ENGINEERING LIGAND-RECEPTOR PAIRS FOR SMALL MOLECULE CONTROL OF TRANSCRIPTION

A Dissertation
Presented to
The Academic Faculty

By

Lauren J. Schwimmer

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

Georgia Institute of Technology

August 2005

ENGINEERING LIGAND-RECEPTOR PAIRS FOR SMALL MOLECULE CONTROL OF TRANSCRIPTION

Approved by:

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

Dr. Katherine L. Seley
Department of Chemistry and Biochemistry
University of Maryland, Baltimore County

Dr. Allen Orville
School of Chemistry and Biochemistry
Georgia Institute of Technology

Dr. Harish Radhakrishna
School of Biology
Georgia Institute of Technology

Dr. Andreas Bommarius
School of Biomedical and Chemical Engineering
School of Chemistry and Biochemistry, Adjunct
Georgia Institute of Technology

Date Approved: June 17th, 2005

To:

Mom, Dad, and Evan

For their constant love, support, and encouragement

ACKNOWLEDGEMENTS

I would like to thank:

Dr. Donald Doyle for his guidance and support as my dissertation advisor. His creative ideas, thoughtful instruction, and helpful insight made the research in this dissertation possible.

Dr. Katherine Seley, Peter O'Daniel, and the Seley laboratory for assistance with the LG335 synthesis

Bahareh Azizi for developing chemical complementation and optimizing the liquid culture quantification method. .

Priyanka Rohatgi for performing the mammalian cell culture experiments, which were a valuable addition to this research.

Terry Watt for help with statistical library analysis and editing of manuscripts and papers.

Kenyetta Johnson and the remaining Doyle lab members for support and critical reading of manuscripts.

The Bommarius lab for use of equipment, support, and critical reading of manuscripts.

Dr. Cameron Sullards and David Bostwick of the Georgia Tech Mass Spectrometry Laboratory for their training, experimental design expertise, and equipment time.

Dr. Ronald Evans for the kind gift of plasmids pCMX-hRXR, pCMX-ACTR, and pCMX-βGal.

Dr. Trent Spencer for his gift of HEK 293 cells.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABBREVIATIONS

3AT	3-amino-1,2,4-triazole
9cRA	9-*cis* retinoic acid
ABC	ATP-binding cassette
ACTR	Human activator for thyroid and retinoid receptors
Ade	Adenine
A	Alanine
ADMET	Absorption, distribution, metabolism, excretion, and toxicology
AmB	Amphotericin B
AUC	Area under the curve
CVFF	Consistent valence forcefield
DBD	DNA binding domain
DTP	Developmental Therapeutics Program
Eff	Efficacy
EMS	Ethyl methane sulfonate
F	Phenylalanine
FEP	Free energy perturbation
GAD	Gal4 activation domain
GBD	Gal4 DNA binding domain
HPLC	High-performance liquid chromatography
I	Isoleucine
L	Leucine
LBD	Ligand binding domain

LC	Liquid chromatography
Leu	Leucine
M	Methionine
MMC	Metropolis Monte-Carlo
MRM	Multiple reaction monitoring
MS	Mass spectrometry
OLRP	orthogonal ligand-receptor pair
ONPG	*o*-nitrophenyl β-D-galactopyranoside
OPBA	γ-oxo-1-pyrenebutyric acid
PBN	Polymyxin B nonapeptide
RE	Response element
RMSD	Root mean squared deviation
RXR	Human retinoid X receptor-α
S	Serine
SA	Simulated annealing
SC	Synthetic complete
T	Threonine
Trp	Tryptophan
V	Valine
WT	wild-type

SUMMARY

Creating receptors for control of transcription with arbitrary small molecules has widespread applications including gene therapy, biosensors, and enzyme engineering. Using the combination of high throughput docking, codon randomization, and chemical complementation, we have created new receptors to control transcription with small molecules. Chemical complementation, a new method of protein engineering, was used to discover retinoid X receptors (RXR) variants that are activated by compounds that do not activate wild-type RXR.

A first library of 32,768 RXR variants was designed for the synthetic retinoid-like compound LG335. The library produced ligand-receptor pairs with LG335 that have a variety of EC_{50}s and efficacies. One engineered variant has essentially the reverse ligand specificity of wild-type RXR and is transcriptionally active at 10-fold lower LG335 concentration than wild-type RXR with 9cRA in yeast. The activity of this variant in mammalian cells correlates with its activity in yeast.

A second library of 262,144 RXR variants was designed for two purposes: (i) to develop a high-throughput chemical complementation method to select variants that have high efficacies and low EC_{50}s; and (ii) to find variants which are activated by small molecules not known to bind RXR variants. Selection conditions were manipulated to find only variants with high efficacies and low EC_{50}s. This library was also selected for variants that activate transcription specifically in response to γ-oxo-1-pyrenebutyric acid (OPBA), which is different from any known RXR ligand. OPBA was chosen as a potential ligand using high-throughput docking with the software program FlexX. Two variants are activated by OPBA with an EC_{50} of 5 μM. This is only ten-fold greater than

the EC_{50} of wild type RXR with its ligand 9cRA (500 nM) in yeast.

An improved method synthesizing LG335 and a method for quantifying intracellular ligand concentrations were developed. Although the LG335 synthetic method has an additional step, the overall yield was improved to 8% from 4% in the original publication. Liquid chromatography and mass spectrometry was used to quantify the intracellular concentration of LG335, which was found to be within four fold of the LG335 concentration in the media.

CHAPTER 1

INTRODUCTION

Orthogonal Ligand-Receptor Pairs

Interactions between small molecules and proteins are ubiquitous in nature. These interactions are involved in biological processes ranging from signal transduction by phosphates to hormones regulating gene expression. Manipulating naturally occurring small molecule-protein interactions can lead to the creation of orthogonal ligand-receptor pairs (OLRP). An orthogonal ligand-receptor pair occurs when a protein is mutated to bind a new, synthetic ligand and no longer binds its natural ligand. Thus, the mutated protein-synthetic ligand pair is orthogonal to the natural protein-natural ligand pair (Figure 1.1).

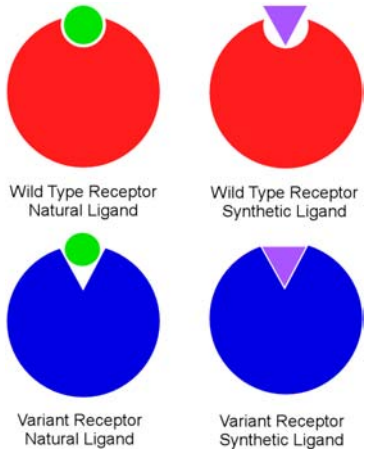


Figure 1.1 Schematic depicting how a naturally occurring ligand-protein interaction is altered to create an OLRP.

Orthogonal ligand-receptor pairs offer the ability to control transcription with small, drug-like molecules. Controlling transcription with exogenous small molecules would provide powerful tools for research and a variety of practical applications. For example, genome projects have revealed tens of thousands of genes with unknown functions. Orthogonal ligand-receptor pairs controlling transcription could be used to study the function of these genes by redirecting the receptor to bind near the promoter of the gene, delivering the gene for the orthogonal receptor to the cell or animal, and activating transcription of the target gene at the desired time by adding the orthogonal ligand. If transcription of a set of genes in a cell line or animal were under the control of multiple ligand-receptor pairs, simply applying the appropriate cocktail of ligands could control multiple genes. Indeed, the levels of expression of each gene could be titrated, not only turned just on or off.

A general method of activating transcription with small molecules would provide useful research tools for creating conditional knockout cell lines and animals. Genetic knockouts can result in embryonic lethality, which does not allow the gene function in a mature organism to be studied. Spatio-temporal control of gene expression through orthogonal ligand-receptor pairs would allow researchers to stop or start the expression of a gene at any point in development. This conditional gene expression would give a better picture of the gene's function and be useful in elucidating signal transduction pathways. For practical applications, orthogonal ligand-receptor pairs controlling transcription have a potentially profound impact for gene therapy. The orthogonal receptor could be directed to activate a gene at its natural locus in the genome that would address the cause of the disease or induce apoptosis to kill the diseased cell. Another possibility for gene

therapy is that the orthogonal ligand-receptor pair can be used to control the expression of another delivered therapeutic gene (e.g. CFTR for cystic fibrosis). When the therapeutic gene is delivered, the gene for the orthogonal receptor would be delivered with it. The receptor would be expressed under a cell-type specific promoter, bind to the synthetic promoter for the therapeutic gene, and express the therapeutic gene to the desired level by the appropriate dose of the orthogonal ligand. This could potentially circumvent a current problem with gene therapy - controlling the amount of expression of the therapeutic protein. Under the described control, if a low efficiency of gene delivery is achieved, then the appropriate amount of protein can be achieved by increasing the dose of the orthogonal ligand (within limits). Alternatively, administering a low ligand concentration would compensate for efficient gene delivery.

Agricultural applications of orthogonal ligand-nuclear receptor pairs include conditional expression of any RNA or protein, including short interfering RNAs to block gene expression, human vaccine proteins, insecticidal proteins, and disease resistance genes. Expressing insecticidal proteins and disease resistance genes only at specific times (rather than constitutively) would likely decrease the occurrence of resistance developing in the environment.

Additionally, OLRPs could be used to activate transcription of a reporter gene and therefore, serve as biosensors for small organic compounds. Arrays of OLRP biosensors that are activated by a variety of small molecules could be used to discriminate between similar compounds based on the pattern of activation (Figure 1.2A). Also, analyte concentration could be determined using arrays of OLRPs that have increasing K_D's for the analyte (Figure 1.2B). Finally, OLRPs could be engineered to bind the desired

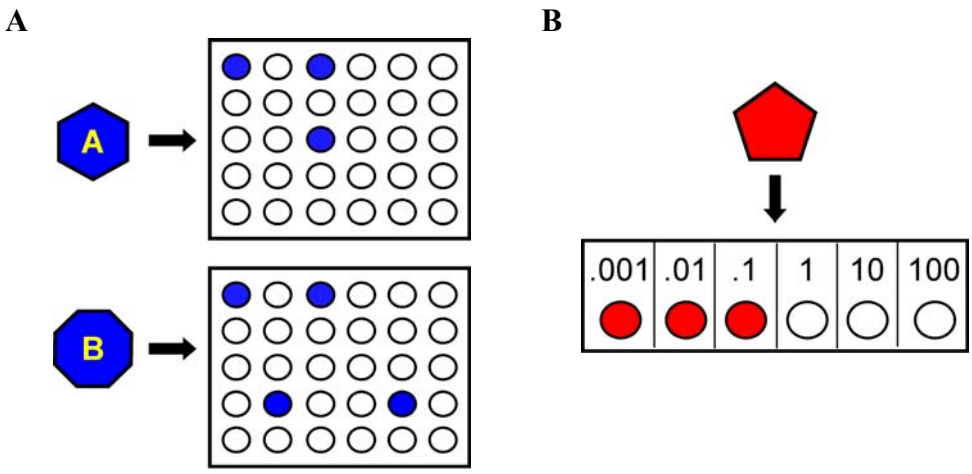


Figure 1.2 OLRP biosensors. (A) Pattern differentiation to identify similar compounds. (B) Analyte concentration determination.

product of an enzymatic reaction. A clone from a library of engineered enzymes that produces the desired product would activate transcription of the reporter gene, thereby identifying the cell that contains the desired enzyme. The gene for this enzyme could then be recovered.

Creation of Orthogonal Ligand-Receptor Pairs

Traditionally, OLRPs are created using site-directed mutagenesis to alter the interactions in the binding pocket. Several strategies have been used including steric complementation, electrostatic interaction reversal, and hydrophobic interaction manipulation [1].

The steric complementation strategy is often referred to as "bumps and holes" [2]. This approach involves mutating a large amino acid in the binding pocket of the receptor to create a "hole", which is complemented by the addition of an additional functional group or "bump" on the ligand. The Schreiber and Shokat groups have used this

approach to engineer cyclophilin / cyclosporin and myosin-Iβ / ATP interactions, respectively [2, 3]. Schreiber and coworkers synthesized a cyclosporin dervative that was able to bind cyclophilin variants with high affinity. These ligand-receptor pairs were created as chemical inducers of dimerization to study protein-protein interactions and signal transduction [2]. Shokat and coworkers synthesized N^6-(2-methylbutyl) ADP, which selectively inhibits the Y61G variant of myosin-Iβ over wild type myosin-Iβ [3]. This ligand-receptor pair was used to determine that myosin-Iβ participates in hair cell adaptation [4]. One major downfall of using bumps and holes is that the modified receptors often retain significant affinity for the wild-type ligand. Although not all applications require that the receptor not bind the wild-type ligand, these receptors cannot be defined as orthogonal.

Researchers have had more success using strategies that alter electrostatic interactions. Hwang and Miller were pioneers in this area. They changed the specificity of *E. coli* elongation factor Tu from guanosine 5'-diphosphate (GDP) to xanthosine 5'-diphosphate (XDP) with a single amino acid change [5]. The mutation of aspartate, which normally forms a salt bridge with the amine at the 2 position of GDP, to asparagine allows hydrogen bonding with the 2-position carbonyl of XDP resulting in the reversal of a hydrogen bond. Because the salt bridge with GDP can no longer occur, orthogonality is introduced.

Charge reversal has also been used to create OLRPs. The Koh group made a serine to aspartic acid mutation in the retinoic acid receptor (RAR), which bound a positively charged retinol derivative and not the negatively charged all-*trans*-retinoic acid (atRA) [6]. However, the transcriptional activation by this variant nuclear receptor was

greatly reduced versus wild type RAR with atRA. Hwang and Warshel suggest that charge reversal will never produce variants that have wild-type-like activity because these interactions are stabilized by more than the amino acids in the immediate vicinity of the salt bridge. Altering only local residues leaves these other interactions intact and does not afford additional stability to the reversed salt bridge [7].

Altering hydrophobic interactions is another strategy for creating OLRPs. Doyle and coworkers used site directed mutagenesis to make a series of single, double, and triple mutations in the binding pocket of the retinoid X receptor (RXR) [8]. These variants were screened against several drug-like compounds. One variant Q275C;I310M;F313I is not significantly activated by 9-*cis* retinoic acid, a ligand for the wild type receptor, but is activated by LG335, a compound that does not activate wild-type RXR and is, therefore, functionally orthogonal.

Site directed mutagenesis and rational design of mutations have generated OLRPs with moderate success. This process, however, is labor intensive with little return on the investment. This dissertation presents a new strategy for creating and discovering OLRPs. Large libraries of RXR variants were created using structure-based codon randomization and chemical complementation, a genetic selection system, was used to find the variants activated by specific small molecules.

Libraries

Libraries of proteins can be constructed to create proteins with altered affinity. In addition to finding new OLRPs, the distribution of amino acids in functional proteins can provide information about the importance of individual residue positions and the

requirement for side chain volume, polarity, and charge. There are three primary methods for creating protein libraries: random mutagenesis, gene shuffling, and structure-based randomization.

Random mutagenesis can be accomplished by both enzymatic and chemical methods. Error prone PCR, an enzymatic method, uses low fidelity DNA polymerases to make random mutations [9]. Varying ion and dNTP concentration, and annealing temperature influences the rate of incorrect base incorporation [10-12]. Chemical methods of random mutagenesis involve incubating the gene of interest with chemicals like hydroxylamine, bisulfate, and ethyl methane sulfonate (EMS) [13, 14]. Chemical methods tend to be less biased than enzymatic methods, which favor AT to GC changes [15-17]. Random mutagenesis methods do not take advantage of any structural or functional information known about the protein. Alternatively, these methods do not require any structural or functional information.

Gene shuffling techniques involve mixing and matching portions of two or more proteins to create libraries. DNA shuffling, developed by Stemmer [18, 19], and its multiple adaptations and improvements [20-24] rely on the same basic steps. First, two or more genes are cut into small pieces. These pieces are then randomly reassembled and PCR is used to amplify the pieces that are the correct length. The pieces are then cloned into an expression vector, expressed, and assayed for function. The set of shuffled genes can range from the same gene with varying mutations, to homologous genes with different functions, to completely unrelated genes. The concept behind directed evolution is to use “Nature's” solutions to find solutions for new problems in a short amount of time. Although the structure and function of the parent protein is considered

when selecting genes to shuffle, this approach does not take advantage of specific information to create proteins with a specific function.

Structure-based randomization, however, does consider structure and function in the design of the library. Using this method, specific amino acids can be randomized to either a small subset of amino acids or to all 20 amino acids [25]. Depending on the experiment, some amino acids may be eliminated due to charge, polarity, hydrophobicity, or side chain volume. Residue positions can be chosen based on location in the protein or potential to alter function and binding affinity. Lim and Sauer used this method to investigate the packing of the hydrophobic core of the λ repressor [26, 27]. After selection of functional variants, the composition of amino acids can be analyzed to determine the amino acid property requirements at each randomized position. This method could also be used to extend alanine scanning and obtain information about the importance of side chain size, polarity, and charge at individual amino acid positions.

Retinoid X Receptor

The human retinoid X receptor α (RXR) is a ligand-activated transcription factor of the nuclear receptor superfamily. RXR plays an important role in morphogenesis and differentiation and serves as a dimerization partner for other nuclear receptors [28]. Like most nuclear receptors, RXR has two structural domains, the DNA-binding domain (DBD) and the ligand-binding domain (LBD), which are connected by a flexible hinge region. The DBD contains two zinc modules, which, when two RXRs homodimerize, bind a direct repeat of six bases with a one base spacer (TAGGTCA A AGGTCA GT AGGTCA A AGGTCA G (CRBPII)) [29, 30]. The LBD binds and activates

transcription in response to multiple ligands including phytanic acid, docosahexanoic acid, and 9-*cis* retinoic acid (9cRA) [31-33] . RXR is a modular protein; the DBD and LBD function independently. Therefore, the LBD can be fused to other DBDs and retain function. A conformational change is induced in the LBD upon ligand binding initiating recruitment of coactivators, which have histone acetyltransferase (HAT) activity [34], and the basal transcription machinery resulting in transcription of the target gene [35].

Nuclear receptors have evolved to bind and activate transcription in response to a variety of small-molecule ligands. The known ligands for nuclear receptors are chemically diverse and include steroid and thyroid hormones, vitamin D, prostaglandins, fatty acids, leukotrienes, retinoids, antibiotics, and other xenobiotics. Evolutionarily closely related receptors (e.g. thyroid hormone receptor and retinoic acid receptor) bind different ligands, whereas some members of distant subfamilies (e.g. RXR and retinoic acid receptor) bind the same ligand [36]. This diversity of ligand-receptor interactions demonstrates the versatility of the fold for ligand binding and suggests that it should be possible to engineer LBDs with a large range of novel selectivities.

Chemical Complementation

Chemical complementation is a three-component genetic selection system in which the survival of *S. cerevisiae* is linked to ligand activation of the nuclear receptor RXR (Figure 1.2A). The PJ69-4A strain has been engineered with an *ADE2* gene under the control of a Gal4 response element (RE) [37]. The Gal4 DNA binding domain (DBD) and the RXR ligand binding domain (LBD) fusion protein binds the Gal4 RE. Another fusion protein comprising ACTR, a nuclear receptor coactivator, and the Gal4

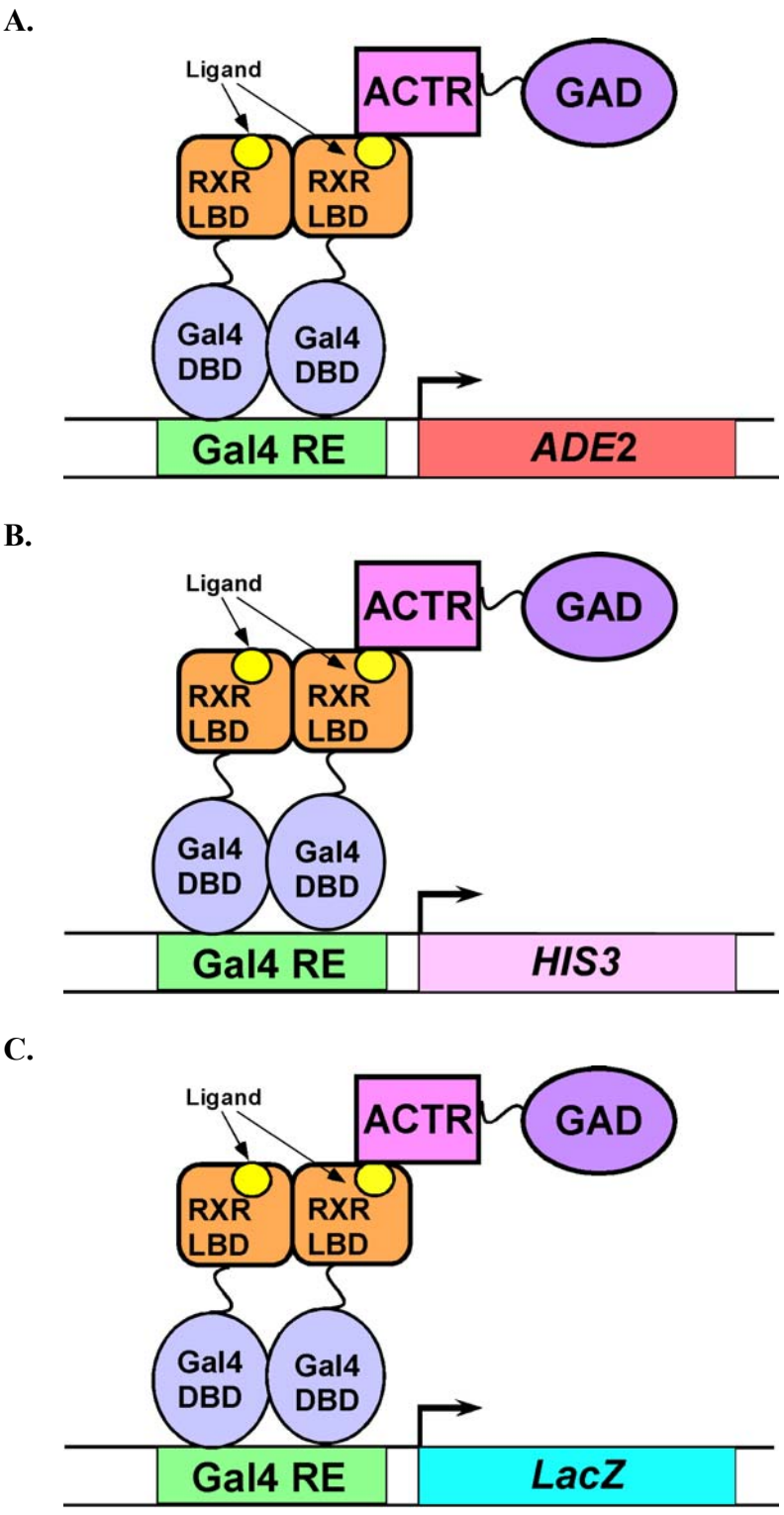


Figure 1.3 Chemical complementation with (A) adenine selection, (B) histidine selection, and (C) β-galactosidase screen. For simplicity only one ACTR:GAD fusion protein is depicted.

activation domain (GAD) is also expressed. Only in the presence of ligand does ACTR bind the RXR LBD. This event allows transcription of the *ADE2* gene, adenine production, and therefore, yeast survival on media lacking adenine [38, 39].

The PJ69-4A strain also has the *HIS3* and *LacZ* genes under the control of Gal4 response elements (Figure 1.3B and C). The *HIS3* gene encodes the enzyme imidazole glycerolphosphate dehydratase, which is inhibited by 3-amino-1,2,4-triazole. This inhibition allows for a variable stringency chemical complementation. The *LacZ* gene encodes β-galactosidase, which hydrolyzes *o*-nitrophenyl-galctopyranoside (ONPG) into galactose and *o*-nitrophenol, a yellow compound. This color change affords an assay to quantify transcriptional activation more directly than the growth assays with the *ADE2* and *HIS3* genes.

Chemical complementation is useful for selecting functional variants from large libraries of receptors. Traditionally, a library is evaluated using a screen in which each member of the library is tested individually for a specific function. This technique requires that each library member be spatially separated before initial evaluation. Because chemical complementation is a genetic selection system, large libraries are evaluated for function in a single experiment before they are spatially separated. The libraries are grown on selective media containing a potential ligand or ligands. The cells that contain a receptor for the ligand are able to survive and form colonies. These colonies can then be spatially separated and further evaluated using additional selection and/or screen assays. Typically, 0.1% or less of the total library survives the initial round of chemical complementation, therefore larger libraries can be evaluated with a fraction of the time, energy, and cost associated with traditional screening methods.

Docking Calculations

Molecular modeling was used to generate hypotheses concerning ligand-receptor interactions. Two software programs were used: Affinity and FlexX.

Affinity®

Affinity® is a module of the Insight II (Accelrys) software family. This program uses a two step method of Monte Carlo simulation followed by simulated annealing to generate possible binding modes of a ligand in a protein's binding pocket. Affinity was chosen to complete *in silico* binding calculations of ligand-receptor pairs for two reasons: (1) it allows flexibility of both the ligand and the amino acids in the binding pocket, and (2) it uses a true force field as opposed to a simpler scoring function.

A general method was developed for docking ligands into the binding pocket of the RXRα LBD crystal structure (1FBY) [40]. After obtaining the coordinates of the crystal structure from the Protein Databank, the structure must be prepared for docking. The crystal structure contains two RXR LBDs; the A chain was used exclusively in these studies*. The A chain and its associated water molecules and 9cRA were unmerged from the total structure. Hydrogen atoms were added to all three molecule sets using a pH of 7.4 to determine the charge state of amino acid side chains. InsightII compares the pH of the "solution" with templates of amino acid charge states and associated pK_a's. If the

* The root mean squared deviation (RMSD) between all atoms of the A and B chains is 0.94 Å. The RMSD for the backbone atoms is 0.60 Å. The only large deviation between the backbone atoms occurs in the loop between helices nine and ten. There are also side chain conformational differences on the exterior of the structure and the B chain is missing residues 224 through 228.

amino acid and pH match a template then the charge state and appropriate hydrogens are assigned. If there is no match, the amino acid is assigned a neutral charge. The crystallographic water was oriented through energy minimization using the steepest descents and conjugate gradient algorithms. The protein was finally "soaked" in a 5 Å layer of water so that the docking is performed in explicit solvent. Soaking is accomplished by placing the protein in the center of an equilibrated box of water followed by removing water molecules that either contact atoms in the protein or are further than 5 Å from the protein.

Potentials are assigned to the atoms using the consistent valence force field (CVFF) [41], which is used to calculate atom-atom interactions. This all-atom force field was parameterized for common protein and small molecule functional groups and was shown to reproduce the crystal structure of dihydrofolate reductase with trimethoprim bound better than a united-atom force field [41]. Atom types are assigned based on hybridization and functional group or amino acid participation. Partial and formal charges are assigned based on CVFF's table of atom types and associated charges.

The functional form of CVFF is:

$$\begin{aligned} E_{pot} = & \sum_b D_b[1-e^{-a(b-b_0)}] + \sum_\theta H_\theta(\theta-\theta_0)^2 + \sum_\phi H_\phi[1+s\cos(n\phi)] \\ & + \sum_\chi H_\chi \chi^2 + \sum_b \sum_{b'} F_{bb'}(b-b_0)(b'-b'_0) + \sum_\theta \sum_{\theta'} F_{\theta\theta'}(\theta-\theta_0)(\theta'-\theta'_0) \\ & + \sum_b \sum_\theta F_{b\theta}(b-b_0)(\theta-\theta_0) + \sum_\phi F_{\phi\theta\theta'}\cos\phi(\theta-\theta_0)(\theta'-\theta'_0) + \sum_\chi \sum_{\chi'} F_{\chi\chi'}\chi\chi' \\ & + \sum \varepsilon[(r^*/r)^{12} - 2(r^*/r)^6] + \sum q_i q_j / \varepsilon r_{ij} \end{aligned} \tag{1.1}$$

The first nine terms describe vibrations of internal coordinates including: (1) bond stretching, (2) angle bending, (3) torsional rotation, (4) out-of-plane bending, and (5-9) combinations of the previous terms. Term ten accounts for van der Waals interactions

and term eleven accounts for electrostatic interactions.

For docking, the ligand is manually placed in the binding site of the protein by superimposing atoms of the ligand to be docked onto atoms of the crystallographic 9cRA. The amino acids within 6 Å of the placed ligand are defined as the "binding site". Therefore, any amino acid with at least on atom within 6 Å of one atom of the ligand is allowed to move. Hydrogen bond donor and acceptor atoms are defined for both the ligand and atoms within 5 Å of the placed ligand. Torsions around which the rotation is allowed are also defined for the ligand.

A Monte Carlo simulation is used to generate possible binding modes of the ligand-protein pair. The placed ligand is first energy minimized using the conjugate gradient method for a defined number of steps. Non-bonding interactions are calculated using a repulsive bounded quartic potential for van der Waals interactions with coulombic interactions set to zero. This function is used for the initial searching because it minimizes the chance that the minimization will terminate when there are many undesirable contacts, while allowing for approximate placement of the ligand in the binding pocket. After initial minimization, the ligand is randomly translated and rotated in the x, y, and z dimensions and the torsions are rotated to generate a new, random ligand placement. The resulting structure is minimized and the energy of the new structure is compared to the previous structure. If the energy of the new structure is lower than the energy of the previous structure, the new structure is accepted. If the energy of the new structure is greater than the energy of the previous structure, it will be accepted if the energy is within a predefined energy range or the Boltzmann factor is greater than a random number between 0 and 1. The Boltzmann factor is calculated using

the following equation:

$$\exp[\frac{-(V_{new}(r^N) - V_{old}(r^N))}{k_B T}] \qquad (1.2)$$

where $V(r^N)$ is the potential energy (V) of each position (r^N) for N particles in the system, k_B is the Boltzmann constant and T is the temperature. Using the Boltzmann factor as the criterion for structure acceptance in a Monte Carlo simulation is termed Metropolis Monte Carlo (MMC) [42]. This procedure permits small uphill moves in which the smaller the uphill move, the greater the probability the move will be accepted. Up to twenty accepted structures are generated and the ten with the lowest energy are kept.

Following the Monte Carlo simulation, simulated annealing (SA) is used to refine each of the best structures. The SA process simulates heating the protein-ligand complex to 500 K and slowly cooling to 300K. At each incremental temperature drop, the system is allowed to reach thermal equilibrium The purpose of simulated annealing is to find the global minimum of the structure by adding potential energy to the system to allow the complex to overcome small energy barriers.

The electrostatic interactions in the SA are calculated using the cell multipole method [43-46]. Using this method, a box containing all the atoms is divided into uniform, cubic cells. The multipole moment is then calculated for each cell as a sum of the moments of all the atoms in the cell. Next, the interactions between each cell and the 26 cells directly surrounding each cell are calculated explicitly using Coulomb's Law. For cells outside the initial 27 cells, the interactions are calculated using a Taylor series multipole expansion. For greater efficiency in calculating the Taylor series coefficients, cells are grouped into larger and larger cells as the distance increase from the initial 27

cells increases. This grouping of cells allows the time required to scale linearly with N, rather than N^2 with the standard Ewald method [47], thereby decreasing the calculation time.

The final step is to calculate the ligand-receptor interaction energies for each accepted and minimized structure. Discover is used to calculate the total energy of the protein-ligand complex, the energies of the protein and the ligand individually, and the interaction energy between the protein and ligand. The protein-ligand interaction energy is separated into the van der Waals and electrostatic contributions. These energies can be used to compare the docked structures. However, the force field does not include desolvation terms. Affinity® also does not allow for any global protein conformational changes. These deficiencies must be considered when comparing energies between different ligand-receptor pairs.

FlexX

FlexX is a software program developed for automatic, high-throughput docking [48]. This program was used to dock libraries of compounds into the RXR LBD binding pocket. FlexX was chosen for this purpose because it is optimized for ligand-protein docking, has high-throughput capability, and allows ligand flexibility. However, FlexX does not allow the protein to be flexible and it uses a scoring function. These were determined to be essential compromises required to decrease docking time.

The A chain of the RXR LBD [40] was prepared using InsightII. The A chain was unmerged from the whole crystal structure and hydrogens were added at a pH of 7.4 with charged C- and N-termini. This structure was saved in .pdb format and used to

create the pocket and surface descriptor files for FlexX using a receptor description file and the associated FlexX commands. The coordinates for each ligand must be in a separate SYBYL mol2 files with three-dimensional coordinates, the correct bond types, and hydrogens added. All of these steps were performed using the InsightII Biopolymer and Builder modules.

FlexX uses a scoring function for estimation of binding free energy of docked ligand-protein complexes. This ΔG value is used to rank the docked solutions. The scoring function is:

$$\begin{aligned}\Delta G = {} & \Delta G_0 + \Delta G_{rot} \times N_{rot} \\ & + \Delta G_{hb} \sum_{\text{neutral H-bonds}} f(\Delta R, \Delta\alpha) \\ & + \Delta G_{io} \sum_{\text{ionic int.}} f(\Delta R, \Delta\alpha) \\ & + \Delta G_{aro} \sum_{\text{aro int.}} f(\Delta R, \Delta\alpha) \\ & + \Delta G_{lipo} \sum_{\text{lipo. cont.}} f^*(\Delta R, \Delta\alpha)\end{aligned} \tag{1.3}$$

and is based on a scoring function developed by Böhm [49]. ΔR and Δα are the deviations of the interactions from the ideal distances and angles. The $f(\Delta R,\Delta\alpha)$ function for hydrogen bonding and ionic interactions was determined by Böhm:

$$\begin{aligned} f(\Delta R, \Delta\alpha) &= f1(\Delta R) f2(\Delta\alpha) \\ f1(\Delta R) &= \begin{cases} 1 & \Delta R \le 0.2 \\ 1 - (\Delta R - 0.2)/0.4 & 0.2 < \Delta R \le 0.6 \\ 0 & \Delta R > 0.6 \end{cases} \\ f2(\Delta\alpha) &= \begin{cases} 1 & \Delta\alpha \le 30° \\ 1 - (\Delta\alpha - 30)/50 & 30° < \Delta\alpha \le 80° \\ 0 & \Delta\alpha > 80° \end{cases} \end{aligned} \tag{1.4}$$

These functions are designed to penalize interactions that are outside the ideal interaction parameters. The FlexX scoring function differs from that defined by Böhm with the

addition of terms for aromatic and lipophilic interactions. The aromatic interaction term uses the same $f(\Delta R,\Delta\alpha)$ as the other interactions, however this function proved to be inadequate for the lipophilic interactions. This term uses the following function:

$$f^*(\Delta R) = \begin{cases} 0 & \Delta R > 0.6 \\ 1 - \dfrac{\Delta R - 0.2}{0.4} & 0.2 < \Delta R \le 0.6 \\ 1 & -0.2 < \Delta R \le 0.2 \\ 1 - \dfrac{-\Delta R - 0.2}{0.4} & -0.6 < \Delta R \le -0.2 \\ \dfrac{\Delta R + 0.6}{0.2} & \Delta R \le -0.6 \end{cases} \qquad (1.5)$$

where ΔR is the difference between the actual atom distances plus 0.3 Å and the ideal atom distance (the sum of the van der Waals radii of both atoms) plus 0.3 Å. Table 1.1 lists default ΔG values.

Table 1.1 FlexX default ΔG values

ΔG	**Value (kJ/mol)**
ΔG_0	5.4
ΔG_{rot}	1.4
ΔG_{hb}	-4.7
ΔG_{io}	-8.3
ΔG_{aro}	-0.7
ΔG_{lipo}	-0.17

FlexX uses an incremental construction algorithm to dock the ligand into the receptor's binding pocket. This method involves three steps: (1) base fragment selection, (2) base fragment placement, and (3) complex construction.

For high-throughput applications base fragment selection is performed automatically [50]. Base fragment selection begins by cutting the ligand at all acyclic rotatable bonds to create components. These components are then combined to create

fragments suitable for placement. A good fragment is defined as one that is both placeable and specific. The authors define placeability of a fragment as having enough interaction sites, but not too many conformations and define specificity of a fragment as having interaction sites that are complementary to the binding pocket. A scoring function is used to choose a set of fragments with the maximum number (up to four) of hydrogen bonds and salt bridges (n_{ric}), the maximum number (up to six) of hydrophobic interactions (n_{ic}), and the minimum number of conformations (n_c):

$$f_{bs}(n_{ric}, n_{ic}, n_c) = 1000 \min(n_{ric}, 4) + 100 \min(n_{ic}, 6) - n_c \qquad (1.6)$$

Fragments must also adhere to the following rules: (1) a base fragment must not be fully contained in another fragment, (2) a component cannot be in more than two base fragments, and (3) a base fragment must be necessary for connectivity or have interaction centers.

The second step in docking is base placement [51]. First, interaction centers and surfaces are defined for the base fragment atoms and the protein atoms in the binding pockets. An interaction center is defined as the center of an atom. An interaction surface shape and distance from the interaction center is defined by the type and hybridization of the atom and its participation in a functional group. Next, three interaction centers from the ligand are superimposed onto three compatible interaction surfaces of the binding pocket and vice versa. Then the root-mean squared (rms) deviation between placements is calculated and fragments are clustered based on a predefined threshold. All fragment placements within the defined rms threshold are merged into one placement. Finally, the fragment placements are ranked using the scoring function.

The final docking step is complex construction [48]. First, a new fragment is

added to all fragment placements in all possible conformations. Next, the new fragment placements are evaluated for overlap with the receptor. A cubic grid is drawn in the three-dimensional space covering the binding pocket and spheres with radii equal to the van der Waals radius of each atom plus the van der Waals radius of the largest atom in the receptor are drawn around the atoms of the ligand. The cubes that intersect the sphere are checked for receptor atoms. If the atoms in the intersecting cubes are too close, the ligand atom will be moved away and if the atoms are too far away, the ligand atoms will be moved closer to optimize the interaction. The optimized fragment placements are then clustered to remove redundant placements. Finally, these new fragment placements are ranked with the scoring function. Complex construction continues by adding another fragment and repeating the optimization until the whole ligand is assembled.

When the docking is complete, FlexX outputs a specified number of placements with their energies and three-dimensional coordinates. It should be noted that although the energies are given in kcal/mole, a scoring function is used to calculate these energies, and therefore this number is not a true free energy of binding.

Literature Cited

1. Koh, J.T., *Engineering selectivity and discrimination into ligand-receptor interfaces.* Chem Biol, 2002. **9**(1): p. 17-23.

2. Belshaw, P.J., J.G. Schoepfer, K.Q. Liu, K.L. Morrison, and S.L. Schreiber, *Rational Design of Orthogonal Receptor-Ligand Combinations.* Angewandte Chemie-International Edition in English, 1995. **34**(19): p. 2129-2132.

3. Gillespie, P.G., S.K.H. Gillespie, J.A. Mercer, K. Shah, and K.M. Shokat, *Engineering of the myosin-I beta nucleotide-binding pocket to create selective sensitivity to N-6-modified ADP analogs.* Journal of Biological Chemistry, 1999. **274**(44): p. 31373-31381.

4. Holt, J.R., S.K. Gillespie, D.W. Provance, K. Shah, K.M. Shokat, D.P. Corey, J.A. Mercer, and P.G. Gillespie, *A chemical-genetic strategy implicates myosin-1c in adaptation by hair cells.* Cell, 2002. **108**(3): p. 371-81.

5. 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.

6. Koh, J.T., M. Putnam, M. Tomic-Canic, and C.M. McDaniel, *Selective regulation of gene expression using rationally-modified retinoic acid receptors.* Journal of the American Chemical Society, 1999. **121**(9): p. 1984-1985.

7. Hwang, J.K. and A. Warshel, *Why ion pair reversal by protein engineering is unlikely to succeed.* Nature, 1988. **334**(6179): p. 270-2.

8. 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.

9. Cadwell, R.C. and G.F. Joyce, *Randomization of genes by PCR mutagenesis.* PCR Methods Appl, 1992. **2**(1): p. 28-33.

10. Zhou, Y.H., X.P. Zhang, and R.H. Ebright, *Random mutagenesis of gene-sized DNA molecules by use of PCR with Taq DNA polymerase.* Nucleic Acids Res, 1991. **19**(21): p. 6052.

11. Vartanian, J.P., M. Henry, and S. Wain-Hobson, *Hypermutagenic PCR involving all four transitions and a sizeable proportion of transversions.* Nucleic Acids Res, 1996. **24**(14): p. 2627-31.

12. Fromant, M., S. Blanquet, and P. Plateau, *Direct random mutagenesis of gene-sized DNA fragments using polymerase chain reaction.* Anal Biochem, 1995. **224**(1): p. 347-53.

13. Lai, Y.P., J. Huang, L.F. Wang, J. Li, and Z.R. Wu, *A new approach to random mutagenesis in vitro.* Biotechnol Bioeng, 2004. **86**(6): p. 622-7.

14. Walton, C.R., K. Booth, and P.G. Stockley, *Random chemical mutagenesis and the non-selective isolation of mutated DNA sequences in vitro*, in *Directed mutagenesis: A practical approach*, M.J. McPherson, Editor. 1991, IRL Press: Oxford. p. 135-162.

15. Cadwell, R.C. and G.F. Joyce, *Mutagenic PCR.* PCR Methods Appl, 1994. **3**(6): p. S136-40.

16. Mutter, G.L. and K.A. Boynton, *PCR bias in amplification of androgen receptor alleles, a trinucleotide repeat marker used in clonality studies.* Nucleic Acids Res, 1995. **23**(8): p. 1411-8.

17. Lueders, T. and M.W. Friedrich, *Evaluation of PCR amplification bias by terminal restriction fragment length polymorphism analysis of small-subunit rRNA and mcrA genes by using defined template mixtures of methanogenic pure cultures and soil DNA extracts.* Appl Environ Microbiol, 2003. **69**(1): p. 320-6.

18. Stemmer, W.P.C., *Rapid Evolution of a Protein in-Vitro by DNA Shuffling.* Nature, 1994. **370**(6488): p. 389-391.

19. Stemmer, W.P.C., *DNA Shuffling by Random Fragmentation and Reassembly - in-Vitro Recombination for Molecular Evolution.* Proceedings of the National Academy of Sciences of the United States of America, 1994. **91**(22): p. 10747-10751.

20. Arnold, F.H., *Design by Directed Evolution.* Accounts of Chemical Research, 1998. **31**(3): p. 125-131.

21. Sieber, V., C.A. Martinez, and F.H. Arnold, *Libraries of hybrid proteins from distantly related sequences.* Nat Biotechnol, 2001. **19**(5): p. 456-60.

22. Lutz, S., M. Ostermeier, G.L. Moore, C.D. Maranas, and S.J. Benkovic, *Creating multiple-crossover DNA libraries independent of sequence identity.* Proc Natl Acad Sci U S A, 2001. **98**(20): p. 11248-53.

23. Lutz, S., M. Ostermeier, and S.J. Benkovic, *Rapid generation of incremental truncation libraries for protein engineering using alpha-phosphothioate nucleotides.* Nucleic Acids Res, 2001. **29**(4): p. E16.

24. O'Maille, P.E., M. Bakhtina, and M.D. Tsai, *Structure-based combinatorial protein engineering (SCOPE).* J Mol Biol, 2002. **321**(4): p. 677-91.

25. Reidhaar-Olson, J.F., J.U. Bowie, R.M. Breyer, J.C. Hu, K.L. Knight, W.A. Lim, M.C. Mossing, D.A. Parsell, K.R. Shoemaker, and R.T. Sauer, *Random mutagenesis of protein sequences using oligonucleotide cassettes.* Methods Enzymol, 1991. **208**: p. 564-86.

26. Lim, W.A. and R.T. Sauer, *The role of internal packing interactions in determining the structure and stability of a protein.* J Mol Biol, 1991. **219**(2): p. 359-76.

27. Lim, W.A. and R.T. Sauer, *Alternative packing arrangements in the hydrophobic core of lambda repressor.* Nature, 1989. **339**(6219): p. 31-6.

28. 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.

29. 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.

30. 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.

31. 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.

32. 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.

33. 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.

34. Xu, J. and Q. Li, *Review of the in vivo functions of the p160 steroid receptor coactivator family.* Mol Endocrinol, 2003. **17**(9): p. 1681-92.

35. Moras, D. and H. Gronemeyer, *The nuclear receptor ligand-binding domain: structure and function.* Current Opinion in Cell Biology, 1998. **10**(3): p. 384-391.

36. Escriva, H., R. Safi, C. Hanni, M. Langlois, P. Saumitou-Laprade, D. Stehelin, A. Capron, R. Peierce, and V. Laudet, *Ligand binding was aquired during evolution of nuclear receptors.* Proc Natl Acad Sci U S A, 1997. **94**: p. 6803-6808.

37. 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.

38. Azizi, B., E.I. Chang, L.J. Schwimmer, and D.F. Doyle, *A sensitive chemical complementation for drug discovery and protein engineering.* Manuscript Submitted.

39. 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.

40. 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.

41. Dauber-Osguthorpe, P., V.A. Roberts, D.J. Osguthorpe, J. Wolff, M. Genest, and A.T. Hagler, *Structure and energetics of ligand binding to proteins: Escherichia coli dihydrofolate reductase-trimethoprim, a drug-receptor system.* Proteins,

1988. **4**(1): p. 31-47.

42. Metropolis, N., A.W. Rosenbluth, M.N. Rosenbluth, A.H. Teller, and E. Teller, *Equation of State Calculations by Fast Computing Machines.* Journal of Chemical Physics, 1953. **21**(6): p. 1087-1092.

43. Greengard, L. and V. Rokhlin, *A Fast Algorithm for Particle Simulations.* Journal of Computational Physics, 1987. **73**(2): p. 325-348.

44. Ding, H.Q., N. Karasawa, and W.A. Goddard, *Atomic Level Simulations on a Million Particles - the Cell Multipole Method for Coulomb and London Nonbond Interactions.* Journal of Chemical Physics, 1992. **97**(6): p. 4309-4315.

45. Ding, H.Q., N. Karasawa, and W.A. Goddard, *The Reduced Cell Multipole Method for Coulomb Interactions in Periodic-Systems with Million-Atom Unit Cells.* Chemical Physics Letters, 1992. **196**(1-2): p. 6-10.

46. Greengard, L., *Fast Algorithms for Classical Physics.* Science, 1994. **265**(5174): p. 909-914.

47. York, D.M., A. Wlodawer, L.G. Pedersen, and T.A. Darden, *Atomic-level accuracy in simulations of large protein crystals.* Proc Natl Acad Sci U S A, 1994. **91**(18): p. 8715-8.

48. Rarey, M., B. Kramer, T. Lengauer, and G. Klebe, *A fast flexible docking method using an incremental construction algorithm.* J Mol Biol, 1996. **261**(3): p. 470-89.

49. Bohm, H.J., *The development of a simple empirical scoring function to estimate the binding constant for a protein-ligand complex of known three-dimensional structure.* J Comput Aided Mol Des, 1994. **8**(3): p. 243-56.

50. Rarey, M., B. Kramer, and T. Lengauer, *Multiple automatic base selection: protein-ligand docking based on incremental construction without manual intervention.* J Comput Aided Mol Des, 1997. **11**(4): p. 369-84.

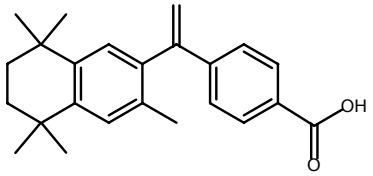
51. Rarey, M., S. Wefing, and T. Lengauer, *Placement of medium-sized molecular fragments into active sites of proteins.* J Comput Aided Mol Des, 1996. **10**(1): p. 41-54.

CHAPTER 2

LG335 SYNTHESIS

Ligand Pharmaceuticals, in the search for RXR selective agonists, synthesized a series of compounds that produced Targretin® (Figure 2.1A). Targretin® is an RXR agonist that is FDA approved to treat cutaneous T-cell lymphoma and AIDS related Kaposi's sarcoma. LG335 (Figure 2.1B) was synthesized as a part of this study and did

A



B

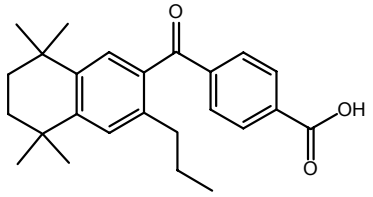
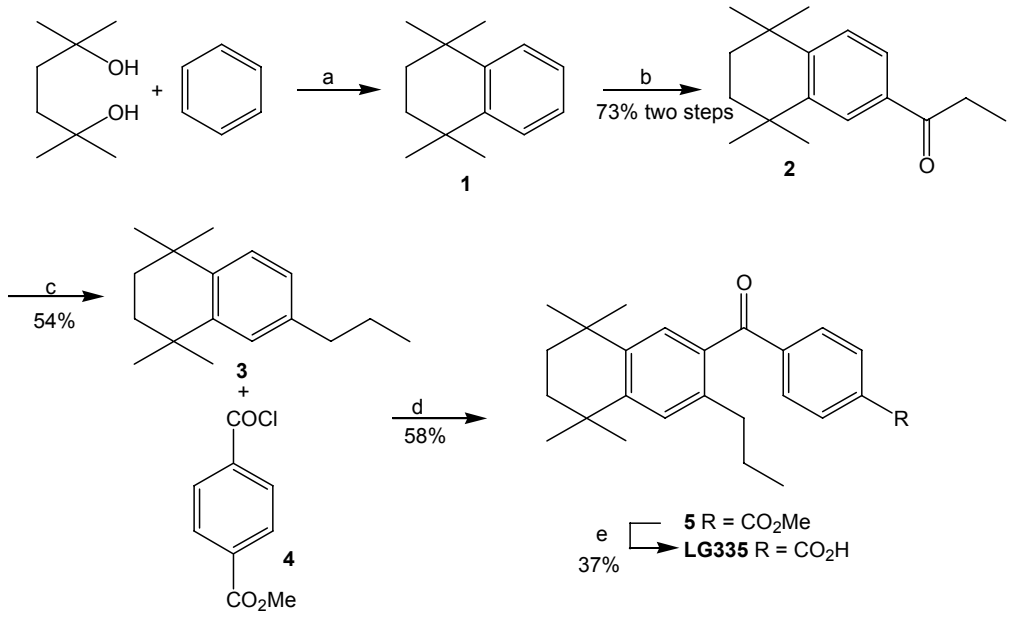


Figure 2.1 A) Targretin and B) LG335

not bind or activate any RXR subtype [1]. Doyle and coworkers found that the RXR variant Q275C;I310M;F313I was activated by LG335, but not by 9cRA [2].

To continue these studies and to find new ligand-receptor pairs, it was necessary to synthesize more LG335. After attempting the published synthesis, the correct product of the second reaction, which has a published yield of 8% [1] could not be isolated. An alternative synthetic scheme was developed (Figure 2.2). First a Friedel-Crafts alkylation



a) $AlCl_3$ b) Propionyl chloride, $AlCl_3$, $C_2H_4Cl_2$ c)Pd/C, MeOH, H_2O, HCl, H_2, 60 psi, heat
d) $FeCl_3$, CH_2Cl_2 e) KOH, MeOH

Figure 2.2 LG335 synthetic scheme

coupled 2,5-dimethylehexanediol and benzene to produce **1** [3]. A Friedel-Crafts acylation with propionyl chloride, produced **2** [4, 5]. The carbonyl was reduced with hydrogenation and the product, **3** [6], was coupled to **4** using Friedel-Crafts acylation. Finally, the ester was hydrolyzed to a carboxylic acid using 5M KOH, producing LG335 [1]. This synthetic route has five steps with an overall yield of 8% [7], which is better than the original published yield of 4% [1]. Optimization could further increase the overall yield of the new synthetic route.

Mass spectrometry, NMR, and elemental analysis were used to verify the identity and purity of LG335. The ^1H NMR ($CDCl_3$) (Figure 2.3) spectrum showed peaks with the following chemical shifts: δ 0.88 (t, 3H, -$CH_2CH_2CH_3$), 1.20 (s, 6H, CH_3), 1.32 (s, 6H, CH_3), 1.55 (dt, 2H, -$CH_2CH_2CH_3$), 1.69 (s, 4H, CH_2), 2.65 (t, 2H, -$CH_2CH_2CH_3$), 7.20 (s, 1H, Ar-CH) 7.23 (s, 1H, Ar-CH), 7.89 (d, 2H, Ar-CH), 8.18 (d, 2H, Ar-CH). The

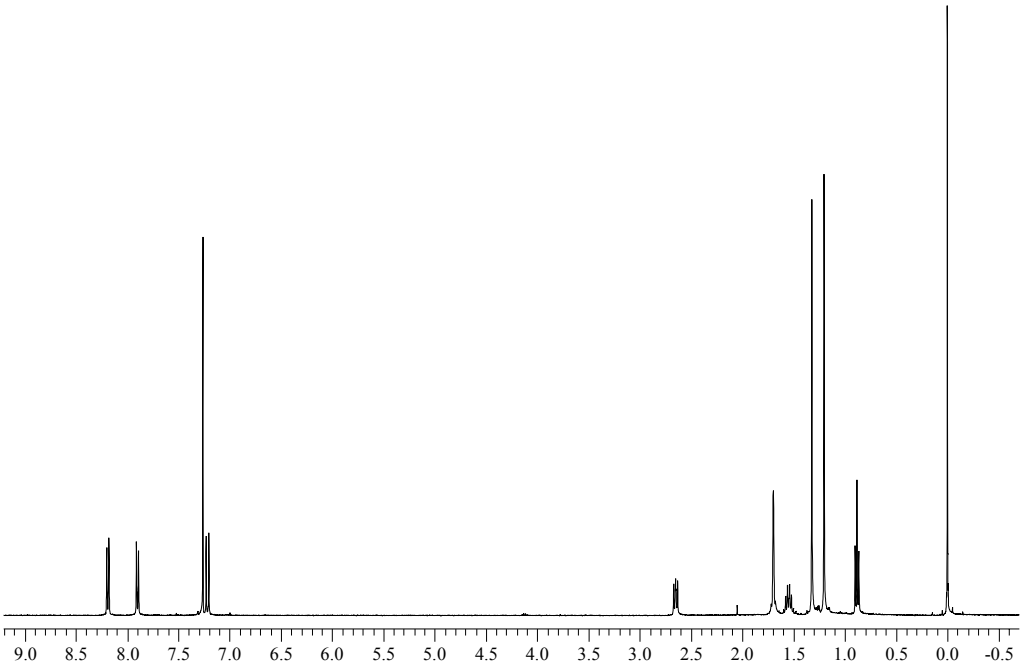


Figure 2.3 NMR spectrum of LG335

electron ionization mass spectrum (Figure 2.4) has major peaks corresponding to the molecular ion and four fragment ions (Table 2.1) consistent with the structure of LG335. High-resolution mass spectrometry found an accurate mass of 378.2195 amu, which differs from the calculated mass of 378.2189 amu by 1.6 ppm. Elemental analysis of LG335 found that the LG335 sample submitted is 79.10% C and 7.96% H, which is consistent with the calculated 79.33% C and 7.99% H.

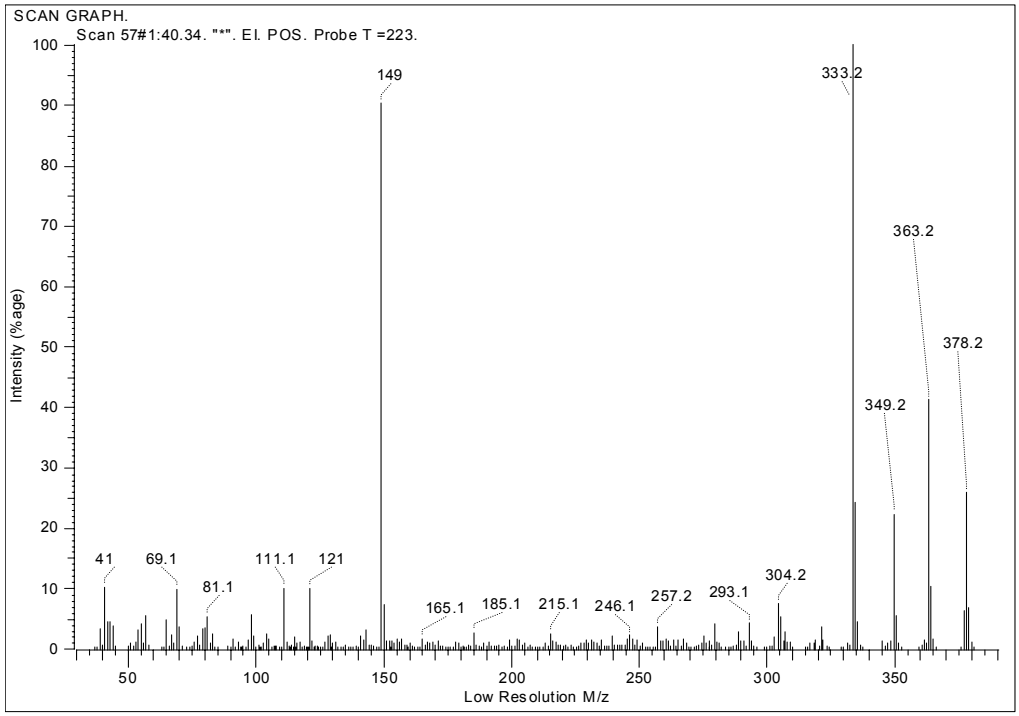
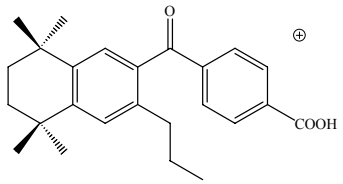
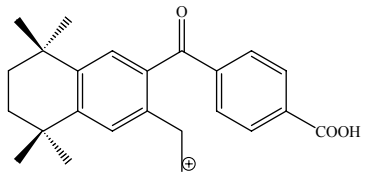
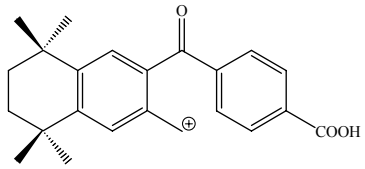
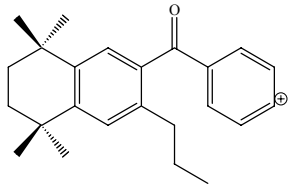
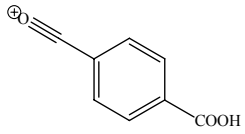


Figure 2.4 Mass spectrum of LG335

Table 2.1 Mass Spectrum Ions

Ion	Molecular Weight
	378
	363
	349
	333
	149

Materials and Methods

3-(1-Carbonyl)propyl-5,5,8,8-tetramethyl-5,6,7,8-tetrahydronapthylene (2)

2,5-dimethyl-2,5,hexanediol (5.0 g, 34 mmol) was dissolved in anhydrous benzene (150 mL). $AlCl_3$ (5.0 g, 38 mmol) was added slowly while the mixture was stirred in an ice bath, followed by stirring at room temperature for 1 hour. Another portion of $AlCl_3$ (5.0 g, 38 mmol) was then added and the reaction was heated to 50 °C and stirred overnight. The brown solution was poured over iced 0.4 M HCl (50 mL) and extracted with ether (3 x 50 mL). The organic layer was then sequentially washed with water, saturated aqueous $NaHCO_3$, and brine (80 mL each) and dried ($MgSO_4$). The solvent was removed in vacuo to afford 6.2 g of a yellow liquid [3].

The crude product was then mixed with propionyl chloride (3.2 mL, 37 mmol) and the resulting solution added dropwise to a mixture of $AlCl_3$ (5.0g, 38 mmol) in dichloroethane (20 mL) while maintaining the temperature between 20 and 25 °C. The mixture was stirred for 2 hours at room temperature, at which point it was quenched by pouring carefully over ice. The reaction mixture was then extracted with methylene chloride (3 x 10 mL). The organic layers were then combined, washed with water and saturated aqueous $NaHCO_3$, and the volatiles removed by rotary evaporation. The product was purified by silica gel column chromatography eluting with hexanes:chloroform (4:1, then 1:1) to yield 6.9 g (28 mmol, 73%) of product as a yellow oil [4, 5].

3-Propyl-5,5,8,8-tetramethyl-5,6,7,8-tetrahydronapthylene (3)

3-(1-Carbonyl)propyl-5,5,8,8-tetramethyl-5,6,7,8-tetrahydronapthylene (1.0 g, 4.1 mmol)

in MeOH (10 mL), H_2O (1 mL), and conc. HCl (3 drops) was treated with 10% Pd/C (144 mg) and subjected to catalytic hydrogenation conditions at 60 psi while heating gently overnight. When the reaction was considered complete (Rf = 0.76, 5% EtOAc in hexanes) it was filtered through a celite pad and rinsed with MeOH (10 mL) and hexane (50 mL). Water (1 mL) was then added to the filtrate and the organic phase separated and washed with brine (2 x 20 mL). The aqueous layer was washed with hexanes (2 x 20 mL). The organic layers were dried (Na_2SO_4), filtered and the volatiles removed by rotary evaporation to produce 510 mg (2.2 mmol, 54%) of a colorless oil [6].

4-[(3-Propyl-5,5,8,8-tetramethyl –5,6,7,8-tetrahydro-2-naphtyl)carbonyl]benzoic Acid (LG335)

3-Propyl-5,5,8,8-tetramethyl-5,6,7,8-tetrahydronapthylene (2.2 g, 9.5 mmol) and chloromethyl terephthalate (2.0g, 10 mmol) were dissolved in dichloroethane (20 mL) and $FeCl_3$ (80 mg, 490 μmol) was added. The reaction mixture was stirred at 75 °C for 24 hours. The reaction was then cooled and MeOH (20 mL) added. The resulting slurry stirred for 7 hours at room temperature, filtered and rinsed with cold MeOH (20 mL) to result in 2.1 g (5.5 mmol, 58%) of white crystals [8].

The crystals (107 mg, 280 μmol) were stirred in MeOH (2 mL), to which 5N KOH (0.5 mL) was added. This mixture was refluxed for 30 minutes, cooled to room temperature and acidified with 20% aqueous HCl (0.5 mL). The MeOH was evaporated and the residue was extracted with EtOAc (2 x 5 mL). The organic layers were combined, dried ($MgSO_4$), and filtered. The filtrate was treated with hexane (10 mL) and reduced in volume to 2 mL. After standing overnight the resulting crystals were

collected to provide 39 mg (103 μmol, 37%) as a white powder [1]. mp 250-252 °C; H^1 NMR ($CDCl_3$) δ 0.88 (t, 3H, -$CH_2CH_2CH_3$), 1.20 (s, 6H, CH_3), 1.32 (s, 6H, CH_3), 1.55 (dt, 2H, -$CH_2CH_2CH_3$), 1.69 (s, 4H, CH_2), 2.65 (t, 2H, -$CH_2CH_2CH_3$), 7.20 (s, 1H, Ar-CH) 7.23 (s, 1H, Ar-CH), 7.89 (d, 2H, Ar-CH), 8.18 (d, 2H, Ar-CH); MS (EI POS) m/z mass for $C_{25}H_{30}O_3$: Calc. 378.2189, Found 378.2195; Anal. for $C_{25}H_{30}O_3$: Calc. C:79.33, H:7.99, Found C:79.10, H:7.96.

Literature Cited

1. 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.

2. 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.

3. Held, P., M.P. Heck, J. Iyer, H. Gronemeyer, L. Lebeau, and C. Mioskowski, *Synthesis of a radiolabelled Retinoid X Receptor (RXR) specific ligand.* Journal of Labelled Compounds & Radiopharmaceuticals, 1997. **39**(6): p. 501-507.

4. Kagechika, H., E. Kawachi, Y. Hashimoto, T. Himi, and K. Shudo, *Retinobenzoic acids. 1. Structure-activity relationships of aromatic amides with retinoidal activity.* J Med Chem, 1988. **31**(11): p. 2182-92.

5. Chandraratna, R.A., *Preparation of 1-cyclopropyl-4-tetrahydronaphthyl-1,3-butadienes having retinoid-like activity*, U.S. Patent 5,344,959, 1994, Allergan, Inc.

6. Alvi, K.A., J. Rodriguez, M.C. Diaz, R. Moretti, R.S. Wilhelm, R.H. Lee, D.L. Slate, and P. Crews, *Protein-Tyrosine Kinase Inhibitory Properties of Planar Polycyclics Obtained from the Marine Sponge Xestospongia Cf Carbonaria and from Total Synthesis.* Journal of Organic Chemistry, 1993. **58**(18): p. 4871-4880.

7. 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.

8. Faul, M.M., A.M. Ratz, K.A. Sullivan, W.G. Trankle, and L.L. Winneroski, *Synthesis of novel retinoid X receptor-selective retinoids.* J Org Chem, 2001. **66**(17): p. 5772-82.

CHAPTER 3

CREATION AND DISCOVERY OF LIGAND-RECEPTOR PAIRS FOR SMALL MOLECULE CONTROL OF TRANSCRIPTION

The crystal structure of RXR bound to 9cRA elucidates important hydrophobic and polar interactions in the LBD binding pocket [2]. There are 20 hydrophobic and polar amino acid side chains within 4.4 Å of the bound 9cRA (Figures 3.1 and 3.2). RXR

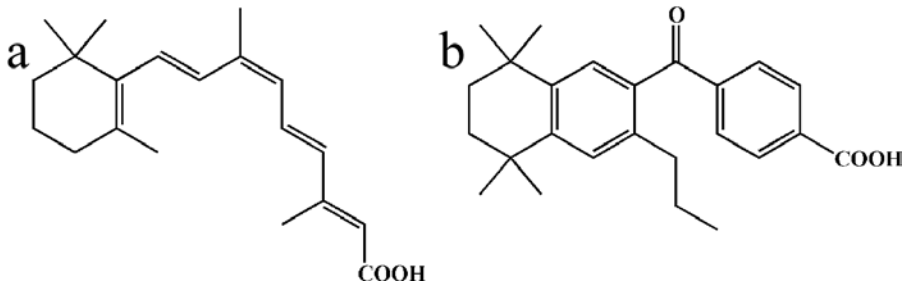


Figure 3.1 (A) 9-*cis* retinoic acid, (B) LG335.

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 [2-5]. A library of RXR LBDs with all 20 amino acids at each of the 20 positions in the ligand-binding pocket screened against multiple compounds could potentially produce many new ligand-receptor pairs. However, the number of possible combinations ($20^{20} \approx 10^{26}$) renders saturation mutagenesis impractical for constructing a complete library.

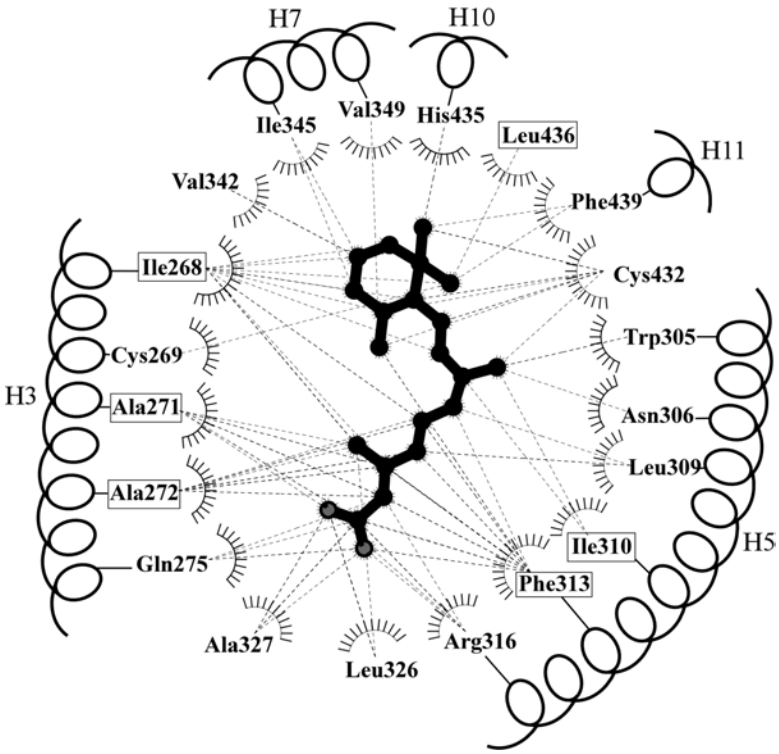


Figure 3.2 Ligplot [1] depiction of hydrophobic interactions between the RXR LBD and 9cRA. Amino acids that are in helices are attached to corresponding coils. The other amino acids are in loop regions. Amino acids mutated in this library are boxed. Coordinates are from 1FBY.pdb [2].

Codon randomization creates protein libraries with mutations at specific sites [6]. For a hydrophobic binding pocket, it is justifiable to eliminate many amino acids from the list of desirable mutations due to size, charge, or hydrophilicity. The bases comprising each codon can then be adjusted to allow only desirable mutations. For example, Sauer and coworkers [7] allowed only T, A, and G at the first position, T at the second position, and C and G at the third position of a codon. This produces 6 possible codons and 5 possible amino acids (L, I, V, F, and M), leading to a manageable library size. A modified version of the Sauer codon randomization method was used to create a library of binding pocket variants in RXR. This library allowed exploration of a large and focused quantity of sequence space in a minimal amount of time.

Chemical complementation is a method which links survival of yeast to the presence of a small molecule [8, 9]. Chemical complementation allows testing for activation of protein variants by specific ligands using genetic selection. Here, LG335, a synthetic retinoid-like compound (Figure 3.1), was used to test the viability of chemical complementation for discovery of ligand-receptor pairs from large libraries. LG335 was previously shown to selectively activate an RXR variant and not activate wild-type RXR [4, 10]. Combining chemical complementation with a large library of protein variants decreases the time, effort, and resources necessary to find new ligand-receptor pairs. Detailed here is my first attempt at large-scale chemical complementation to engineer ligand-receptor pairs.

Results

Library Design

The binding pocket of the RXR LBD is composed of primarily hydrophobic side chains plus several positively charged residues that stabilize the negatively charged carboxylate group of 9cRA [2]. The target ligand, LG335, contains an analogous carboxylate group, so the positively charged residues were left unchanged. The library was designed based on the hypothesis that binding affinity arises from hydrophobic contacts and that specificity arises from binding pocket size, shape, hydrogen bonding, and electrostatics. The randomized amino acids were chosen based on their proximity to the bound 9cRA as observed in the crystal structure and the results of site directed mutagenesis (Figure 3.2) [2, 4]. The electrostatic interactions were held constant while the size, shape, and potential hydrogen bonding interactions were varied to find optimum contacts for LG335 binding. A library of RXRs with mutations at six positions was created. At three of the positions (I268, A271, and A272) are four possible amino acids (L, V, A, and P) and at the other three positions (I310, F313, and L436) there are eight possible amino acids (L, I, V, F, M, S, A, and T). The combination of six positions and number of encoded amino acids allowed testing of the library construction while keeping the library size (32,768 amino acid combinations and 2,985,984 codon combinations) within reasonable limits. Proline was included in the library as a negative control. Residues 268, 271, and 272 are in the middle of helix 3, which would be disrupted by the inclusion of proline. Therefore, proline residues should appear at these positions only in unselected variants and not in the variants that activate in response to ligand. The substitutions at positions 268, 271, and 272 were restricted to small amino acids allowing

access to the positively charged residues at this end of the pocket.

To eliminate contamination of the library with unmutated, wild-type RXR the gene was modified to create a non-functional gene, RXR:3Stop. Forty base pairs were deleted at three separate sites producing three stop codons in the coding region to create this nonfunctional gene. The deletions correspond to regions in the RXR gene where randomized codons are designed. This plasmid, pGBDRXR:3Stop, was cotransformed into yeast with the library of insert cassettes containing full-length RXR LBD genes with randomized codons at positions 268, 271, 272, 310, 313, and 436. The insert cassettes and the plasmid contain homologous regions enabling the yeast to homologously recombine the cassette into the plasmid. Recombination repairs the deletions in the RXR:3Stop gene to make full-length genes with mutations at the six specific sites.

Library selection

To limit the number of variants to be screened, the library was subjected to chemical complementation [9] . Chemical complementation exploits the power of genetic selection to make the survival of yeast dependent on the presence of a small molecule. Plasmids created through homologous recombination in PJ69-4A express the Gal4 DBD fused with a variant RXR LBD (GBD:RXR). A plasmid expressing ACTR, a nuclear receptor coactivator, fused with the Gal4 activation domain (ACTR:GAD), was also transformed into PJ69-4A. If a ligand causes a variant RXR LBD to associate with ACTR, transcription of the *ADE2* gene is activated. Expression of *ADE2* permits adenine biosynthesis and therefore, yeast survival on media lacking adenine.

A small amount of the yeast library was plated onto media (SC -Leu -Trp)

selecting only for the presence of the plasmids pGAD10BAACTR (expressing ACTR:GAD and containing a leucine selective marker) and mutant pGBDRXR (expressing variant GBD:RXR and containing a tryptophan selective marker). The majority of the yeast cells transformed with the RXR library were plated directly onto SC -Ade -Trp -Leu media containing 10 μM LG335, selecting for adenine production in response to the compound LG335. The transformation efficiency of this library into yeast strain PJ69-4A was 3.8×10^4 colonies per μg DNA. This number includes both the efficiency of transforming the DNA into the cells and the homologous recombination efficiency. Of the approximately 380,000 transformants, approximately 300 grew on SC -Ade -Trp -Leu + 10 μM LG335 selective media.

Library Characterization

Twenty-one plasmids were rescued from yeast colonies: nine from non-selective plates (SC -Trp -Leu) and twelve from selective plates (SC -Ade -Trp -Leu + 10 μM LG335). The relevant portion of plasmid DNA from these colonies was sequenced to determine the genotype (Table 3.1). All nine of the plasmid sequences from the non-selective plates contained at least one deletion and are non-functional genes. Of the twelve plasmids that grew on the selective media, all contain full-length RXR LBDs with designed mutations. To evaluate the efficiency of library creation and selection a binary approach is taken – either the sequence is or is not a designed sequence. The following equation is the relevant binomial distribution for statistical evaluation of the libraries:

$$P = \frac{(N-1)!}{(k-1)!(N-k)!} p^k (1-p)^{N-k} \qquad (3.1)$$

Table 3.1 Genotypes of variants from unselected and selected libraries

Unselected Library

	I268	A271	A272	I310	F313	L436
1	deleted			deleted		deleted
2	deleted			deleted		deleted
3	GTA(V†)	CCT(P)	CCT(P)		TCG(S)	deleted
4	deleted			deleted		deleted
5	deleted			deleted		GCG(A)
6	deleted			deleted		deleted
7	deleted			deleted		deleted
8	deleted			deleted		deleted
9	deleted			deleted		TTC(F)

Selected Library

	I268	A271	A272	I310	F313	L436
1	GTG(V)		GTG(V)	TTG(L)	ATG(M)	TTG
2	GTG(V)		GCA	GTG(V)	TCC(S)	TTG
3	CTA(L)	GCT	GCA	ATG(M)	GTG(V)	TTG
4	GCG(A)		GCA	TCC(S)	GTG(V)	TTC(F)
5	GCT(A)	GCT	GCA	GCC(A)	GCG(A)	TTC(F)
6	GCT(A)	GCT	GTT(V)	GCC(A)	GCG(A)	TTC(F)
7	CTT(L)	GCT	GCT	GTC(V)	ATC(I)	TTG
8	CTG(L)	GTG(V)	GCG	TTG(L)	TTG(L)	TTG
9	GTG(V)	GTG(V)	GCG	TTG(L)	GTG(V)	TTG
10	GTA(V)		GTG(V)	ATG(M)	TCC(S)	ATG(M)
11	GCG(A)	GCG	GCA	ATG(M)	GCG(A)	ACG(T)
12	GCG(A)	GCT	GCG	TCG(S)	GTC(A)	TTC(F)

Columns are organized to indicate randomized codons in the same region of the insert cassette. Sequenced codons are indicated followed by the encoded amino acid in parentheses. Where no codon is indicated, the sequence corresponds to the wild-type RXR codon. Where no amino acid is indicated, the codon change did not cause an amino acid change. The presence of an unmutated 3Stop deletion background cassette is indicated by "deleted".

†Single letter amino acid codes. See abbreviations for full names.

In this equation, N is the number of sequenced plasmids; k is the number of background or designed plasmids; p is the frequency of the occurrence of either background or designed plasmid; and P is the measure of certainty. Applying the equation to the libraries, it can be concluded that with 95% certainty that the unselected library is at least 72% background and the selected library is at least 78% designed sequences.

Variant Characterization in Yeast

The twelve plasmids rescued from the selective plates were retransformed [11] into PJ69-4A to confirm that their phenotype is plasmid linked. The strain PJ69-4A was engineered to contain a Gal4 response element controlling expression of the *LacZ* gene, in addition to the *ADE2* gene. Both selection and screening were used to determine the activation level of each variant by 9cRA and LG335. The selection assay quantifies yeast growth occurring through transcriptional activation of the *ADE2* gene, while the screen quantifies β-galactosidase activity occurring though transcriptional activation of the *LacZ* gene. The selection assay (Figures 3.3) is ~10-fold more sensitive than the screen (Figure 3.4) as growth is observed at 10-fold lower ligand concentrations than β-galactosidase activity. However, the selection assay does not quantify activation level (efficacy) as well as the screen. In the selection assay, there is either growth or no growth, whereas the screen more accurately quantifies different activation levels at various concentrations of ligand (Figures. 3.3 and 3.4).

Three plasmids were used as controls in the screen and selection assays. The plasmids pGBDRXRα and pGBT9Gal4 were used as positive controls to which the activation level of the variants can be compared. pGBDRXRα expresses the gene for the

"wild-type" GBD:RXR, which grows and is activated by 9cRA but not by LG335. pGBT9Gal4 expresses the gene for the ligand-independent yeast transcription factor Gal4 [12], which is constitutively active in the presence or absence of either ligand. The plasmid pGBDRXR:3Stop serves as a negative control. pGBDRXR:3Stop carries a non-functional *RXR LBD* gene; therefore, yeast transformed with this plasmid does not grow in the selection assay nor show activity in the screen. This plasmid provides a measure of background noise in both the selection and screen assays.

Both the selection and screen assays show that ten of the twelve variants are selectively activated by LG335. Results of these assays are shown in Figures 3.3 and 3.4. Table 3.2 summarizes the transcriptional activation profiles of all twelve variants in response to both 9cRA and LG335 compared to wild-type RXR.

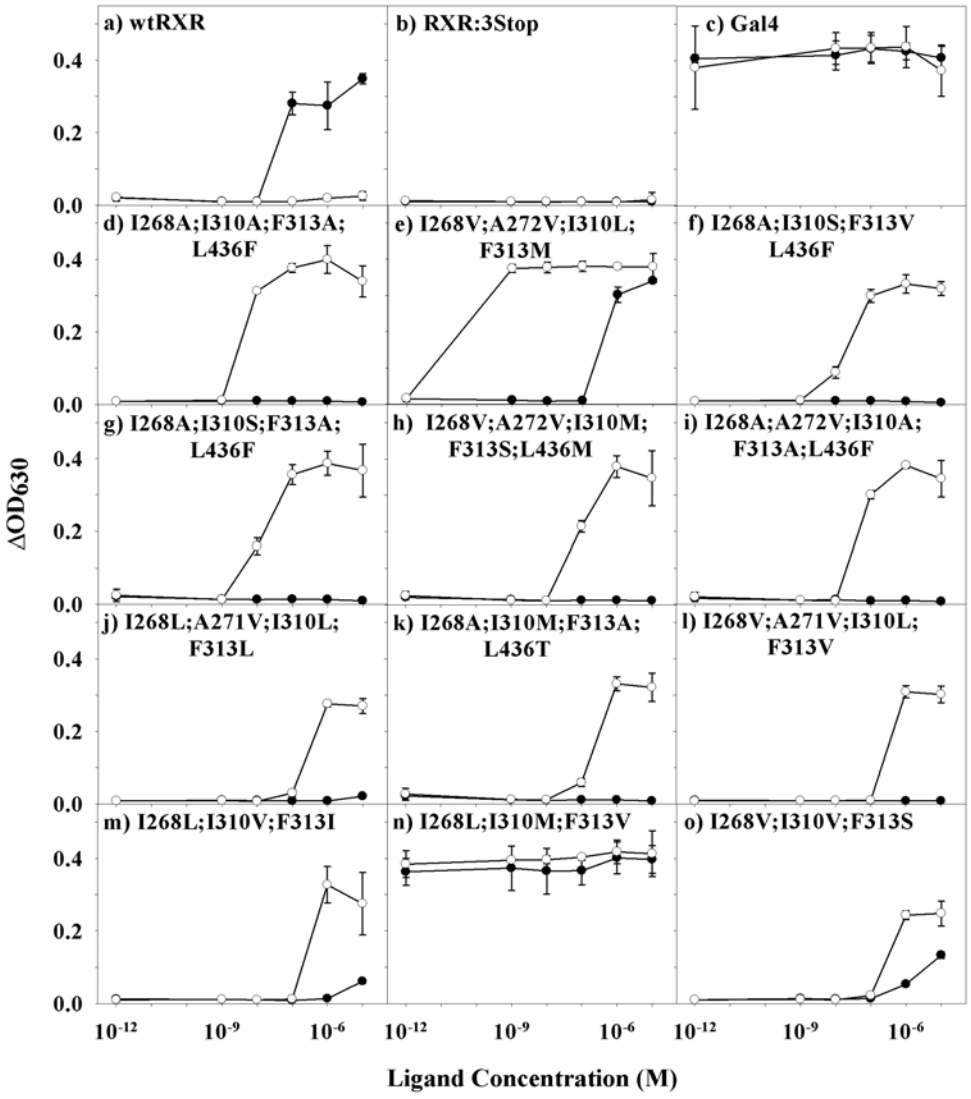


Figure 3.3 Selection assay (SC -Ade -Trp -Leu plus ligand medium) for yeast growth in the presence of 9cRA (closed circles) and LG335 (open circles) for 43 hours. The 10^{-12} M point is actually 0 M ligand concentration.

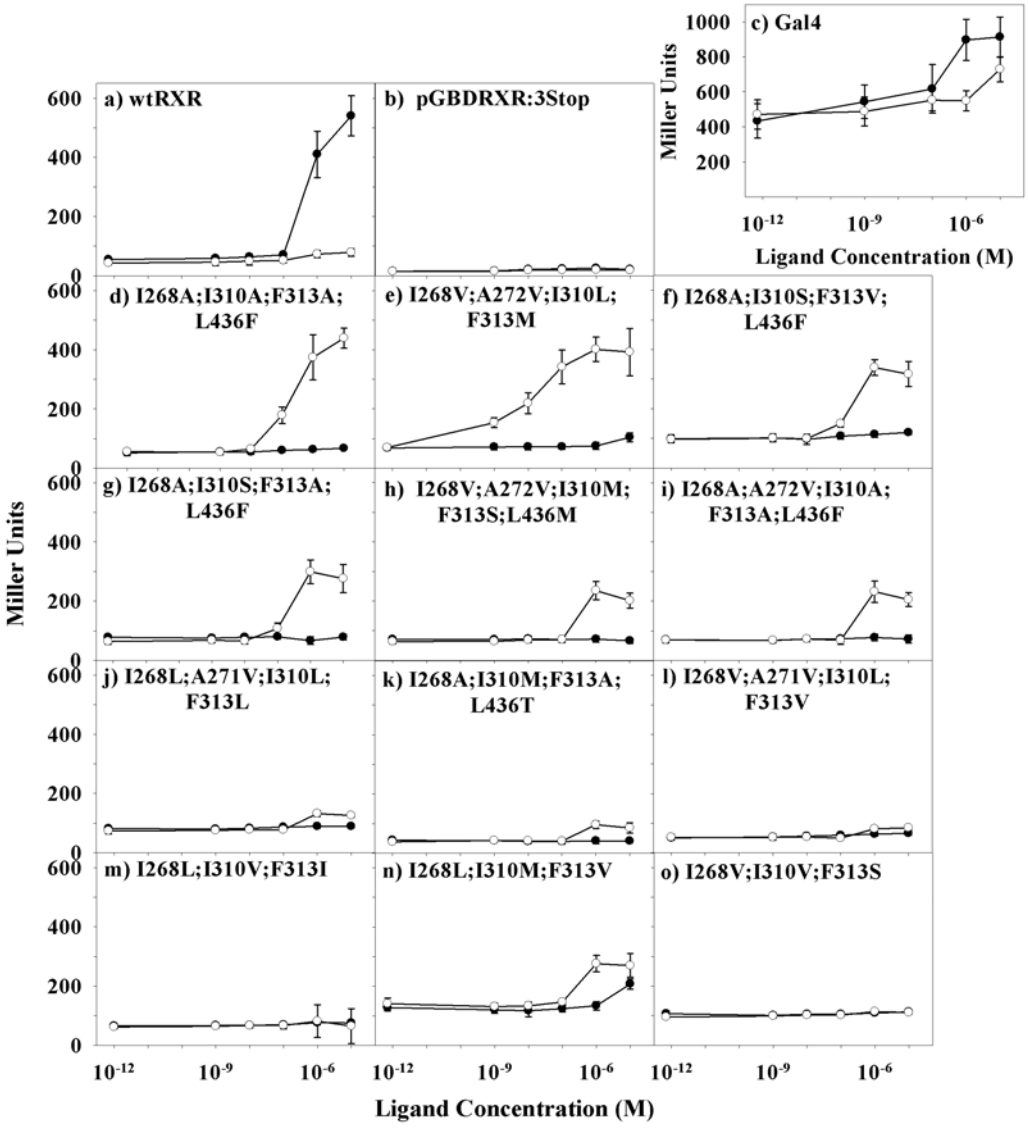


Figure 3.4 Screen assay (SC -Trp -Leu plus ligand for β-galactosidase activity with *o*-nitrophenyl β-D-galactopyranoside (ONPG) in the presence of 9cRA (closed circles) and LG335 (open circles). Miller units normalize the change in absorbance at 405 nm for the change in OD at 630 nm, which reflects the number of cells per well. The 10^{-12} M point is actually 0 M ligand concentration.

Table 3.2 EC_{50} and Efficacy in yeast and HEK 293 cells for RXR Variants

	9cRA				LG335			
	Yeast		HEK 293		Yeast		HEK 293	
Variant	EC_{50}[‡] (nM)	Eff[‖] (%)	EC_{50} (nM)	Eff (%)	EC_{50} (nM)	Eff (%)	EC_{50} (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		

[‡] EC_{50}, concentration of ligand producing half maximal activity. Values represent averages of two screen experiments in quadruplicate for yeast and in triplicate for HEK 293

[‖] Eff, efficacy, maximum increase in activation relative to the increase in activation of wild type with 10 μM 9cRA. Values represent averages of two screen experiments in quadruplicate for yeast and in triplicate in HEK 293.

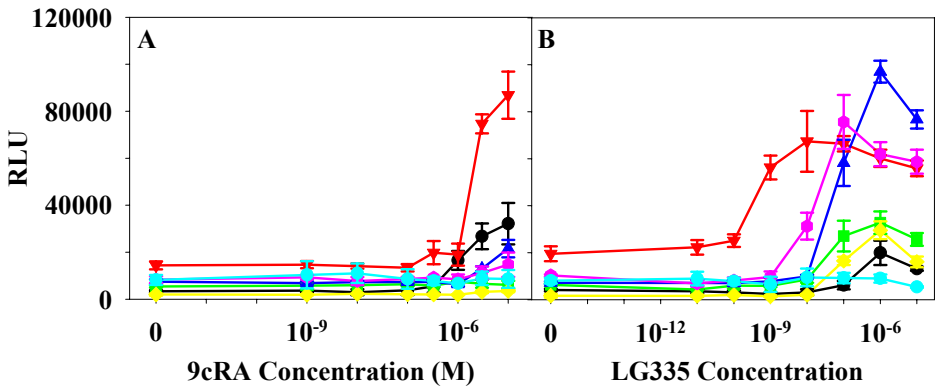


Figure 3.5 Activation of a leuciferase reporter in HEK 293 cells by RXR variants in response to 9cRA (A) and LG335 (B). wt RXR (black circle), I268V;A272V;I310L;F313M (red inverted triangle), I268A;I310M;F313A;L436T (green square), I268A;I310S;F313A;L436F (yellow diamond), I268V;A272V;I310M;F313S;L436M (blue triangle), I268A;A310A;F313A;L436F (magenta octagon), and RXR:3Stop (cyan circle).

Mammalian Correlation

Five variants were chosen for testing in mammalian cell culture for comparison of the activation profiles (I268A;I310A;F313A;L436F, I268V;A272V;I310L;F313M, I268A;I310S;F313A;L436F, I268V;A272V;I310M;F313S;L436M, and I286A;I310M;F313A;L436T). The genes for these variants were removed from yeast expression plasmids and ligated into mammalian expression plasmids (Figure 3.5).

Discussion

The chemical complementation system presented here focuses on one small molecule target ligand and utilizes the power of genetic selection to reveal proteins within the library that bind and activate transcription in response to that small molecule. Functional receptors were isolated from a large pool of non-functional variants, even

from a non-optimized library. Of the nine plasmids from the unselected library, all nine contained sequences that were at least partially background. Therefore, the majority of the library contains at least partial sequence from the non-functional gene (RXR:3Stop) used in the recombination process. Of the 12 plasmids from the selected library, all contain designed sequences. Chemical complementation is very efficient at finding functional receptors within a large collection of nonfunctional receptors.

Importantly, this research demonstrates that ligand-receptor pairs that have various activation properties can be created and discovered. Figures 3.3 and 3.4 depict examples of the differing activation profiles, which are summarized in Table 3.2. Some of the variants approach wild-type and Gal4 levels of activation (> 50% Eff., Figure 3.4a and c-f), while some are activated only to low (< 25% Eff., Figure 3.4j-l) or moderate levels (between 25% and 50% Eff.) (Figure 3.4g-i). Most of the ligand-receptor pairs act like switches (i.e., either on or off) (Figure 3.4h); they are fully activated at one concentration, but show no activation at the next lower concentration. However, some act as rheostats (Figure 3.4e); transcription levels can be varied by changing ligand concentration over several orders of magnitude. These variants can be used to titrate levels of transcription by varying the concentration of ligand. The ability to titrate transcription levels should be useful in gene therapy to control transcription of therapeutic genes and in agriculture to control transcription of pesticidal or other genes.

Although I268L;I310M;F313V is constitutively active in the selection assay (Figure 3.3n) and has high basal activity in the screen assay, both 9cRA and LG335 increase activity at micromolar concentrations (Figure 3.4n). This variant may be in an intermediate conformation with weakly activated transcription that can be improved by

ligand binding. The high basal activation could also be due to a change in the conformation equilibrium with a shift towards the active conformation when ligand is not present.

I268V;I310V;F313S is constitutively active on solid media (data not shown), but shows no activation in the screen (0% Eff., Table 3.2, Figure 3.4o) and only grows in the liquid media selection after two days (Figure 3.3o). The basal activation level may be below the threshold of detection for the liquid media assays. However, it is also possible that agar, which is not present in the liquid assays, contains some small molecule that activates the receptor.

Activation levels and EC_{50}s correlate in yeast and HEK 293 cells (Figure 3.5 and Table 3.2). For the majority of the variants 9cRA shows little or no activation in yeast or mammalian cells. Variant I268V;A272V;I310L;F313M is activated slightly by 9cRA in yeast, but in mammalian cells is activated to the same level as with both 9cRA and LG335 (Figures 3.3, 3.4 and 3.5). With one exception, all variants tested have EC_{50}s within 10-fold in yeast and mammalian cells. However, the EC_{50}s in mammalian cells are generally lower than in yeast. This shift may be due to increased penetration of LG335 into mammalian cells versus yeast.

Subtle differences in binding pocket shape can have a drastic effect on specificity. For example, the I268V;A272V;I310L;F313M variant is activated to high levels by LG335 (60% Eff., Table 3.2), and is only slightly activated by 10 μM 9cRA in yeast (Figure 3.4e), yet the amino acid changes are extremely conservative. The volume difference between phenylalanine and methionine side chains is only ~ 4 $Å^3$ [13] and their polarity difference is minimal (hydration potentials of the methionine and

phenylalanine side chains are –0.76 kcal mol^{-1} and –1.48 kcal mol^{-1}, respectively) [14]. The other mutations redistribute methyl groups within the binding pocket, with a net difference of one methyl group (~ 18 $Å^3$) [13] .

The LG335-I268V;A272V;I310L;F313M ligand-receptor pair also represents a 25-fold improvement in EC_{50} over our previous best LG335 receptor, Q275C;I310M;F313I (40 nM vs. 1 μM in yeast) [4]. The Q275C;I310M;F313I variant was created using site directed mutagenesis. The subtle changes in the I268V;A272V;I310L;F313M variant would not have been predicted to produced a better ligand receptor pair than the Q275C;I310M;F313I variant. This conclusion is consistent with the observation that nuclear receptors bind ligands through an induced-fit mechanism [15]. With current knowledge about protein-ligand interactions it is not possible to design ligand-receptor pairs with specific activation profiles. Libraries and chemical complementation are a new way to circumvent this problem and obtain functional variants with a variety of activation profiles.

Molecular modeling was used to generate hypotheses about the structural basis of ligand specificity for the variants discovered in the library. First, mutations to smaller or more flexible side chains at positions 310 and 313 are essential to provide space for the propyl group of LG335. All variants activated by LG335 have mutations at these two positions. Second, mutations to amino acids with larger side chains at position 436 stericly clash with the methyl group at the 9 position of 9cRA. This interaction may prevent helix 12 from closing properly and therefore prevent activation by 9cRA. The only variant significantly activated by 9cRA (I268V;A272V;I310L;F313M) does not contain a mutation at position 436. Third, tight packing in the binding pocket may lead to

lower EC_{50}s. The docking results for I268V;A272V;I310L;F313M with LG335 show that the methionine and leucine side chains pack tightly against the propyl group of LG335, which may result in tighter binding and consequently a lower EC_{50}.

In this study, both structural information and functional data from previous site directed mutations were utilized for the library design. In the absence of functional data, chemical complementation may be used to test more hypotheses about the function of particular residues than would be possible through site directed mutagenesis. By making a library of changes at a single site, additional information could be obtained about the importance of side chain size, polarity, and charge over just the traditional mutation to alanine that is often used to explore single residue importance. In the absence of structural information, it is possible to make large libraries using error prone PCR or gene shuffling. Chemical complementation could also be used to select active variants from these types of libraries.

Chemical complementation allows high-throughput testing of large libraries. Hundreds of thousands of variants can be assayed in one experiment without the spatial resolution necessary for traditional screening methods (i.e., no need for one colony per well). Yeast can be spread on solid media and, through the power of genetic selection, cells expressing active variants will grow into colonies. Survivors can then be spatially resolved (i.e. transferred to a microplate, one colony per well) for further characterization, decreasing the time and effort required to find new ligand-receptor pairs.

Chemical complementation identifies nuclear receptors with a variety of responses to a specific ligand. Nuclear receptors that activate transcription in response to

targeted molecules and not to endogenous compounds have several additional potential applications. The ability to switch a gene on and off in response to any desired compound can be used to build complex metabolic pathways, gene networks, and to create conditional knockouts and phenotypes in cell lines and animals. This ability should also be useful in gene therapy and in agriculture to control expression of therapeutic or pesticidal genes. A variety of responses would be useful in engineering biosensor array. An array of receptors with differing activation profiles for a specific ligand could provide concentration measurements and increased accuracy of detection. The ability to engineer proteins that activate transcription in response to any desired compound with a variety of activation profiles will provide a general method of engineering enzymes. Receptors that bind the product of a desired enzymatic reaction can be used to select for enzymes that perform this reaction. The stringency of the assay can be adjusted by using ligand-receptor pairs with lower or higher EC_{50}s. The lack of a general system for genetic selection is currently the limiting step for directed evolution of enzymes [16, 17].

Materials and Methods

Ligands

9-*cis* retinoic acid (MW=304.44 g/mol) was purchased from ICN Biomedicals. LG335 [10] was synthesized [18-22] (Chapter 2).

Expression Plasmids

pGAD10BAACTR [9, 23] , pGBT9Gal4, pGBDRXRα [8], pCMX-hRXR [24],

and pCMX-βGAL [3] have been described. pGBDRXRα was cut with *SmaI* and *NcoI,* filled in, and blunt-end ligated to eliminate 153 amino acids of the RXR DBD. A *HindIII* site in the tryptophan selectable marker was silently deleted and the sole remaining *HindIII* site was cut, filled in, and blunt-end ligated to remove the restriction site. Unique *HindIII* and *SacI* sites were inserted into the RXR LBD gene and *MfeI* and *EcoRI* sites were removed from the plasmid using QuikChange Site-Directed Mutagenesis (Stratagene, La Jolla, CA) to create pGBDRXRαL-SH-ME. pCMX-hRXR mutants were cloned from pGBDRXR vectors using *SalI* and *PstI* restriction enzymes and ligated into similarly cut pCMX-hRXR vectors. pLuc_CRBPII_MCS 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 (MCS), controlling transcription of the firefly luciferase gene. All plasmids have been confirmed through sequencing.

Plasmid Construction

The zero background plasmid, pGBDRXR:3Stop, was constructed using QuikChange Site-Directed Mutagenesis with pGBDRXRαL-SH-ME as the template and the 3Stop insert cassette (described below) as primers.

The 3Stop insert cassette was synthesized using PCR from eight oligonucleotides (Figure 3.6). All PCRs were done using 2.5 U Pfu Polymerase (Stratagene, LaJolla, CA), 1x Pfu buffer, 0.8 mM dNTPs, 50 ng of pGBDRXRαL-SH-ME as a template, 125 ng of primers and sterile water to make 50 μL. First, four small cassettes were synthesized in reactions containing the following primers: Cassette 1, F

(5’-CGGAATTTCC CATGGGC-3’), BPf (5’-CTCGCCGAAC GACCCGGTCA CCGCATGCCA CTAGTGG-3’), and BPr (5’-CCGCTTGGCC CACTCCACTA GTGGCATGCG GTGACC-3’); Cassette 2, BPf, BPr, SEf (5’-CGGGCAGGCT GGAATGAGCT CCTCGACGGA ATTCTCC-3’), and SEr (5’-CAGCCCGGTG GCCAGGAGAA TTCCGTCGAG GAGCTC-3’); Cassette 3, SEf, SEr, AMf (5’-CTCTGCGCTC CATCGGGCTT AAGTGCCCAC CAATTGACAC-3’), and AMr (5’-CTCCAGCATC TCCATAAGGA AGGTGTCAAT TGGTGGGCAC TTAAGC-3’); Cassette 4, AMf, AMr, and R (5’-CAAAGGATGG GCCGCAG-3’). The cassettes were cleaned with either the DNA Clean and Concentrator-5 (Zymo Research, Orange, CA) or the Zymoclean Gel DNA Recovery Kit (Zymo Research, Orange, CA) depending on product purity. The four cassettes were used to make the final 3Stop insert cassette in a PCR that contained each cassette, primers F and R, dNTPs, Pfu Polymerase, and sterile water to a final volume of 50 μL. The 3Stop cassette was cleaned using the Zymoclean Gel DNA Recovery Kit.

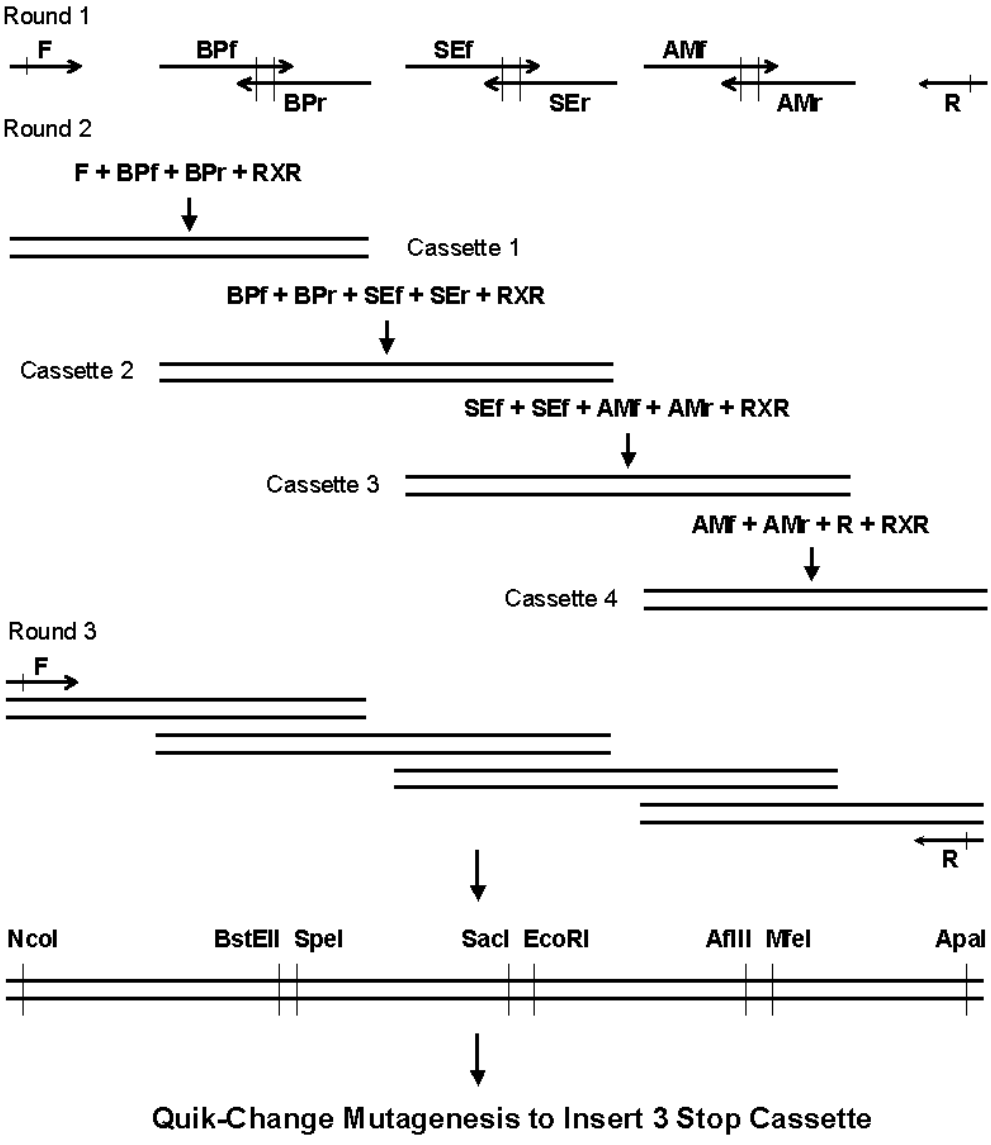


Figure 3.6 Construction of pGBDRXR:3 stop.

Insert Cassette Library Construction

The library of insert cassettes with randomized codons was constructed in a similar manner as above. The four cassettes (FBP, BPSE, SEAM and AMR) were made in the following ways (Figure 3.7).

For the FBP cassette, oligos BP1 (5'-GGCAAACATG GGGCTGAACC CCAGCTCGCC GAACGACCCG GTCACC-3'), BP2 (5'-GCCCACTCCA CTAGTGTGAA AAGCTGTTTG TC (A, C, or T)(A or G)(C or G)(A, C, or T)(A or G)(C or G)TT GGCA(A, C, or T)(A or G)(C or G)GTT GGTGACCGGG TCGTTCG-3'), BP3 (5'-CTTTTCACAC TAGTGGAGTG GGCCAAGCGG ATCCCACACT TCTCAGAG-3'), and BP4 (5'-GGGGCAGCTC TGAGAAGTGT GGGATCCG-3') were mixed with TE containing 100 mM NaCl to bring the total volume to 50 μL. The mixture was heated to 95 °C for 1 minute, then slowly cooled to 10 °C. The annealed mixture was combined with EcoPol Buffer, dNTPs, ATP, Klenow (NEB, Beverly, MA), T4 DNA ligase (NEB, Beverly, MA) and sterile water to 200 μL, and kept at 25°C for 45 min before heat inactivation at 75°C for 20 minutes. The product was cleaned with DNA Clean and Concentrator-5 to make the BP cassette. Next, BP cassette was combined with Pfu Buffer, pGBDRXR:3Stop, oligo F, dNTPs, Pfu polymerase, and sterile water to make 50 μL for a PCR. The final FBP product (300bp) was purified using the Zymoclean Gel DNA Recovery Kit.

BPSE was made in two consecutive PCRs. First, SE1 (5'-GCAGGCTGGA ATGAGCTCCT C(A, G, or T)(C or T)(G or C)GCCTCC (A, G, or T)(C or T)(G or C)TCCCACC GCTCCATC-3') and SE2 (5'-CCGGTGGCCA GGAGAATTCC GTCCTTCACG GCGATGGAGC GGTGGG-3') were combined with Pfu buffer,

dNTPs, Pfu polymerase, and sterile water to make 50 μL. After 5 PCR cycles, pGBDRXR:3Stop and BP were added to the reaction and the PCR was continued for 30 cycles. The product (240 bp) was purified using the Zymoclean Gel DNA Recovery Kit.

SEAM was constructed in a similar way to BPSE. SE1 and SE2 were mixed with Pfu Buffer, dNTPs, Pfu polymerase, and sterile water to 25 μL. Simultaneously, AM1 (5'-GGCTCTGCGC TCCATCGGGC TTAAGTGCCT GGAACAT(A, G, or T)(C or T)(G or C) TTSCTTCTTC AAGCTCATCG GGG-3') and AM2 (5'-GCATCTCAAT AAGGAAGGTG TCAATTGTGT GTCCCCGATG AGCTTGAAGA A-3') were combined with Pfu Buffer, dNTPs, Pfu polymerase, and sterile water to 25 μL. After 5 cycles, these two reactions were mixed and pGBDRXR:3Stop was added. The PCR was continued for 30 cycles. The PCR product (460 bp) was purified using the Zymoclean Gel DNA Recovery Kit.

The AMR cassette was made similarly to FBP. AM1 and AM2 were mixed with TE containing 100 mM NaCl to make 50 μL, heated to 95°C for 1 minute, then slowly cooled to 10°C. The annealed mixture was combined with EcoPol Buffer, dNTPs, Klenow, and sterile water to 200 μL, and kept at 25°C for 45 min before heat inactivation at 75°C for 20 minutes. The product (AM) was precipitated with isopropanol. Next, AM and R were combined with Pfu buffer, pGBDRXR:3Stop, dNTPs, Pfu Polymerase, and sterile water to make 50 μL for a PCR. The product (140 bp) was purified using the Zymoclean Gel DNA Recovery Kit.

The four cassettes (FBP, BPSE, SEAM, and AMR) were combined in a PCR to make the library of randomized insert cassettes (6mutIC). The library was cleaned using Bio-Spin 30 columns (Bio-Rad Laboratories, Hercules, CA).

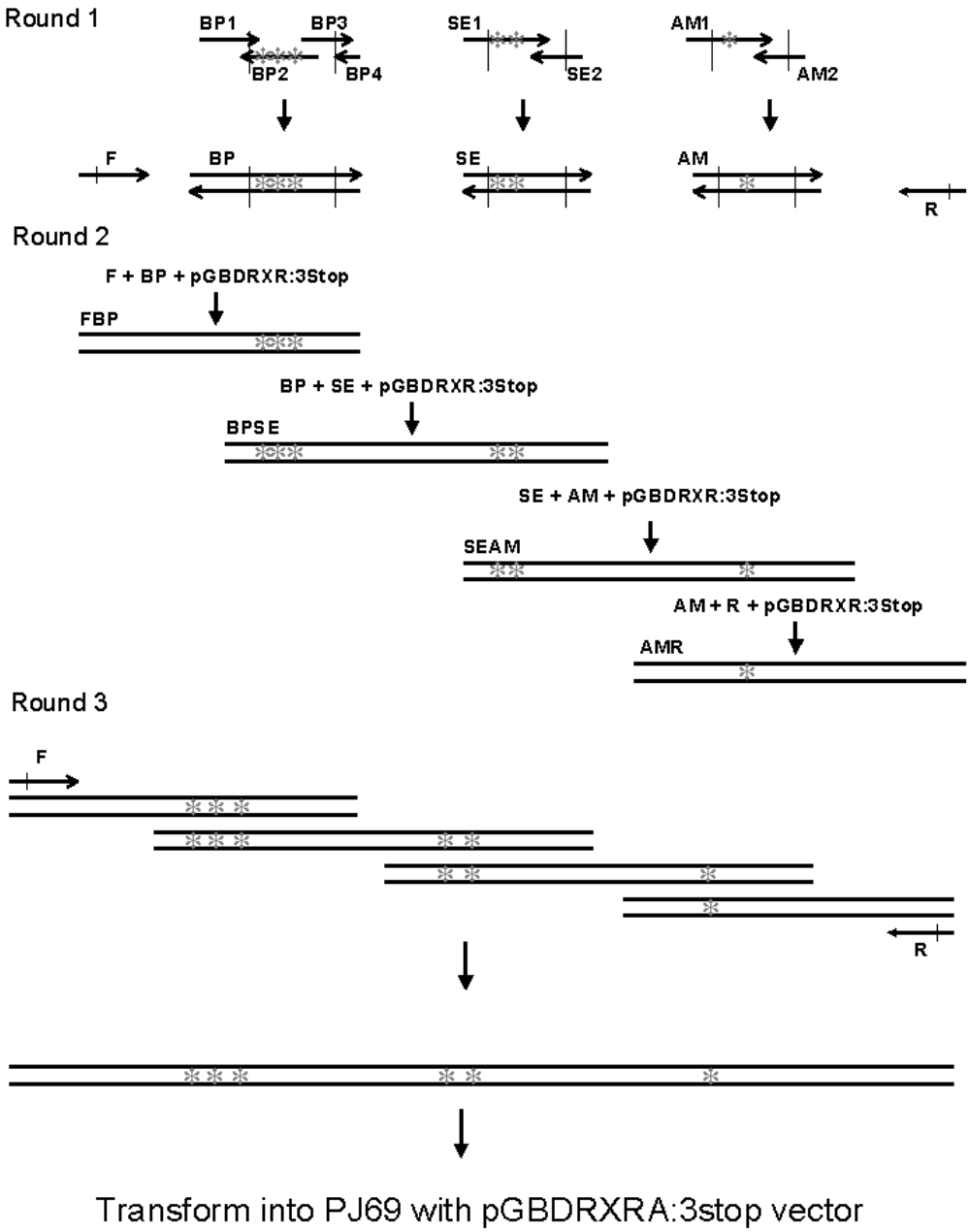


Figure 3.7 Construction of insert cassette library. Asterisks indicate location of randomized codons.

Yeast selection plates and transformation

Synthetic complete (SC) media and plates were made as previously described [8]. Selective plates were made without tryptophan (-Trp) and leucine (-Leu) or without adenine (-Ade), tryptophan (-Trp) and leucine (-Leu). Ligands were added to the media after cooling to 50 °C.

The randomized cassette library was homologously recombined into the pGBDRXR:3Stop plasmid using the following method. pGBDRXR:3Stop was first digested with *BssHII* and *EagI* (NEB, Beverly, MA), and then treated with calf intestinal phosphatase (NEB, Beverly, MA), to make a vector cassette. Vector cassette (1 μg) and 6mutIC (9 μg) were transformed according to Geitz's transformation protocol [11] on a 10X scale into the PJ69-4A yeast strain, which had previously been transformed with a plasmid (pGAD10BAACTR) [9] expressing the nuclear receptor coactivator ACTR fused to the yeast Gal4 activation domain. Homologous regions between the vector cassette and the insert cassette allow the yeast to homologously recombine the insert cassette with the vector cassette forming a circular plasmid with a complete RXR LBD gene. The transformation mixture (1 mL) was spread on each of 10 large plates of SC -Ade -Trp -Leu media containing 10 μM LG335. The transformation mixture (2 and 20 μL) was also spread on SC -Trp -Leu media. These plates were grown for 4 days at 30 °C.

Genotype Determination

Plasmids were rescued using either the Powers method (www.fhcrc.org/labs/gottschling/yeast/yplas.html) or the Zymoprep Kit (Zymo Research, Orange, CA). The plasmids were then transformed into Z-competent (Zymo Research,

Orange, CA) XL1-Blue cells (Stratagene, La Jolla, CA). The QIAprep Spin Miniprep Kit (Qiagen, Valencia, CA) was used to purify the DNA from the transformants. These plasmids were sequenced.

Quantification Assays

Solid Media

The rescued plasmids were transformed [11] into PJ69-4A containing the pGAD10BAACTR plasmid and plated on SC -Trp -Leu media. These plates were grown for 2 days at 30 °C.

Colonies were streaked onto the following media: SC, SC -Trp -Leu, SC -Ade -Trp -Leu, SC -Ade -Trp -Leu plus increasing concentration of LG335 or 9cRA from 1 nM to 10 μM.

Liquid Media

The method used for quantitation in the paper was modified from a method developed by Miller [25]. Briefly, SC -Trp -Leu media was inoculated with PJ69-4a containing the pGBDRXR plasmid of choice and the pGAD10BA:ACTR plasmid and grown overnight at 30 °C with shaking. The cultures were centrifuged at 5000 rpm for 5 minutes and washed twice with sterile water. The cultures were then used to inoculate 100 μL of either SC -Trp -Leu (screen) or SC -Ade -Trp -Leu (selection) media in 96-well plates. The OD_{630} was taken with an OpsysMR plate reader (Dynex Technologies, Chantilly, VA). These were grown for 1 to 2 days at 30 °C with moderate shaking. For selection experiments, the OD_{630} was recorded after 1 and 2 days. For screens, the OD_{630} was

taken after 24 hours. The yeast was pelleted and the supernatant was removed. The cells were lysed with yeast lytic enzyme and Triton X-100. After lysis, the cell debris was pelleted by centrifugation and the supernatant moved to new 96-well plates. ONPG was added to the wells and the OD_{405} was recorded. The plates were incubated at 37 °C for 30 to 60 minutes. The reaction was quenched with sodium carbonate and the OD_{405} was recorded again. Miller units were calculated using Equation 4.2 where *v* is the culture volume and *t* is the time from the addition of ONPG to the addition of sodium carbonate.

$$MU = \frac{1000 \times \Delta OD_{405}}{v \times t \times \Delta OD_{630}} \qquad (3.2)$$

Mammalian Luciferase Assay

HEK 293 cells were cultured in DMEM supplemented with 10% calf bovine serum. Twenty-four hours before transfection, 8×10^4 cells per well were plated in a 48 well plate. pCMX-hRXR mutants (20ng) were cotransfected with pLucCRBPII (40ng) and pCMX-βGal (40ng) using lipofectamine cationic lipid according to the protocol recommended by the manufacturer (Gibco BRL/Life Technologies) . After 10 h of transfection, dilutions of ligands 9cRA and LG335 were prepared in super-stripped media and added to the plate in triplicate. Cells were harvested after 36 h and assayed for luciferase and β–galactosidase activity.

Molecular Modeling

Docking of LG335 in to modified binding pockets was done using the InsightII module Affinity®. The wild type RXR with 9cRA crystal structure [2] was modified using the Biopolymer module residue replace tool to make mutations in the binding

pocket that corresponded to the mutations in variants I268;I130A;F313A;L436F, I268V;A272V;I310L;F313M, and I268A;I310S;F313A;L436F. The ligand was placed in the binding pocket by superimposing the carboxylate carbon and two carbons in the tetrahydronapthalene ring of LG335 onto corresponding carbons of 9cRA in the crystal structure. A Monte Carlo simulation was performed first, followed by Simulated Annealing of the best docked conformations. See chapter 1 for a more complete description of the method.

Literature Cited

1. Wallace, A.C., R.A. Laskowski, and J.M. Thornton, *LIGPLOT: a program to generate schematic diagrams of protein-ligand interactions.* Protein Eng, 1995. **8**(2): p. 127-34.

2. 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.

3. 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.

4. 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.

5. 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**(2): p. 206.

6. Reidhaar-Olson, J.F., J.U. Bowie, R.M. Breyer, J.C. Hu, K.L. Knight, W.A. Lim, M.C. Mossing, D.A. Parsell, K.R. Shoemaker, and R.T. Sauer, *Random mutagenesis of protein sequences using oligonucleotide cassettes.* Methods Enzymol, 1991. **208**: p. 564-86.

7. Lim, W.A. and R.T. Sauer, *The role of internal packing interactions in determining the structure and stability of a protein.* J Mol Biol, 1991. **219**(2): p. 359-76.

8. 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.

9. Azizi, B., E.I. Chang, L.J. Schwimmer, and D.F. Doyle, *A sensitive chemical complementation for drug discovery and protein engineering.* Manuscript Submitted.

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. Geitz, R.D. and R.A. Woods, *Transformation of yeast by the LiAc/SS carrier DNA/PEG method.* Methods Enzymol, 2002. **350**: p. 87-96.

12. Johnston, S.A., M.J. Zavortink, C. Debouck, and J.E. Hopper, *Functional domains of the yeast regulatory protein GAL4.* Proc Natl Acad Sci U S A, 1986. **83**(17): p. 6553-7.

13. Doyle, D.F., *Experimental and Theoretical Stability Studies on Cytochrome c*. 1996, University of North Carolina: Chapel Hill.

14. Wolfenden, R., L. Andersson, P.M. Cullis, and C.C. Southgate, *Affinities of amino acid side chains for solvent water.* Biochemistry, 1981. **20**(4): p. 849-55.

15. Egea, P.F., A. Mitschler, and D. Moras, *Molecular recognition of agonist ligands by RXRs.* Mol Endocrinol, 2002. **16**(5): p. 987-97.

16. Zhao, H. and F.H. Arnold, *Combinatorial protein design: strategies for screening protein libraries.* Curr Opin Struct Biol, 1997. **7**(4): p. 480-5.

17. Firestine, S.M., F. Salinas, A.E. Nixon, S.J. Baker, and S.J. Benkovic, *Using an AraC-based three-hybrid system to detect biocatalysts in vivo.* Nat Biotechnol, 2000. **18**(5): p. 544-7.

18. Held, P., M.P. Heck, J. Iyer, H. Gronemeyer, L. Lebeau, and C. Mioskowski, *Synthesis of a radiolabelled Retinoid X Receptor (RXR) specific ligand.* Journal of Labelled Compounds & Radiopharmaceuticals, 1997. **39**(6): p. 501-507.

19. Kagechika, H., E. Kawachi, Y. Hashimoto, T. Himi, and K. Shudo, *Retinobenzoic acids. 1. Structure-activity relationships of aromatic amides with retinoidal activity.* J Med Chem, 1988. **31**(11): p. 2182-92.

20. Chandraratna, R.A., *Preparation of 1-cyclopropyl-4-tetrahydronaphthyl-1,3-butadienes having retinoid-like activity*, U.S. Patent 5,344,959, 1994, Allergan, Inc.

21. Alvi, K.A., J. Rodriguez, M.C. Diaz, R. Moretti, R.S. Wilhelm, R.H. Lee, D.L. Slate, and P. Crews, *Protein-Tyrosine Kinase Inhibitory Properties of Planar Polycyclics Obtained from the Marine Sponge Xestospongia Cf Carbonaria and from Total Synthesis.* Journal of Organic Chemistry, 1993. **58**(18): p. 4871-4880.

22. Faul, M.M., A.M. Ratz, K.A. Sullivan, W.G. Trankle, and L.L. Winneroski, *Synthesis of novel retinoid X receptor-selective retinoids.* J Org Chem, 2001. **66**(17): p. 5772-82.

23. Chen, H., R.J. Lin, R.L. Schiltz, D. Chakravarti, A. Nash, L. Nagy, M.L. Privalsky, Y. Nakatani, and R.M. Evans, *Nuclear receptor coactivator ACTR is a novel histone acetyltransferase and forms a multimeric activation complex with P/CAF and CBP/p300.* Cell, 1997. **90**(3): p. 569-80.

24. 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.

25. Miller, J.H., *Experiments in Molecular Genetics*. 1972, Cold Spring Harbor, NY: Cold Spring Harbor Laboratory.

CHAPTER 4

VARIABLE STRINGENCY CHEMICAL COMPLEMENTATION LIQUID ASSAYS

Chemical complementation is a powerful method to find receptors that are activated by a specific small molecule. However, when large libraries (>100,000 variants) of receptors are expressed in yeast, the number of hits (receptors activated by the desired ligand) may be too numerous to practically analyze. Many of the variants may only slightly activate transcription or have high EC_{50}s (in the μM range). It is desirable to have a versatile selection method with which the receptors that have high efficacy and low EC_{50}s can be identified. To this end, a chemical complementation selection method with variable stringency was developed. With variable stringency genetic selection, chemical complementation can identify not only receptors that are activated by the targeted small molecule, but also those that have the desired property of high activation levels. The *HIS3* gene encodes the enzyme imidazole glycerolphosphate dehydratase, which is inhibited by 3-amino-1,2,4-triazole (3AT). Using the *HIS3* gene as a selective marker, the concentration of 3AT can be varied to adjust the stringency of chemical complementation. The concentration of ligand can also be varied to identify receptors with low EC_{50}s. Hits with desired properties originally identified through chemical complementation on solid medium can be characterized using variable stringency chemical complementation.

Results

Library Design

A library was designed based on the results of the previous library [2] (Chapter 3) with mutations at amino acid positions 268, 271, 272, 310, 313, and 436. All six positions are randomized to L, I, V, F, M, A, S, and T creating a new library with 262,144 amino acid combinations (2,985,984 codon combinations). As with the previous library, the plasmid pGBDRXR:3Stop encoding a non-functional RXR gene is used to eliminate wild-type background.

The library of full-length *RXR LBD* genes with randomized codons at positions 268, 271, 272, 310, and 313 and doubly cut pGBDRXR:3Stop vector were co-transformed into yeast. The synthetic library cassettes and the cut vector contain homologous regions enabling the yeast to insert the cassette into the vector by homologous recombination, thereby creating full length *RXR LBD* genes with mutations at the desired positions. A 30x transformation with 3 μg vector and 27 μg library cassette was performed using the Gietz method [3].

Library Selection

Chemical complementation was used to select RXR variants that are activated by LG335. After transformation, the yeast were plated onto medium (SC -Trp -Leu) selecting for two plasmids: pGBDRXR and pGAD10BAACTR. The transformation / recombination efficiency of this library into the yeast strain PJ69-4A was 1.4×10^5 colonies per μg of vector with a total of ~450,000 transformants. These plates were replica plated onto SC -Ade -Trp -Leu medium plus 10 μM LG335. Approximately 1000

colonies grew on the selective plates. These colonies were picked and stored in 96-well plates at -80 °C in SC -Trp -Leu medium with 15% glycerol.

Variable Stringency Selection

The goal for developing a variable stringency selection method is to raise the thranscription activation threshold required for yeast growth (Figure 4.1). The amount of

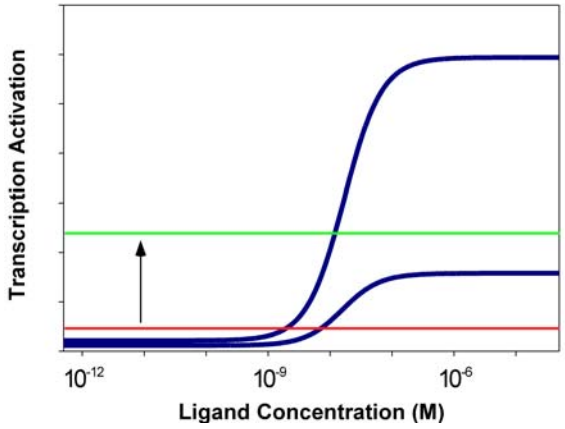


Figure 4.1 Schematic depicting increasing the activation threshold for chemical complementation. Chemical complementation with adenine selection has a low activation threshold that allows growth (red line). A higher activation threshold (green line) is desired to increase the stringency of genetic selection, selecting against low efficacy RXR variants.

activation necessary to allow yeast growth using the adenine selective marker with chemical complementation is approximately 10% or less of wild type activation with 9*c*RA based on the activation and growth profiles of previously characterized variants (Chapter 3). The variants with maximum activation levels of ~10% are not very useful for practical applications of ligand receptor pairs. To remove these low efficacy variants, a selection system with a high activation threshold for growth is desirable.

The PJ69-4A yeast strain has a *HIS3* selective marker under the control of a Gal4 response element. The *HIS3* gene encodes imidazole glycerolphosphate dehydratase, an essential enzyme in the histidine biosynthetic pathway. This enzyme is inhibited by 3AT. By adjusting the concentration of 3AT in SC -His -Trp -Leu medium the stringency of

the selection can be tuned.

Three concentrations of 3AT (20, 40, and 100 mM) were tested to determine the oeffective concentration for selecting against transformants with low efficacy receptors (data not shown). Using 100 mM 3AT in SC -His -Trp -Leu liquid medium plus 10 μM LG335, of the 1000 transformants tested, 35 transformants (deemed a reasonable number for further analysis) grew to a ΔOD_{630} greater than 0.300. Plates grown with 20 and 40 mM 3AT had 10 and 2 fold more transformants growing to a ΔOD_{630} greater than 0.300, respectively.

The 35 transformants were then assayed for low EC_{50}s. Each transformant was grown in SC -His -Trp -Leu medium plus 10 μM or 10 nM LG335 and 100 mM 3AT. SC -His -Trp -Leu plus 100 mM 3AT, and SC -Trp -Leu medium were used as controls. Of the 35 transformants, eight grew with 10 nM LG335 (Figure 4.2). The plasmids in these transformants were rescued and sequenced. For two of the plasmids (2.B1 and 8.D7) complete sequencing was not obtainable for unknown reasons. The plasmids for which complete sequencing was obtained were retransformed to ensure a plasmid-linked phenotype.

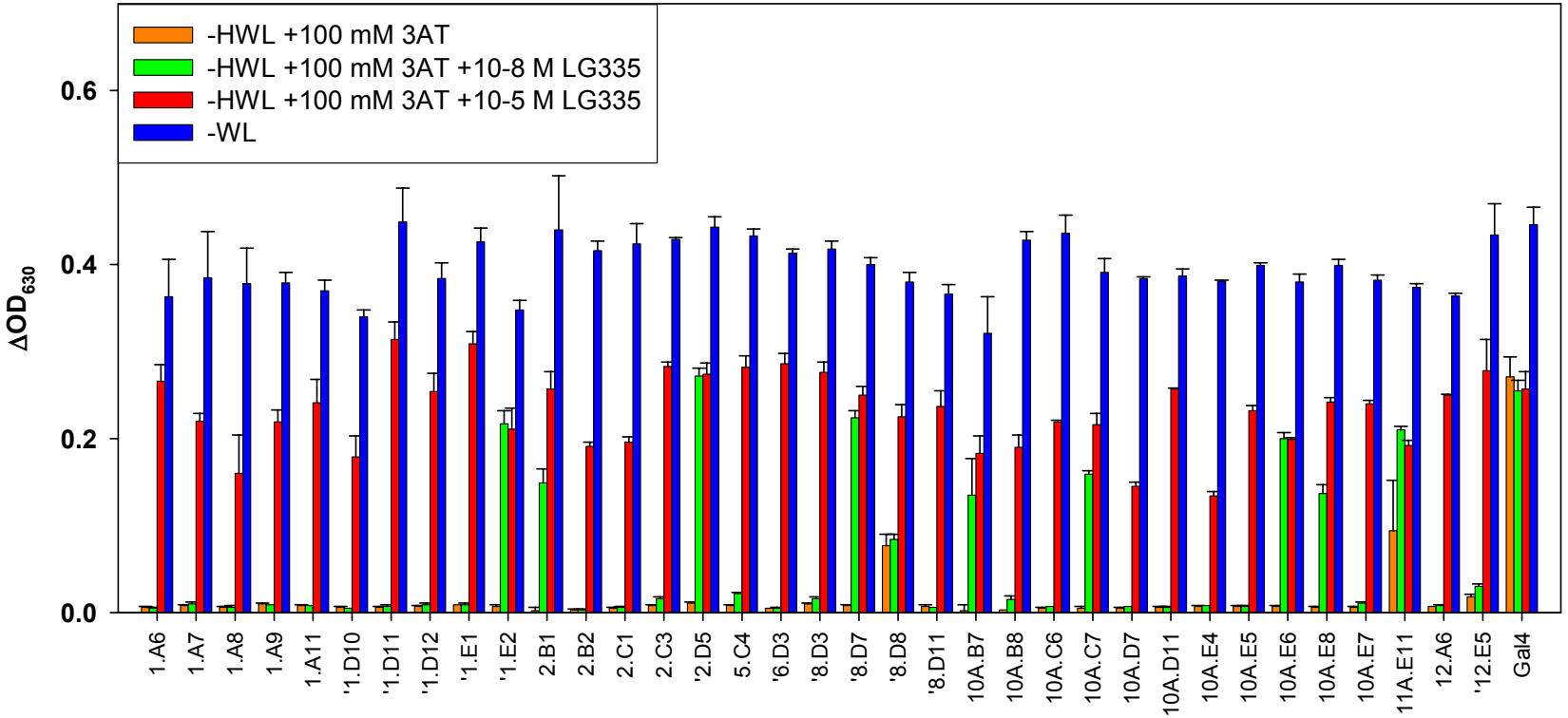


Figure 4.2 Growth of transformants in variable stringency media.

Library Characterization

Ten plasmids from the naïve library were also sequenced. Since this library was constructed in the same manner as the library created in Chapter 3, it was expected to contain mostly non-functional background. Of the ten plasmids sequenced from the naïve library, one plasmid contained the designed sequence, four plasmids were pGBDRXR:3Stop, one plasmid had one deleted cassette, and four plasmids had two deleted cassettes. All of the plasmids from the selected library that could be completely sequenced contained the designed library. With 95% certainty, the naïve library is at least 62% background and the selected library is at least 61% designed sequences as calculated using a binomial distribution (Chapter 3). At most, this library covers 6% of the total possible DNA sequences.

Variant Characterization

Both chemical complementation selection (adenine selection) and screen (β-galactosidase activity) assays were performed to characterize the six variants after retransformation (Figures 4.3 and 4.4 and Table 4.1). Wild-type RXR and Gal4, a ligand independent yeast transcription factor, were used as controls. Five of the six variants are selectively activated by 9cRA. The sixth variant (I310L;F313M;L436M) is activated by a lower concentration and to a higher level with LG335 versus 9cRA. Table 4.1 summarizes the transcriptional activation profiles of all six variants in response to 9cRA and LG335.

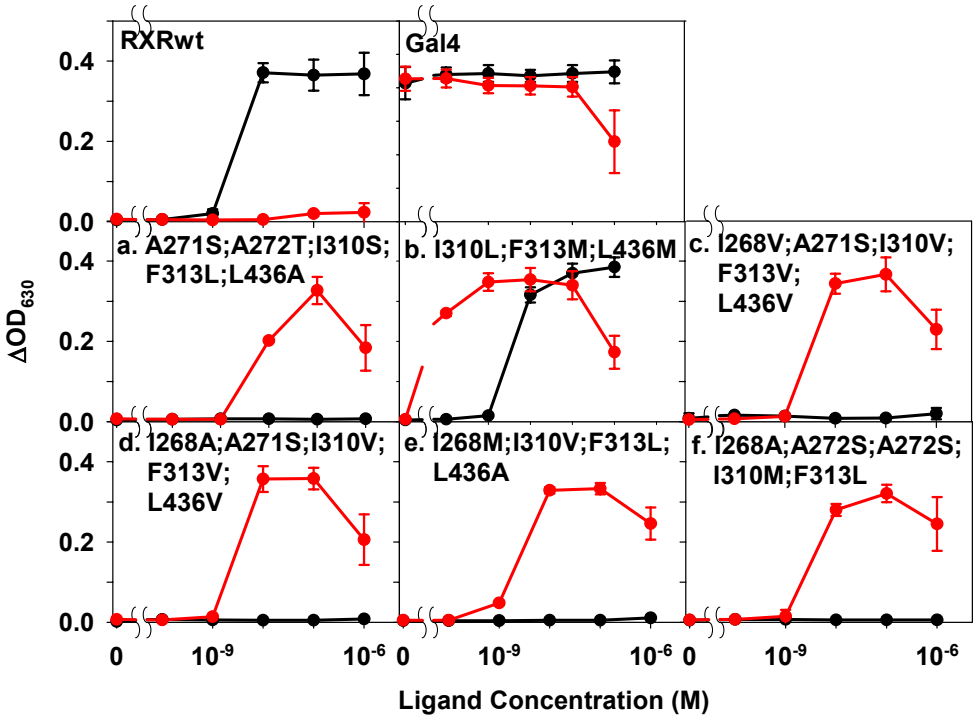


Figure 4.3 Selection assay (SC -Ade -Trp -Leu plus ligand medium) for yeast growth in the presence of 9cRA (black) and LG335 (red) for 48 hours.

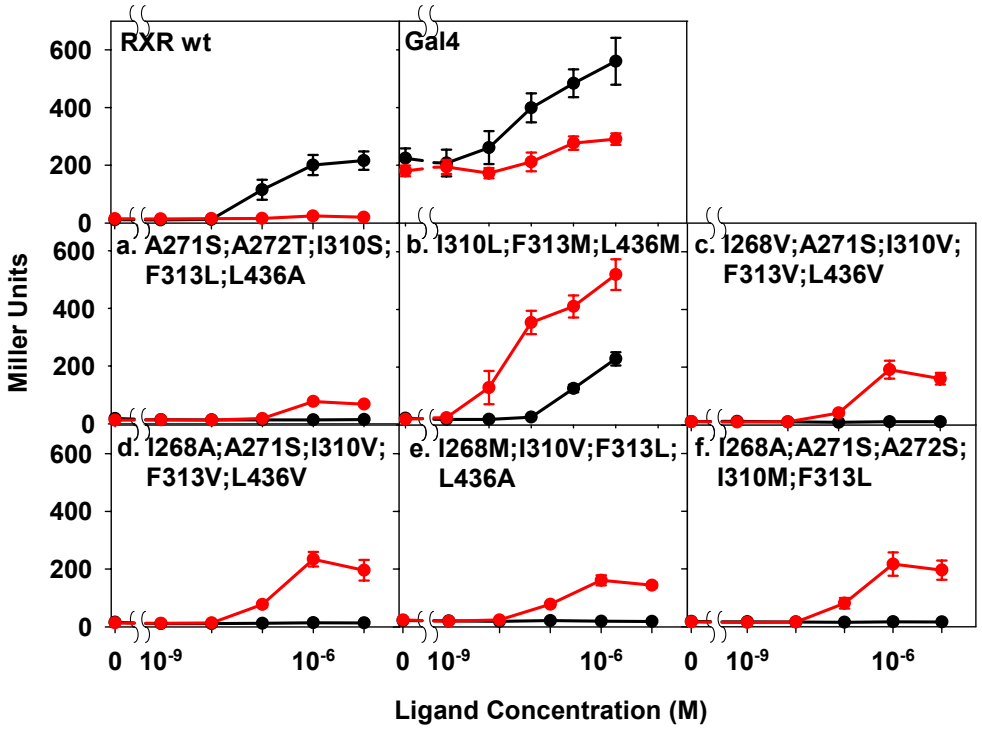


Figure 4.4 Screen assay (SC -Trp -Leu plus ligand medium) for β-galactosidase activity with *o*-nitrophenyl β-D-galactopyranoside (ONPG) in the presence of 9cRA (black) and LG335 (red). Miller units normalize the change in absorbance at 405 nm for the change in optical density at 630 nm, which reflects the number of cells per well.

Table 4.1 EC_{50} and Efficacy for RXR Variants

Variant	Variant Code	9cRA		LG335	
		EC_{50}[ξ] (nM)	Eff[£] (%)	EC_{50} (nM)	Eff (%)
WT		100	100	>10,000	10
I310L;F313M;L436M	2.D5	600	90	20	190
I268V;A271S;I310V;F313V;L436V	10A.B7	>10,000	0	200	80
I268A;A271S;A272S;I310M;F313L	10A.E8	>10,000	0	100	100
I268A;A271S;I310V;F313V;L436V	10A.C7	>10,000	0	300	140
I268M;I310V;F313L;L436A	10A.E6	>10,000	0	60	70
A271S;A272T;I310S;F313L;L436A	1.E2	>10,000	0	200	40

[ξ] EC_{50}, concentration of ligand producing half maximal activity. Values represent averages of three screen experiments in quadruplicate.
[£] Eff, efficacy, maximum increase in activation relative to the increase in activation of wild type with 10 μM 9cRA. Values represent averages of three screen experiments in quadruplicate.

Docking

Using the Insight II docking module Affinity (Accelrys), LG335 was docked into each of the six RXR variants discovered using variable stringency selection as well as twelve RXR variants previously discovered (Chapter 3) [2]. The variants were created *in silico* starting with the RXR-9cRA crystal structure [4] and using the Biopolymer module's residue replacement feature. The side chains (past the β-carbon) of the replaced residues were energy minimized. Docking was accomplished using Metropolis Monte Carlo simulation followed by simulated annealing.

Discussion

Variable Stringency Selection

Under normal conditions, chemical complementation using the *HIS3* gene as a selective marker is less stringent that *ADE2* selection [5]. 3AT is an inhibitor for the enzyme imidazole glycerolphosphate dehydratase encoded by the *HIS3* gene. Adjusting the concentration of 3AT in the selection assay alters the stringency of selection; as the concentration of 3AT increased, the number of transformants that survived the selection decreased. To survive selection with high concentrations of 3AT, a large quantity of enzyme must be present. The amount of enzyme in a cell is related to the transcriptional activation of that particular gene. Therefore, for a transformant to survive selection with high concentrations of 3AT in chemical complementation, the receptor must be able to activate transcription effectively enough for the transformant to produce enough imidazole glycerolphosphate dehydratase to overcome inhibition by 3AT.

A variable stringency selection is desirable when a great number of transformants survive the initial round of chemical complementation. In this case, 1000 transformants grew on solid medium in response to LG335. It was not practical to screen all 1000 transformants for activity to discover the variants with the highest activation and lowest EC_{50}s. Using a more stringent chemical complementation allowed selection of the best six variants in two high-throughput experiments. The variability of this selection system allows the user to adjust the stringency of the selection to meet the needs of the experiment.

Docking

LG335 was docked into each of 18 variants that are activated by LG335. The goal was to compare the energies and structures of the docked complexes to the EC_{50} and efficacy values for each of the 18 variants to create hypotheses about binding and activation properties to explain the activities of these variants and for use in designing future libraries. Each calculation produced between two and five docked structures for each variant. The force field used for this docking (CVFF) is similar to typical docking force fields in that there are no terms to calculate changes in desolvation energy or loss of conformational entropy for the ligand. Since the ligand is constant for this set of docking calculations, these values are identical for each set of calculations and therefore it is justifiable to directly compare the results. However, we must assume that the solvation and conformational entropy of the protein does not change significantly among variants.

A problem that was not anticipated arose during the evaluation of the docking results. Because the protein is "mutated", the energy of the protein changes based on the mutations made. Although this energy change is small compared to the overall energy of the total protein, enough uncertainty is introduced between variants that when the interaction energy between the ligand and the protein is calculated, the uncertainty is greater than the final interaction energy values. Therefore, although the energies of the docked structures of the same variant can be compared to determine the best position of the ligand, it is not appropriate to compare the energies of different variants. To overcome this problem free energy perturbation (FEP) could be used to transform wild-type RXR into each of the variants followed by docking calculations. The FEP calculations would give an energy difference between wild-type RXR and each variant

that could be used to normalize the docking calculation energies. This solution exceeds the scope of this study.

The docked structure with the lowest total energy correlates to the expected docked structure in 17 of the 18 variants (Figure 4.5). These docked structures have the carboxylate of LG335 in a similar position to the carboxylate of 9cRA in the crystal structure. In this position, the carboxylate is within 2.5 Å of the amide backbone hydrogen of A327 and the guanidinium group of R316 (Figure 4.5 A and B). Half of the docked structures are also within 2.5 Å of the side chain amide of Q275, which in the crystal structure is facing away from the binding pocket and toward the solution (Figure 4.5 C). Because of the position of Q275, it is reasonable that the side chain of this residue could rotate inward and provide additional stability through a hydrogen bond to the carboxylate of the bound LG335.

The docked structure with the lowest total energy for A271S;A272T;I310S;F313L;L436A did not correlate with the expected docked structure (Figure 4.5C). In the structure with the lowest total energy, the carboxylate is within 2.5 Å of the hydroxide hydrogen of serines 314 and 317 and the n-propyl group is directed toward the back of the pocket. The positively charged end of the pocket is unoccupied.

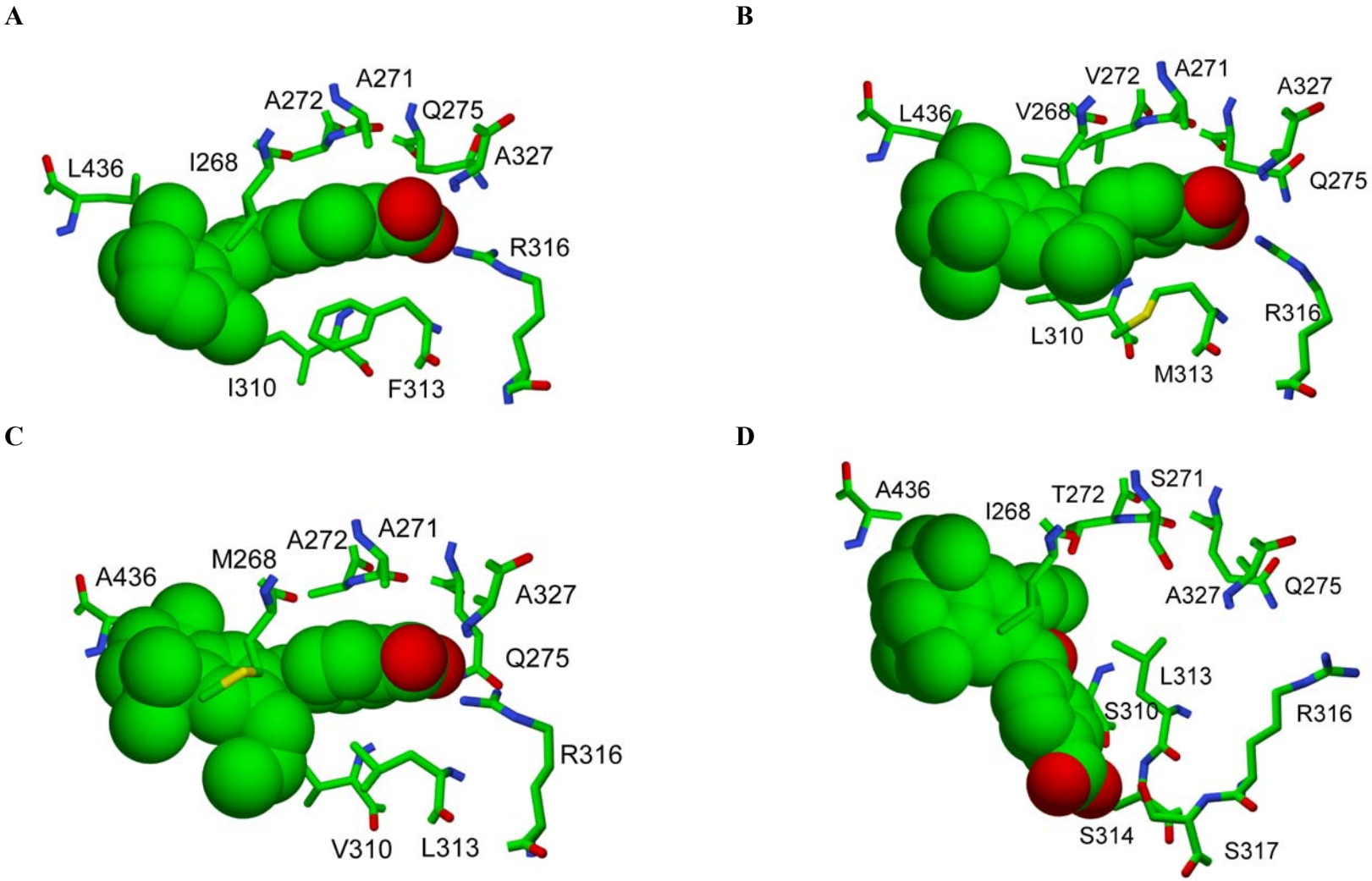


Figure 4.5 (A)Wild type RXR crystal structure with 9cRA and LG335 docked into (B) I268V;A272V;I310L;F313M, (C) I268M;I310V;F313L;L436A, and (D) A271S;A272T;I310S;F313L;L436A.

Table 4.2 LG335 activated RXR variants

	Lib.[Ŧ]	I268	A271	A272	I310	F313	L436	Volume Diff. (Å^3)[†]	EC_{50} (nM)	Eff. (%)
1	b				L	M	M	4	20	190
2	a	V		V	L	M		-12	40	60
3	b	M			V	L	A	80	60	70
4	b	A	S	S	M	L		48	100	100
5	b	V	S		V	V	V	71	200	80
6	b		S	T	S	L	A	78	200	40
7	a	A			A	A	F	165	220	70
8	b	A	S		V	V	V	109	300	140
9	a	A			S	A	F	157	430	50
10	a	V			V	S		92	440	10
11	a	A			S	V	F	120	470	60
12	a	L	V		L	L		-25	530	20
13	a	A		V	A	A	F	127	530	30
14	a	L			M	V		26	610	20
15	a	A			M	A	T	145	610	10
16	a	V	V		L	V		8	650	10
17	a	V		V	M	S	M	31	680	30
18	a	L			V	I		28	2000	10

[Ŧ] Library source for each variant. a indicates library created in Chapter 3 and b indicates library created in Chapter 4.

[†] The difference in volume of the binding pocket in the variant versus the wtRXR binding pocket, calculated by subtracting the volume of the amino acid side chain in the variant from the volume of the with type side chain [1]. Values greater than 0 indicate a larger binding pocket.

Variant Analysis

Table 4.2 contains a summary of the variants activated by LG335. Several trends are observed upon analysis. First, every variant has mutations at positions 310 and 313 indicating that the residues at these positions in wild-type RXR, isoleucine and phenylalanine, must interact poorly with LG335. Based on the docking results we hypothesize that the propyl group of LG335 sterically clashes with the wild-type side chains at these positions.

A second observation invloves positions 271 and 272. For the first library (Chapter 3), mutations at these two positions were restricted to L, V, A, or P. For the current library, the possibilities were expanded to include L, I, V, F, M, A, S, or T. However, other than the wild type amino acid alanine, the only residues observed at these two positions are V, S, and T. We hypothesized that large side chains at 271 or 272 would disrupt access to the positively charged end of the binding pocket and therefore would not lead to functional receptors. Based on the results of these two libraries, we find that this hypothesis is most likely correct.

Finally, of the 18 variants, 13 contain an additional polar amino acid in the binding pocket. The threonine, serine, and methionine residues were included in the library to add possible hydrogen bonding sites. However, these residues do not seem to be participating in hydrogen bonding. It is interesting to note that although these amino acids are not participating in hydrogen bonds, the best ligand-receptor pairs would not have been found if these amino acids had not been included in the library. These residues may be decreasing the hydrophobicity of the binding pocket, thereby making the binding pocket environment more favorable for the less hydrophobic ligand LG335 and less

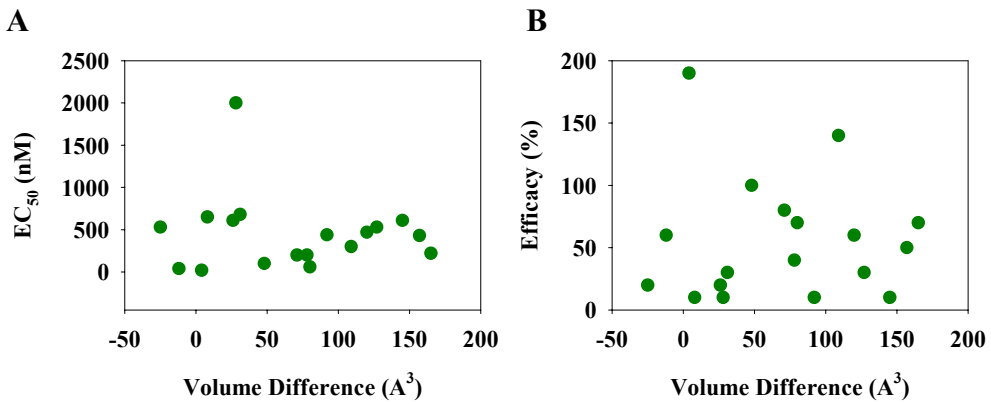


Figure 4.6 No correlation is found between binding pocket volume and (A) EC_{50} or (B) efficacy.

favorable for the more hydrophobic ligand 9cRA.

Table 4.2 contains the changes in binding pocket volume, efficacy, and EC_{50} for each variant activated by LG335. No statistical correlation is evident between binding pocket size change and either EC_{50} or efficacy (Figure 4.6)

Materials and Methods

Compounds

9cRA (molecular weight of 300.4 g/mol) was purchased from BIOMOL (Plymouth Meeting, PA). LG335 [6] was synthesized as described [2] (Chapter 2). 3-Amino-1,2,4-triazole (3-AT) was purchased from Sigma (St. Louis, MO).

Expression Plasmids

pGBDRXR:3Stop [2], pGAD10BAACTR [5, 7], pGBT9Gal4, and pGBDRXRαL-SH-ME [8] have been described.

Randomized Cassette Library Construction

The insert cassette 6mutIC was made as previously described [2] (Chapter 3), but oligo BPrev2cor (5'-GCCCACTCCACTAGTGTGAAAAGCTGTTTGTC (G or C) (A or G) (A, C, or T) (G or C) (A or G) (A, C, or T) GTTGGTGACCGGGTCGTTCG-3') was used instead of BP2. The randomized codons code for L, I, V (2), F, M, A (2), S (2), and T (2).

Transformation

The randomized cassette library was homologously recombined into the pGBDRXRa:3Stop plasmid using to following method. pGBDRXRα:3Stop was first digested with BssHII and *EagI* (NEB), followed by digestion with calf intestinal phosphatase (NEB), to make a vector cassette. 3 μg of this vector cassette and 27 μg of 6mutIC were transformed according to Geitz's Two Hybrid System TRAFO Protocol [3] on a 30X scale into the PJ69-4A yeast strain, which had previously been transformed with a plasmid expressing the coactivator ACTR. 2.5 mL of the transformation mixture was spread on each of twelve large, square (24 cm × 24 cm) plates of SC -Trp-Leu medium. These plates were grown for 4 days at 30 °C.

These twelve plates were then replica plated onto SC -Ade -Trp -Leu and SC -Ade -Trp -Leu plus 10 μM LG335. The plates were grown at 30 °C for 3 days. Colonies that grew on the plates with LG335, but not on plates without ligand were picked by hand and inoculated into wells on 96-well plates containing SC -Trp -Leu liquid medium. The 96-well plates were grown at 30 °C with shaking at 150 rpm overnight. To each well, 50 μL of 75% glycerol in water was added and mixed. The

plates were frozen at -80 °C for storage.

Variable Stringency Selection

The 96-well plates were replicated into SC -Trp -Leu medium using a 96-pin plate replicator (Boekel). These plates were grown at 30 °C with shaking. After 48 hours of growth, these plates were then replicated into 96-well plates containing SC -Trp -Leu, SC -His -Trp -Leu + 3AT (20, 40 or 100 mM) and SC -His -Trp -Leu + 3AT (20, 40 or 100 mM) + 10 μM LG335. These plates were grown at 30 °C with shaking for 24 hours. Each well was evaluated for growth by collecting the optical density at 630 nm with an OpsysMR plate reader (Dynex Technologies, Chantilly, VA). Yeast from wells that had a change in optical density greater than 0.300 were streaked onto solid SC -Trp -Leu medium and grown for 3 days at 30 °C.

Low EC_{50} Selection

The colonies that passed the increased stringency selection with 100 mM 3AT were picked and grown overnight at 30 °C with shaking (300 rpm). These cultures were washed twice with sterile water and diluted to 1.5×10^7 cells/mL. 96-well plates were filled with 80 μL per well of SC -Trp -Leu, SC -His -Trp –Leu + 100 mM 3-AT, SC -His -Trp –Leu + 100 mM 3-AT + 10 μM LG335, and SC -His -Trp –Leu + 100 mM 3-AT + 10 nM LG335. To three wells per plate, 20 μL of each diluted yeast culture was added. These plates were grown for two days at 30 °C with shaking (150 rpm). Growth was monitored at 24 and 48 hours by recording the OD at 630 nm.

Liquid Quantification Assay

The method used for quantification was modified from a method developed by Miller [9]. Briefly, SC -Trp -Leu medium was inoculated with PJ69-4A containing the pGBDRXR plasmid of choice and the pGAD10BAACTR plasmid and grown overnight at 30 °C with shaking. The cultures were centrifuged for 15 minutes and washed twice with sterile water. The cultures were then used to inoculate 80 μL of either SC -Trp -Leu (screen) or SC -Ade -Trp -Leu (selection) medium with increasing concentrations of ligand in 96-well plates. The OD_{630} was taken with an OpsysMR plate reader (Dynex Technologies, Chantilly, VA). These were grown for 1 to 2 days at 30 °C with moderate shaking. For selection experiments, the OD_{630} was recorded after 1 and 2 days. For screens, the OD_{630} was taken after 24 hours. The yeast was pelleted and the supernatant was removed. The cells were lysed with 10 mg mL^{-1} yeast lytic enzyme (MP Biomedicals) and 0.1% Triton X-100 (Sigma). After lysis, the cell debris was pelleted by centrifugation and the supernatant moved to new 96-well plates. ONPG was added to the wells and the OD_{405} was recorded. The plates were incubated at 37 °C for 30 to 60 minutes. The reaction was quenched with sodium carbonate and the OD_{405} was recorded again. Miller units were calculated using the following equation:

$$MU = \frac{1000 \times \Delta OD_{630}}{vt\Delta OD_{405}} \qquad (4.1)$$

where v is the reaction volume and t is the development time.

Genotype Determination

Plasmids were rescued using the Zymoprep Kit (Zymo Research, Orange, CA). The plasmids were then transformed into Z-competent (Zymo Research, Orange, CA) XL1-Blue cells (Stratagene, La Jolla, CA). The QIAprep Spin Miniprep Kit (Qiagen, Valencia, CA) was used to purify the DNA from the transformants. Plasmids were confirmed through sequencing.

Docking

The A chain and the associated crystallographic waters of the RXR-9cRA crystal structure [4] was used for this docking. A 5 Å layer of explicit water was added to the exterior of the protein. All water molecules were removed from the binding pocket. Variants were created using the residue replacement tool in the Biopolymer module of Insight II (Accelrys). After replacement, the side chains (not including the α- or β-carbons) were energy minimized to a maximum derivative of 0.001 using the steepest descents and conjugate gradient methods.

LG335 was superimposed upon the crystallographic 9cRA for use as the starting structure. Affinity (Accelrys) was used for docking LG335 into each of 18 variants. Metropolis Monte Carlo simulation was first implemented to find alternative docked structures. The initial temperature of 500 K was used for the metropolis acceptance filter. If no structure was found, the temperature was increased to 700 K and 1000 K in successive trials. The program was set to find twenty structures and accept the best ten, however, the most structures found was five.

Each of the accepted structures was further optimized using simulated annealing.

Simulated annealing was preformed from 500 K to 300 K in fifty 100 fs intervals. The resulting structures were analyzed for total energy of the system, energy of the ligand, energy of the protein, and interaction energy between the ligand and protein using a BTCL script [10] in Discover (Accelrys).

Literature Cited

1. Doyle, D.F., *Experimental and Theoretical Stability Studies on Cytochrome c*. 1996, University of North Carolina: Chapel Hill.

2. 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.

3. Agatep, R., R.D. Kirkpatrick, D.L. Parchaliuk, R.A. Woods, and R.D. Gietz, *Transformation of Saccharomyces cerevisiae by the lithium acetate/single-stranded carrier DNA/poluethylene glycol (LiAc/ss-DNA/PEG) protocol.*, in *Technical Tips Online*. 1998. p. PO1525.

4. 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.

5. Azizi, B., E.I. Chang, L.J. Schwimmer, and D.F. Doyle, *A sensitive chemical complementation for drug discovery and protein engineering.* Manuscript Submitted.

6. 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.

7. Chen, H., R.J. Lin, R.L. Schiltz, D. Chakravarti, A. Nash, L. Nagy, M.L. Privalsky, Y. Nakatani, and R.M. Evans, *Nuclear receptor coactivator ACTR is a novel histone acetyltransferase and forms a multimeric activation complex with P/CAF and CBP/p300.* Cell, 1997. **90**(3): p. 569-80.

8. 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.

9. Miller, J.H., *Experiments in Molecular Genetics*. 1972, Cold Spring Harbor, NY: Cold Spring Harbor Laboratory.

10. *Affinity*. 1998, San Diego: Molecular Simulations Inc.

CHAPTER 5

DISCOVERING RECEPTORS FOR γ-OXO-1-PYRENEBUTYRIC ACID

Discovering a variety of receptors that are responsive to different small molecules is important for the development of biosensors, gene therapy, and enzyme engineering. For biosensors, arrays of receptors could be used to detect explosives, biotoxins, and chemical weapons. The pattern of activation on the array could precisely identify the agent detected. For gene therapy, receptors that activate transcription in response to different drugs could be used to control expression of multiple genes. The specific dose of each drug could be adjusted based on the patient's needs. Finally, for enzyme engineering, receptors that bind the desired product of a reaction, but not the precursors, could aid in identification of enzymes with novel catalytic abilities.

The ultimate goal of this research is to be able to find receptors for arbitrary ligands. To this end, an approach combining computational screening, library design, and chemical complementation was used to identify receptors for a specific compound, γ-oxo-1-pyrenebutyric acid (OPBA), that is not known to be bound by RXR or any other nuclear receptor.

Results

High Throughput Docking

To identify potential ligands, the software program FlexX (BioSolveIT) [2] was used to screen large databases for compounds that would fit in the binding pocket of RXR. FlexX was designed to automatically dock small molecules into proteins in a high-throughput manner (see Chapter 1 for a full description).

The Developmental Therapeutics Program (DTP) database from the National Cancer Institute was the first target library of potential ligands. This database contains three-dimensional coordinates for 127,000 compounds, most of which are freely available upon request. Biological data is also available for some compounds.

Before docking, the DTP database was filtered to remove non-drug-like compounds. Using the software program Filter (OpenEye Software), compounds were chosen that met the conditions in Table 5.1 and were within the default parameters set for elements and functional groups (see Appendix A). Approximately 23,000 compounds

Table 5.1 Filter conditions

Property	**Min.**	**Max.**
Molecular Weight	100	900
Rings	2	12
Carbons	5	None
Rotatable Bonds	0	8
H-bond Donors	0	10
H-bond Acceptors	2	10
Formal Charge (absolute value)	0	2
Partition Coefficient[t]	0	8
Polar Surface Area	0	140

[t] XlogP value calculated as the octanol/water partition [1].

met these criteria.

Each of the 23,000 remaining compounds was docked into the binding pocket of the RXR LBD crystal structure [3], in which position 268, 271, 272, 310, 313, and 436 were changed to alanines. This is the largest possible binding pocket (229 $Å^3$ larger than the wild-type pocket [4]) for a library with the standard L, I, V, F, M, A, S, and T amino acid possibilities at each of the six positions. Of the 23,000 compounds docked, 16,490 fit in the binding pocket with energy scores ranging from +11 to -25 kcal mol^{-1}. The docking calculation took an average of 30 seconds of CPU time per compound, totaling 8 days of CPU time on one 195 MHz MIPS R10000 processor of an SGI Octane.

The compounds with the 100 lowest scores were evaluated. From this set, OPBA was chosen as the first target compound (Figure 5.1). OPBA is a commercially available

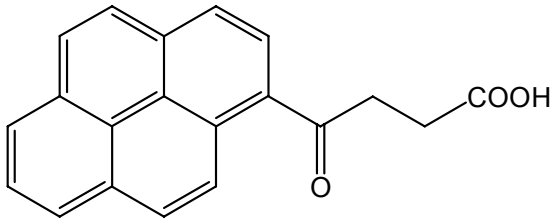


Figure 5.1 γ-oxo-1-pyrenebutyric acid (OPBA).

compound that has a carboxylate analogous to the 9cRA carboxylate. The pyrene ring system makes this compound fluorescent, which should be useful for future binding studies. OPBA is also not known to bind any member of the nuclear receptor family.

Like 9cRA and LG335, OPBA has a negatively charged carboxylate. However, OPBA is a flat molecule with the aromatic pyrene conjugated to the carbonyl, while 9cRA and LG335 are both bent out of plane. The 9cRA-RXR crystal structure [3] shows

that the *cis* bond forces the conjugated alkene chain out of plane with the β-ionone ring (Figure 4.5). For LG335, the docking calculations presented in Chapter 4 show that the two phenyl rings are forced out of plane due to steric clash with the n-propyl group (Figure 4.5). OPBA is also more hydrophobic than both 9cRA and LG335. The logP value for OPBA is 4.24 compared to 6.83 and 8.56 for 9cRA and LG335, respectively (as calculated by LogP DB (Advanced Chemistry Development)).

Library Design and Selection

The library used to find receptors for OPBA is the same library discussed in Chapter 4. In addition to SC -Ade -Trp -Leu plus 10 μM LG335 medium, approximately 110,000 (three 24 cm × 24 cm plates) of the 450,000 transformants were replica-plated onto SC -Ade -Trp -Leu plus 10 μM OPBA medium and grown at 30 °C for 3 days. Forty-seven colonies grew on these three plates. These colonies were picked and stored in 96-well plates at -80 °C in SC -Trp -Leu with 15% glycerol.

High Throughput Screening

For initial screening for activation by OPBA, a subset of 167 transformants were selected from the 1000 LG335 activated transformants (Chapter 4) and the 47 transformants that grew on the SC -Ade -Trp -Leu plus 10 μM OPBA medium (see above) for initial screening for activation by OPBA. This is 17% of the transformants from this library that grew on solid selective media. These transformants were inoculated into 96-well plates containing SC -Ade -Trp -Leu medium with 10 μM OPBA. Yeast growth in each well was monitored by change in optical density at 630 nm between 0 and

39 hours.

Based on the results of the initial high-throughput screen, ten transformants were chosen for additional screening for OPBA activation. These ten transformants had an ΔOD_{630} of greater than 0.300 in SC -Ade -Trp -Leu medium with 10 µM OPBA, but did not show significant growth in SC -Ade -Trp -Leu medium. Of the ten transformants, four were originally from LG335 selective plates and six were from OPBA selective plates. These transformants were grown in SC -Ade -Trp -Leu liquid medium without ligand and with 0.1, 1, and 10 µM OPBA at 30 °C. After 39 hours of growth three transformants proved to be constitutively active. The other seven transformants proved to be activated by OPBA, two of which grow with 1 µM OPBA (Figure 5.2). The plasmids from these two transformants were rescued, sequenced, and retransformed to ensure plasmid-linked phenotypes.

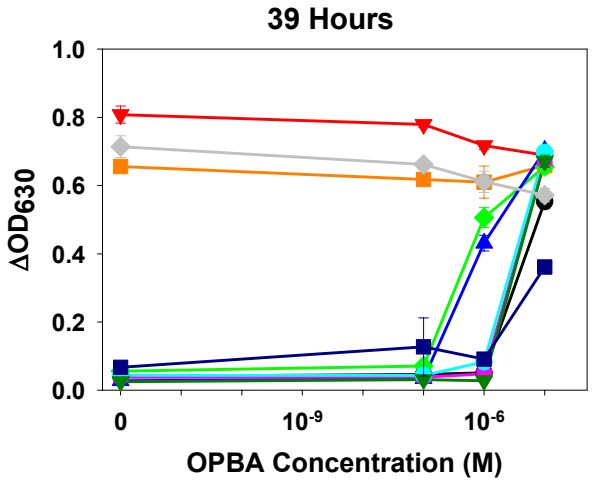


Figure 5.2 Activation of 10 RXR variants by OPBA

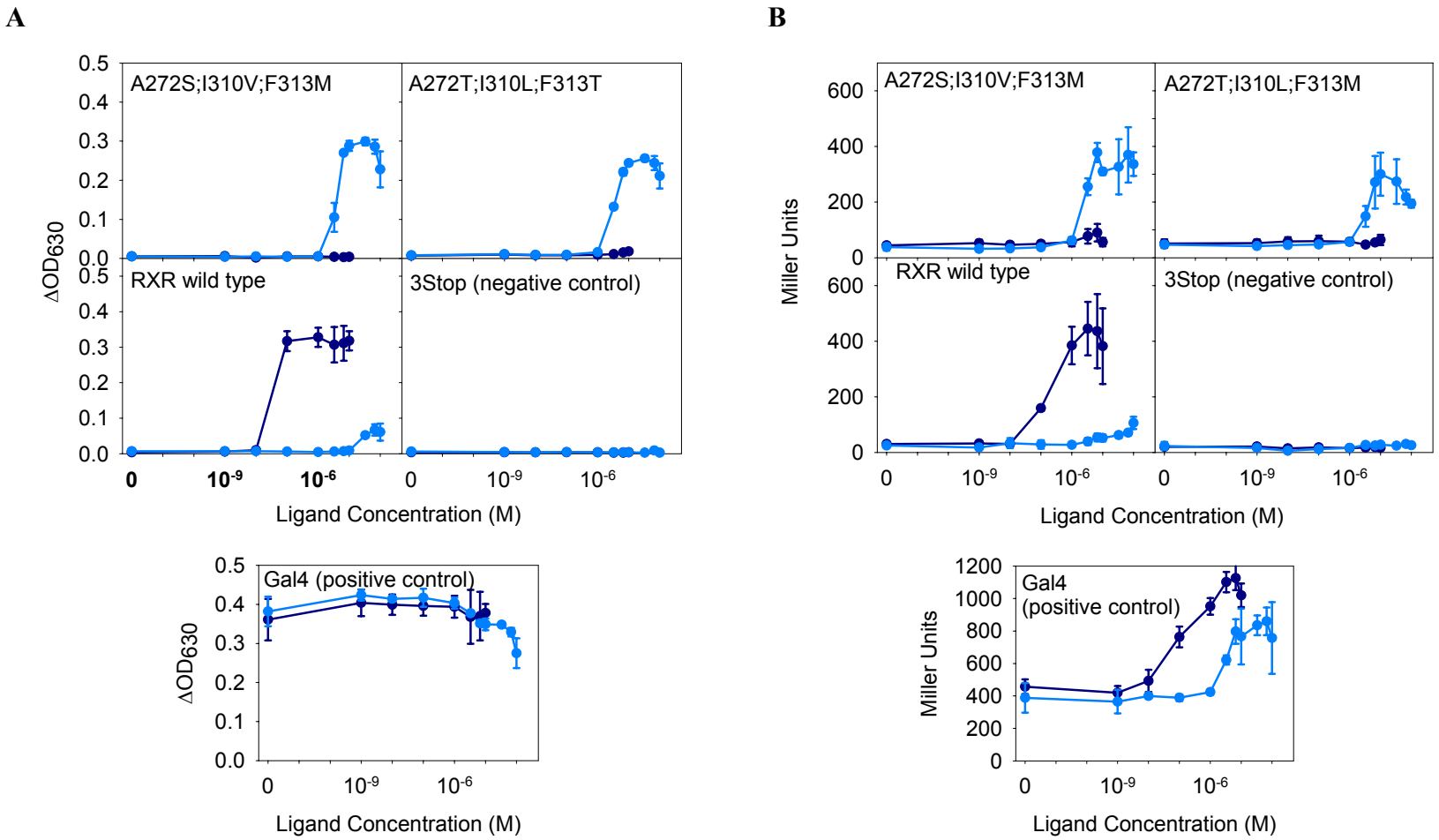


Figure 5.3 Selection (A) and screening (B) of RXR variants A272S;I310V;F313M and A272T;I310L;F313T. The variants were grown in SC -Ade -Trp -Leu (A) or SC -Trp -Leu (B) medium with increasing concentrations of 9cRA (dark blue) or OPBA (light blue).

Table 5.2 EC_{50} and Efficacy for RXR Variants

Variant	9cRA		OPBA	
	EC_{50}[§] (μM)	Eff[**] (%)	EC_{50} (μM)	Eff (%)
WT	0.1	100	70	20
A272S;I310V;F313M	>100	10	2	100
A272T;I310L;F313T	>100	0	3	80

[§] EC_{50}, concentration of ligand producing half maximal activity. Values represent averages of two screen experiments in quadruplicate.

[**] Eff., efficacy, maximum increase in activation relative to the increase in activation of wild type with 10 μM 9cRA. Values represent averages of two screen experiments in quadruplicate

Variant Characterization

Chemical complementation selection (adenine selection) and screening (β-galactosidase assay) were used to characterize the activation profiles of these RXR variants (Figures 5.3 and Table 5.2). Wild-type RXR and Gal4, a ligand independentyeast transcription factor, were used as controls. Both variants A272S;I310V;F313M and A272T;I310L;F313T are activated by OPBA, but not significantly activated by 9cRA.

Chemical complementation selection (adenine selection) and screening (β-galactosidase assay) were used to characterize the activation profiles of these RXR variants (Figures 5.3 and Table 5.2). Wild-type RXR and Gal4, a ligand independent yeast transcription factor, were used as controls. Both variants A272S;I310V;F313M and A272T;I310L;F313T are activated by OPBA, but not significantly activated by 9cRA.

Docking

OPBA was computationally docked into the variants using RXR LBD-9cRA crystal structure [3] and the InsightII module Affinity. The two variants were created *in silico* with the Biopolymer module's residue replacement feature. The side chains (past the β-carbon) of the replaced residues were energy minimized. Docking was accomplished using Monte Carlo simulation. Simulated annealing was performed on the resulting structures to optimize the docked ligand placements.

Discussion

From a library of 450,000 RXR variants, two variants were discovered that are able to bind and be activated by OPBA, a compound not previously known to activate any nuclear receptor. The combination of chemical complementation selection and screening assays on both solid and liquid media proved to be an effective method for finding active receptors by zeroing in on the most active and sensitive receptors.

Both variants A272S;I310V;F313M and A272T;I310L;F313T have wild-type like efficacies in response to OPBA (Table 5.2 and Figure 5.3). However, the EC_{50}s of the variants are 20 and 30 fold greater than the EC_{50} of wild-type RXR with 9cRA. While EC_{50}s greater than wild-type are not ideal, these variants still represent an advance in engineering receptors to bind new small molecules. The hydrophobic pyrene group is readily bound in cell membranes [5] and fatty acid derivatives including 1-pyrenebutyric acid are often used as fluorescent membrane probes. Therefore, finding receptors that are more sensitive to OPBA may be challenging.

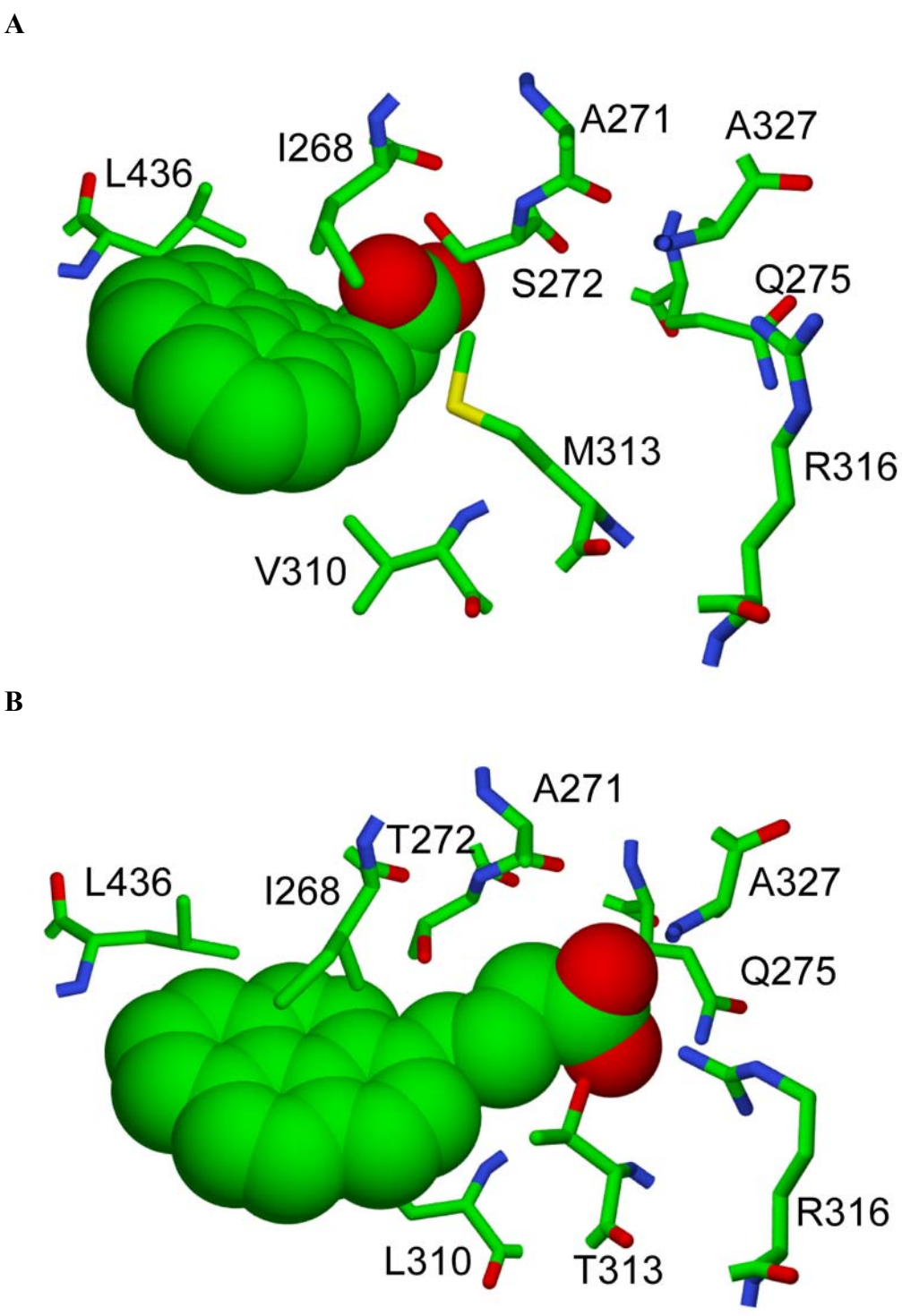


Figure 5.4 Structures generated by docking calculations for (A) A272S;I310V;F313M and (B) A272T:I301L;F313T.

The docking calculations indicate possible binding modes for OPBA in the two RXR variants (Figure 5.4). The butyl chain of OPBA is 3 carbons shorter that 9cRA between the methyl group on carbon 9 and the carboxylate. A large amino acid side chain at position 313 could sterically clash with the pyrene moiety of OPBA. For A272S;I310V;F313M, the methionine at position 313 may be having this effect. However for this variant, the serine at position 272 may interact with the OPBA carboxylate in an alternate binding mode as shown in Figure 5.4A. For A272T;I310L;F313T, a small amino acid at position 313 seems to allow enough space for the pyrene group to position the OPBA carboxylate close to R316, Q275, and the backbone hydrogen of A327 (Figure 5.4B). The hydroxyl group of T313 may also be involved in stabilization of the OPBA carboxylate.

High-throughput docking calculations to identify potential ligands were effective for selecting a compound that eventually became a ligand. This method could be useful for selecting near-drugs with desirable ADMET properties when developing ligand-receptor pairs for gene therapy. Compounds that do not interfere with other biological processes and that can reach the desired site of action could be compiled into a database. Compounds from this database could then be selected as potential ligands using *in silico* screening based on their compatibility with the library of receptors. Chemical complementation could then be used to select the receptor that best fits the ligand.

Materials and Methods

Compounds

9cRA was purchased from BIOMOL (Plymouth Meeting, PA). LG335 [6] was synthesized as described [7] (Chapter 2). OPBA was purchased from Sigma (St. Louis, MO).

Expression Plasmids

pGBDRXR:3Stop [7], pGAD10BAACTR [8, 9], pGBT9Gal4, and pGBDRXRαL-SH-ME [10] have been described.

Randomized Cassette Library Construction, Transformation, and Genotype Determination

The randomized insert cassette was constructed, transformed into PJ69-4a, and genotyped as described in Chapter 4.

High-Throughput Selection

A set of 164 transformants and yeast containing the plasmids pGBDRXRαL-SH-ME and pGBDT9Gal4 (all with pGAD10BAACTR) were grown in wells of 96-well plates with 200 μL of SC -Trp -Leu medium for 22 hours at 30 °C with shaking. From these plates 20 μL of culture from each well was used to inoculate 96-well plates containing 180 μL of SC -Ade -Trp -Leu medium with 10 μM of 9cRA, LG335, or OPBA or SC -Trp -Leu medium. These plates were incubated at 30 °C for 39 hours. The change in optical density at 630 nm was monitored at 18, 24, and 39 hours.

Dose Response Quantification

Ten transformants were selected from the high-throughput selection assay for further analysis. SC -Trp -Leu medium was inoculated with each transformant and incubated at 30 °C with shaking. The yeast were washed twice with sterile water and 20 µL was used to inoculate 180 µL of SC -Ade -Trp -Leu medium in 96-well plates with 0.1, 1, and 10 µM OPBA or without ligand in triplicate. These plates were grown at 30 °C with shaking (150 rpm). Optical density at 630 nm was used to monitor the yeast growth over 39 hours.

Liquid Quantification Assay

The protocol for the liquid quantification selection and screen assays is described in Chapter 4.

Docking Calculations

All calculations were run on an SGI Octane MXI with dual 195 MHz MIPS R10000 processors and 1 GB of RAM.

FlexX

The 3-D DTP database was obtained from the National Cancer Institute (dtp.nci.nih.gov). Drug-like compounds were selected using Filter (OpenEye Scientific Software). Compounds were selected if they met the criteria in Table 5.1 and the program's default elemental and functional group criteria (Appendix A). The structures were then separated into individual .mol2 files using InsightII (Accelrys).

The RXR LBD-9cRA crystal structure [3] was modified at positions 268, 310, 313, and 436 to alanine using the residue replace function of the InsightII module Biopolymer. The protein's coordinates were saved in .pdb format and the 9cRA structure was saved in .mol2 format. FlexX was used to define the binding pocket atoms and surface using the coordinates of the carboxylate carbon as the probe location. The 23,000 structures were docked using the dock_list batch file available in the FlexX user guide. See Chapter 1 for a complete description.

Affinity

Docking of OPBA into modified binding pockets was done using the InsightII module Affinity®. The wild type RXR with 9cRA crystal structure [3] was modified using the Biopolymer module residue replace tool to make mutations in the binding pocket that corresponded to the mutations in variants A272S;I310V;F313M and A272T;I310l;F313T. The ligand was placed in the binding pocket by superimposing the OPBA onto 9cRA in the crystal structure. A Monte Carlo simulation with an energy range acceptance of 10 kcal mol^{-1} was performed first, followed by Simulated Annealing of the best docked conformations. See Chapter 1 for a more complete description of the method.

Literature Cited

1. Wang, R.X., Y. Fu, and L.H. Lai, *A new atom-additive method for calculating partition coefficients.* Journal of Chemical Information and Computer Sciences, 1997. **37**(3): p. 615-621.

2. Rarey, M., B. Kramer, T. Lengauer, and G. Klebe, *A fast flexible docking method using an incremental construction algorithm.* J Mol Biol, 1996. **261**(3): p. 470-89.

3. 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.

4. Doyle, D.F., *Experimental and Theoretical Stability Studies on Cytochrome c*. 1996, University of North Carolina: Chapel Hill.

5. Martins, J. and E. Melo, *Molecular mechanism of lateral diffusion of py(10)-PC and free pyrene in fluid DMPC bilayers.* Biophys J, 2001. **80**(2): p. 832-40.

6. 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.

7. 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.

8. Chen, H., R.J. Lin, R.L. Schiltz, D. Chakravarti, A. Nash, L. Nagy, M.L. Privalsky, Y. Nakatani, and R.M. Evans, *Nuclear receptor coactivator ACTR is a novel histone acetyltransferase and forms a multimeric activation complex with P/CAF and CBP/p300.* Cell, 1997. **90**(3): p. 569-80.

9. Azizi, B., E.I. Chang, L.J. Schwimmer, and D.F. Doyle, *A sensitive chemical complementation for drug discovery and protein engineering.* Manuscript Submitted.

10. 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 6

QUANTIFICATION OF INTRACELLULAR LIGAND CONCENTRATIONS

Yeast can serve as a good model system for mammalian cells. Yeast are easy and inexpensive to culture and there are multiple selective markers and reporter genes available. Plasmids can also be easily transformed into and rescued from yeast. However, unlike mammalian cells, yeast have a cell wall, which, in conjunction with a cell membrane enriched with the steroid ergosterol decrease the penetrability of small, drug-like molecules.

Several attempts have been made to increase the permeability of yeast. Disruption of two endogenous genes has been used to influence intracellular ligand concentration. First, yeast express ATP-binding cassette (ABC) transporters. A member of this class, Pdr5p, is known to selectively transport steroid hormones out of yeast, thereby decreasing the intracellular hormone concentration [1]. When Pdr5p is knocked-out or inhibited by FK506 [2], the concentration of hormone within the yeast cells is consequently increased. Second, yeast strains have been engineered with an *ERG6* knockout. *ERG6* is an essential gene in the ergosterol biosynthetic pathway and without this gene, the permeability of the yeast cell membrane by small molecules is enhanced [3].

Small molecules and peptides like amphotericin B (AmB) and polymixin B

nonapeptide (PBN) have also been shown to increase the permeability of yeast to a variety of antibiotics including rifampicin. AmB is hypothesized to interact with ergosterol in the yeast cell membrane to form a cage [4]. These cages disorder the cell membrane and may create pores, thereby increasing permeability of the cell membrane to cationic antibiotics. PBN, a circular peptide, induces a similar increased sensitivity to some antibiotics [5]. However, not all of the antibiotics tested by Boguslawski had increased efficacy, indicating that the effect of this peptide is not a general increase in membrane permeability.

The EC_{50}s of the ligand-receptor pairs discussed in this thesis are expressed in terms of the ligand concentration in the media. Because of permeability concerns with the yeast, a method was developed to determine the intracellular ligand concentration. This method uses LC-MS/MS to determine the concentration of ligand in cell lysates. When compared to a standard curve the intracellular concentration can be determined.

The yeast were grown with and without 10 μM LG335 overnight (see Materials and Methods for details). The cells were washed, then lysed with boiling MeOH and sonicated. The cell debris was removed by centrifugation. The supernatant and the final cell wash were lyophilized, and the residue was rehydrated in fifty times the yeast volume of 95:5 H_2O:MeOH with 1% (v/v) formic acid. High-performance liquid chromatography (HPLC) was coupled by electrospray ionization with multiple reaction monitoring (MRM) using a triple quadrupole mass spectrometer for detection. The first quadrupole allows only ions with a mass of 379 amu (M+H) to pass. The second quadrupole is used as a fragmentation chamber. The third quadrupole is used to select ions with a mass of 149 amu (the most abundant fragment) of LG335. Thus, only

compounds with a specific molecular weight that fragment to produce a specific product ion are detected. Standards from 100 nM to 10 μM LG335 were used to generate a standard curve (Figure 6.1). Samples (cell lysates and the final wash) and standards were injected in triplicate and the experiment was duplicated. The limit of detection of LG335 using this technique is 100 nM. Peaks are detected if they are greater than ten counts above baseline and the area under the curve (AUC) is greater than five. Below 100 nM LG335, not peak was detected. The average AUC of three injections was used to calculate a concentration, which was then multiplied by 50 to reflect the actual intracellular concentration. The intracellular concentration of LG335 was found to be 20 and 40 μM in the two trials, respectively. No ligand was detected in the final wash, signifying that the LG335 detected is from inside the yeast cells. The results of these experiments show that the intracellular concentration of LG335 is within four-fold of the ligand concentration in the media.

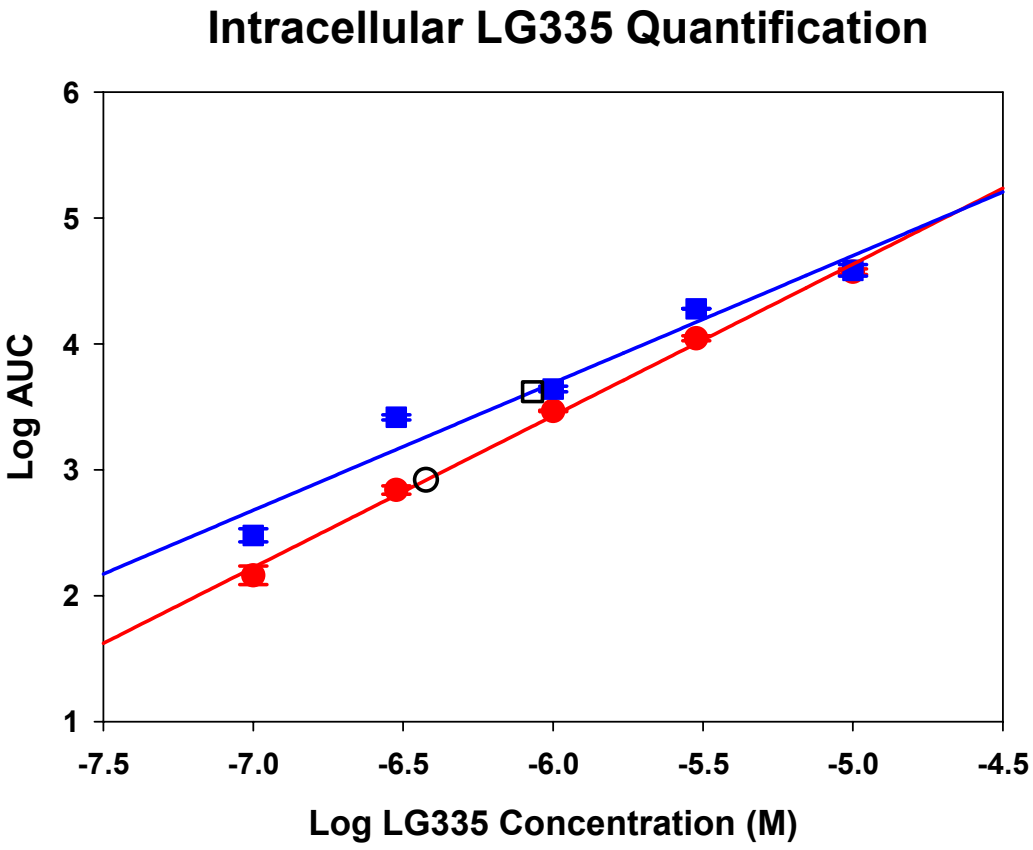


Figure 6.1 LG335 quantification standard curve. Trial 1 (red circles) and Trial 2 (blue squares) were fitted with linear regression lines (red and blue respectively). Trail 1 sample (open circle) and Trial 2 sample (closed square).

Materials and Methods

Lysate Preparation Method

Cultures of PJ69-4A were grown in 5 mL of synthetic complete (SC) media with and without 10 μM LG335 for 24 hours at 30º with shaking. The cultures were then centrifuged and the media was removed. The cells were washed four times with 1:10 TE. The number of yeast cells was determined by measuring optical density at 600 nm with a BioPhotometer (Eppendorf) after resuspension in 1:10 TE for the final time. The conversion factor of 3×10^6 cells per 0.1 OD_{600} was used to determine the total number of cells. The total cell volume was determined by using the average cell volume of 70 μm^3 [6]. The final wash was saved. The cells were then lysed by boiling in MeOH for 10 minutes followed by 10 minutes of sonication [7, 8]. The cell debris was removed by centrifugation at 3500 g for 10 minutes. The supernatant and the final 1:10 TE washes were evaporated by vacuum centrifugation. Each sample and the corresponding washes were resuspended in 50 times the total cell volume.

Liquid Chromatography Method

High-performace liquid chromatography was performed on an Agilent 1100 HPLC with binary pumps. The separations were done on a 5 cm × 1 mm Discovery® HS F5 3 μm column (Supelco). Solvent A was 95:5 H_2O:MeOH with 1% CHOOH and solvent B was 5:95 H_2O:MeOH with 1% CHOOH . The gradient consisted of 2 minutes with 100% solvent A, an 8 min gradient to 100% solvent B, 5 minutes at 100% solvent B, and reequilibration to 100% solvent A for 5 minutes with a flow rate of 150 μL per

minute. The samples were injected via autosampler in 50 μL volumes.

Mass spectrometry method

The HPLC was coupled to a Quattro LC (Micromass) triple quadrupole mass spectrometer. The mass spectrometer was operated with an electrospray ionization source in positive mode (ESI^+) using multiple reaction monitoring (MRM). The transition monitored for LG335 was 379 → 149 amu. The desolvation and ion source temperatures were 250 ºC and 100 ºC respectively with a cone voltage of 25.9 eV. Fragmentation occurred in the second quadrupole with collision energy of 35 eV and 2.9 $\times 10^{-4}$ mBar of argon.

Literature Cited

1. Kralli, A., S.P. Bohen, and K.R. Yamamoto, *LEM1, an ATP-binding-cassette transporter, selectively modulates the biological potency of steroid hormones.* Proc Natl Acad Sci U S A, 1995. **92**(10): p. 4701-5.

2. Kralli, A. and K.R. Yamamoto, *An FK506-sensitive transporter selectively decreases intracellular levels and potency of steroid hormones.* J Biol Chem, 1996. **271**(29): p. 17152-6.

3. Clark, D.D. and B.R. Peterson, *Analysis of protein tyrosine kinase inhibitors in recombinant yeast lacking the ERG6 gene.* Chembiochem, 2003. **4**(1): p. 101-7.

4. Brajtburg, J., W.G. Powderly, G.S. Kobayashi, and G. Medoff, *Amphotericin B: current understanding of mechanisms of action.* Antimicrob Agents Chemother, 1990. **34**(2): p. 183-8.

5. Boguslawski, G., *Effects of polymyxin B sulfate and polymyxin B nonapeptide on growth and permeability of the yeast Saccharomyces cerevisiae.* Mol Gen Genet, 1985. **199**(3): p. 401-5.

6. Sherman, F., *Getting Started with Yeast*, in *Guide to Yeast Genetics and Molecular Biology*, C. Guthrie and G.R. Fink, Editors. 1991, Academic Press, Inc.: San Diego. p. 17.

7. Morozov, V.N., T.Y. Morozova, K.L. Johnson, and S. Naylor, *Parallel determination of multiple protein metabolite interactions using cell extract, protein microarrays and mass spectrometric detection.* Rapid Commun Mass Spectrom, 2003. **17**(21): p. 2430-8.

8. Gonzalez, B., J. Francois, and M. Renaud, *A rapid and reliable method for metabolite extraction in yeast using boiling buffered ethanol.* Yeast, 1997. **13**(14): p. 1347-55.

CHAPTER 7

CONCLUSIONS

The goal of this research was to develop a strategy for engineering receptors that bind arbitrary small molecules. Using a combination of molecular modeling, library design, and chemical complementation, receptors that bind and are selectively activated by two new compounds, LG335 and OPBA, were created and discovered. In the course of this research several techniques and methods were developed including a new synthesis route for LG335, ligand-protein docking protocols for two software programs, protein library design and construction, and high-throughput and variable stringency chemical complementation.

Ligand binding to RXR causes specific conformational changes that allow recruitment of coactivators and the transcription machinery. The interactions in the binding pocket between the ligand and protein atoms facilitate the conformational change and therefore, must be precise. Because the ligand-protein interactions required for the correct conformational change are not yet understood, direct design of receptors that are activated by specific ligands is extremely difficult. Creating a library of proteins with mutations at specific positions in the binding pocket, and using chemical complementation to select active variants allows the protein to determine the best fit. Libraries were designed based on structural and functional knowledge of RXR. By

randomizing hydrophobic amino acids in the binding pocket, the shape and size of this pocket were altered allowing for selective binding of new ligands. This process was used to find receptors for two ligands, LG335 and OPBA.

LG335 is similar in structure the known RXR agonist bexarotene (Targretin®). The synthetic yield of LG335 was improved to 8% from 4% in original publication by altering the synthetic route [1, 2]. The inclusion of polar amino acids in the library proved essential for discovering receptors with high efficacy and low EC_{50}s for LG335. This may be due to an alteration of the hydrophobicity of the binding pocket, which makes binding of the less hydrophobic molecule LG335 more favorable than the more hydrophobic wild-type ligand 9cRA.

OPBA was chosen as a potential ligand using high-throughput docking. Two receptors for this ligand were discovered. These receptors have wild-type-like efficacies, but the EC_{50}s are 20 to 30 fold greater than wild-type RXR with 9cRA. Both receptors have mutations to polar amino acids in the center of the binding pocket. These amino acids may be required to stabilize the carboxylate because the butyl chain may not be long enough to bring the carboxylate in contact with the positively charged end of the pocket.

Chemical complementation [3, 4] is a versatile method that can be used both for genetic selection and as a functional screen. The genetic selection capability of chemical complementation on solid media can be used for initial selection of functional variants. Receptors among the surviving pool that have desired properties can then be analyzed using chemical complementation in liquid media. Variable stringency chemical complementation uses the histidine selective marker and 3AT to tune the activation

threshold level required for yeast growth in response to ligand. This method allows targeting of variants with high efficacies. Decreasing the concentration of ligand present in selective media allows targeting of variants with low EC_{50}s. High-throughput chemical complementation can be used to identify variants that are activated by specific ligands. Replicating 96-well plates in liquid medium containing different ligands can quickly differentiate receptors that show selectivity in activation versus variants that are promiscuous (activated by multiple ligands) or constitutively active.

As a functional selection and screen, chemical complementation can be used to determine efficacy and EC_{50}. Chemical complementation in selective liquid medium responds as a switch; once transcription of the selective marker (*HIS3* or *ADE2*) reaches the growth threshold, full growth occurs. A better reporter of transcription, and therefore a better indicator of efficacy and EC_{50}, is the β-galactosidase assay. This assay measures the activity of β-galactosidase, which correlates to the transcription level of the *LacZ* gene.

When characterizing individual RXR variants, EC_{50}s and efficacies are calculated according to the concentration of ligand in the medium. These values are only correct if the intracellular ligand concentration is equal to the concentration of ligand in the medium. An LC-MS/MS method was developed to quantify intracellular concentrations. The intracellular concentration of LG335 was determined to be within four fold of the LG335 concentration in the medium.

In silico ligand-protein docking is a useful tool for generating hypotheses concerning the placement and interaction of specific ligands with specific RXR variants. However, due to insufficiencies in the force field and docking method, the energies

between different receptor variants and/or ligands do not correlate with experimental data.

Future work

During the course of this research nine constitutively active RXR variants were sequenced that had base insertions or deletions resulting in shifted reading frames. The frame shift for these nine variants results in truncation of the protein due to an in-frame stop codon at position 319. These proteins are terminated between helix 5 and sheet 1 with a protein that is 143 amino acids shorter than the full RXR LBD. Since this new protein does not contain a complete ligand-binding pocket or the activation function domain 2 (AF2) normally necessary for RXR to be active, the activity of these truncated proteins is unexpected. It is possible that this new protein is able to interact with the yeast transcription machinery independently of ligand or ACTR. To test this hypothesis, each of the nine variants could be transformed into PJ69-4A without pGAD10BAACTR. These variants could also be tested in mammalian cell culture to determine if the constitutive activity of these variants is specific to yeast. A subset of these proteins should also be expressed, purified, and the molecular weight and amino acid sequence determined to ensure that these proteins are truly truncated.

A variant (I268A;A271V;G304V;F313M;Δ13) in an early library also had base deletions that resulted in a frame shift termed Δ13. A one base deletion followed by 37

```
RXRwt -     HLFFFKLIGDTPIDT
RXRΔ13 -    HTSSSSSSGTHQL T
```

Figure 7.1 Sequence changes between wild-type RXR and I268A;A272V;G304V;F313M;Δ13

bases and a two base deletion results in 13 amino acid changes (Figure 7.1). These changes occur from helix 11 to helix 12 preceding the AF2 domain, which remains intact. This variant is activated by LG335 with an EC_{50} of approximately 200 nM (Figure 7.2).

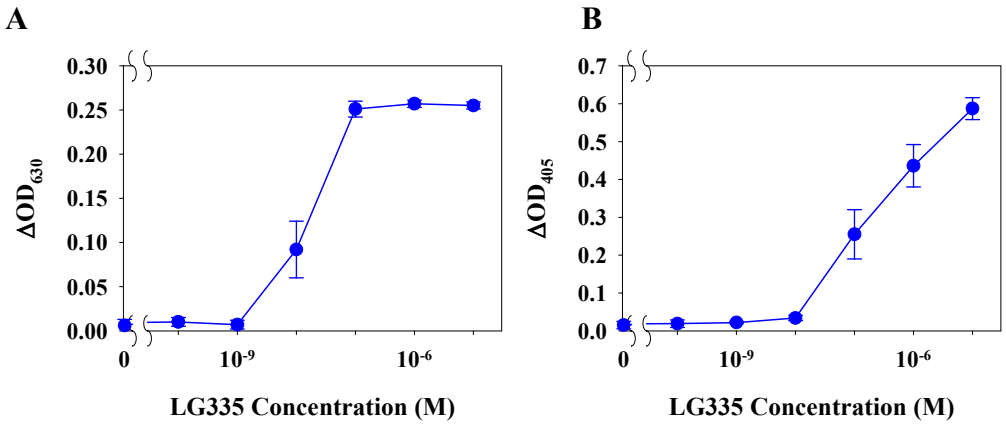


Figure 7.2 A) Selection and B) screening experiments for frame shift RXR variant I268A;A271V;G304V;F313M;Δ13

Further investigation of the mutations in this variant is warranted. First, the Δ13 set of mutations could be incorporated into wild-type RXR independent of the other mutations to determine if these changes have any effect of the activation of wild-type RXR with 9cRA and/or LG335. The G304V mutation may also be interesting. The glycine at position 304 causes a break between helices 4 and 5. A valine at this position may stiffen the connection between helix 4 and 5 resulting in a slight repositioning of helix 5. Helix 5 is one of the helices that define the RXR binding pocket and a change in the position of this helix may affect ligand binding.

The overall yield of the new synthesis route for LG335 could be improved by optimizing the final deprotection step. This reaction has a previously published yield of 93% [1]. If a yield of 93% could be obtained for this final reaction, the overall yield would improve from 8% to 20%.

Literature Cited

1. 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.

2. 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.

3. 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.

4. Azizi, B., E.I. Chang, L.J. Schwimmer, and D.F. Doyle, *A sensitive chemical complementation for drug discovery and protein engineering.* Manuscript Submitted.

APPENDIX A

FILTER DEFAULT PARAMETERS

```
#This file defines the rules for filtering multi-structure files based on
#properties and substructure patterns.
#NOTE: this rule file is suitable for drug-like filtering and may be
#       too restrictive for reagent filtering.
#These rules include the contributions of countless chemists & modellers including:
#Rule of 5:
#       Lipinski, C, et al., Adv.Drug Deliv. Rev., 23:3, 1997.
#General Filtering:
#       Oprea, T., JCAMD 14:251-264, 2000.
#XLogP:
#       Wang, R, Ying, Fu, & Lai, Luhua, J.Chem. Inf. Comput. Sci., 37:615-621, 1997.
#Polar Surface Area
#       Peter Ertl, Bernhard Rohde, & Paul Slzer, J. Med. Chem. 43:3714-3717, 2000.
#       David E. Clark, Journal of Pharmaceutical Sciences, 88(8):807-814, 1999.

MIN_MOLWT                   100     "Minimum molecular weight"
MAX_MOLWT                   900     "Maximum molecular weight"

MIN_SSSR_RINGS              2       "Minumum number of SSSR rings"
MAX_SSSR_RINGS              7       "Maximum number of SSSR rings"

MAX_RING_SIZE               12      "Maximum size of any SSSR ring"

MIN_CARBONS                 5       "Minimum number of carbons"
MIN_HETEROATOMS             0       "Minimum number of heteroatoms"
MIN_Het_C_Ratio             0       "Minimum heteroatom to carbon ratio"
MAX_Het_C_Ratio             1.0     "Maximum heteroatom to carbon ratio"


#count ring degrees of freedom = (#BondsInRing) - 4 - (RigidBondsInRing) -
(BondsSharedWithOtherRings)
#must be >= 0, from JCAMD 14:251-265,2000.
ADJUST_ROT_FOR_RING true                    "BOOLEAN for weather to estimate
degrees of freedom in rings"

MIN_ROT_BONDS               0       "Minimum number of rotatable bonds"
```

```
MAX_ROT_BONDS                   8       "Maximum number of rotatable bonds"

MIN_RIGID_BONDS                 8       "Minimum number of rigid bonds"
MAX_RIGID_BONDS                 50      "Maximum number of rigid bonds"

MIN_HBOND_DONORS                0       "Minimum number of hydrogen-bond
donors"
MAX_HBOND_DONORS                10      "Minimum number of hydrogen-bond
donors"

MIN_HBOND_ACCEPTORS             2       "Minimum number of hydrogen-bond
acceptors"
MAX_HBOND_ACCEPTORS             10      "Minimum number of hydrogen-bond
acceptors"

MIN_COUNT_FORMAL_CRG            0       "Minimum number formal charges"
MAX_COUNT_FORMAL_CRG            3       "Maximum number of formal charges"
MIN_SUM_FORMAL_CRG              0       "Minimum sum of formal charges"
MAX_SUM_FORMAL_CRG              2       "Maximum sum of formal charges"

MIN_XLOGP                       0.0     "Minimum XLogP"
MAX_XLOGP                       8.0     "Maximum XLogP"

MIN_2D_PSA                      0.0     "Minimum 2-Dimensional (SMILES) Polar
Surface Area"
MAX_2D_PSA                      140.0   "Maximum 2-Dimensional (SMILES) Polar
Surface Area"

ALLOWED_ELEMENTS    H,C,N,O,F,S,P,Cl,Br,I

#acceptable molecules must have <= instances of each of the patterns below

#specific undesirable functional groups
RULE 0      Carbazides
RULE 0      Acid_anhydrides
RULE 0      Pentafluorophenyl_esters
RULE 0      Paranitrophenyl_esters
RULE 0      HOBT_esters
RULE 0      Triflates
RULE 0      Lawesson_s_reagent
RULE 0      Phosphoramides
RULE 0      Aromatic_azides
RULE 0      Beta_carbonyl_quart_nitrogen
RULE 0      Acylhydrazide
RULE 0      Quarternary_C_Cl_I_P_or_S
RULE 0      Phosphoranes
```

```
RULE 0      Chloramidines
RULE 0      Nitroso
RULE 0      P_S_Halides
RULE 0      Carbodiimide
RULE 0      Isonitrile
RULE 0      Triacyloxime
RULE 0      Cyanohydrins
RULE 0      Acyl_cyanides
RULE 0      Sulfonyl_cyanides
RULE 0      Cyanophosphonates
RULE 0      Azocyanamides
RULE 0      Azoalkanals
RULE 0      Polyenes
RULE 0      Saponin_derivatives
RULE 0      Cytochalasin_derivatives
RULE 0      Cycloheximide_derivatives
RULE 0      Monensin_derivatives
RULE 0      Cyanidin_derivatives
RULE 0      Squalestatin_derivatives

#functional groups which often eliminate compounds from consideration
RULE 0      acid_halide
RULE 0      aldehyde
RULE 0      alkyl_halide
RULE 0      anhydride
RULE 0      azide
RULE 0      azo
RULE 0      di_peptide
RULE 0      long_aliphatic_chain                //(>7 atoms)
RULE 0      michael_acceptor
RULE 0      beta_halo_carbonyl
RULE 0      nitro
RULE 0      peroxide
RULE 0      phosphonic_acid
RULE 0      phosphonic_ester
RULE 0      phosphoric_acid
RULE 0      phosphoric_ester
RULE 0      sulfonic_acid
RULE 0      sulfonic_ester
RULE 0      triphenyl_phosphene
RULE 0      unbranched_chain                    //(>4_atoms)
RULE 0      epoxide
RULE 0      hetero_hetero
RULE 0      sulfonyl_halide
RULE 0      halopyrimidine
RULE 0      perhalo_ketone
```

RULE 0 methyl_ketone
RULE 0 aziridine
RULE 0 imine
RULE 0 oxalyl

#the dye group includes a set of patterns which describe all cpds with colors in their names from the ACD98.2

RULE 0 dye

#functional groups which are allowed, but may not be wanted in high quantities
#common functional groups

RULE 6 alcohol
RULE 4 alkene
RULE 4 amide
RULE 4 amino_acid
RULE 4 amine
RULE 4 primary_amine
RULE 4 secondary_amine
RULE 4 tertiary_amine
RULE 4 carboxylic_acid
RULE 1 halide
RULE 1 iodine
RULE 4 ketone
RULE 4 phenol

#other functional groups

RULE 4 alkyne
RULE 4 aniline
RULE 4 aryl_halide
RULE 4 carbamate
RULE 4 ester
RULE 4 ether
RULE 4 hydrazine
RULE 4 hydrazone
RULE 4 hydroxylamine
RULE 4 nitrile
RULE 4 sulfide
RULE 4 sulfone
RULE 4 sulfoxide
RULE 4 thiourea
RULE 4 thioamide
RULE 4 thiol