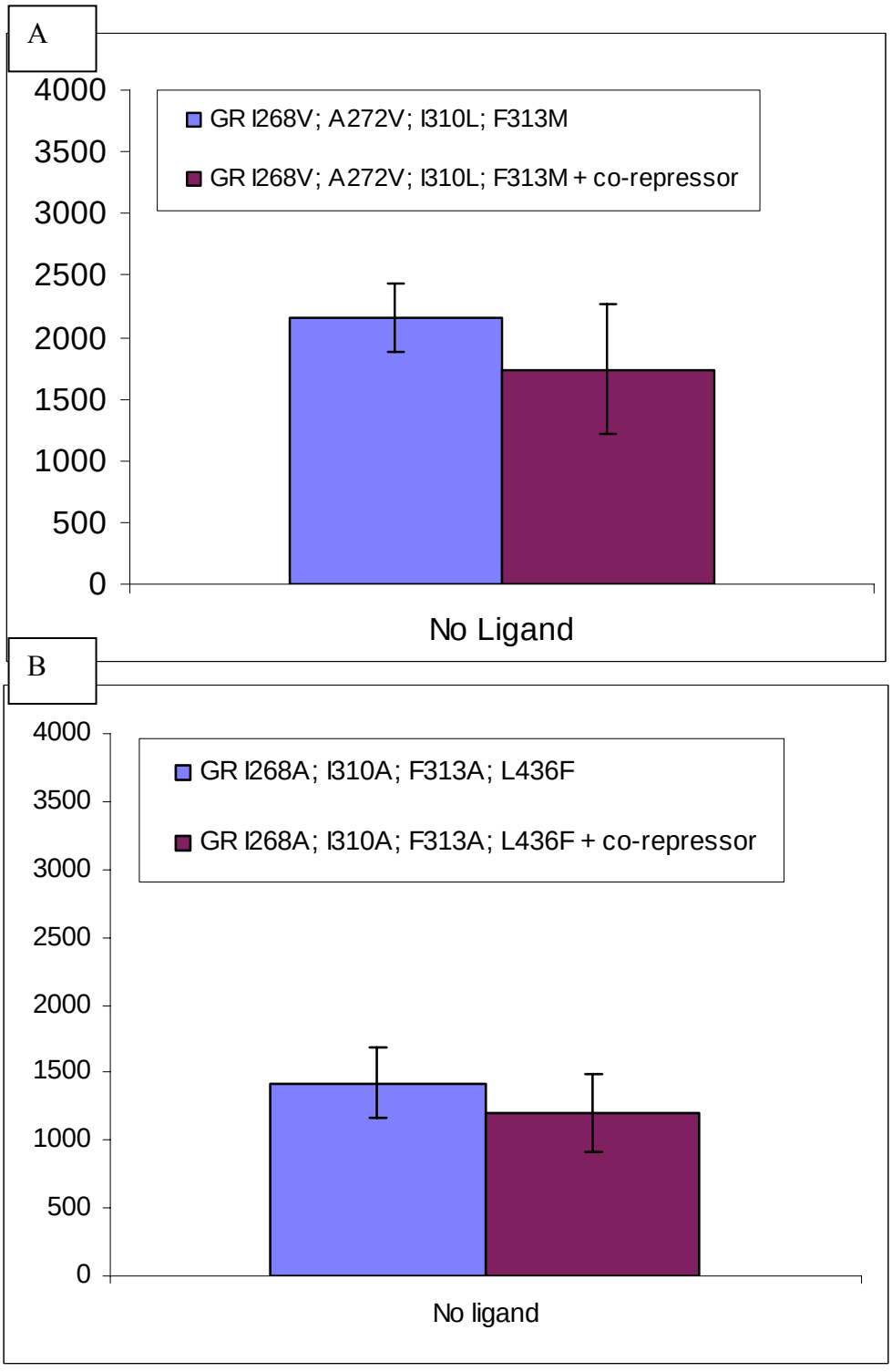


Figure 4.11 Effect of corepressor SMRT on the basal activity of A) GR I268V; A272V; I310L; F313M and B) GR I268A; I310A; F313A; L436F in the absence of ligand

vector, pMSCV (Figure 4.6). The retroviral vectors containing the GR variants, pMSCV GR ('A' or 'B' or 'C') were transfected in HEK293T cells with a reporter plasmid containing Gal4 RE driven firefly luciferase, p17*4 Tata Luc. The GR variants ('A' or 'B' or 'C') in the retroviral vector, pMSCV, induced luciferase expression similar to when they were present in the mammalian vector, pCMX GR ('A' or 'B' or 'C') (Figure 4.5). To further investigate the high basal activity, the pMSCV GR ('A' or 'B' or 'C') were tested with reporter plasmids containing Gal4 RE driven GFP, p17*4 Tata GFP. All the GR variants show negligible expression of GFP in the absence of ligand (Figure 4.12, 4.13, 4.14). The variant GR 'B' induces expression of GFP at 1 μM concentration of LG335 (Figure 4.13), GR 'A' induces expression at 0.1 μM concentration of LG335 (Figure 4.14) and GR'C' induced expression at 0.1 μM concentration of LG335 (Figure 4.12). The difference in basal expression of luciferase and GFP could imply that the leaky expression from the luciferase is still lower than the threshold required to observe a significant physiological change to express GFP.

Retroviral vectors containing both the GR variants and the reporter gene

To use these variants as molecular switches in mice, each GR variant 'A', 'B', 'C' was cloned into a retroviral vector along with the reporter gene, pMSCVGR ('A' or 'B' or 'C') _ reporter. The GR variant is under the control of LTR promoter and the reporter gene was under the control of Gal4 RE and minimal promoter (Figure 4.6). Similar luminescence was observed from the retroviral vectors, pMSCV GR ('A' or 'B' or 'C')_Luc as from the GR variant and reporter being in two different plasmids, pMSCV GR ('A' or 'B' or 'C') and p17*4 Tata Luc (Figure 4.5, 4.15). All the pMSCVGR ('A' or 'B' or 'C') _Luc are slightly

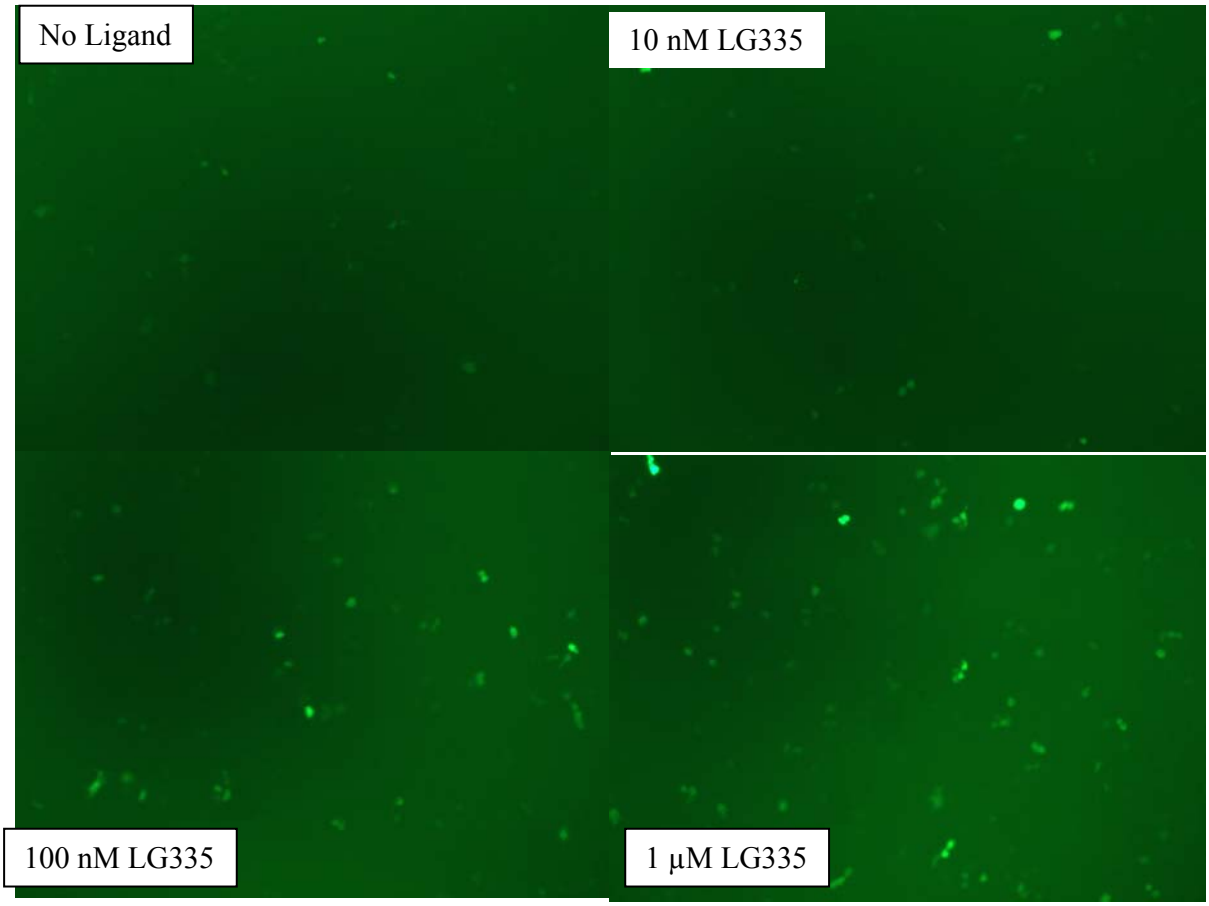


Figure 4.12 Detection of fluorescence from HEK 293T cells transfected with retroviral vector containing the GR variant Q275C; F310M; F313I and mammalian expression plasmid containing Gal4RE driven GFP. Fluorescence is observed is at four different concentrations of LG335

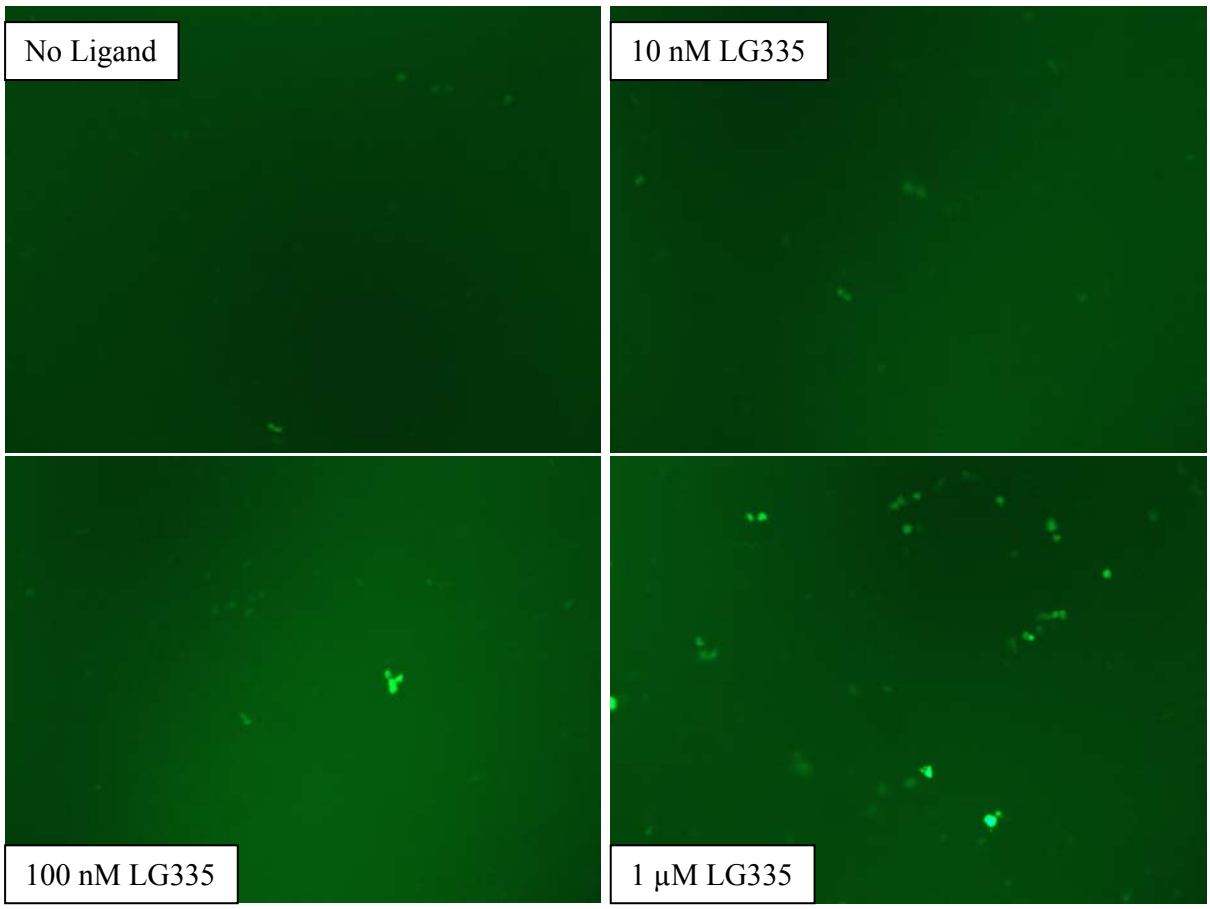


Figure 4.13 Detection of fluorescence from HEK 293T cells transfected with retroviral vector containing the GR variant I268V; A272V; I310M; F313S; L436M and mammalian expression plasmid containing Gal4RE driven GFP. Fluorescence is observed is at four different concentrations of LG335

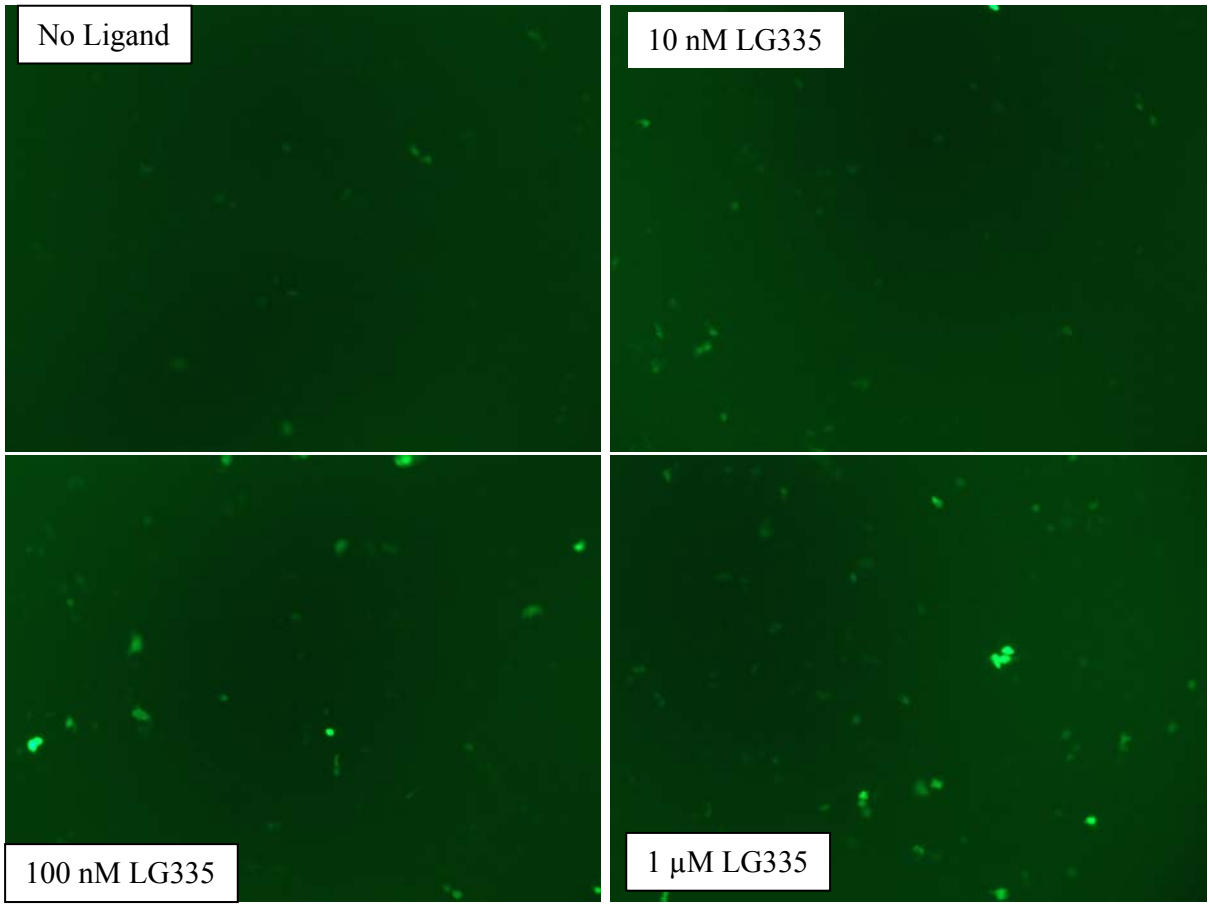
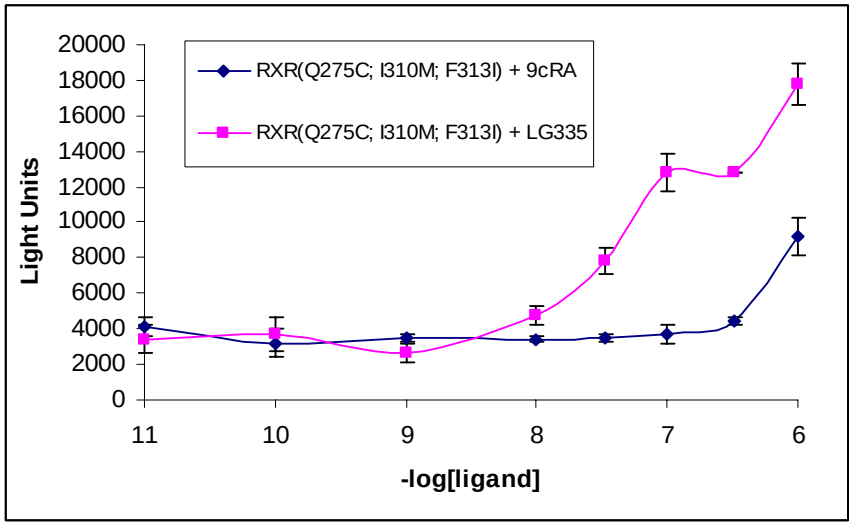
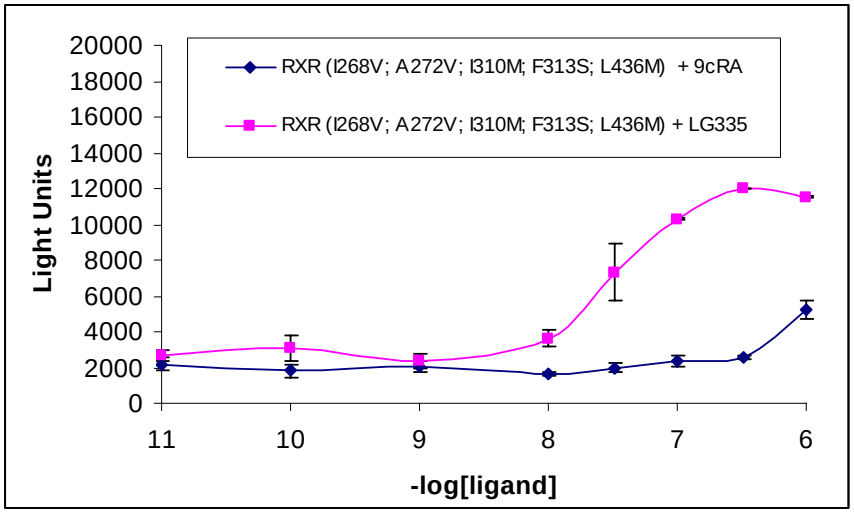


Figure 4.14 Detection of fluorescence from HEK 293T cells transfected with retroviral vector containing the GR variant I268A; I310A; F313A; L436F and mammalian expression plasmid containing Gal4RE driven GFP. Fluorescence is observed is at four different concentrations of LG335





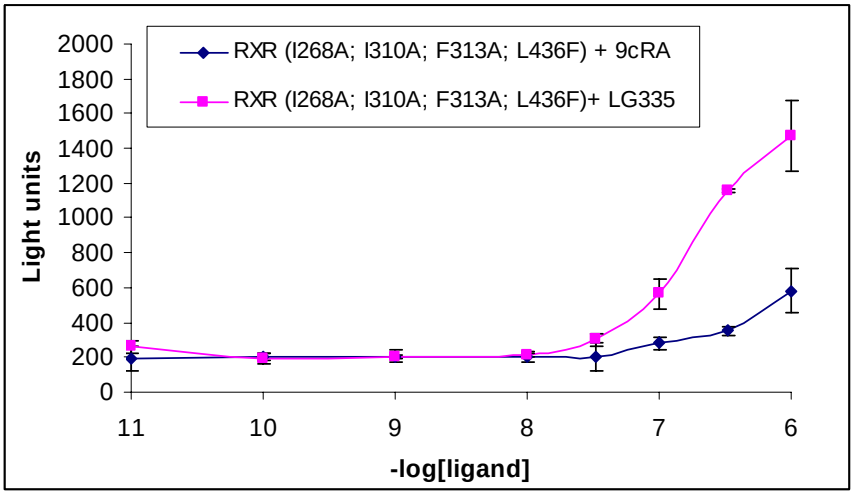


Figure 4.15 Activation profile of retroviral construct containing both the GR variants and Gal4 RE driven firefly luciferase in response to 9cRA and LG335.

activated by 9cRA and their EC_{50} for LG335 ranged from 50 nM – 100 nM. In addition, the induction varied from 6 to 10 fold.

The low induction level is due to the high basal activity of luciferase in the absence of ligand. To use these variants as molecular switches they should not induced in the absence of ligand. To further investigate the high basal activity the luciferase gene was be replaced by GFP in the retroviral vector. Retroviral vectors pMSCVGR ('A' or 'C') _GFP were transfected in HEK293T cells. In the absence of LG335, no fluorescence was observed from pMSCVGR 'A' _GFP and pMSCVGR 'C' _GFP, indicating the absence of basal activity (Figure 4.16, 4.17). At .1 μM concentration of LG335, fluorescence was detected by both variants pMSCVGR 'A' _GFP and pMSCVGR 'C'_GFP. With increasing concentration of the ligand, fluorescence from GFP increased (Figure 4.16, 4.17). The difference in basal expression of luciferase and GFP could mean that the leaky expression from the luciferase is still lower than the threshold required to observe a significant physiological change to express GFP. The results from GFP indicate that GR (I268A; I310A; F313A; L436F) 'A' and GR Q275C; I310M; F313I 'B' can potentially be used as molecular switches as it does not show a leaky expression and have comparatively low EC_{50}. The next step would be to virally transduce these switches and regulate the expression of HOXB4 gene, to differentiation hematopoietic stem cells into blood and immune cells [Figure 4.18]. Efforts will be made to modify these switches such that it can compete with other inducible systems.

Summary

In chapter 3, a molecular switch was engineered using RXR and LG335. This

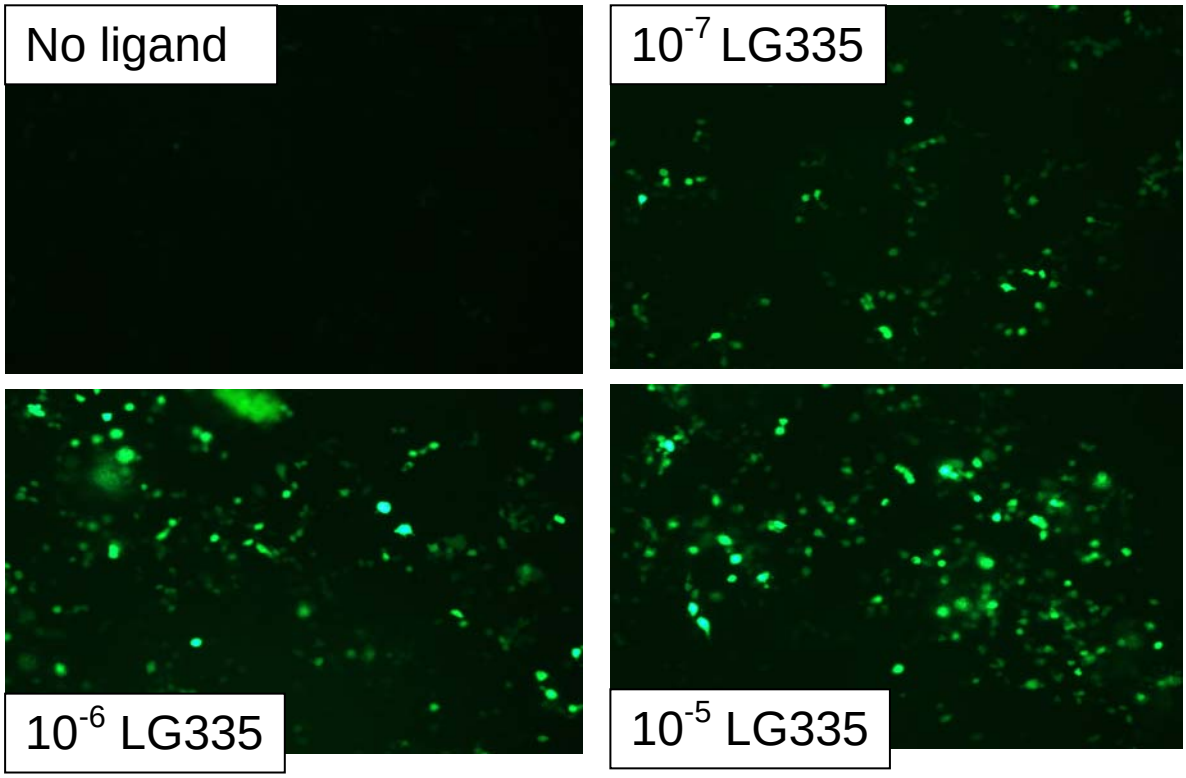


Figure 4.16 Detection of fluorescence from of HEK293T transfected with retroviral vector containing GR I268A; I310A; F313A; L436F and Gal4 RE controlled GFP in the same vector.

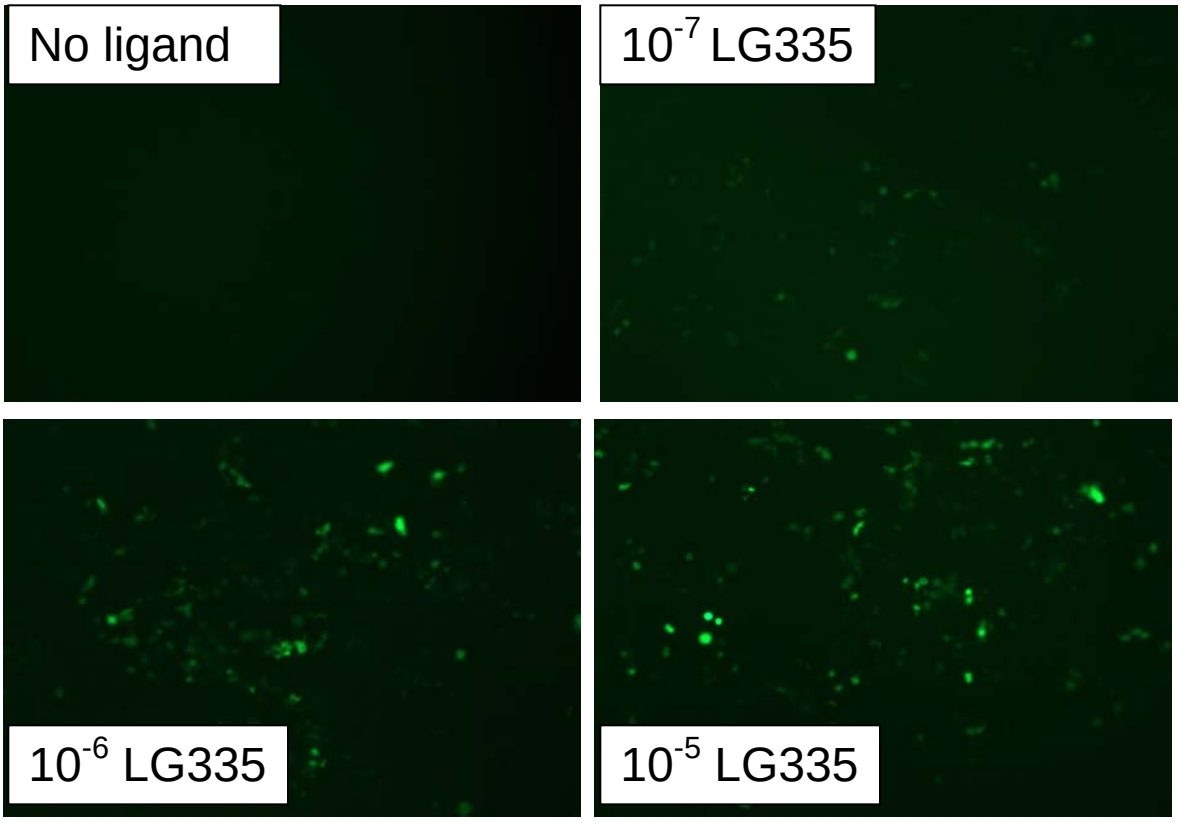


Figure 4.17 Detection of fluorescence from HEK293T transfected with retroviral vector containing both GR Q275C; I310M; F313I and Gal4 RE controlled GFP in the same vector.

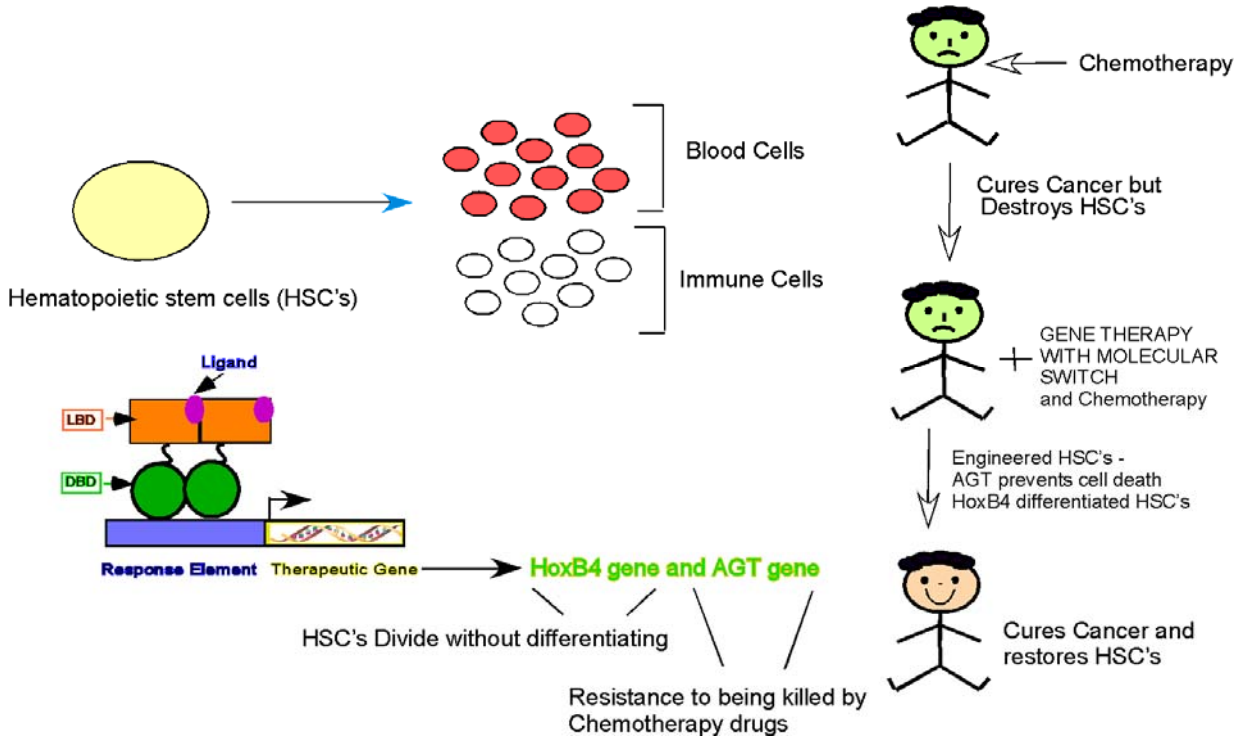


Figure 4.18 Schematic diagram showing the future application of the molecular switch on hematopoietic stem cells

molecular switch was created by mutating three residues in the RXR LBD via site-directed mutagenesis and replacing the DBD with the yeast transcription factor, Gal4 DBD. To engineer a sensitive switch that has nanomolar affinity for LG335, structure-based approach was applied to generate a library of RXR variants. The substitutions were made on the basis of size, charge or hydrophobicity at six positions (I268, A271, A272, I310, F313, and L436). To discover the functional variants within the libraries, a genetic selection assay method called chemical complementation was used in yeast. The library produced ligand-receptor pairs with LG335 that have a variety of EC_{50}s (40nM to > 2 μM) and activation levels (10% to 80% of wild-type RXR with 9cRA). The five most sensitive RXR variants I268A; I310A; F313A; L436F, I268V; A272V; I310L; F313M, I268A; I310S; F313A; L436F, I268V; A272V; I310M; F313S; L436M, and I268A; I310M; F313A; L436T were tested in mammalian cell culture. Out of the five, three variants I268V; A272V; I310L; F313M, I268A; I310A; F313A; L436F and I268V; A272V; I310M; F313S; L436M were activated by LG335 to a much higher level then wild type RXR. They had low EC_{50} for LG335 of 1 nM, 30 nM, and 180 nM respectively and an EC_{50} for 9cRA of > 1 μM. These variants were then made orthogonal by replacing the RXR DBD by Gal4 DBD (GR variants). These orthogonal receptors GR I268V; A272V; I310L; F313M, I268A; I310A; F313A; L436F and I268V; A272V; I310M; F313S; L436M had similar activation profiles as with RXR variants containing the RXR DBD, with the exception of high basal activity in the uninduced state which results in a low induction ratio. In an attempt to reduce the basal activity, co-repressor SMRT was used. Addition of co-repressor had no effect on decreasing the basal activity. To introduce the GR variants in *vivo,* they were cloned into retroviral vectors along with a Gal4 RE controlled reporter genes, firefly luciferase or green fluorescence protein (GFP). One of

the variants I268V; A272V; I310L; F313M expressed GFP and luciferase at high levels in the absence of ligand. Other variants GR I268A; I310A; F313A; L436F, I268V; A272V; I310M; F313S; L436M and Q275C; I310M; F313I expressed luciferase gene but no fluorescence from GFP was not detected in the absence of ligand. The variants had relatively low EC_{50} with LG335 and 8 to 10 induction ratio. The results suggest that GR I268A; I310A; F313A; L436F and Q275C; I310M; F313I could be used as potential molecular switches.

Future Work

The reduction in fold induction of GR variants is due to an elevated level of basal expression in the absence of ligand. The molecular switches can be improved by addressing the issue of high basal expression. Presently, full length Gal4 DBD (residue 1-147) is utilized to construct the molecular switches. GeneSwitch® inducible system, utilizes a chimeric regulator composed of VP16 activation domain, a truncated version of Gal4 DBD (residues 2-93) and mutated progesterone receptor. As an effort to decrease the basal activity a truncated version of Gal4 DBD can be utilized [17].

Higher basal expression implies that the GR variants are binding the Gal4 sites and are partially activating the transcription of the target gene in the absence of ligand. One strategy to reduce basal expression is to introduce autogeneous regulatory circuit [17]. For autogenous regulatory circuit the constitutive CMV promoter of the GR variant would be replaced by an autoinducible promoter consisting of four copies of Gal4 response elements linked to a minimal thymidine kinase (tk) promoter. The minimal tk promoter will provide a low level of expression of the GR protein in the absence of ligand. When the ligand is added,

the GR variant present at a low level would get activated and bind to Gal4 RE in the autoinducible promoter. This would induce the synthesis of more GR variant that would in return activate the target gene also. Hence, introducing the autoinducible promoter may provide a reduction in the basal expression.

Materials and Methods

Constructing RXR library and plasmids

The RXR library was constructed and screened by Dr Lauren Schwimmer, the details are mentioned in her thesis, from page 53-62 [14, 18]. The RXR variants that were activated with LG335 were cloned from yeast vector pGBD to mammalian vector pCMX. pCMXRXR variants expressing RXR DBD- RXR LBD (variant) was constructed, by cloning only the LBD of RXR variant from pGBD vectors using Sal I and Pst I restriction sites. pCMXGR variants expressing Gal4DBD fused to RXR LBD variants was constructed by cutting the RXR LBD variant from pGBD vector and inserting it into pCMXGRwt vector using Sal I and Pst I restriction enzymes. Single and triple mutants from GR I268V; A272V; I310L; F313M were made by either eliminating a mutation from the quadruplet to make a triple mutant variant or by introducing a mutation in wild type pCMXGR using oligonucleotides.

The retroviral vector pMSCV was a kind gift from Dr Trent Spencer (Emory University). pMSCVGR variants were cloned from pCMXGR by amplifying Gal 4 DBD-RXR LBD gene using primers and inserting it via Avr II and Not I restriction sites. pMSCVGR variant_Luc was constructed by inserting the PCR amplified fragment of Gal4

RE _luciferase from p17*4 Tata luc into pMSCVGR via Not I and Sac II. p17*4 Tata GFP was constructed from p17*4 Tata luc by replacing luciferase gene via Sac II and Not I restriction sites and inserting in GFP gene from pEGFP (Clonetech). pMSCVGR variant_GFP was constructed by introducing GFP gene after the Gal4 RE via Sac II and Hind III restriction sites in pMSCV GR variant.

Transfection protocol

Transient transfection protocol is same as mentioned in chapter 3.

Literature Cited

1. Wurm, F.M., K.A. Gwinn, and R.E. Kingston, *Inducible overproduction of the mouse c-myc protein in mammalian cells.* Proc Natl Acad Sci U S A, 1986. **83**(15): p. 5414-8.

2. Kothary, R., S. Clapoff, S. Darling, M.D. Perry, L.A. Moran, and J. Rossant, *Inducible expression of an hsp68-lacZ hybrid gene in transgenic mice.* Development, 1989. **105**(4): p. 707-14.

3. Filmus, J., J. Remani, and M.H. Klein, *Synergistic induction of promoters containing metal- and glucocorticoid-responsive elements.* Nucleic Acids Res, 1992. **20**(11): p. 2755-60.

4. Rinsch, C., E. Regulier, N. Deglon, B. Dalle, Y. Beuzard, and P. Aebischer, *A gene therapy approach to regulated delivery of erythropoietin as a function of oxygen tension.* Hum Gene Ther, 1997. **8**(16): p. 1881-9.

5. Wang, Y., B.W. O'Malley, Jr., S.Y. Tsai, and B.W. O'Malley, *A regulatory system for use in gene transfer.* Proc Natl Acad Sci U S A, 1994. **91**(17): p. 8180-4.

6. Wang, Y., J. Xu, T. Pierson, B.W. O'Malley, and S.Y. Tsai, *Positive and negative regulation of gene expression in eukaryotic cells with an inducible transcriptional regulator.* Gene Ther, 1997. **4**(5): p. 432-41.

7. Nordstrom, J.L., *Antiprogestin-controllable transgene regulation in vivo.* Curr Opin Biotechnol, 2002. **13**(5): p. 453-8.

8. Favre, D., V. Blouin, N. Provost, R. Spisek, F. Porrot, D. Bohl, F. Marme, Y. Cherel, A. Salvetti, B. Hurtrel, J.M. Heard, Y. Riviere, and P. Moullier, *Lack of an immune response against the tetracycline-dependent transactivator correlates with long-term doxycycline-regulated transgene expression in nonhuman primates after intramuscular injection of recombinant adeno-associated virus.* J Virol, 2002. **76**(22): p. 11605-11.

9. Latta-Mahieu, M., M. Rolland, C. Caillet, M. Wang, P. Kennel, I. Mahfouz, I. Loquet, J.F. Dedieu, A. Mahfoudi, E. Trannoy, and V. Thuillier, *Gene transfer of a chimeric trans-activator is immunogenic and results in short-lived transgene expression.* Hum Gene Ther, 2002. **13**(13): p. 1611-20.

10. No, D., T.P. Yao, and R.M. Evans, *Ecdysone-inducible gene expression in mammalian cells and transgenic mice.* Proc Natl Acad Sci U S A, 1996. **93**(8): p. 3346-51.

11. Karns, L.R., A. Kisielewski, K.M. Gulding, J.M. Seraj, and D. Theodorescu, *Manipulation of gene expression by an ecdysone-inducible gene switch in tumor xenografts.* BMC Biotechnol, 2001. **1**: p. 11.

12. Palli, S.R., M.Z. Kapitskaya, M.B. Kumar, and D.E. Cress, *Improved ecdysone receptor-based inducible gene regulation system.* Eur J Biochem, 2003. **270**(6): p. 1308-15.

13. Subbarayan, V., M. Mark, N. Messadeq, P. Rustin, P. Chambon, and P. Kastner, *RXRalpha overexpression in cardiomyocytes causes dilated cardiomyopathy but fails to rescue myocardial hypoplasia in RXRalpha-null fetuses.* J Clin Invest, 2000. **105**(3): p. 387-94.

14. Schwimmer, L.J., P. Rohatgi, B. Azizi, K.L. Seley, and D.F. Doyle, *Creation and discovery of ligand-receptor pairs for transcriptional control with small molecules.* Proc Natl Acad Sci U S A, 2004. **101**(41): p. 14707-12.

15. Doyle, D.F., D.A. Braasch, L.K. Jackson, H.E. Weiss, M.F. Boehm, D.J. Mangelsdorf, and D.R. Corey, *Engineering orthogonal ligand-receptor pairs from "near drugs".* J Am Chem Soc, 2001. **123**(46): p. 11367-71.

16. Schmitz, M.L. and P.A. Baeuerle, *The p65 subunit is responsible for the strong transcription activating potential of NF-kappa B.* Embo J, 1991. **10**(12): p. 3805-17.

17. Abruzzese, R.V., D. Godin, M. Burcin, V. Mehta, M. French, Y. Li, B.W. O'Malley, and J.L. Nordstrom, *Ligand-dependent regulation of plasmid-based transgene expression in vivo.* Hum Gene Ther, 1999. **10**(9): p. 1499-507.

18. Schwimmer, L.J., *Engineering ligand receptor pairs for small molecule control of transcription.* Thesis, Georgia Tech, 2005.

CHAPTER 5

DEVELOPING CHEMICAL COMPLEMENTATION IN MAMMALIAN CELLS

Introduction

Genetic selection is a powerful tool used in many aspects of molecular biology such as protein and enzyme engineering, analyzing large protein libraries, and evaluating macromolecular interactions[1-8]. The idea behind genetic selection is that the host cells will survive only if the desired function of a protein is present in the cell. Hence the signal to noise ratio is high which makes it efficient to evaluate protein libraries. Chemical complementation is a method that links a small molecule to genetic selection where the small molecule allows the survival of the cell. In other words, survival of cells is dependent on the ability of the ligand to bind and activate the nuclear receptor which in turn leads to the expression of a selective marker. This can be extended to various applications, including drug discovery and designing molecular switches for gene therapy and protein engineering. Recently chemical complementation was developed in our lab in *S. cerevisiae* using the strain PJ69-4A[9],[4]. This strain consists of the Gal4 response element (Gal4 RE) controlling expression of the *ADE2* gene which encodes for phosphoribosylaminoimidazole carboxylase which catalyzes an essential step in adenine and histidine biosynthesis respectively. PJ69-4A is transformed with a library of RXR variants fused to the Gal4 DBD and a nuclear receptor coactivator fused to the Gal4 activation domain (Gal4 AD). In the presence of an agonist, the small molecule binds to

the RXR LBD, recruiting the coactivator Gal4 AD fusion protein initiating transcription of either the *ADE2* or *HIS3* gene. Therefore, in media deficient in adenine, yeast cells will only grow in the presence of ligand. The RXR variant that binds the ligand is rescued from the yeast cell and sequenced. This method is positive selection, the presence of the ligand induces cell growth. A negative selection method was also developed to eliminate the constitutively active variants that are able to turn on transcription even in the absence of ligand. A derivative of the PJ69-4A strain was developed called BAPJ69 [10]. This strain contains Gal4 REs controlling the expression of *ADE2*, *HIS3*, *lacZ*, and *URA3* genes (Figure 5.1). The *URA3* gene codes for the enzyme orotidine-5'-phosphate decarboxylase (OMP) that plays a role in the uracil biosynthetic pathway. The advantage of using the *URA3* gene is that it can function as both a positive and a negative selective marker. The *URA3* gene acts as a positive selective marker when the yeast cells are grown in uracil deficient media in the presence of ligand. The *URA3* gene can also be used in negative selection with the use of the compound, 5'-fluoroorotic acid (5'-FOA). Yeast cells expressing the *URA3* gene in media containing FOA, produce the enzyme OMP which converts the compound FOA into 5'-fluorouracil (5'-FU). 5'-FU is a toxin, and the yeast cells producing 5'-FU die. Therefore, this method is termed "negative chemical complementation". This technique is useful to eliminate constitutively active receptors from protein libraries.

The goal here is to develop chemical complementation in mammalian cells to screen or genetically select for ligand-activated RXR variants. Developing chemical complementation in mammalian cells would eliminate the extra step of creating and testing protein libraries in *S. cerevisiae* before testing them in mammalian cells. Also, it

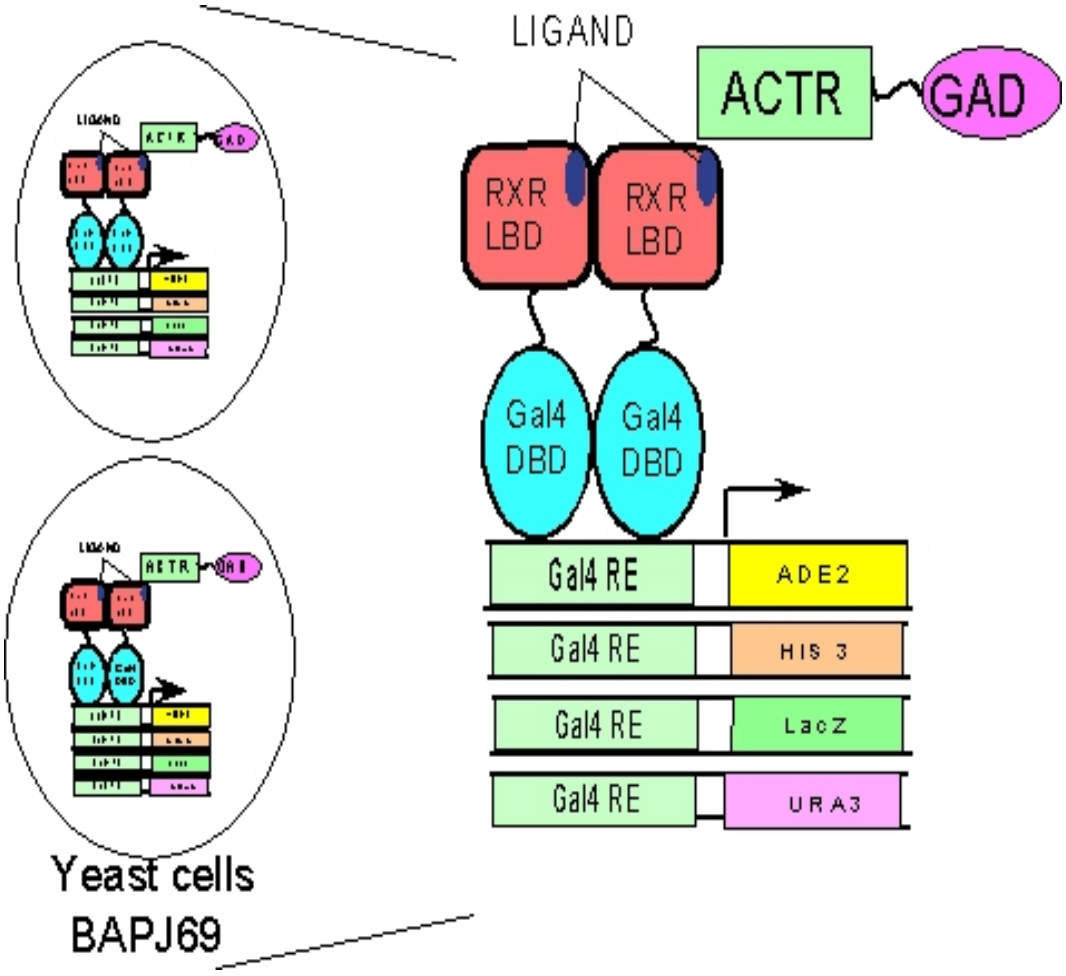


Figure 5.1 Positive and negative chemical complementation in *S. cerevisiae*

would eliminate the experimental discrepancy and variations observed between *S. cerevisiae* and mammalian cells.

To develop chemical complementation in mammalian cells, a modified HeLa cell line has to be engineered to perform positive and negative genetic selection and screen for ligand activated RXR variants. To achieve this goal, HeLa cells have to be modified to contain a Gal4 response element controlling the expression of a triple fusion target gene (Figure 5.2). The target gene is a fusion of positive and negative selective markers and the reporter gene. The positive selective marker is a neomycin resistant gene (Neo), which allows the cells expressing Neo to survive in media containing geneticin [11]. The negative selective marker is the thymidine kinase gene (Ttk) from the herpes simplex virus, which sensitizes cells to the toxic effect of the antiviral drug ganciclovir [12, 13]. The reporter gene is *Renilla* luciferase (Hrl), this gene can be assayed for luciferase activity and can be used to quantify the expression level.

The modified HeLa cell line when transfected with ligand activated RXR variants will initiate transcription of the triple fusion gene, Neo-Hrl-Ttk (NHT). Transfected HeLa cells cultured in media containing geneticin will survive only if the RXR variant is ligand activated and initiates transcription of the Neo gene. The transfected cells that survive the positive selection must contain an RXR variant that is either constitutively active or is activated by the agonist. To eliminate constitutively active RXR variants, the cells will be subjected to negative selection (i.e. negative chemical complementation). For negative selection, transfected cells will be cultured in media containing ganciclovir, constitutively active receptors that bind to an endogenous small molecule and activate the transcription

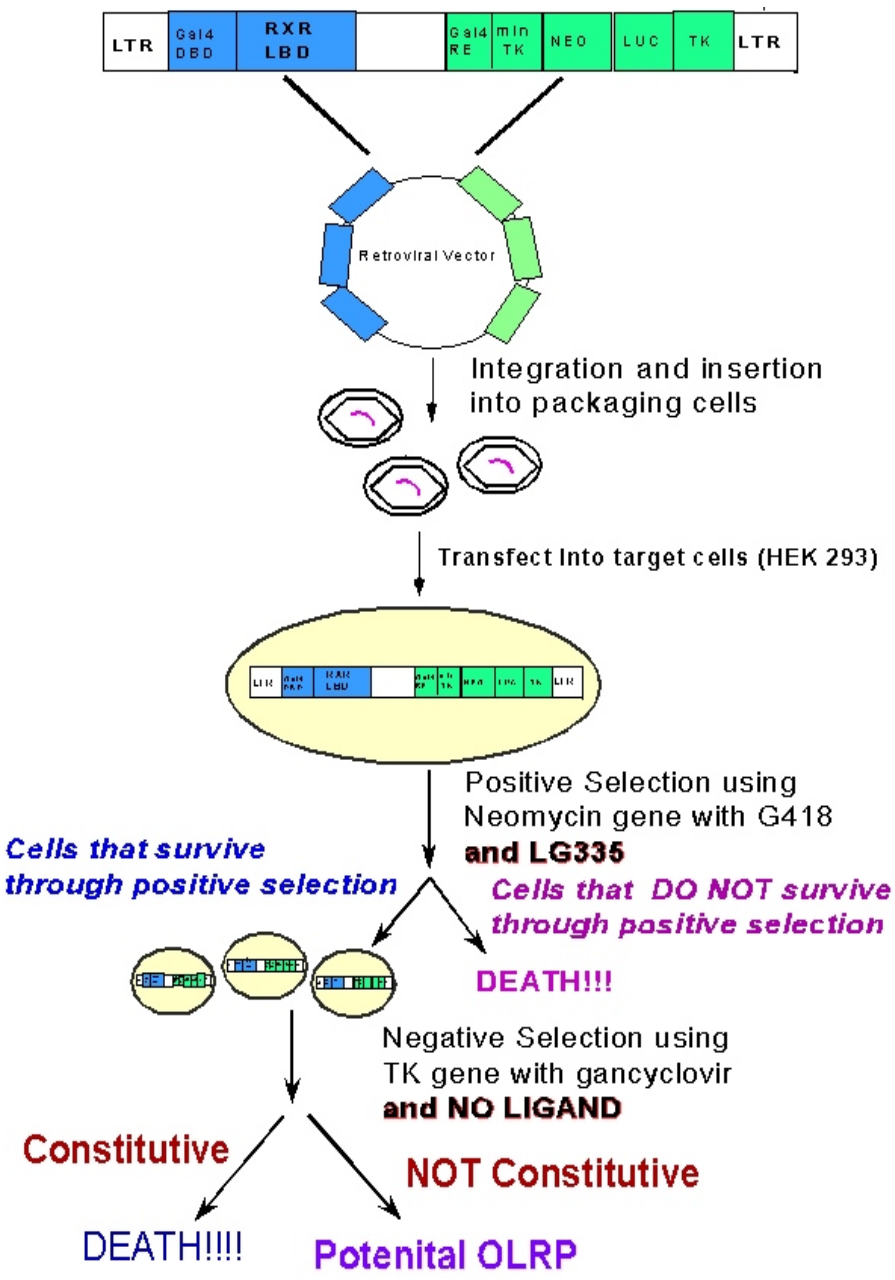


Figure 5.2 Schematic depiction of developing chemical complementation in mammalian cells.

of thymidine kinase gene will die. Chemical complementation will leave only those cells containing ligand activated RXR variants or orthogonal ligand receptor pairs (Figure 5.2).

Result and Discussion

Constructing and testing the triple fusion gene

To perform positive and negative chemical complementation in mammalian cells the triple fusion gene Neo, Hrl and Ttk (NHT) was initially cloned in a commercial vector pBind from Stratagene, namely pBindNHT. Details about the plasmid construction and cloning are described in the "Materials and Methods" section. The NHT gene is under the control of a strong promoter such that it is constitutively expressed and the functional of the Neo gene could be examined. To test the functionality of Neo gene in the triple fusion construct, pBindNHT was stably transfected in HeLa cells in the presence of geneticin. Transfected cells expressed the Neo gene and survived in media containing 400 µg/ml geneticin whereas the control i.e. non transfected cells died. This result confirmed that the Neo gene is functional in the triple fusion gene.

The next step was to test if the Hrl gene is functional in the NHT fusion construct. The NHT gene was cloned into a vector containing the TATA box and four tandem repeats of 17 bp Gal4 RE, p17*4 Tata NHT. The expression of NHT gene is controlled via a ligand activated RXR variant. To test the functional of the Hrl gene, HEK293T cells were transiently transfected with RXR variant Q275C; I310M; F313I containing the Gal4 DBD (GR QCIMFI) and p17*4 Tata NHT. In the presence of LG335, GR QCIMFI

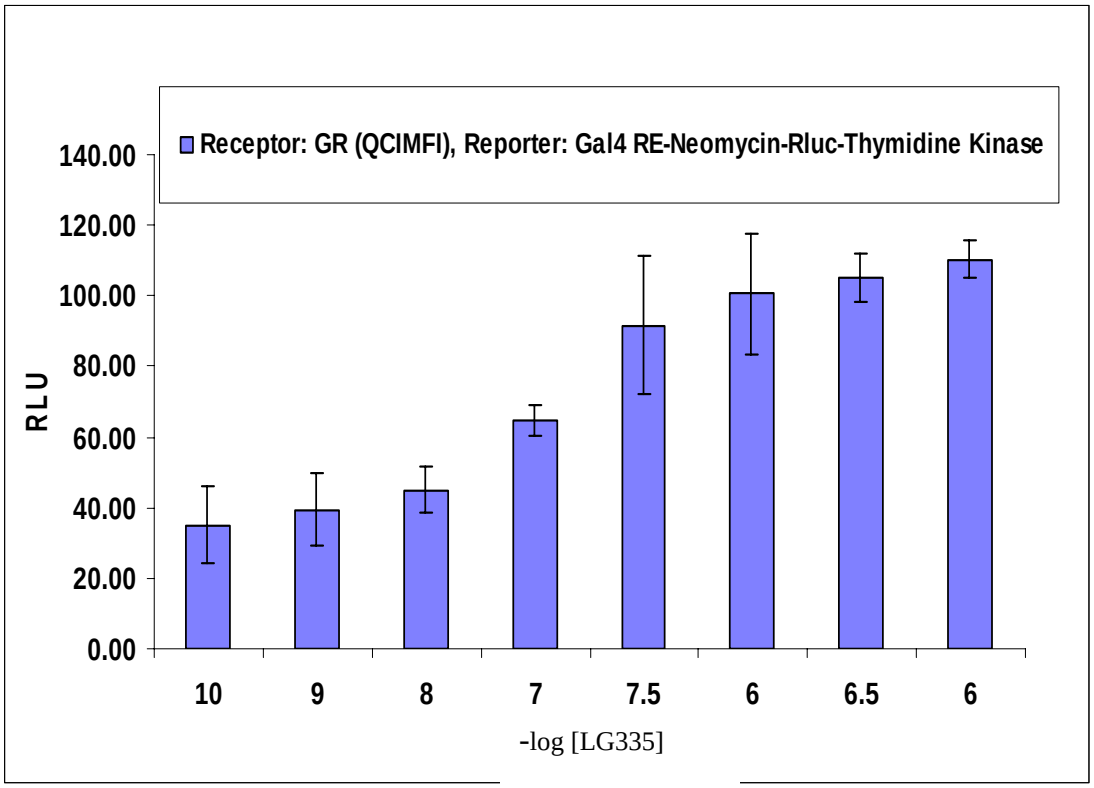


Figure 5.3 Dose response of GR Q275C; I310M; F313I and the triple fusion gene construct of neomycin, luciferase and thymidine kinase (NHT) with LG335 in HEK293T cells.

underwent a conformational change and initiated transcription of the NHT gene. When assayed for *Renilla* luciferase activity, a slight induction was observed in the presence of LG335 (Figure 5.3). The fold activation of GR QCIMFI and *Renilla* luciferase in NHT fusion protein (p17*4 Tata NHT) is four fold. Whereas, the fold activation of GR QCIMFI and *Renilla* luciferase alone i.e. p17*4 Tata Hrl is nine fold (Figure 3.5). This difference in the fold induction of the NHT gene versus the Hrl gene alone is due to the high basal activity of the NHT gene in the absence of ligand (Figure 5.3). The reason for the low expression level of the *Renilla* luciferase gene could also be the improper folding of the protein. Two research groups, Oh [14] and Strathdee [15] have also constructed triple fusion genes with positive and negative selective markers. Strathdee and coworkers constructed a triple fusion gene between the hygromycin gene, GFP and thymidine kinase gene [15]. Oh and coworkers constructed a triple fusion gene between thymidine kinase, *Renilla* luciferase and neomycin. The *Renilla* luciferase activity observed in the triple fusion protein was only 5% of the *Renilla* luciferase activity when it is expressed alone [14]. This may mean that *Renilla* luciferase gene does not express very well being in the centre of the triple fusion gene due to the improper protein folding and hence requires more molecular freedom. In the future, the triple fusion gene would be constructed with *Renilla* luciferase gene being at the N- or C- terminus but for initial testing of chemical complementation the same triple fusion gene NHT was utilized.

Testing chemical complementation in mammalian cells

The long term goal is to perform positive and negative chemical complementation in mammalian cells to screen or select protein libraries such as a library of RXR variants for creating ligand activated molecular switches. As a proof of principle, positive chemical complementation was examined by using a RXR variant that is activated by the ligand LG335. The RXR variant I268V; A272V; I310L; F313M containing the Gal4 DBD (GR I268V; A272V; I310L; F313M) as mentioned in chapter 4, has a nano molar affinity with ligand LG335. This variant GR I268V; A272V; I310L; F313M was employed to test chemical complementation in mammalian cells.

A retroviral vector containing the variant GR I268V; A272V; I310L; F313M followed by the Gal4 RE and the triple fusion gene NHT was constructed. The retroviral vector was stably transfected in HeLa cells by using 400 µg/ml geneticin and 1 µM LG335 in the media. Theoretically, the HeLa cells would survive only if the GR I268V; A272V; I310L; F313M and NHT genes are integrated in the cell's genome and if GR I268V; A272V; I310L; F313M is ligand activated and initiates transcription of the NHT gene. After 5 days, the non transfected cells incubated with geneticin in the media died, whereas the transfected HeLa cells formed colonies. This result suggests successful integration of the genes and expression of Neo gene. To test the functional of the *Renilla* luciferase gene and the fold induction of GR I268V; A272V; I310L; F313M the cells were lysed and assayed for *Renilla* luciferase activity. There was no luminescence observed, suggesting that the Hrl gene in the triple fusion gene is not expressing. This could be because of improper folding of Hrl gene or due to integration of a truncated version of NHT gene in the cell's genome.

Summary

In the present study, the triple fusion gene NHT was cloned in a mammalian vector and was tested if the Neo and Hrl genes are functional. HeLa cells were stably transfected with the triple fusion gene NHT, to show that the Neo gene is working. HEK293T cells were transiently transfected with a plasmid containing GR variant Q275C; I310M; F313I and a plasmid containing a Gal4 RE driven NHT gene to show the that the Hrl gene is functional. Chemical complementation was tested using the retroviral vector containing GR variant I268V; A272V; I310L; F313M and Gal4 RE driven triple fusion gene. HeLa cells were stably transfected with the retroviral vector. The GR variant I268V; A272V; I310L; F313M activated the triple fusion gene in the presence of ligand and expressed the neomycin gene, but *Renilla* luciferase activity was not observed.

Future Work

To troubleshoot chemical complementation in mammalian cells the first step is to confirm the integration of Hrl gene in the stably transfected HeLa cells. This can be achieved via PCR amplification of each component of the triple fusion gene from the cell's genome. The next step would be to investigate the best possible way for constructing the triple fusion gene between the neomycin resistant gene, *Renilla* luciferase gene and thymidine kinase gene, to improve the expression of *Renilla* luciferase. Presently, *Renilla* luciferase gene is in the centre of NHT gene with approximately twenty amino acid spacer on both ends. Different spacer lengths and

rearrangement of the genes can be tried such that Hrl gene is at the N or C terminal of the triple fusion gene and have more molecular freedom to fold properly and have better expression. In addition, other reporter genes such as GFP can be used instead of *Renilla* luciferase gene.

Materials and Methods

Plasmid construction

Construction of plasmids expressing the RXR variants Q275C; I310M; F313I and I268V; A272V; I310L; F313M fused to Gal4 DBD, pCMXGR Q275C; I310M; F313I and pCMXGR I268V; A272V; I310L; F313M is described in chapter 3 and 4 respectively. The Hrl (933 bp) and Ttk (993) genes were a kind gift from Dr Gambhir (Stanford University). These two genes Hrl-Ttk were cloned into a commercial vector pDrive (Qiagen) using restriction enzymes NheI and NotI, named pDrive HrlTtk. The spacer between the Hrl and Ttk gene is ctcgagaattctcacgcgtctgcaggatatcaagcttgcggtaccgcgggcccgggatccgccacc. The Neo gene (801) was PCR amplified from pEGFPNeo (Clonetech) and inserted into pDrive HrlTrk using restriction enzymes SalI and NheI, named pDrive NHT (Figure 5.4). The spacer between the Neo and Hrl gene is ggcacagtcaagctcggagctagcgccacc. To test the functionality of Neo in the triple fusion gene, NHT was PCR amplified from pDrive NHT and inserted into a commercial mammalian vector pBind from Stratagene, using

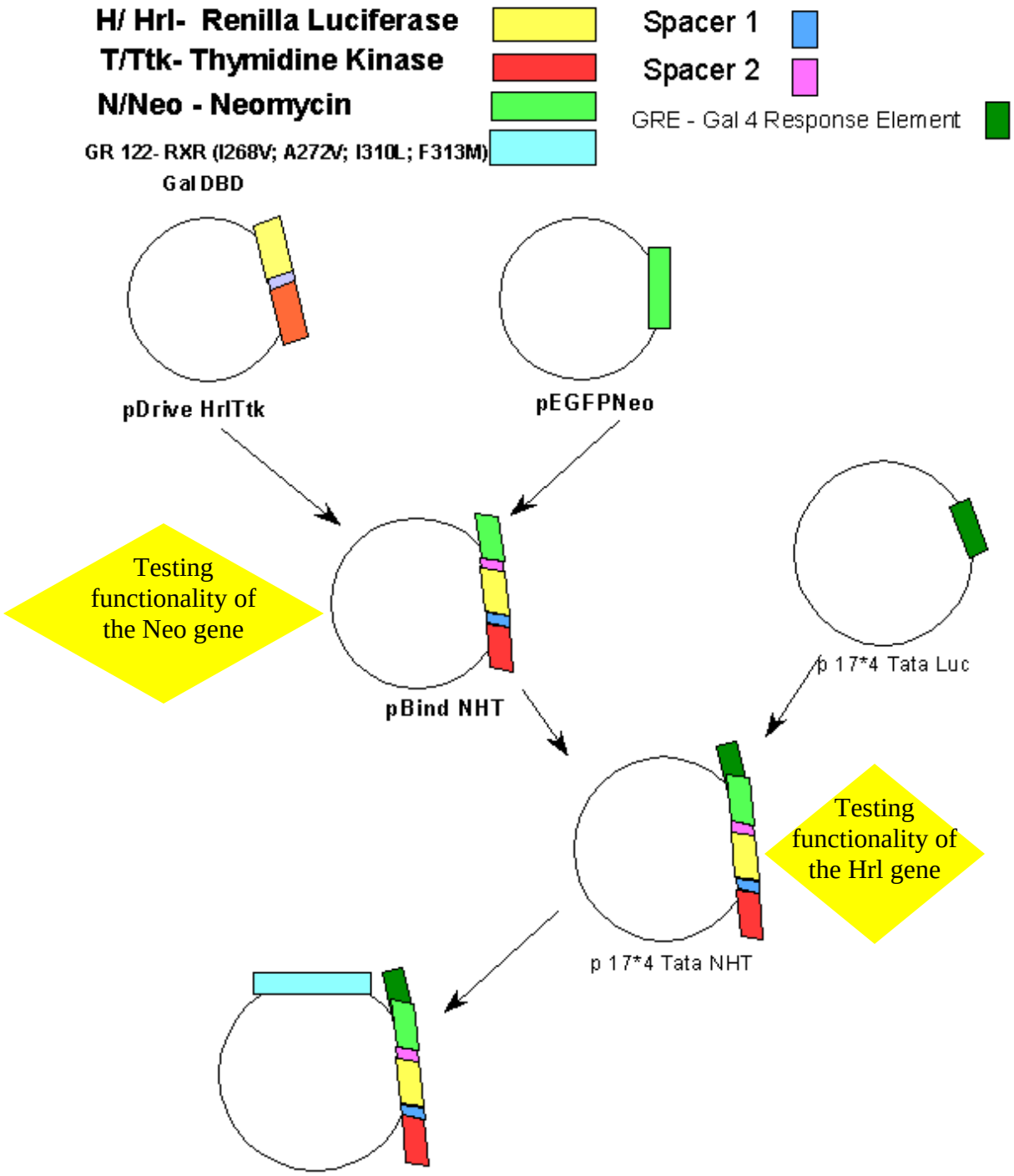


Figure 5.4 Gene fusion between neomycin resistant gene, *Renilla* luciferase and thymidine kinase.

restriction enzymes AvrII and NotI. To test the function of the Hrl gene in the triple fusion gene, NHT was cloned in a mammalian vector containing four repeats of 17 base pair Gal4 RE and a TATA box using restriction enzymes AvrII and NotI, namely p17*4TataNHT. To test chemical complementation in mammalian cells, a retroviral vector was employed, pMSCV. The GR variant I268V; A272V; I310L; F313M was cloned from pCMXGR I268V; A272V; I310L; F313M into the pMSCV vector using restriction enzymes AvrII and SalI. Four copies of the Gal4 RE and the NHT gene was PCR amplified from p17*4Tata NHT and inserted in the pMSCV vector. This vector wasnamed pMSCV GR I268V; A272V; I310L; F313M_ NHT (Figure 5.4).

Testing the triple fusion gene NHT

To test the functional of the Neo gene, pBind NHT was stably transfected in HeLa cells by using 400 µg/ml geneticin in the media. To test the functionality of *Renilla* luciferase, HEK 293T cells were transiently transfected with p17*4 Tata NHT and pCMXGR Q275C, I310M, F313I. The transfected cells were incubated with varying concentration of LG335 and assayed for *Renilla* luciferase activity. To test for chemical complementation in mammalian cells, HeLa cells were stably transfected with the retroviral vector pMSCV GR I268V; A272V; I310L; F313M_ NHT in the presence of 400 µg/ml geneticin and 1 µM LG335. When the variant GR I268V; A272V; I310L; F313M and the NHT gene are integrated in the HeLa cells, the cells are tested for *Renilla* luciferase activity.

Stable Transfection Protocol

Each cell line has a different sensitivity towards geneticin. To determine the optimal concentration of geneticin needed for HeLa cells, different concentrations of geneticin were tested. The lowest concentration of drug that begins to give massive cell death in 3 days and kills all the cells within two weeks was choosen. For HeLa cells the optimal concentration was 400 µg/ml geneticin.

After determining the optimal concentration needed for selection, HeLa cells were grown to 80% confluence in complete media. Cells were transfected with pBind NHT or pMSCV GR I268V; A272V; I310L; F313M_ NHT using Lipofectamine 2000®. After 24 to 48 hours of transfection, the cells were split and cultured in media containing 400 µg/ml geneticin. Cell growth was observed every 2 to 3 days and fresh medium with geneticin was added. After one week, cells start forming colonies and only cells expressing neomycin resistant gene survive.

Literature Cited

1. Arnold, F.H. and G. Georgiou, *Directed enzyme evolution: Screening and selection methods.* Methods in molecular biology. 2003. **230**.

2. Woycechowsky, K.J. and D. Hilvert, *Deciphering enzymes. Genetic selection as a probe of structure and mechanism.* Eur J Biochem, 2004. **271**(9): p. 1630-7.

3. Taylor, S.V., P. Kast, and D. Hilvert, *Investigating and Engineering Enzymes by Genetic Selection.* Angew Chem Int Ed Engl, 2001. **40**(18): p. 3310-3335.

4. Azizi, B., E.I. Chang, and D.F. Doyle, *Chemical complementation: small-molecule-based genetic selection in yeast.* Biochem Biophys Res Commun, 2003. **306**(3): p. 774-80.

5. Tao, H. and V.W. Cornish, *Milestones in directed enzyme evolution.* Curr Opin Chem Biol, 2002. **6**(6): p. 858-64.

6. Lin, H. and V.W. Cornish, *Screening and selection methods for large-scale analysis of protein function.* Angew Chem Int Ed Engl, 2002. **41**(23): p. 4402-25.

7. Fields, S. and O. Song, *A novel genetic system to detect protein-protein interactions.* Nature, 1989. **340**(6230): p. 245-6.

8. Schwimmer, L.J., P. Rohatgi, B. Azizi, K.L. Seley, and D.F. Doyle, *Creation and discovery of ligand-receptor pairs for transcriptional control with small molecules.* Proc Natl Acad Sci U S A, 2004. **101**(41): p. 14707-12.

9. James, P., J. Halladay, and E.A. Craig, *Genomic libraries and a host strain designed for highly efficient two-hybrid selection in yeast.* Genetics, 1996. **144**(4): p. 1425-36.

10. Azizi, B. and D.F. Doyle, *Unpublished data.*

11. Southern, P.J. and P. Berg, *Transformation of mammalian cells to antibiotic resistance with a bacterial gene under control of the SV40 early region promoter.* J Mol Appl Genet, 1982. **1**(4): p. 327-41.

12. Wei, C.M., M. Gibson, P.G. Spear, and E.M. Scolnick, *Construction and isolation of a transmissible retrovirus containing the src gene of Harvey murine sarcoma virus and the thymidine kinase gene of herpes simplex virus type 1.* J Virol, 1981. **39**(3): p. 935-44.

13. Smith, K.O., K.S. Galloway, W.L. Kennell, K.K. Ogilvie, and B.K. Radatus, *A new nucleoside analog, 9-[[2-hydroxy-1-(hydroxymethyl)ethoxyl]methyl]guanine, highly active in vitro against herpes simplex virus types 1 and 2.* Antimicrob Agents Chemother, 1982. **22**(1): p. 55-61.

14. Strathdee, C., M. McLeod, and T. Underhill, *Dominant positive and negative selection using luciferase, green fluorescent protein and beta-galactosidase reporter gene fusions.* Biotechniques, 2000. **28**(2): p. 210-214.

15. Oh, S.-C., S.-Y. Nam, H.-C. Kwon, C.-M. Kim, J.-S. Seo, R.H. Seong, Y.-J. Jang, Y.-H. Chung, and H.-Y. Chung, *Generation of fusion genes carrying drug resistance, green fluorescent protein, and herpes simplex virus thymidine kinase genes in a single cistron* Molecules and c ells, 2001. **11**(2): p. 192-197.

CHAPTER 6

IN VIVO LABELING OF PROTEINS IN MAMMALIAN CELLS

Introduction

Identifying and characterizing protein-protein interactions are a prerequisite to better understand cellular mechanisms and functions. To study the dynamics, movement, and interactions of a protein inside living cells various techniques have been developed. Traditionally this has been done by in vitro labeling of proteins with fluorescent dyes and other molecular probes[1]. These dyes lack selectivity toward a particular protein and hence were attached to the protein of interest by means of antibodies to the protein of interest. Recent advances in genetic engineering have made it possible to track protein movement and interactions by fusing green fluorescent protein (GFP) and its variants to the protein of interest[2, 3]. However, there are major drawbacks of fluorescent proteins including their large size, ~27 KD, oligomerization which can affect the biological activity of the fused protein, photo bleaching, and low signal to background on the single molecule level.

In the past decade there has been development in methodologies for studying interaction on the single-molecular level[4-7]. These techniques have many advantages over conventional ones: high sensitivity of a single molecule, the ability to show real time dynamics of cellular processes, and the ability to carry out assays with low quantities and concentrations which correspond to the natural cellular level[7-11]. Presently, there are two main approaches; atomic force microscopy (AFM) and fluorescence techniques.

AFM characterizes the forces involved in protein interaction using force spectroscopy[4, 12]. Among the fluorescence techniques fluorescence correlation spectroscopy (FCS) is a powerful method for studying interaction dynamics of proteins in solution but *in vivo* its success is limited. A more robust approach is fluorescence cross-correlation of proteins labeled with two fluorophores. This technique has been used to monitor interactions within living cells [13]. Another approach for imaging interactions at interfaces on the single molecule level is total internal reflection fluorescence[5]. This technique has been applied to visualize signal transduction on the molecular level[14], and ligands binding to their receptors [9]. A major limitation of using the fluorescence-based single molecule imaging, mentioned above, is the fast photo bleaching of organic fluorophores which limits the observation time to a few seconds. A new technique has emerged which uses fluorescent quantum dots technology[7, 15]. Quantum dots are inorganic nanometer sized nanocrystals that contain CdSe or a CdTe core and ZnS shell. They fluoresce at sharp and discrete wavelengths, they have high extinction coefficients (10 to 100 times those of small fluorophores and FPs), and have good quantum yields. When coated, these quantum dots become water soluble and can be conjugated to protein targeting molecules such as antibodies[6, 7, 15].

Another way to create ultra small molecular biolabels is via non-toxic hydrophilic dendrimer-encapsulated noble metal (DNM) nanocluster. Gold and silver nanodots have single molecule raman emission and show extremely strong and size-dependent single molecule fluorescence[8]. Here a linkage method between the DNM and the protein of interest is being developed. DNM will be attached to protein of interest via an amide bond between a cysteine residue (Cys) in the protein and the dendrimer. To test the

linkage method in mammalian cells, the short term goal is to use fluorescein instead of the DNM. Fluorescein will be linked with the protein of interest to observe the movement of the protein using Fluorescence Resonance Energy Transfer (FRET).

The target protein is the nuclear receptor, retinoid X receptor (RXR) fused to cerulean fluorescence protein (CFP). RXR predominantly localizes in the nucleus of the cell and cerulean is expressed throughout the cell. The fusion protein between RXR and CFP should theoretically localize in the nucleus. This protein will be attached to the fluorescein via N-terminal Cys. To generate the N-terminal Cys on the CFP-RXR construct, intein splicing technology in mammalian cells was applied.

An intein is an internal segment of a precursor protein that is excised by a self catalytic mechanism, known as protein splicing, followed by ligation of the flanking protein regions, known as exteins[16]. To generate the N-terminal Cys an intein was fused to the target protein. The C-terminal of the intein will code for an Aspargine (Asn) and a Cys. Once expressed in mammalian cells excision will occur between the Asn and Cys generating an N terminal Cys on the target protein as shown in Figure 6.1[17].

Incubation of the cells with a thioester-containing fluorescein allows it to efficiently penetrate through the cell membrane into the cell. A chemo-selective ligation reaction occurs between the thioester of the fluorescein and the N-terminal Cys of the protein, giving rise to an amide bond between the fluorescein and the protein as shown in the Figure 6.2 and 6.3. After the protein of interest had been labeled, the next step was to test the linkage between the fluorescein and target protein by observing the movement of the CFP-RXR fusion protein using FRET.

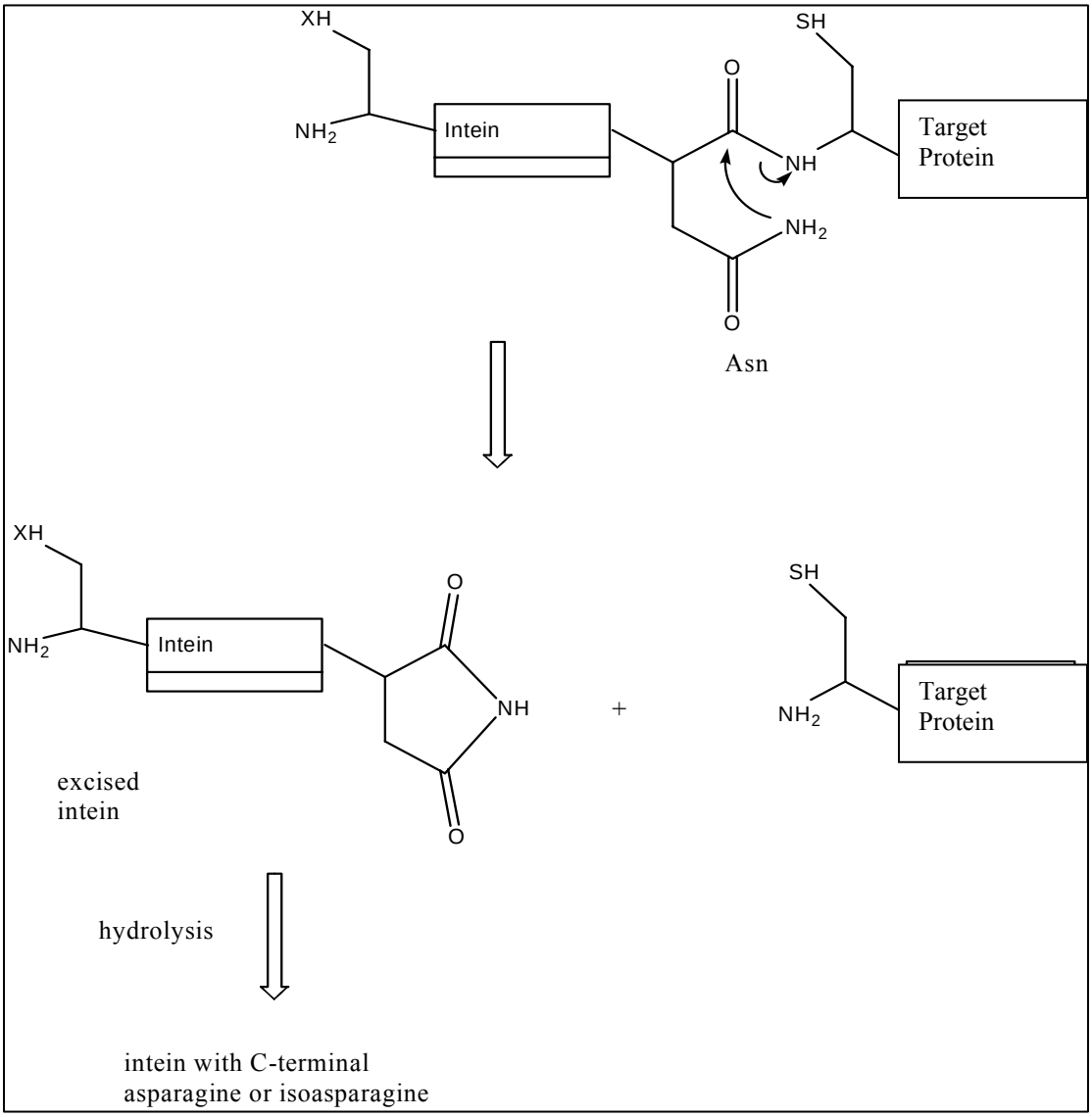


Figure 6.1 Schematic diagram showing the mechanism of intein splicing at the C-terminal junction of the intein with aspargine as the last amino acid of the intein and cysteine as the first residue of the target protein. Self cleavage generates an N-terminal cysteine to the protein of interest.

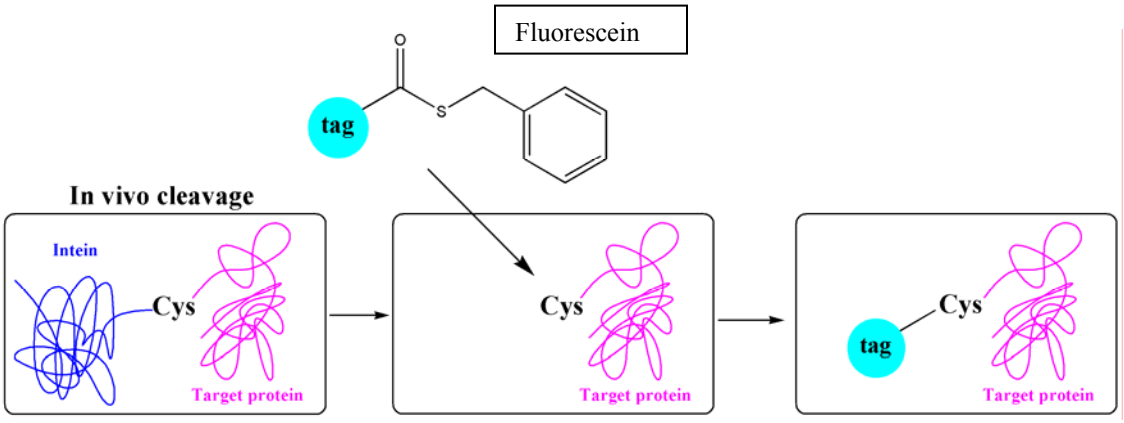


Figure 6.2 Schematic diagram showing the strategy of covalent labeling of the protein with fluorescein (thioester tag).

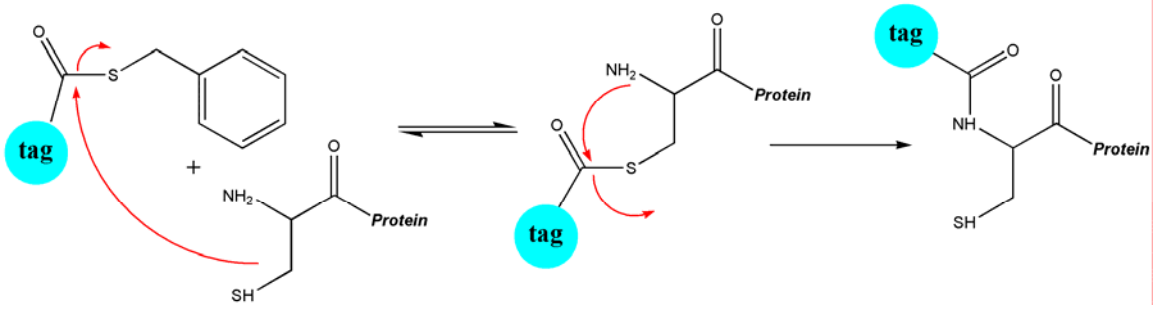


Figure 6.3 Schematic diagram showing the mechanism of chemo-selective ligation between the thioester of the fluorescein and the target protein.

Non-FRET fluorescence occurs when a fluorophore absorbs energy at one wavelength, called excitation frequency and re-emits that energy at a different wavelength, the emission frequency. FRET imaging measures interactions between two proteins. Two different fluorophores are fused to the proteins of interest and each fluorophore has a two-peaked spectrum. The first peak is the excitation peak, and the second is the emission. The emission peak of the donor should overlap with the excitation peak of the acceptor (Figure 6.4). The advantage of FRET technology is that it has excellent resolution and its efficiency is dependent on the inverse sixth power of the distance between molecules[18]. Thus, FRET only occurs when the two fluorophores, in our case cerulean and fluorescein, are within 20-100Å of each other, which means that the fluorophores must be brought together via very close protein-protein interactions. If the fluorophores are over 200Å apart, no signal will be observed.

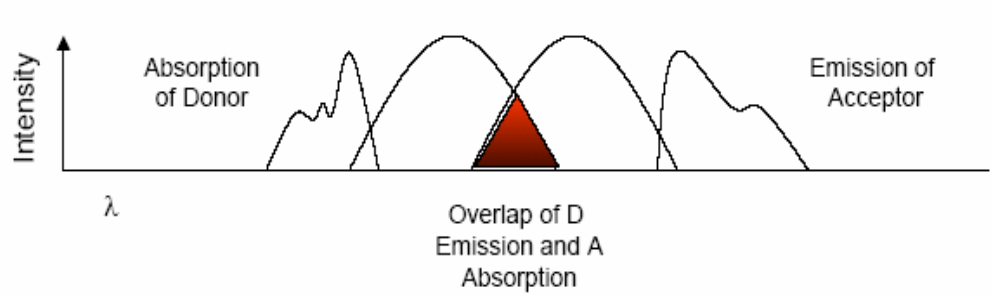


Figure 6.4 Pictorial diagram illustrating the absorption and emission spectrum in FRET www.depts.washington.edu/chemcrs

CFP has an excitation wavelength of 430 nM and an emission wavelength of 480 nm (http://www.olympusfluoview.com/applications/fpcolorpalette.html). Fluorescein has excitation wavelength of 480 nm and emission wavelength is of 520 nm. If fluorescein is indeed linked to the CFP-RXR fusion protein via the N-terminal cysteine, it will then be excited at the emission wavelength of CFP, proving that a probe of DNM can be linked to a target protein via this method. Using this technique of DNM biolabelling we will be able to observe the movement of RXR or other receptor such as progesterone receptor from the cytoplasm to the nucleus.

Results and Discussion

Here a linkage method between the DNM and the protein of interest is being developed for *in vivo* imaging. As a proof of principle, the linkage method is tested using fluorescein instead of DNM and the N-terminal Cys CFP-RXR fusion protein. If fluorescein is attached to the CFP-RXR fusion protein via the amide bond, allowing them to be in close proximity, then FRET will be observed between them. Ideally, to observe FRET in transfected cells incubated with fluorescein the following should happen 1) At 430 nm, only CFP should get excited, 2) At 480 nm, only CFP should emit and fluorescein should be excited 3) At 520 nM, only fluorescein should emit fluorescence.

Inherently NIH3T3 and HEK293T cells do not have background fluorescence. Cells transfected with the control plasmid containing only CFP gene (mCerulean from Dr Piston lab), emit fluorescence through out the cell when excited with the mercury lamp (Figure 6.5 A). To create the N-terminal Cys via intein splicing an intein, CFP and RXR

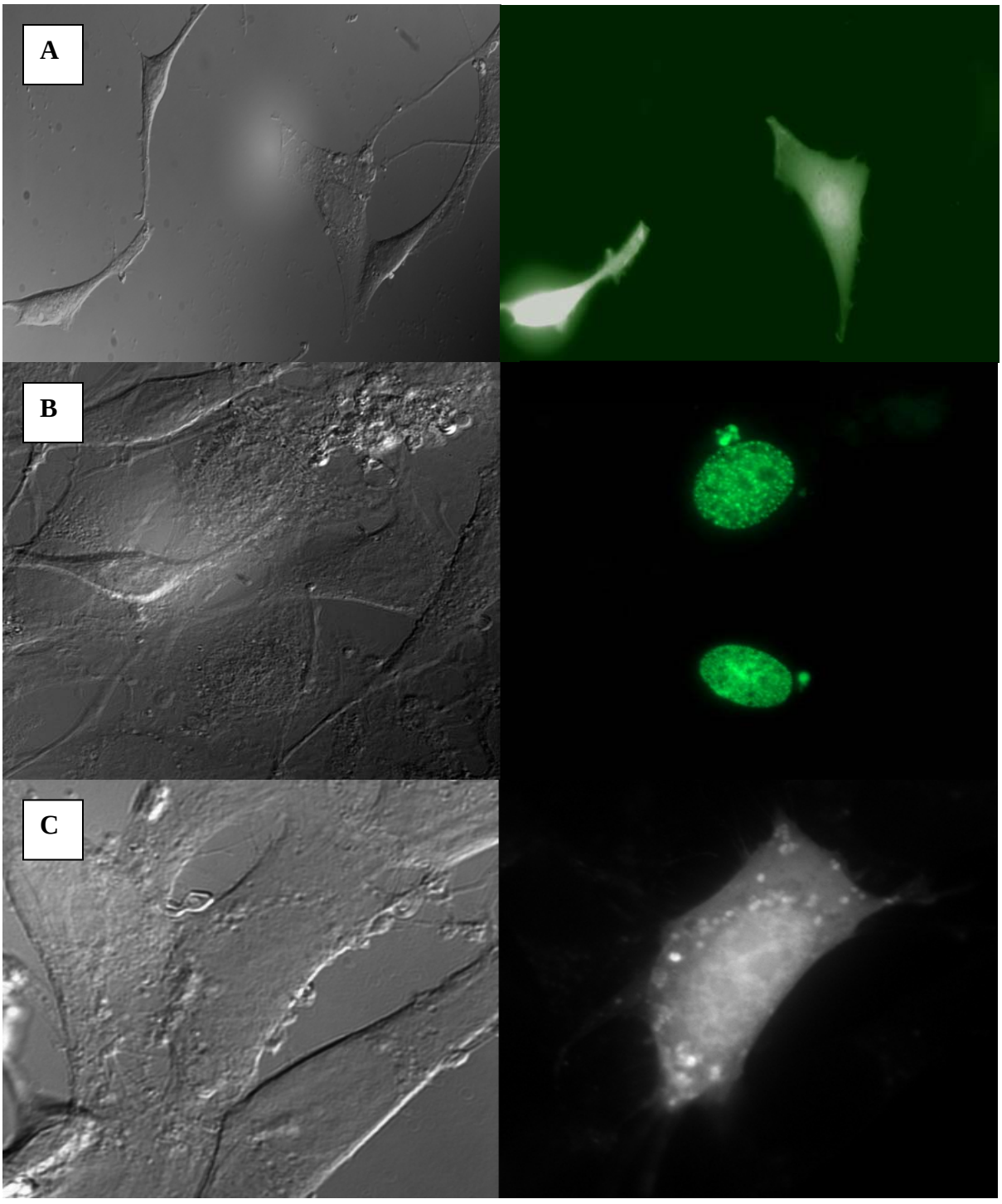
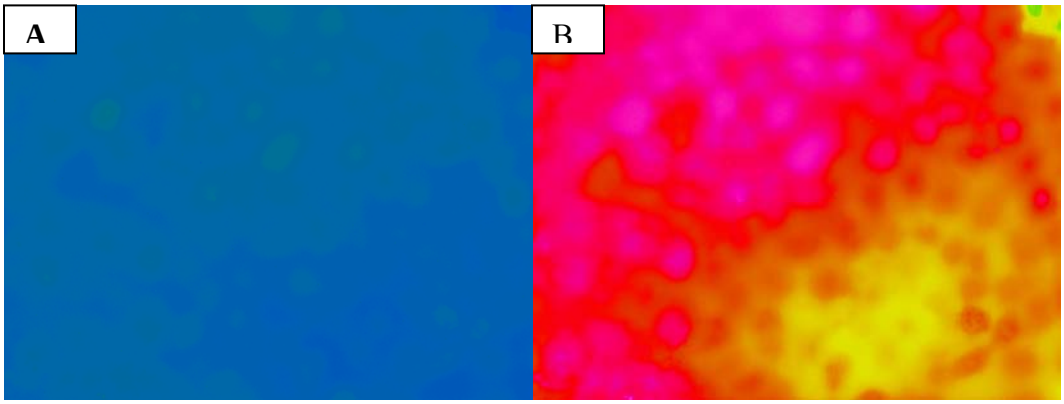


Figure 6.5 Fluorescence images of NIH3T3 cells
A. Transfected with mCerulean (control) showing fluorescence throughout the cell.
B. RXR-Cerulean fusion gene showing fluorescence in the nucleus of the cell.
C. NIH3T3 cells incubated with fluorescein showing fluorescence through out the cell.

fusion gene were inserted into a mammalian expression plasmid called pBind ICR (Stratagene). To test the functionality of CFP in the fusion gene, NIH 3T3 cells were transfected with pBind ICR. When excited with the mercury lamp, CFP fluorescence was observed only in the nucleus of the cells (Figure 6.5 B). This result indicates that the CFP part of the fusion gene is functional and also RXR is expressing in the fusion gene as CFP-RXR localizes in the nucleus (Figure 6.5 B). The transfection efficiency of the pBind ICR in NIH3T3 cells was 10%. To overcome the low transfection efficiency, different transfection reagents such as Polyfect®, Genejuice®, Perfectin®, Lipofectamine®, and Lipofectamine 2000® were tested. None of the transfection reagents increased the transfection efficiency of the fusion gene CFP-RXR in NIH3T3 cells, indicating its poor expression. HEK293T cells were also tested with all the above transfection reagents and the best result was obtained by using lipofectamine 2000 with 40% transfection efficiency.

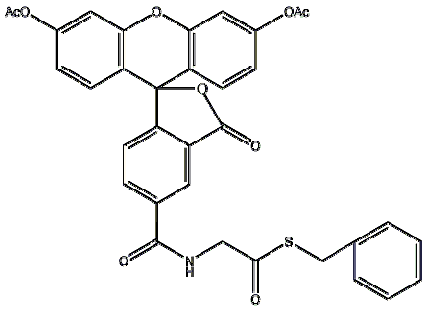
The next step was to test the emission fluorescence of cerulean and fluorescein at specific wavelengths. Non transfected HEK293T cells incubated with fluorescein were excited at two wavelengths 430 nm and 480 nm by using the excitation monochromator. At 430 nm, fluorescein is not excited and no fluorescence was observed between 514 nm to 540 nm band pass as expected (Figure 6.6). Also, at 480 nm fluorescein emits bright fluorescence between 514 nm to 540 nm band pass (Figure 6.6). The emission spectrum of fluorescein was taken using emission monochromator and the spectra looks similar to the theoretical spectra with peak at 514 nm (Figure 6.7).

To test the emission fluorescence of CFP, cells were transfected with pBind ICR and were excited at 430 nm by using an excitation monochromator. Cerulean



No fluorescence

Maximum fluorescence



Fluorescein

Figure 6.6 Fluorescence from fluorescein when excited at wavelength A) 430 nm and B) 480 nm.

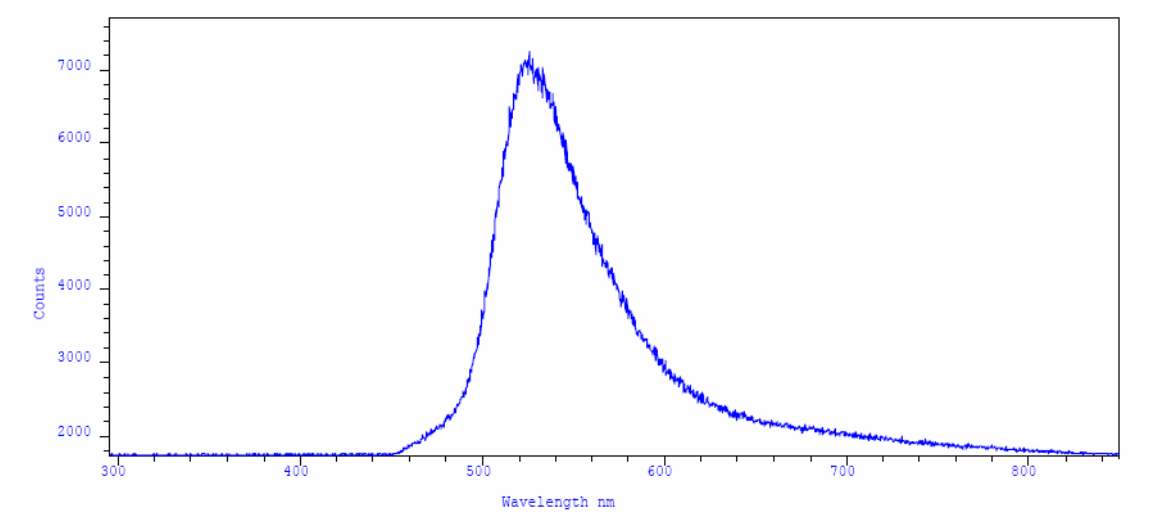


Figure 6.7 Emission spectrum of fluorescein.

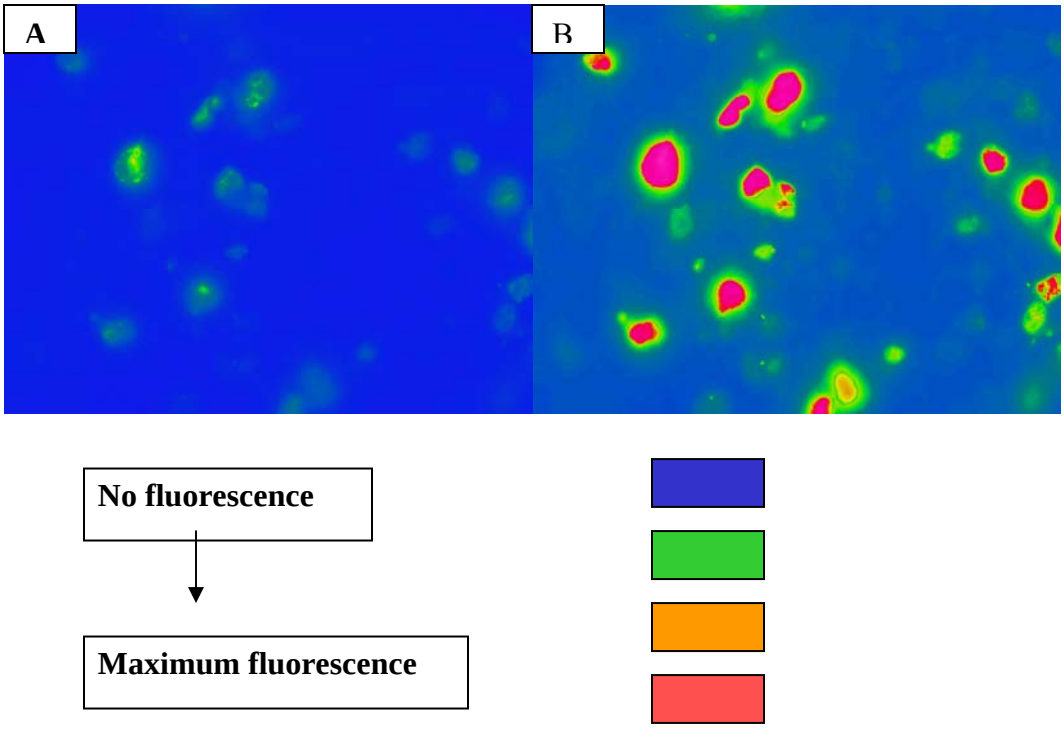


Figure 6.8 Fluorescence from cerulean when excited at 430 nm and emission at A) 480 nm and B) 514 nm.

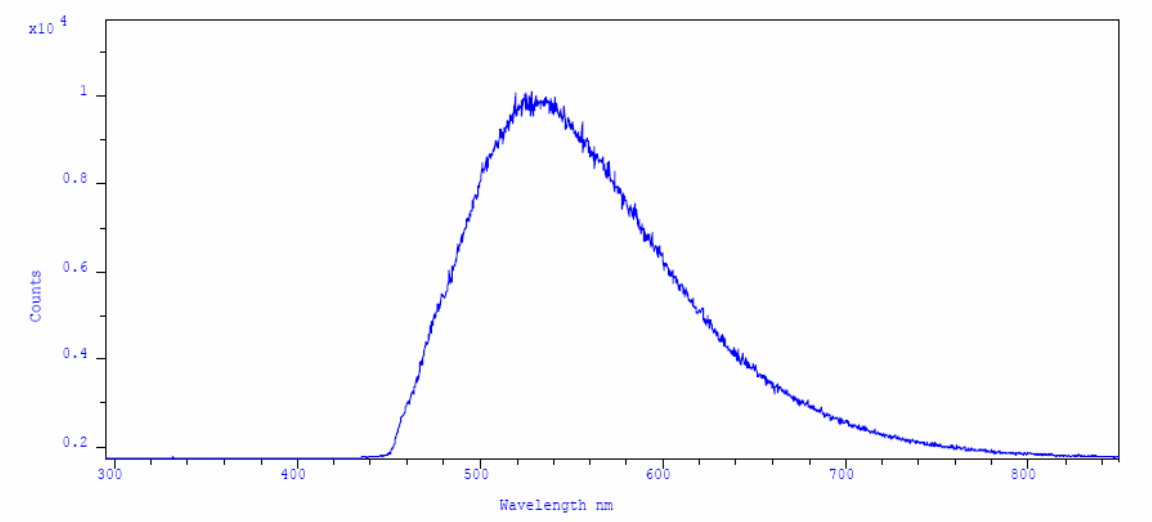


Figure 6.9 Emission spectrum of cerulean fluorescence protein (CFP).

fluorescence images were observed using band filters with wavelength 480 nm and 514 nm. Fluorescence from CFP at 514 nm was brighter then at 480 nm (Figure 6.8) which is opposite to what was expected. Theoretically cerulean emission spectra should peak at 480 nm and fade around 500 nm (http://www.olympusfluoview.com/applications/fpcolorpalette.html). Using an emission monochromator the emission spectrum of cerulean was taken and an emission shift was observed with peak at 520nm (Figure 6.9). To observe FRET the emission spectra of the two fluorophores should not overlap. Emission shift of CFP from 480 nm to 520 nm needs to be investigated and rectified such that the pair of CFP and fluorescein can be utilized to test the link between them.

Materials and Methods

Construction of expression plasmid

Retinoid X receptor was cloned in pBind plasmid (Stratagene). CFP and Ssp DnaB intein were PCR amplified from mCerulean (Piston Lab) and pTWIN1 vectors (NEB, USA) respectively, and fused to RXR, resulting in pBind Intein-CFP-RXR (pBind ICR). The two genes were inserted in frame after the C-terminal of Ssp DnaB intein, with the first amino acid of the CFP-RXR fusion protein to be cysteine residue. The last amino acid of the intein was aspargine (codon AAC).

In vivo labeling in mammalian cells

Human NIH3T3 cells were grown in Dulbecco's modified Eagle's medium (DMEM, Sigma) supplemented with 10% calf bovine serum (Cellgro, USA) at 37°C with 5% CO_2. Cells were seeded at 10^6 cells per six well plate on a cover slip. After overnight incubation, cells were transiently transfected with pBindICR or mCerulean using Lipofectamine 2000 (Invitrogen, USA). After 36 hours of protein expression, the cells were incubated with fluorescein at 5 μM concentration for five minutes. The cells were washed twice with 1 X PBS for five minutes, fixed with 3.6 % paraformaldehyde and mounted on slides using ProLong® Gold antifade reagent (Invitrogen, USA).

Literature Cited

1. Haugland, R.P., The handbook- A guide to fluorescent probes and labeling technologies. 2005.

2. Griffin, B.A., S.R. Adams, and R.Y. Tsien, *Specific covalent labeling of recombinant protein molecules inside live cells.* Science, 1998. **281**(5374): p. 269-72

.

3. Chalfie, M. and S. Kain, *GFP: Green fluorescence protein strategies and applications.* 1998.

4. Yang, Y., H. Wang, and D.A. Erie, *Quantitative characterization of biomolecular assemblies and interactions using atomic force microscopy.* Methods, 2003. **29**(2): p. 175-87.

5. Mashanov, G.I., D. Tacon, A.E. Knight, M. Peckham, and J.E. Molloy, *Visualizing single molecules inside living cells using total internal reflection fluorescence microscopy.* Methods, 2003. **29**(2): p. 142-52.

6. Michalet, X., F.F. Pinaud, L.A. Bentolila, J.M. Tsay, S. Doose, J.J. Li, G. Sundaresan, A.M. Wu, S.S. Gambhir, and S. Weiss, *Quantum dots for live cells, in vivo imaging, and diagnostics.* Science, 2005. **307**(5709): p. 538-44.

7. Gao, X., L. Yang, J.A. Petros, F.F. Marshall, J.W. Simons, and S. Nie, *In vivo molecular and cellular imaging with quantum dots.* Curr Opin Biotechnol, 2005. **16**(1): p. 63-72.

8. Lee, T.H., J.I. Gonzalez, J. Zheng, and R.M. Dickson, *Single-molecule optoelectronics.* Acc Chem Res, 2005. **38**(7): p. 534-41.

9. Sako, Y., S. Minoghchi, and T. Yanagida, *Single-molecule imaging of EGFR signalling on the surface of living cells.* Nat Cell Biol, 2000. **2**(3): p. 168-72.

10. Murakoshi, H., R. Iino, T. Kobayashi, T. Fujiwara, C. Ohshima, A. Yoshimura, and A. Kusumi, *Single-molecule imaging analysis of Ras activation in living cells.* Proc Natl Acad Sci U S A, 2004. **101**(19): p. 7317-22.

11. Ishijima, A. and T. Yanagida, *Single molecule nanobioscience.* Trends Biochem Sci, 2001. **26**(7): p. 438-44.

12. Nevo, R., V. Brumfeld, M. Elbaum, P. Hinterdorfer, and Z. Reich, *Direct discrimination between models of protein activation by single-molecule force measurements.* Biophys J, 2004. **87**(4): p. 2630-4.

13. Bacia, K., I.V. Majoul, and P. Schwille, *Probing the endocytic pathway in live cells using dual-color fluorescence cross-correlation analysis.* Biophys J, 2002. **83**(2): p. 1184-93.

14. Sako, Y. and T. Yanagida, *Single-molecule visualization in cell biology.* Nat Rev Mol Cell Biol, 2003. **Suppl**: p. SS1-5.

15. Pinaud, F., X. Michalet, L.A. Bentolila, J.M. Tsay, S. Doose, J.J. Li, G. Iyer, and S. Weiss, *Advances in fluorescence imaging with quantum dot bio-probes.* Biomaterials, 2006. **27**(9): p. 1679-87.

16. Perler, F.B. and E. Adam, *Protein splicing and its applications.* Curr Opin Biotechnol, 2000. **11**(4): p. 377-83.

17. Yeo, D.S., R. Srinivasan, G.Y. Chen, and S.Q. Yao, *Expanded utility of the native chemical ligation reaction.* Chemistry, 2004. **10**(19): p. 4664-72.

18. Stryer, L. and R.P. Haugland, *Energy transfer: a spectroscopic ruler.* Proc Natl Acad Sci U S A, 1967. **58**(2): p. 719-26.