Use of computational methods to identify new ligands of RORγ

by

Anthony DiProperzio

B.S., Manhattan College, 2016

A THESIS

submitted in partial fulfillment of the requirements for the degree

MASTER OF SCIENCE

Department of Biochemistry and Molecular Biophysics
College of Arts and Sciences

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2021

Approved by:

Major Professor
Ho Leung Ng

Copyright

Abstract

Retinoid-related orphan receptor γ (RORγ) is called an orphan receptor because for a while it was not known whether this nuclear receptor had any endogenous ligands, but as of now, a myriad of sterols as RORγ natural ligands have been discovered. It is the lineage-specific master transcription factor that is also the only nuclear receptor established to enhance T helper 17 differentiation and restrain T regulatory cell differentiation. T helper 17 cells are the major proinflammatory cells that produce cytokines that cause responses such as tissue inflammation and clearance of extracellular pathogens. T helper 17 cells can also produce a particular cytokine called Interleukin-17 which induces autoimmune responses. For the past 10 years, many new synthetic RORγ agonists and antagonists have been developed.

There are still remaining questions to understand RORγ. To increase understanding of the RORγ signaling pathway and to speed the drug design of RORγ agonists, computational methods were used to study the structure of RORγ and its agonist drug design. These computational methods included virtual screening, molecular docking, and fragment-based drug discovery to formulate a chemical starting point for the identification of new ligands of RORγ. More information on the activated state of this transcription factor was obtained by studies of these agonists and RORγ. The results of this research consisted of chemical fragments that bound different areas of the active site of the ligand binding domain of RORγ. In addition, new compounds were discovered that have different scaffolds compared to known RORγ bioactives and are currently awaiting experimental testing in the laboratory.

Table of Contents

List of Figures

List of Tables

List of Equations

Acknowledgements

I thank Professor Ho Leung Ng for his continued guidance, support, and helpful discussions during my time at Kansas State University. Your constant optimism about life, science, etc. is keeping me going to pursuing my career goals. Without your help I could not imagine where I would be.

I also thank Professor Gerald Reeck for continuously giving me guidance and help during my graduate life. Your assistance in helping me pinpoint my research interests and constant encouragement are gratefully acknowledged.

I also thank my committee members Professor Michal Zolkiewski and Professor Om Prakash. Thank you for your advice in my research.

I would like to thank Professor Erika Geisbrecht, Professor Brian Geisbrecht, Professor Gregory Finnegan, Professor Phillip Klebba, Professor Lawrence Davis, Professor Anna Zolkiewska, Professor John Tomich, Professor Michael Kanost, and Professor Timothy Durrett. Without your help I would not have been able to make so much progress since I started the program.

I would like to thank Dr. Micheal Barnett. I had such a wonderful time working with you as a teaching assistant.

Helpful discussion with Dr. Samson Souza, Mr. John Ewalt, Ms. Mian Huang, Ms. Ye Zou, and Dr. Bowen Tang are all gratefully acknowledged.

I also want to acknowledge all of the graduate students in this department, in particular Ashish Kumar, Taihao Yang, Jacob Weber, and Patrick Corn for helping me to become a better teaching assistant. I will always remember the fun we had working together. You guys taught me how to get students to get more interested in the experiments and to teach more effectively.

I would also like to thank Ms. Linah Alkotami. You gave me a lot of advice on how to prepare for my Master's defense and suggestions for my writing.

I would also like to thank Mr. Nitin Mishra. I will never forget you inviting me to have dinner together which made my graduate life more fun.

I would like to thank St. Isidore's Catholic Church for helping me to find a home in and fall in love with Manhattan, KS.

Dedication

I dedicate this thesis to my parents and grandparents for continuous love. I also dedicate this thesis to my siblings Joseph, May, Anne Marie, Catherine, Lisa, Vincent, and John Paul for all the childhood memories we shared together and everything that happened between us for all those years.

To my friend George Goodman for being such a committed and dedicated friend for so many years, and in memory of all the amazing adventures we shared together including that unforgettable road trip.

Chapter 1 - Introduction (Literature Review)

ROR Family Lineage and Function

Nuclear hormone receptors are a superfamily of ligand-dependent transcription factors which contain steroid hormones, retinoids, thyroid hormones, and eicosanoid metabolites. The cloning of several steroid hormone receptors in the 1980s led to an intense search by many laboratories for additional members of the steroid hormone superfamily. This resulted in the identification of orphan receptors, including members of the retinoid-related orphan receptor (ROR) subfamily in the 1990s. RORs are a subfamily of thyroid hormone receptors, which are one of the 48 types of transcription factors in humans. Thyroid hormone receptors are a subfamily of the nuclear hormone receptor superfamily. Figure 1.1 shows the relationship of nuclear hormone receptors, thyroid hormone receptors, and retinoid-related orphan receptors. The three isoforms of retinoid-related orphan receptors in humans are RORα, RORβ, and RORγ. (1)

As transcription factors, retinoid-related orphan receptors (RORs) control the rate of transcription of genetic information from DNA to messenger RNA, by binding as monomers to a specific DNA sequence (Figure 1.2). (1) These DNA response elements consist of the consensus 5'-AGGTCA-3' core motif, and are preceded by an A/T-rich sequence in the regulatory region of the target genes. This allows retinoid-related orphan receptors to regulate many physiological processes and have roles in several pathologies.

The Structure of Retinoid-Related Orphan Receptors

RORs as typical nuclear receptors are composed of an N-terminal domain (residues 1-33), a DNA binding domain (residues 31-95), a hinge domain (residues 96-265), a ligand binding

domain, and a C-terminal domain (residues 506-518). The N-terminal domain contains Activation Function 1 (AF1), which cooperates with Activation Function 2 (AF2) in the ligand binding domain to regulate gene transcription (Figure 1.3), and confers DNA binding specificity to the various retinoid-related orphan receptor isoforms. The ligand-binding domain, which is at the C-terminus, is multifunctional and plays important roles in ligand binding and nuclear localization. It also interacts with coactivator and corepressor proteins. The DNA binding domain is the most conserved domain in all nuclear receptors. It interacts with DNA response elements to induce transcription. (1)

The ligand binding domain of RORγ is where the entire focus of discovery and design of small molecule modulators is directed (Figure 1.4). (2) The structure of the RORγ ligand binding domain consists of a globular structure with 12 α-helices and a tiny bit of sheet as shown in Figure 1.4 A. Helix 12 in the ligand binding domain is essential for facilitating the recruitment of the coactivator protein to the ligand binding domain surface and activating transcription. Helix 12 is disordered in the ligand binding domain when no ligand is present. Figure 1.4 A shows the crystal structure of RORγ. In the red circle, it shows how the benzyloxytricyclic agonist binds to RORγ. The ligand is shown in magenta. In Figure 1.4 A, agonist binding induces helix 12 to become a more ordered helix. Figure 1.4 B shows a hydrogen bond between residues His479 and Tyr502 (yellow dashed line), which locks helix 12 in the correct position, facilitating coactivator recruitment. All this information indicates that RORγ is very important to drug design.

Retinoid-Related Orphan Receptors in Cell Survival, Differentiation, and Diseases

RORγ is essential for the development of secondary lymphoid tissues (Figure 1.5). Lymphoid tissue inducer cells are derived from fetal liver hematopoietic stem cells upon

induction of RORγ and the DNA-binding protein inhibitor (Id-2). Lymphoid tissue inducer cells are recruited from circulation by mesenchymal organizer cells. This is mediated by multiple receptor-ligand interactions, which involve several positive feedback loops and amplifying mechanisms. Lymph node organogenesis is initiated by the interaction of lymphotoxin α1β2 at the surface of lymphoid tissue inducer cells with their respective receptors on mesenchymal organizer cells. This leads to the induction of several adhesion molecules, and the production of homeostatic chemokines by mesenchymal organizer cells. Lymphoid tissue inducer cells are required for the development lymph nodes, Peyer's patches, cryptopatches, and isolated lymphoid follicles, which are thought to be derived from cryptopatches after the colonization of the intestine by bacteria. RORγ is required for the generation and survival of lymphoid tissue inducer cells.

RORγ plays a critical role in the regulation of thymopoiesis (Figure 1.6). (3) Thymopoiesis describes the process which turns thymocytes into mature T cells. During thymopoiesis, T cell precursors differentiate successfully via two intermediate stages. A T cell is a type of lymphocyte that can be distinguished by the presence of a T cell receptor on its surface. These cells then differentiate via immature single positive cells into double positive thymocytes (immune cells before they undergo transformation into T cells). After successful T cell receptor gene rearrangement, double positive cells expressing the T cell receptor undergo a careful selection process to eliminate thymocytes expressing nonfunctional or autoreactive T cell receptors. Figure 1.6 shows that the positively selected double positive thymocytes mature into single positive $CD4^{+}CD8^{-}$ cytotoxic T cells that then colonize the secondary lymphoid organs. RORγ promotes the differentiation of immature single positive cells into double positive cells, and regulates the expression of the anti-apoptotic gene B-cell lymphoma-extra large (Bcl-X_L),

which enhances cell survival and promotes T cell receptor rearrangements. Lack of RORγ inhibits the differentiation of immature single positive cells into double positive cells. This results in reduced expression of Bcl-X_L in double positive thymocytes. This results in accelerated apoptosis, reduced lifespan of double positive thymocytes, and impaired T cell receptor rearrangements.

In 2007 and 2008, Littman, Cua, and their colleagues were the first to report that RORγ is required for the differentiation of immature $CD4^+$ T cells into T helper 17 cells (Figure 1.7). (4) RORγ is the lineage-specific master transcription factor that is the only nuclear receptor established to enhance T helper 17 differentiation and restrain T regulatory cell differentiation. (1) Differentiation into different effector T cell lineages is initiated through interaction of dendritic cells with uncommitted cluster of differentiation 4 positive T helper cells. These T helper cell lineages are characterized by their synthesis of specific cytokines and their immuno-regulatory functions. T helper 17 and T regulatory cells produce cytokines that play roles in several inflammatory responses, including immune responses to various pathogens. Proinflammatory T helper 17 cells coordinate host defense. T regulatory cells are involved in the downregulation and contraction of an immune inflammatory response.

T helper 17 cells have been implicated in some autoimmune diseases and inflammatory conditions in humans, including lupus, arthritis, multiple sclerosis, etc. (5) Interleukin-17 proteins induce the secretion of proinflammatory mediators and promote the recruitment of neutrophils (white blood cells). It has been shown that mice lacking Interleukin-17 cytokines are resistant to collagen-induced arthritis and experimental autoimmune encephalomyelitis (EAE). Neutralizing Interleukin-17 cytokines with antibodies suppresses central nervous system autoimmune inflammation. (6) Another study using the adoptive transfer model of colitis

showed that transfer of RORγ-null T cells failed to increase mucosal Interleukin-17 cytokine levels and did not induce colitis, while treatment with Interleukin-17 restored colitis after transfer of RORγ-null T cells. (7) By controlling T helper 17 differentiation and Interleukin-17 expression, RORγ plays a critical role in the regulation of inflammatory responses, and can function as a potential drug target for therapeutic intervention of autoimmune disease.

Recent studies suggest that RORγ is involved in various cancers, including lymphoma, melanoma, and lung cancer. RORγ is sometimes upregulated in cancer tissues compared with normal tissues, but other times is not. (8) The expression of RORγ in tumor cells is either increased or decreased depending on the cancer type and signaling pathway. Increased expression of RORγ in lung cancer, prostate cancer, and gastric cancer results in a protumor effect, while decreased expression of RORγ in breast cancer and melanoma could induce an antitumor effect. As the main transcription factor in Interleukin-17 expressing immune cells, RORγ has been investigated in various cancer cells and tumor-infiltrating cells, indicating that it might be a promising prognostic factor in lung and breast cancer, and a potential therapeutic target in prostate cancer. High RORγ expression can enhance antitumor immunity by increasing Interleukin-17 expression, granulocyte-macrophage colony-stimulating factor secretion, and decreasing programmed cell death protein 1 expression. (9) and (10) T helper 17 cells and Interleukin-17 cytokines play mixed protumor roles in the development of colorectal cancer. (11) High RORγ activity can enhance antitumor immunity by increasing Interleukin-17 expression. (12)

The roles that ROR family members play in tumorigenesis vary in different cancers and depend on different producing cells in the tumor microenvironment. (8) Concentration on the relationships between RORs and tumorigenesis should be meticulously organized, and should

deeply explore the clinical significance and the underlying mechanisms. Since different ROR isoforms present different biological functions, research on therapeutic targets of RORs in cancer should identify all isoforms of specific RORs. Agonists and antagonists for RORγ could efficiently inhibit tumor growth and progression through activation or inactivation so that their ligands and targets become valid or invalid. Another promising new strategy for anticancer therapy might involve directly targeting tumor cells with RORγ specific modulators due to the correlations between high or low expression of RORγ and tumor progression. RORs are sometimes produced by immune cells in tumor microenvironments and then induce anti-tumor or protumor activity by regulating tumor related cytokines or chemokines. Therapies targeting RORs producing immune cells could be novel treatments for certain cancers.

Under the engagement of pharmaceutical companies, including Merk, Takeda, and Biogen, in developing novel RORγ-targeting drugs, hundreds of small-molecule modulators binding in the active site of the ligand binding domain have been developed and structurally investigated. (2) Many highly active compounds were discontinued or suspended for further development in the stages of preclinical trials or clinical trials, owing to off-target effects, toxicity, and poor therapeutic efficacy. (13) and (14) No compound has yet passed Phase II clinical trials. No full-length three-dimensional structure of RORγ is currently available. The ligand binding domain is the only domain that has been successfully co-crystallized with ligands. Over 100 crystal structures of RORγ ligand binding domain-ligand complexes are published in the Protein Data Bank, including 69 antagonists and 18 agonists. Many of the released modulators have been classified by chemotype and described in a recent review. (15)

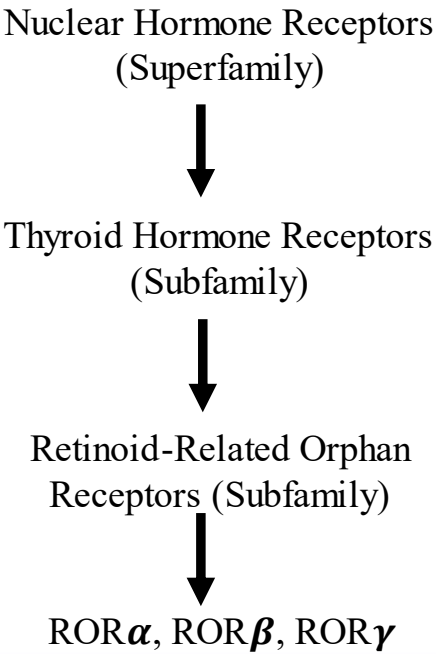


Figure 1.1. Family Lineage of RORγ.

Retinoid-related orphan receptors are a subfamily of the nuclear hormone receptor superfamily. They fall under the category of thyroid hormone receptors, one of the 48 types of transcription factors in humans.

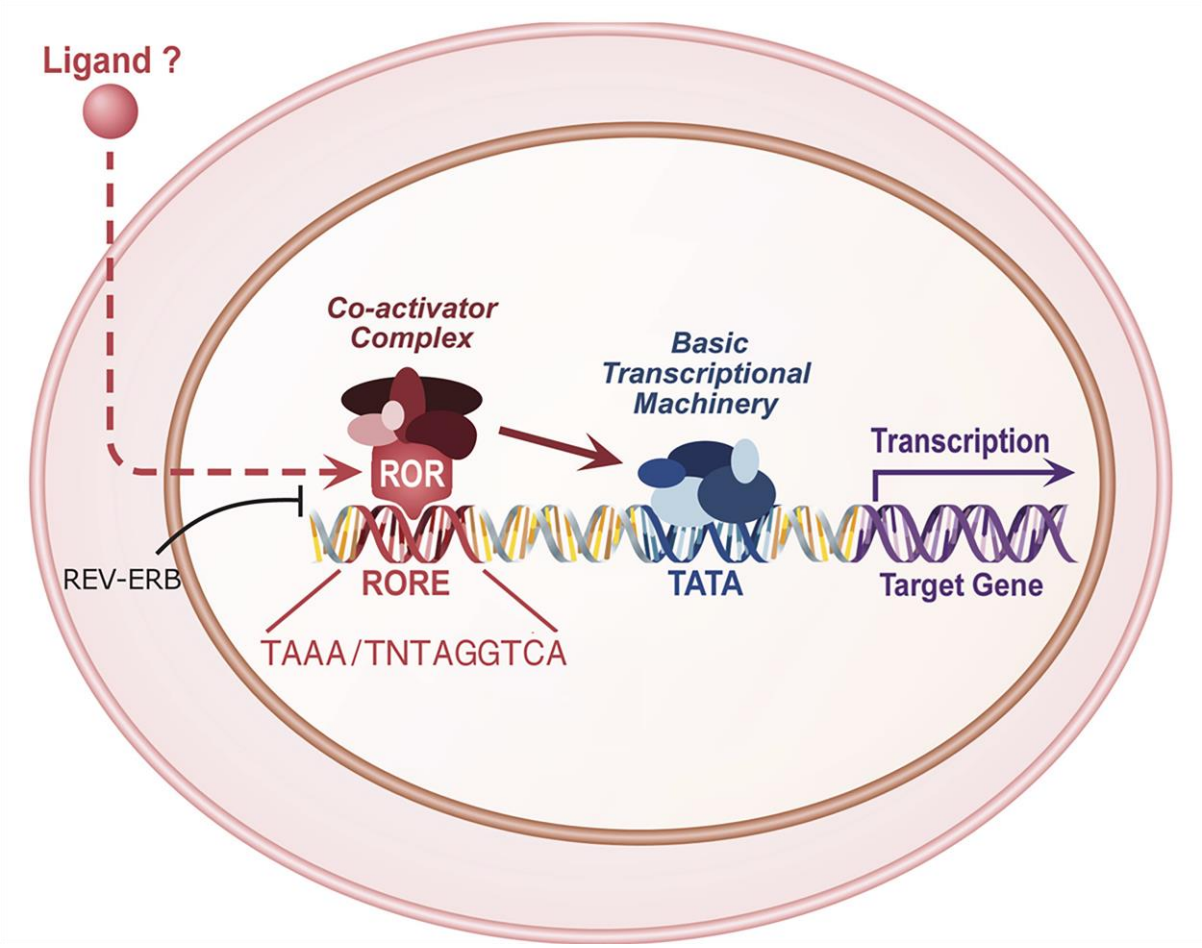


Figure 1.2. ROR Mechanism of Action.

As transcription factors, RORs bind as monomers to DNA with a specificity consisting of the 5'-AGGTCA-3' core motif preceded by an A/T rich region. They can interact with either coactivator or corepressor proteins to regulate gene transcription positively or negatively. This figure was modified from Anton Jetten (Jetten, 2009).

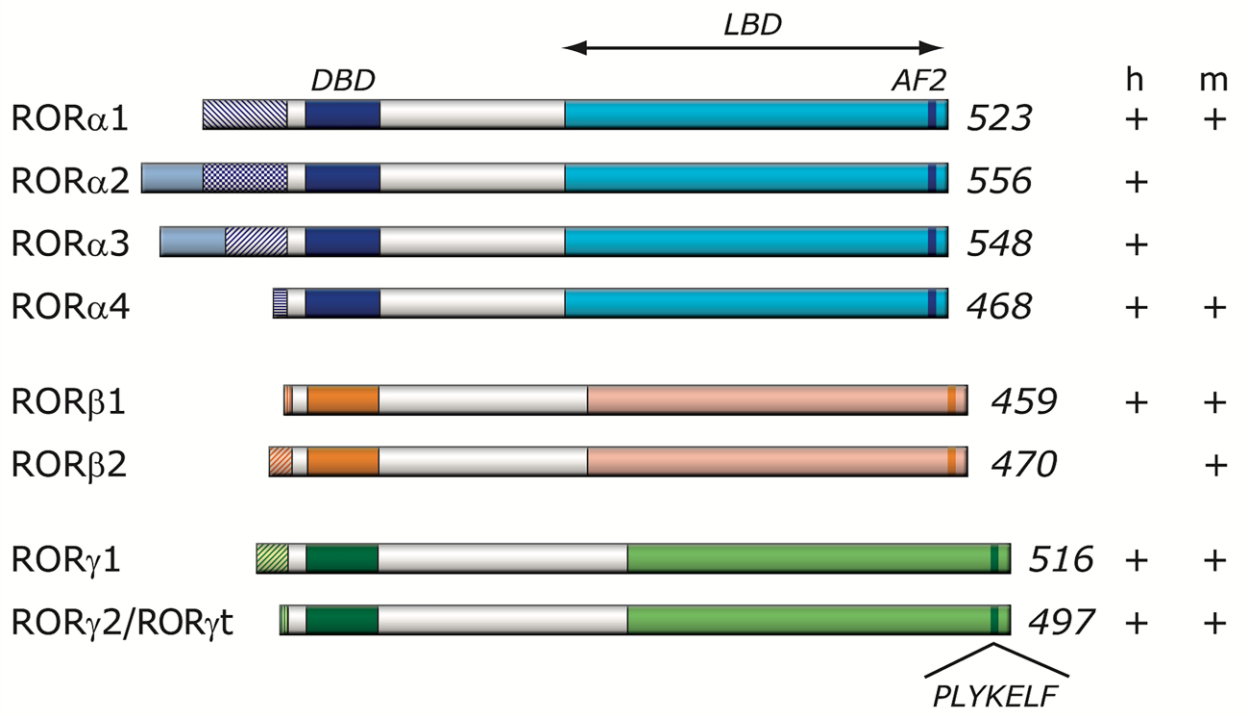


Figure 1.3. Retinoid-Related Orphan Receptor Family Members.

These are schematic structures of the various human (h) and mouse (m) ROR isoforms. The DNA binding domain (DBD), ligand binding domain (LBD), and Activation Function 2 (AF2) are indicated. The variable regions at the N-terminus of each ROR are indicated by patterned boxes on the left. The numbers on the right refer to the total number of residues in each ROR isoform. The different ROR isoforms identified in human and mouse are shown on the right (+/-). This figure was taken from Anton Jetten (Jetten, 2009).

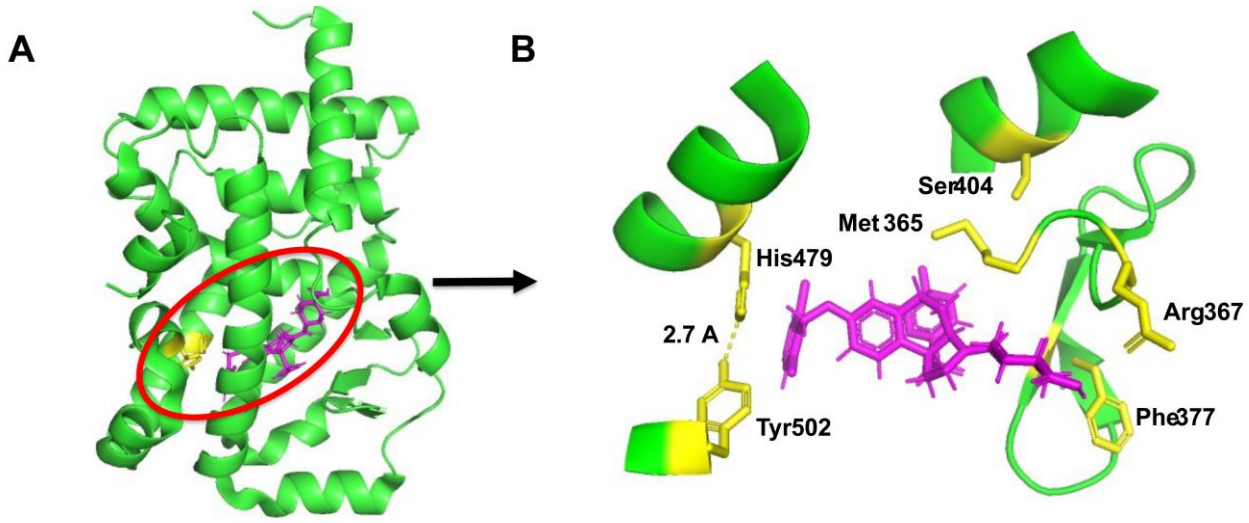


Figure 1.4. Crystal Structure of RORγ Ligand Binding Domain (PDB Id: 6XAE).

A) shows the crystal structure of RORγ ligand binding domain bound to a benzyloxytricyclic compound (magenta) (PDB ID: 6XAE). This compound is an agonist of RORγ. B) shows the hydrogen bond (yellow dashed line) between residues His479 and Tyr502, which stabilizes the ligand binding domain in the active conformation. The importance of the other labeled residues will be explained later.

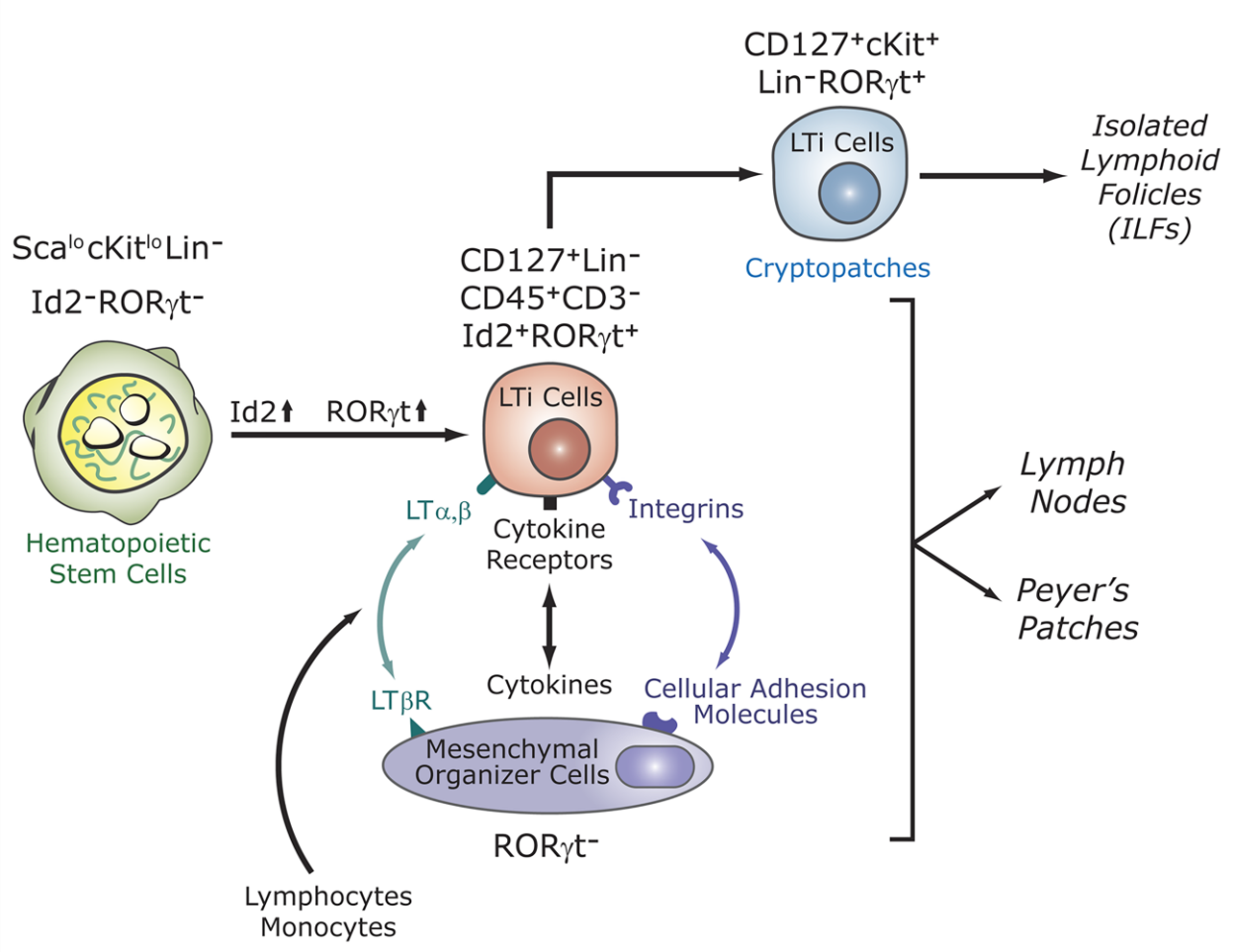


Figure 1.5. Role of RORγ in the Development of Secondary Lymphoid Tissues.

Lymphoid tissue inducer (LTi) cells are derived from fetal liver hematopoietic stem cells. The differentiation is accompanied by the transcription factors DNA-binding protein inhibitor 2 (Id2) and RORγ. Lymphoid tissue inducer cells are recruited from circulation by mesenchymal organizer cells. This is mediated by multiple receptor-ligand interactions. Lymphoid tissue inducer cells are required for the development of lymph nodes, Peyer's patches, cryptopatches, and isolated lymphoid follicles. RORγ is required for the generation and survival of lymphoid tissue inducer cells. The absence of lymphoid tissue inducer cells in RORγ-deficient mice is

responsible for the lack of lymph nodes, Peyer's patches, etc. This figure was taken from Anton Jetten (Jetten, 2009).

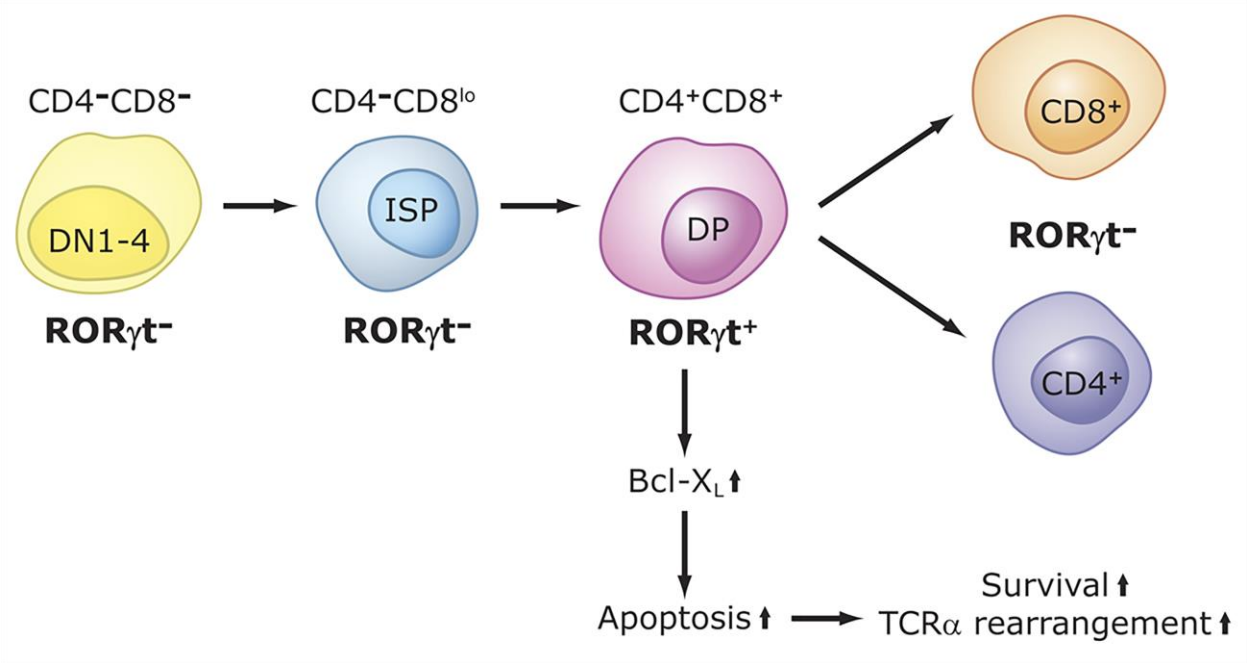


Figure 1.6. Role of RORγ in Thymopoiesis.

The double negative cells $CD4^{-}CD8^{-}$ (CD= cluster of differentiation) differentiate into immature single positive cells $CD4^{-}CD8^{low}$. These cells differentiate into double positive thymocytes called $CD4^{+}CD8^{+}$. RORγ and B-cell lymphoma-extra large are induced during the transition from immature single positive cells to double positive cells, and are down-regulated during the differentiation of double positive cells $CD4^{+}$ $CD8^{+}$ into $CD4^{+}$ and $CD8^{+}$. RORγ promotes the differentiation of immature single positive cells into double positive cells, and positively regulates the expression of the anti-apoptotic B-cell lymphoma-extra large. The latter enhances cell survival that subsequently promotes T cell receptor rearrangements. Lack of RORγ expression inhibits the immature single positive to double positive transition and reduces expression of B-cell lymphoma-extra large in double positive thymocytes. The latter results in accelerated apoptosis, reduced lifespan of double positive thymocytes, and impaired T cell receptor rearrangements. This figure was taken from Anton Jetten (Jetten, 2009).

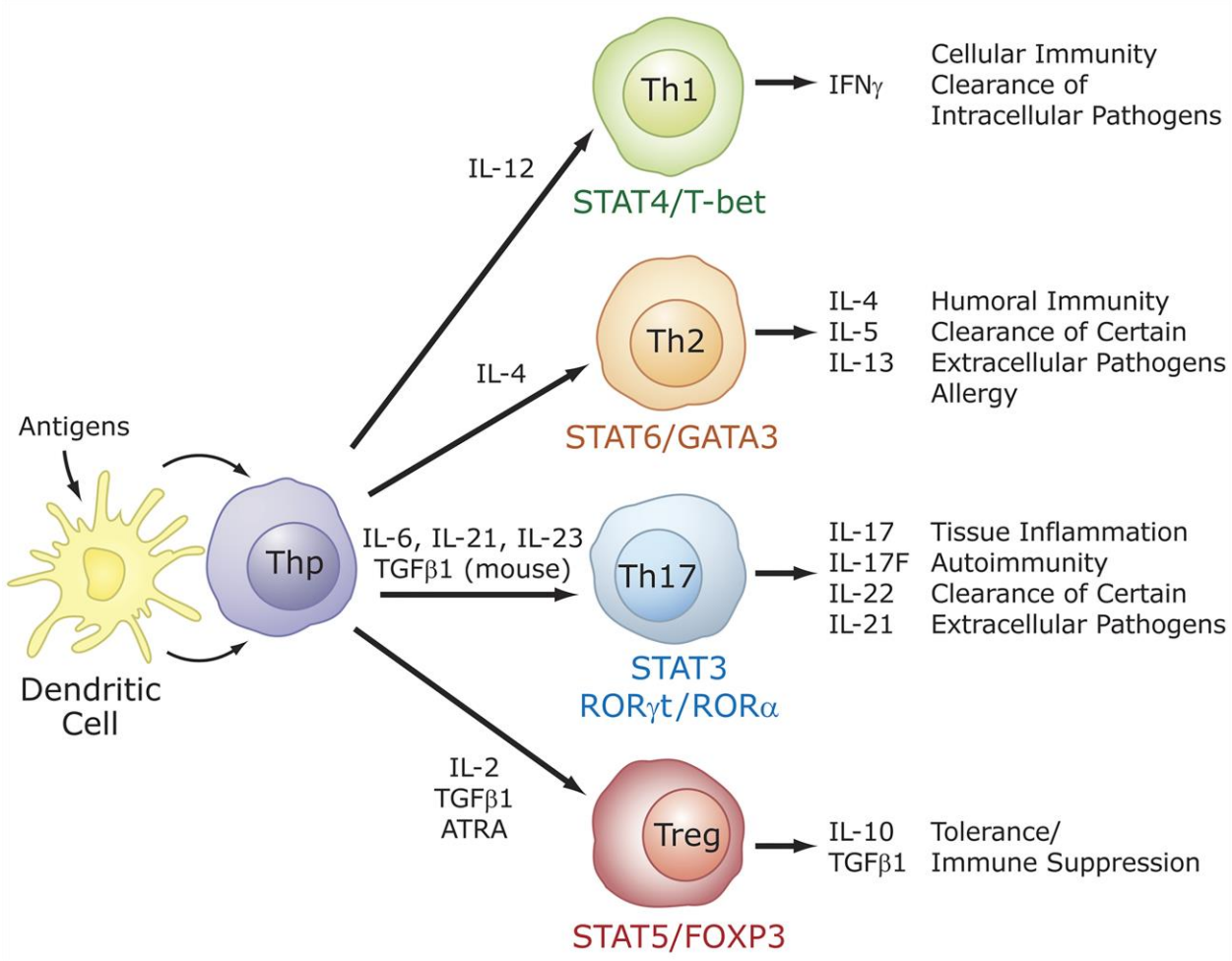


Figure 1.7. Specific Role for RORγ in T Cell Lineage Specification.

Differentiation into different effector cluster of differentiation 4 positive ($CD4^+$) T cell lineages, T helper 1 (Th1), T helper 2 (Th2), T helper 17 (Th17), and T regulatory (Treg) cells is initiated through interactions of dendritic cells with uncommitted cluster of differentiation 4 positive T helper cells (Thp). The effector cell types are characterized by their synthesis of specific cytokines and their immuno-regulatory functions. The differentiation along different lineages involves different cytokines and the activation of distinct signaling cascades and transcription factors that result in the induction of additional cyto/chemokines and cyto/chemokine receptors, which may be part of positive and negative feedback loops. These cytokines and transcription

factors may favor one cell lineage, while inhibiting another. RORγ is induced during differentiation of uncommitted cluster of differentiation 4 positive T helper cells into the T helper 17 lineage and is critical for this lineage specification. This figure was taken from Anton Jetten (Jetten, 2009).

Chapter 2 - Computational Methods

Introduction to Computational Methods

Virtual screening has aided in the discovery of new ligands of RORγ accomplishing the searching process *in silico* by docking a library of drug-like small molecules into a structural model and scoring each compound by mathematical algorithms. It saves time and cost in the initial screening stages of narrowing down the chemical pool. In 2014, the Xu group conducted virtual screening on a 220,000-compound database against the active site binding pocket in an antagonist ligand binding domain crystal model, resulting in 115 hits. (16) In 2016, they did another virtual screening to a five-times larger database against the same model, resulting in 28 hits. (17) Most of the hits showed moderate or strong activity in biological and biochemical assays, but the atomic structures for the best hits were only modeled. It is feasible to apply virtual screening in the discovery of RORγ ligand-binding domain targeting modulators. A feature of virtual screening is that the programs usually dock the compounds against a rigid model to speed up the screening process. It seems indifferent to the search for new agonists since the crystal models before and after ligand binding are almost identical. Some antagonists can be omitted by the program if the docked model is different from the real shape of the pocket. The flexibility of the pocket needs to be considered in virtual screening to avoid missing important hits. The state-of-the-art docking methods incorporate multiple model templates and machine learning scoring functions to improve the effectiveness of predicting ligand activity and the quantitative predictions of binding affinity.

Fragsite Method

The Harigua-Souiai method is a branch of fragment-based drug discovery. (18) This is a method that involves docking small organic molecules called fragments that are usually less than

300 Daltons to a receptor. The purpose is to obtain information on the landscape of the binding pocket through weak interactions between the fragments and the receptor. The fragments are then grown or combined to produce a compound with high affinity. In the Harigua-Souiai method, a software called mkgrid is used to detect cavities in the protein target. Then fragments from a library are docked to the protein target. This results in lots of poses in which the lowest energy poses for each fragment are chosen. In this method, a self-organizing map is used to analyze where most of the fragments have docked, and these regions are distinguished using different colors. Atomic coordinates of the docked molecules are clustered to determine the preferential binding positions of the ligands on the protein surface. These positions constitute consensual clusters that define the favorable binding sites of the probes.

This project involves the use of the Fragsite method. The Fragsite method is a name that Dr. Ng invented for a method the laboratory is still developing and has not yet published. It is similar, but not the same as the method published by Harigua-Souiai. The Harigua-Souiai paper was published while Dr. Ng was working on the Fragsite method. Dr. Ng is working to improve the Harigua-Souiai method by using machine-learning optimization against a fragment database. The Fragsite and the Harigua-Souiai methods are not interchangeable. The Fragsite method was used to dock fragments from the Maybridge Fragment Library (30,000 chemical fragments) to the ligand binding domain of ROR$\boldsymbol{\gamma}$ (PDB Id: 6XAE). This crystal structure was chosen because it performed better in docking than all the other crystal structures.

The Smina Scoring Function

Smina is a machine-learning scoring function that uses a linear regression model. It computes the fitness of protein-ligand binding by summing up the contributions of several

individual terms. They consist of physically meaningful terms that are parametrized to reproduce binding affinities or binding poses. (19)

$$\Delta G = \sum_i W_i G_i$$

Equation 2.1. **Smina Scoring Function** (20)

As shown in Equation 2.1, smina decomposes the overall binding free energy into several energetic terms. (20) The coefficients W_i for various energy terms G_i are obtained from linear regression analysis using experimentally determined binding free energies and structural information. G_i may represent different energy terms such as hydrogen bonds, hydrophobicity, electrostatics, etc. The respective coefficients W_i are determined by fitting the known binding affinity values of a set of protein-ligand complexes (training set) with available three-dimensional structures information. The coefficients are assigned using linear regression methods that fit the experimentally determined affinity values with the computed binding score. The main limitation of smina is its dependence on the experimental data set used to perform regression analyses and fitting. (20) Based on the limiting training set, it is difficult to establish a comprehensive and consistent description of all possible weighting factors in protein-ligand binding. In other words, there is no guarantee that smina can predict the binding affinity of the ligands that are structurally different from those used in the training set.

RF-Score-VS

Random Forest (RF)-Score Virtual Screening (VS) is another machine learning technique that has been shown to be effective as a re-scoring function and can be used for virtual screening. This algorithm characterizes each protein-ligand structure as a set of features that are relevant for

binding affinity. Random forests are an ensemble learning method for classification, regression, and other tasks that operate by constructing a multitude of decision trees at training time and outputting the class that is the mode (middle) of the classes (classification) or average prediction (regression) of the individual trees. (21) RF-Score-VS also merges K_d and K_i measurements into a single binding constant because this increments the amount of data that can be used to train the machine-learning algorithm. A random forest is an ensemble of many different decision trees randomly generated from training data, with predictions calculated by consensus over all trees. (22) As the learning ability of the ensemble of trees improves with the diversity of trees, the random forest algorithm promotes diverse trees by introducing some modifications in the tree training. A new set of complexes is randomly selected with replacement from the training complexes, so that each tree grows to learn a closely related but slightly different version of the training data. Instead of using all the features, the random forest algorithm selects the best split at each node of the tree from a small number of randomly chosen features. This subset changes at each node, but the same number of features is used for every node of each of the trees in the ensemble. Predictions are made by averaging the number of individual predictions of all the trees in the forest.

Using more than one scoring function to evaluate and rank ligands from chemical libraries is now standard practice in virtual screening. Smina assumes a linear relationship between the features characterizing the protein-ligand complex and its predicted binding affinity. RF-Score-VS assumes a non-linear relationship between the features characterizing the protein-ligand complex and its predicted binding affinity. It is this difference in modelling that allows RF-Score-VS to improve and other scoring functions like smina to stagnate as the number of training complexes increases (Figure 2.1). The linear regression models are unable to assimilate

large amounts of data because they are provided in a way that does not permit changing the linear regression model. Non-linear based machine-learning algorithms can assimilate large amounts of data because they are not restricted to this linear regression model. This allows them to learn, grow, and improve with new data.

It is firmly established that non-linear based scoring functions can outperform linear-based scoring functions and can lead to more accurate binding affinity predictions. This is attested to by higher hit rates and Pearson correlation coefficients. (23) Non-linear based scoring functions can perform much better in identifying active molecules in the top percentages of compounds in target screens. Linear-based scoring functions do not account well for conformational entropy or solvation energy contributions. Non-linear based scoring functions can implicitly capture binding effects that are hard to model explicitly. The validation of non-linear based scoring functions is also much more rigorous compared to linear based scoring functions.

Pharmit Platform

The Pharmit platform provides an online, interactive environment for the virtual screening of large compound databases. (24) A Pharmit session is automatically initialized using any complex in the Protein Data Bank (PDB) by inputting the corresponding PDB accession code on the Pharmit main page (Figure 2.2). The software will then identify the interactions that it thinks are most important between the receptor and the ligand. These interactions are also known as features and include hydrogen bond acceptors and donors, negative and positive charges, aromatic groups, hydrophobic interactions, etc. The features can also be adjusted manually in the software by the user. This ensemble of features, which is known as a pharmacophore, is necessary to ensure the optimal supramolecular interactions with a specific

biological target structure and to trigger or to block its biological response. A pharmacophore does not represent a real molecule or a real association of functional groups, but is a purely abstract concept that accounts for the common molecular interaction capacities of a group of compounds towards their target structure. (25)

The Pharmit platform then searches library compounds that match that pharmacophore, which are built into the software. These libraries include MolPort, Chemspace, etc. Compounds match if they can be positioned so that their corresponding features match those designated in the search query. The selected database is searched for compounds that match the specified pharmacophore. Results are aligned to the pharmacophore to minimize the root mean square deviation (RMSD) between the query features and the hit compound features. Results are sorted in decreasing or increasing order based on root mean square deviation (RMSD), molecular weight, etc. Energy minimization of the results can also be done to optimize both the pose and conformation of identified hits with respect to the provided receptor using the AutoDock Vina scoring function and smina. Minimized results may be sorted and filtered to eliminate poses with unfavorable binding energies.

This software was chosen because it has distinct advantages including high search speed on the order of seconds and an expansive suite of search features that allow the user to identify hits more rapidly compared to existing browser-based virtual screening servers. (24) The Pharmit platform democratizes structure-based computer aided drug discovery by offering free and open access, making it ideal for research and education. Pharmit can also be deployed locally by the user, ensuring the complete privacy of queries and results. It anticipates the needs of the medicinal chemist by facilitating direct comparison between a search ligand and hits. Overall,

Pharmit enables the user to perform structure-based virtual screening in a manner that is fast and intuitive.

The Tanimoto Coefficient

The Tanimoto coefficient (Jaccard Index Coefficient) is defined as the size of the intersection of sets A and B divided by the size of the union of the sets A and B. (26) It is a statistic used for gauging the similarity and diversity of sample sets. It is the most widely used measure of intermolecular similarity in both the clustering and searching of databases, and in comparing the similarity of two or more compounds. (27) It uses a two-dimensional description to address the problem of quantifying chemical similarity based on some count of shared features. These features include atoms or element types, bonds, topological torsions, or bits set in a bit strong.

Fragments within a molecular structure map to bits set within a bit string (fingerprint). (27) The fingerprint is defined as a binary bit string that indicates the presence or absence of potential pharmacophores in a molecule (Figure 2.3). The value of the bit is set to one if the pharmacophore is present in a molecule or zero if it is absent. Different compounds have different fingerprint patterns. The Tanimoto coefficient can be used to measure the similarity of the fingerprints. Compounds are generally deemed similar if they have a chemical similarity of 0.9 or better.

The Fragsite method was used to dock all 30,000 fragments from the Maybridge Fragment Library to the active site of the ligand binding domain of RORγ (PDB Id: 6XAE). This crystal structure was chosen because it performed better than the others in docking. It also has an agonist bound to it (EC_{50} = 7.0 nm), and the goal of this project is to find new agonists of RORγ. It is also important to note that the Maybridge Fragment Library denotes fragments as having a

molecular weight of less than 350 Daltons. The smina platform was then used to dock the fragments to the active site of the ligand binding domain. The highest scoring fragments were kept. These fragments had predicted binding affinities to the ligand binding domain of RORγ of less than -9.0 kcal/mol. These fragments were then re-scored using RF-Score-VS. The top 20 fragments that had the highest random forest scores were then further analyzed during which four were thrown away because they had floppy tails.

The 16 fragments were used to map out pharmacophores for receptor binding to determine hotspot interactions and favorable chemotypes that could guide hit discovery and lead optimization. This was done by eyeballing for clusters among the fragments. As shown in Table 2.1, three different clusters were identified among the fragments. Each cluster is denoted with a different feature. The coordinates for the fragments were then averaged together. The average coordinates for each feature were then programmed into the Pharmit platform. To have a drug candidate, the molecular weight and other physical properties must be within an acceptable drug-like range. Lipinski's "Rule of five" was applied to define that range. Both the MolPort and Chemspace libraries were screened because they are the two largest libraries of purchasable compounds. All compounds were selected for further screening that had a minimized binding energy of less than -6.0 kcal/mol. Smina was then used to dock all the selected compounds from Pharmit to the ligand binding domain of RORγ. Compounds were kept for further docking if they had a predicted binding affinity of less than -9.0 kcal/mol. These fragments were then rescreened using the random forest scoring function, which yielded 20 fragments.

The next step was to determine if these 20 fragments were chemically similar to any known bioactive compounds of RORγ. This was done by using a combination of RDKit and python to calculate chemical similarity using the Tanimoto coefficient. The databases that were

searched were ChEMBL and BindingDB. In addition, Ms. Mian Huang has a list of known bioactive compounds of RORγ, and I also compared the compounds I obtained with virtual screening to that list as well. The goal was to find compounds that are not chemically similar to any known bioactive of RORγ. A molecule with a high Tanimoto score to a known bioactive compound would be too closely related and considered less interesting.

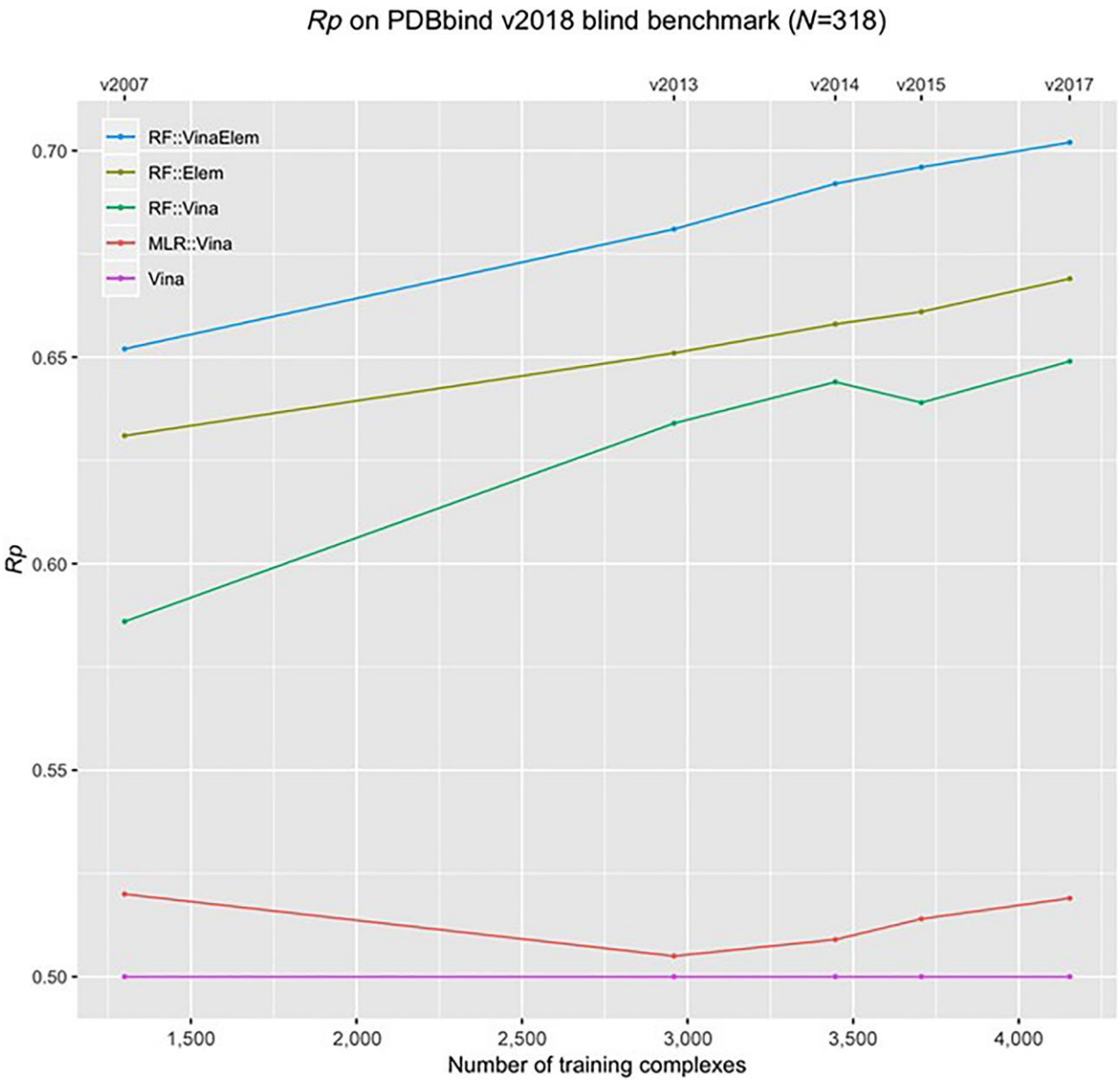


Figure 2.1. Non-linear Versus Linear Based Machine-Learning Scoring Functions.

Test set performance in terms of Pearson Correlation Coefficient (R_p) grows with more training data when using the random forest scoring function (blue, green, gold), but stagnates with machine-learning scoring functions that use linear regression models (purple and red). This figure was taken from Hongjian Li (Li, 2020).

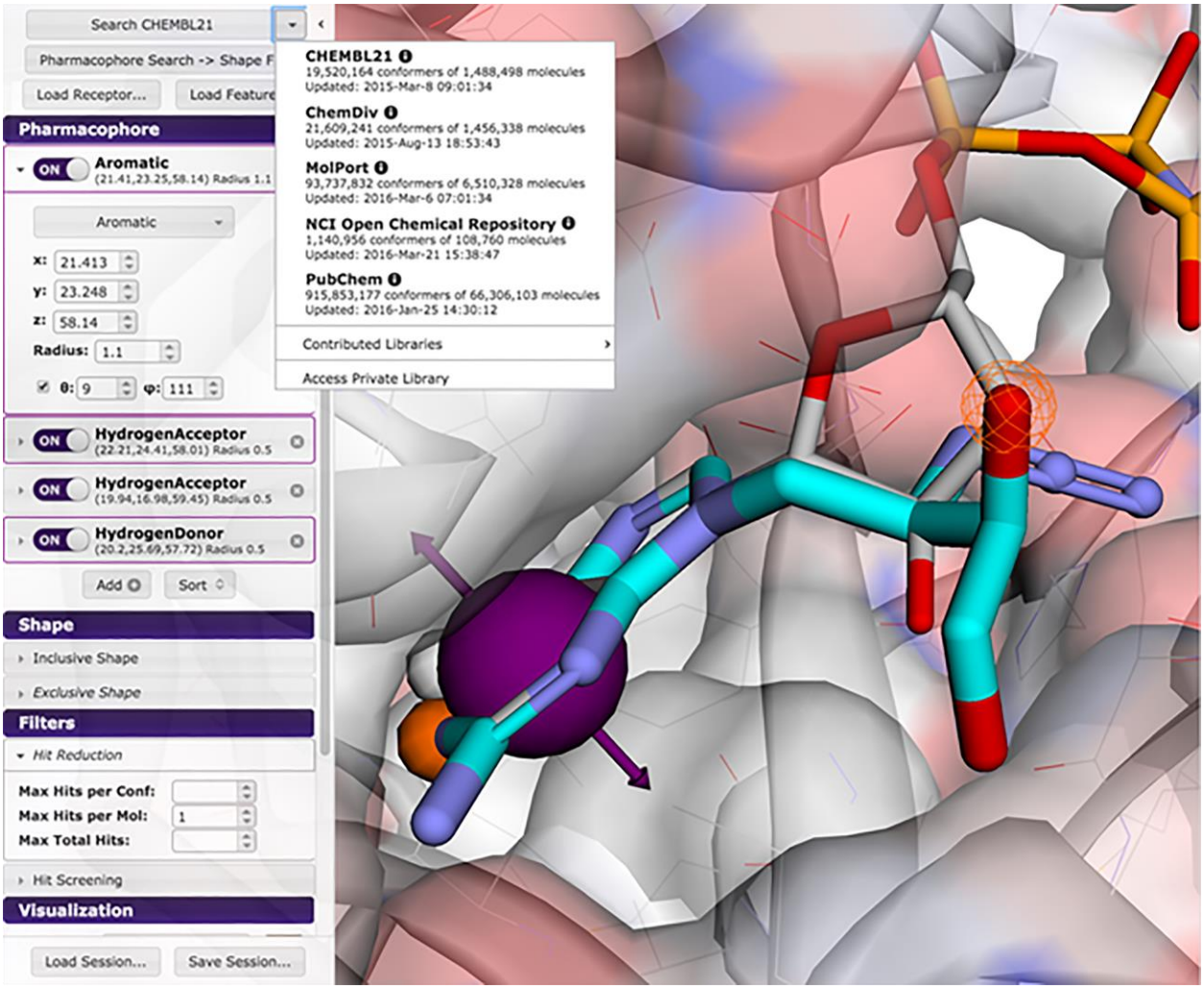


Figure 2.2. The Pharmit Platform.

Pharmit provides an online interactive environment for the virtual screening of large compound databases. It works by loading a protein-ligand complex from PDB into the platform, putting in the coordinates and their corresponding features as well as other filters (e.g., Lipinski's rule of 5), selecting the proper chemical libraries, and screening for hit compounds. This figure was taken from Jocelyn Sunseri (Sunseri, 2016).

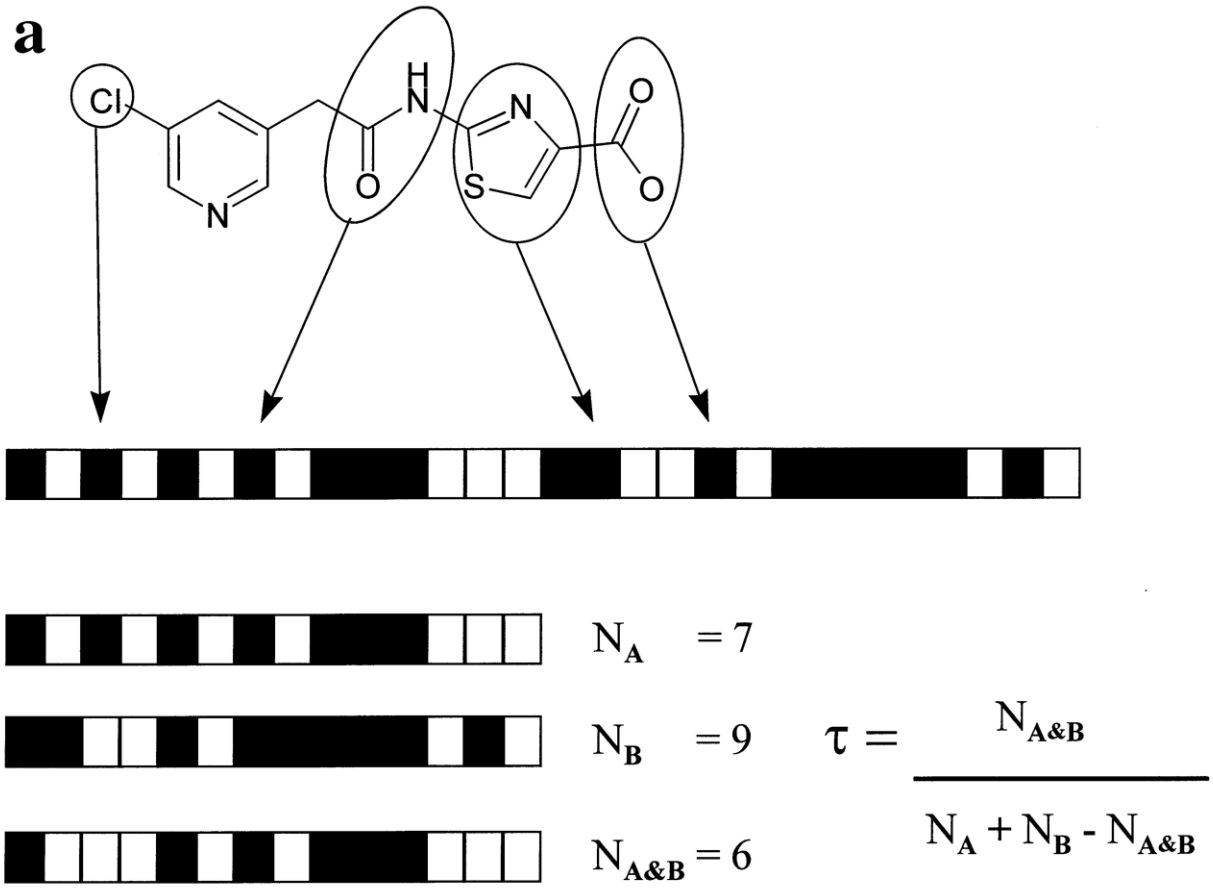


Figure 2.3. Encoding a Chemical Structure as a Bit String.

This figure is taken from Darren R. Flower (Flower, 1998).

Table 2.1. Common Features and Coordinates of the Compounds.

Cluster	Feature	Coordinates
1	Hydrogen Acceptor	(-23.0, -4.46, -24.3)
2	Hydrophobic and Aromatic	(-24.0, 1.35, -23.8)
3	Hydrophobic and Aromatic	(-28.4, -7.63, -23.9)

Chapter 3 - Results

Choosing a Protein Crystal Model for Virtual Screening

The first step was to determine what protein crystal model to choose for virtual screening. Three crystal structures of the ligand binding domain of RORγ (PDB Ids: 4WPF, 6XAE, and 6E3G) from the Protein Data Bank were compared. All three structures have agonists bound to them because the goal is to find new agonists of RORγ. The agonists from the three models were compared to see what functional groups and pharmacophores they had in common. The MolPort and Chemspace libraries were screened on Pharmit for all three of the structures. These libraries are two of the largest databases of purchasable compounds. Three lists of compounds were made for each crystal structure of about 100 molecules. The three lists were combined into one list of the top 100 molecules. The goal was then to determine which crystal structured performed the best. This was done by ranking the minimization scores. It was determined that the compounds coming from the virtual screening against PDB Id: 6XAE generally scored better than the compounds from the other two crystal structures.

Identification of Fragments

Through docking against the cleaned protein model by the Fragsite, smina, and random forest methods, 16 fragments were identified to bind potentially the active site of the protein (Figure 3.2). Based on their molecular weights, it may raise the question as to why these chemical entities are being called fragments instead of compounds (Table 3.1). This is because most of them have molecular weights within 350 Daltons, which is the limit for the definition of fragment based on the Maybridge fragment library. In addition, protein-ligand binding is comprised of widely varying contributions from different atoms and functional groups on the ligand and the protein. That is, certain atoms, groups, or residues can contribute more toward the

free energy of binding than others.(28) This phenomenon, described as an activity cliff, has motivated significant study and method development. In 2016, Xue et al. mapped experimentally using cocktail crystallography the ligand binding pocket and identified four distinct hotspots (Figure 3.1). (29) These hotspots are Met365 and Ser404 (site 1), His479 and Tyr502 (site 2), Arg367 (site 3), and Phe377 (site 4). According to this study, these residues contribute the most to the free energy of binding. Ligands which bind with high affinity to the ligand binding domain of RORγ could be exploiting optimal interactions with these residues by properly orienting them in the correct positions.

The Pymol and LigPlot diagrams show both the 3-dimensional and 2-dimensional ligand docking predictions of the 16 fragments to the ligand binding domain of RORγ (Figures 3.3 and 3.4). In addition, Harikrishnan et al. identified two parts of the binding pocket of the ligand binding domain of RORγ.(30) They are the hydrophobic pocket (residues Trp317, Cys320, Met358, Leu391, Ile397, Ile400, His479, and Tyr502) and the polar pocket (backbone amino groups of Leu287 and Gln286 and the side chains of Arg364 and Arg367). The docking predictions of the 16 fragments show that they bind with different conformations and energy scores while also interacting with different residues.

Fragment 1 has the best predicted smina score and ligand efficiencies out of all the fragments. As shown in Table 3.1, it is highly lipophilic (log P= 4.76) and exhibits low solubility. (31) This may explain why fragment 1 is predicted to interact with residues only in the hydrophobic portion of the ligand binding domain including with the site 2 hotspot residues His479 and Tyr502, which is also termed the “business end” of the binding region. On the contrary, fragment 2 binds mostly in the polar portion of the RORγ ligand binding pocket. It forms hydrophobic interactions with Arg364, Arg367 (hotspot 3), Leu287, and Met365 (site 1).

Fragment 2, in contrast with fragment 1, is not very lipophilic (log P= -2.46) and is much more soluble even though its polar surface area is only slightly larger than that of fragment 1. This could be explained by the presence of a nitro group that shares a hydrogen bond with the backbone amino group of Gln286.

Fragment 3 binds in both the hydrophobic and the hydrophilic areas of the RORγ ligand binding pocket. In the hydrophobic portion, it shares direct interactions with the residues Cys320, Ile400, and Ile397. In the hydrophilic pocket, its phenol group shares hydrogen bonds with the side chain of Arg367 (hotspot 3 residue) and the backbone amide and carbonyl of Leu287. It also shares a direct interaction with Met365 (hotspot 1 residue). It has a positive log P value (i.e., 2.46), meaning it is more likely to dissolve in non-polar solvent as opposed to polar solvent. It also has a polar surface area that is almost twice as large as those of fragments 1 and 2. Fragment 4 also extends to both the hydrophobic and hydrophilic binding pockets of the RORγ ligand binding domain. In the hydrophobic portion of the pocket, it shares hydrophobic interactions with residues Cys320 and Ile397. In the hydrophilic binding pocket, its nitro group shares hydrogen bond interactions with the side chains of Arg364 and Arg367 (hotspot 3 residues) as well as with the backbone amide and carbonyl groups of Leu287. It is slightly less soluble than fragment 3 and has a slightly larger polar surface area. Both fragments 2 and 4 have a charged nitro group and may not be appropriate starting points for drug development. It is interesting that the sulfur atom in fragment 3 does not share any hydrophobic interactions with the protein, but it does in fragment 4. Fragments 2, 3, and 4 possess an amide group which could improve potency and reduce lipophilicity. (15)

Fragment 5 is localized in the hydrophobic pocket and shares only hydrophobic interactions with Tyr502 and His479 (hotspot 2 residues), Ile397, Ile400, and Cys320. It has a

slightly larger partition coefficient than fragment 1 and could exhibit poor solubility. It is interesting that the trifluoromethyl (CF_3) moiety does not share any interactions with the protein. Fragment 6 is also localized in the hydrophobic portion of the binding pocket and shares hydrophobic interactions with Tyr502, His479, Ile400, Ile397, Met365 (hotspot 1), Met358, Csy320, Trp317, and Ala321. Its large network of six-membered rings makes it the most planar of all the fragments and very insoluble. Fragment 7 is also only in the hydrophobic part of the binding site where it interacts with residues Trp317, Ala321, His479, Tyr502, Ile400, Ile397, Met358, and Cys320. It possesses a fluorine atom in the para position on one of its six-membered rings. On the other six-membered ring, it has two chains that are both three carbon atoms long. It is the most insoluble out of all the fragments with a log P value equal to 5.79.

Fragment 8 binds in the hydrophobic portion of the binding pocket where it shares interactions with residues Ile400, Cys320, Tyr502, Trp317, Ala321, Leu391, Ile397, and Ile400. It shares a hydrogen bond with the side chain of Tyr502 and the backbone carbonyl group of Trp317. It may not be a good starting point because of the carboxyl group. Fragment 9 also binds in the hydrophobic portion of the binding pocket and shares hydrophobic interactions with residues Tyr502, Trp317, Cys320, Ile400, and Ile397. It also shares a hydrogen bond with the backbone carbonyl group of Cys320. It possesses an ortho halogen, which could be important for potency. (15) Fragment 10 is also only in the hydrophobic portion of the binding pocket and shares hydrophobic interactions with Cys320, Ala321, Trp317, lle400, and Ile397. Like fragment 9, it shares a hydrogen bond with the backbone carbonyl group of Cys320.

Fragment 11 binds in the hydrophilic portion of the binding pocket and shares hydrophobic interactions with Arg367, Met365, and Arg364. It also shares a hydrogen bond with the backbone of Arg367. Fragment 12 is localized in the hydrophobic portion of the binding

pocket and shares hydrophobic interactions with Ile400, Trp317, Tyr502, Cys320, Ile397, and Ile400. It shares a hydrogen bond with the side chain of Tyr502. It possesses a sulfone group, which could be critical for activity. (15) Fragment 13 is in both the hydrophobic portion and the hydrophilic portions of the binding pocket and shares hydrophobic interactions with residues Ile400, Ile397, and Cys320. In the hydrophilic pocket it shares interactions with the residues Arg364, Arg367, and Leu287. It also shares a hydrogen bond with the side chain of Arg367 and the backbone of Leu287. It also shares a hydrogen bond with the backbone carbonyl group of Phe377 (hotspot residue 4). Fragment 14 binds to the hydrophobic portion of the binding pocket and shares hydrophobic interactions with Cys320, Trp317, Ile400, and Ile397. Fragment 15 binds in only the hydrophilic portion of the binding pocket and shares hydrophobic interactions with Arg364, Arg367, Gln286, and Leu287. The nitro group shares hydrogen bonds with the side chains of Arg364 and Arg367. It also shares a hydrogen bond with the backbone amino group of Leu287. It also shares a hydrogen bond with the backbone amino group of Glu379. Fragment 15 has the highest RF-Score, but the lowest smina score out of all the fragments. Fragment 16 binds in the hydrophobic portion of the binding pocket and shares hydrophobic interactions with the residues Cys320, Met358, Leu391, Ile400, and Ile397, but not hotspot residues.

It is interesting that fragments 12 and 8 share a hydrogen bond with the side chain of Tyr502. This is significant because these compounds could interfere with the crucial hydrogen bond between Tyr502 and His479 which is required to stabilize the secondary structure of helices 11 and 12 and induce coactivator recruitment. (32) These fragments could disrupt this interaction and induce an antagonist response. In addition, halogenated fragments 5, 7, 9, 10, and 14 and the aminothiazole 15 should be regarded with suspicion. (33) They may serve as interesting molecular probes, but they may not be ideal starting points for hit-to-lead chemistry.

There is potential they can exhibit substantial promiscuity. In addition, almost half the compounds interact Trp317. The conformation adopted by this residue is critical for agonist activity. All the compounds together show good coverage of chemical space. No compounds interact with Ser404, which is unique to RORγ among the ROR isoforms. Lead compounds with specific interactions to Ser404 may confer isoform selectivity. The fragments that contain nitro groups, which are typically frowned upon in medicinal chemistry practice, can serve as important molecule recognition functions in the early phases of drug discovery by securing hydrogen bonds to both neutral and charged side chains. (34) Fragment 16 is a seven-membered ring and it could be well-suited to a fragment-based context because of its peculiar conformational preferences and projected substitution vectors which represent interesting design features that are well-suited to a fragment-based context. In addition, the fragments that form hydrogen bonds with Arg367 and Arg364 could be crucial for potency and selectivity. (15)

Identification of Potential Ligands of RORγ

The challenge of fragment-based drug discovery is to use the obtained fragments to find new compounds that could bind and have some sort of bioactivity on the protein. The fragments were used to identify pharmacophores, which were used to identify larger, more drug-like molecules using the Pharmit platform. Some of these compounds are shown in Figure 3.5. Comparison of these compounds to known bioactive compounds of RORγ in the BindingDB (9,811 RORγ bioactive compounds) and ChEMBL (6,071 RORγ bioactive compounds) databases as well as a personalized list compiled by Ms. Mian Huang (118 RORγ bioactive compounds) yielded low Tanimoto coefficients (i.e., ~0.6) for all the compounds. These compounds are not chemically similar to any known bioactive compounds of RORγ. If the compounds had high Tanimoto scores (i.e., > 0.9) to any known bioactive compounds, they

would be too closely related and considered less interesting. In addition, these compounds have similar chemical scaffolds. It would be nice to find compounds that have different chemical scaffolds. This would increase the chances of identifying a new bioactive compound of RORγ.

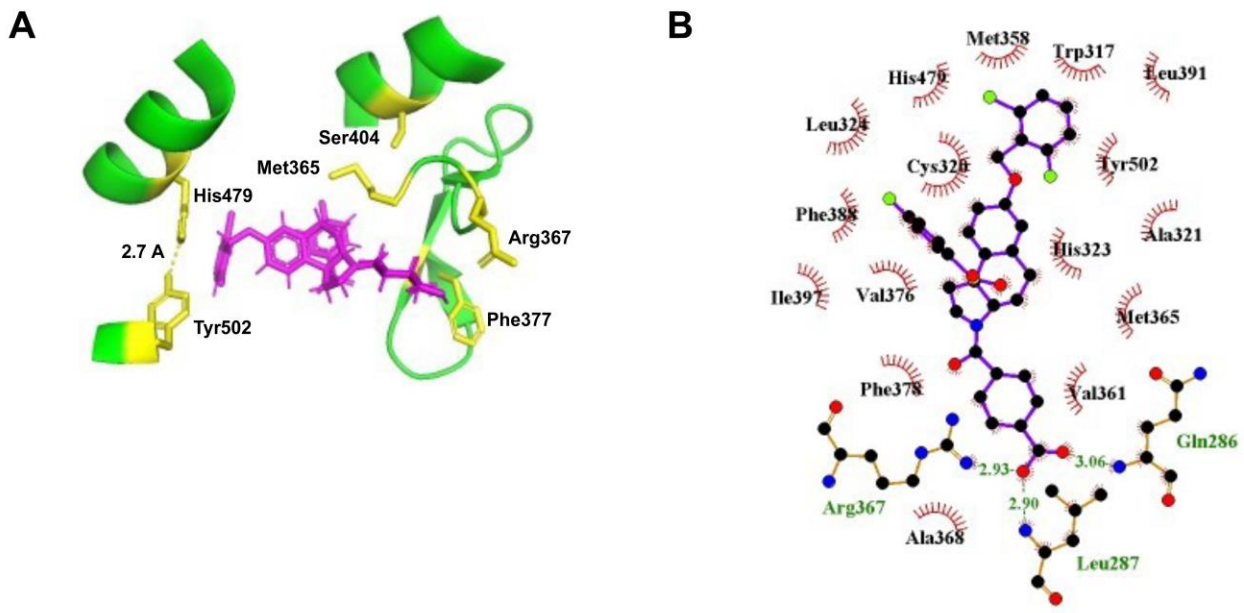
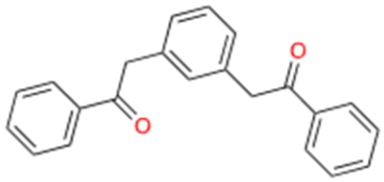
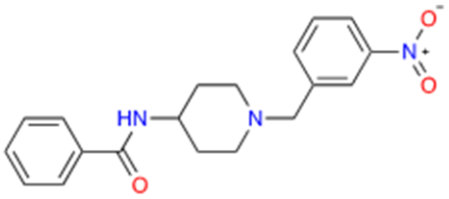


Figure 3.1. A Complex of the RORγ Ligand Binding Domain Bound with a Benzyloxytricyclic Agonist (PDB ID 6XAE).

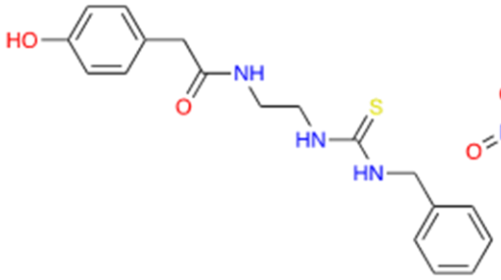
A) A co-crystal model of the benzyloxytricyclic agonist bound to the ligand binding domain of RORγ (PDB Id: 6XAE). The protein is presented in ribbons with green color, and the residues interacting with the ligand (magenta) are represented by yellow sticks. B) The 2D-interaction diagram of the protein-ligand interaction. The ligand is presented in sticks with purple bonds, and other residues hydrogen-bonding against the ligand are with green bonds. The hydrogen bond is presented in green dash lines, and the hydrophobic interactions are presented with red scattered lines.



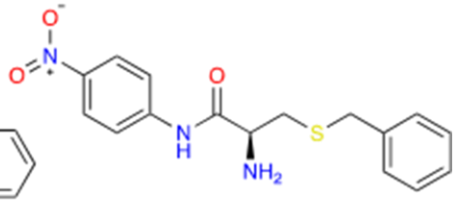
1



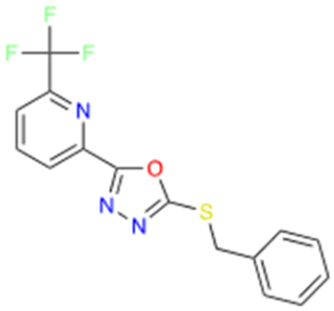
2



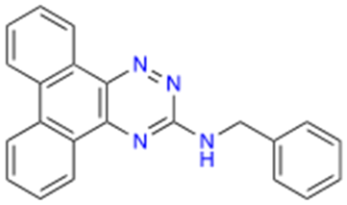
3



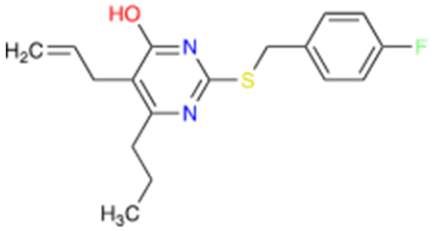
4



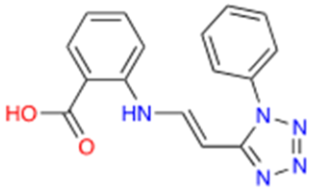
5



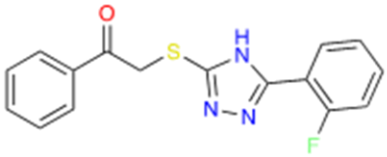
6



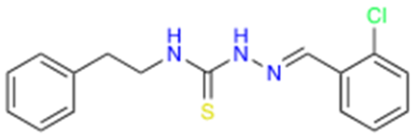
7



8



9



10

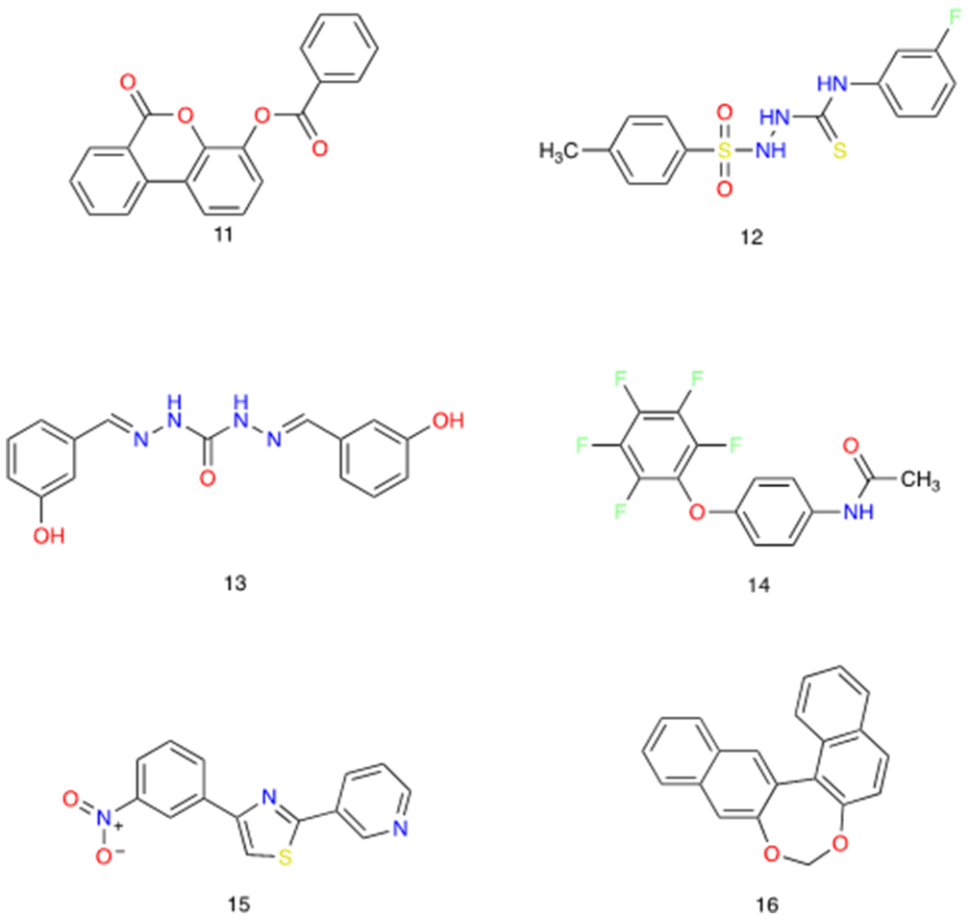
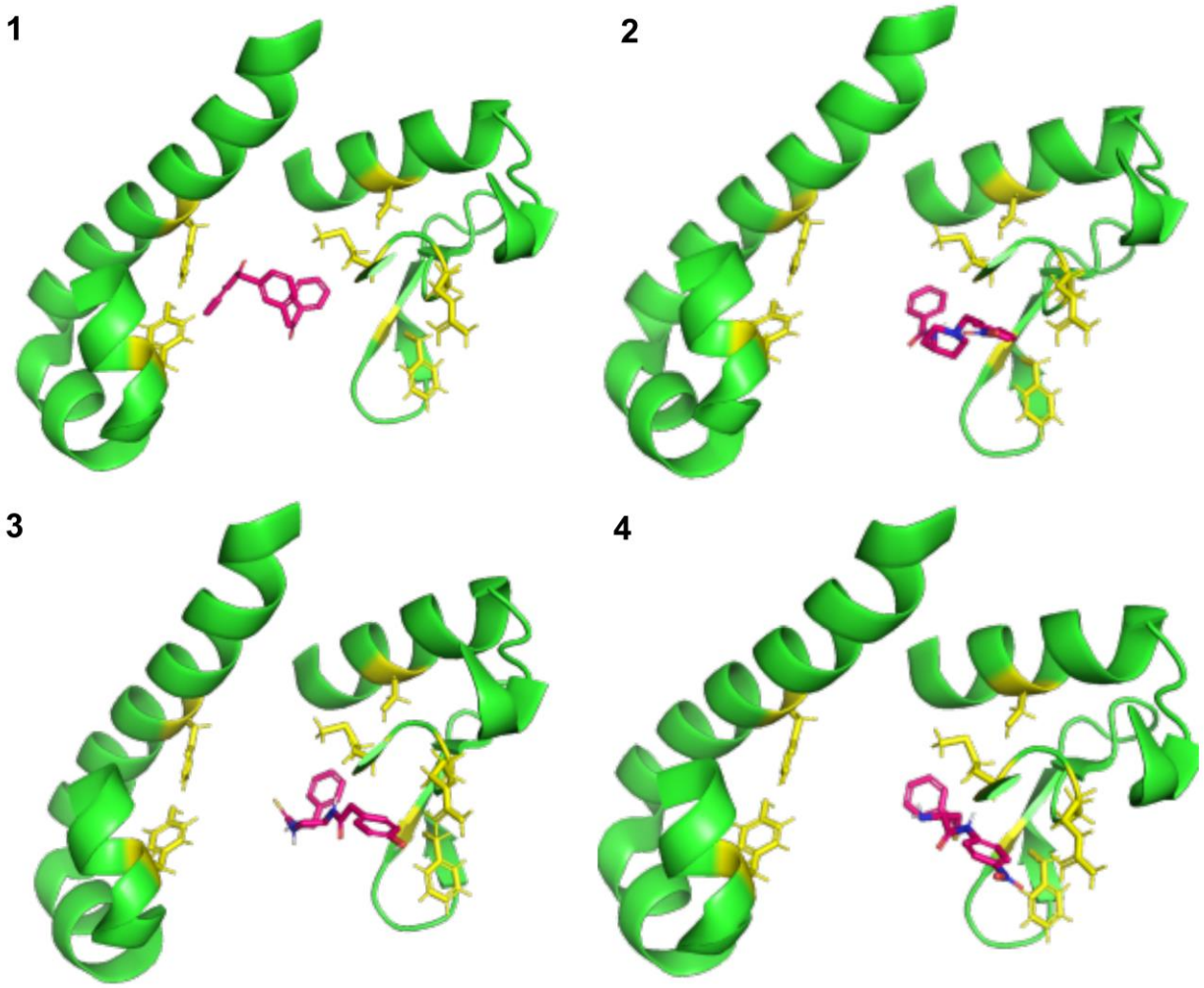
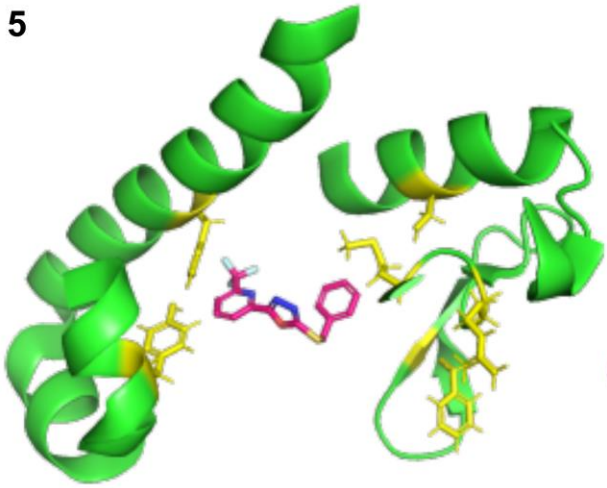


Figure 3.2. Top Fragments.

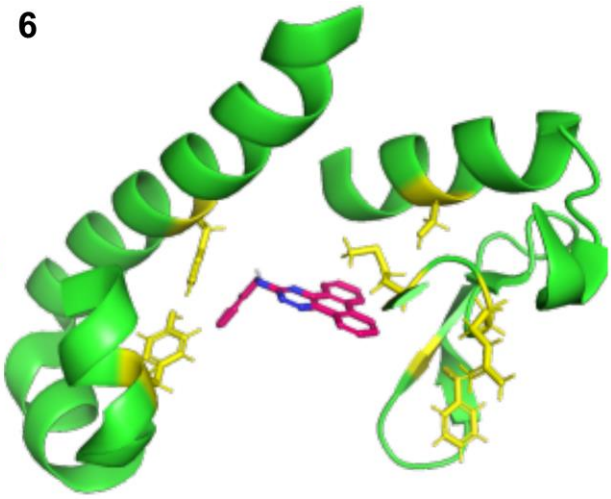
Through docking against the cleaned protein model by the Fragsite, smina, and random forest methods, 16 fragments were identified to bind potentially the active site of the protein.


1
2
3
4

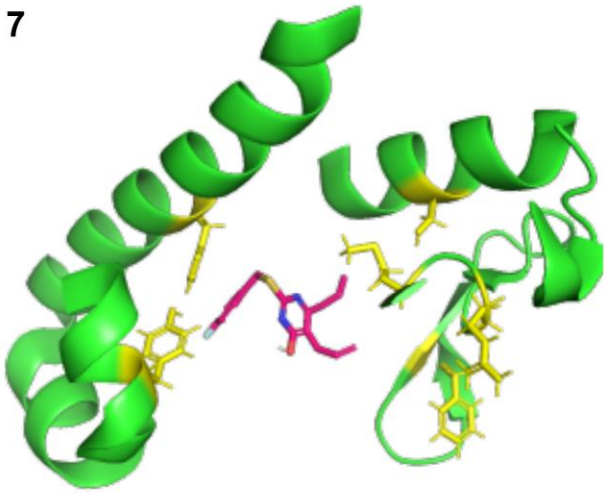
5



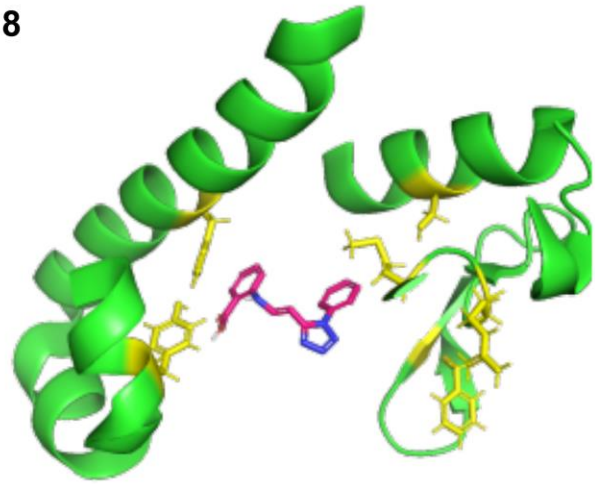
6



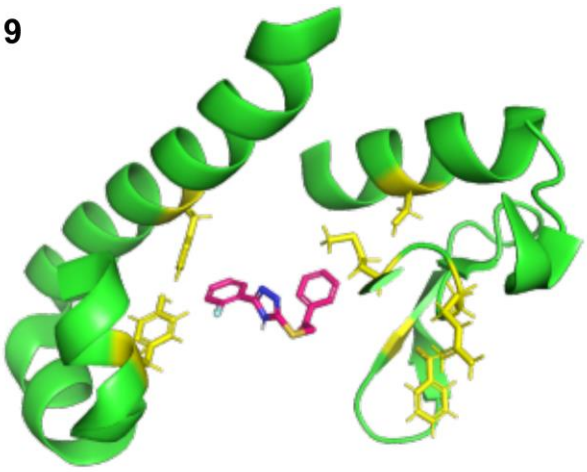
7



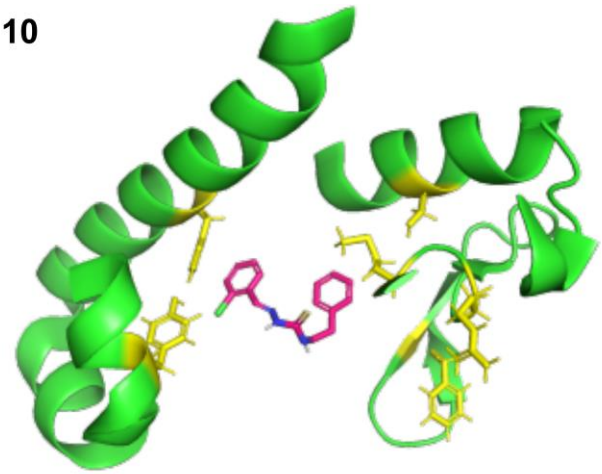
8



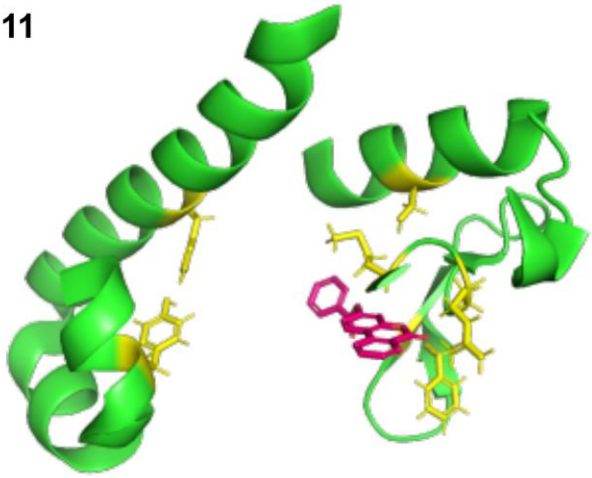
9



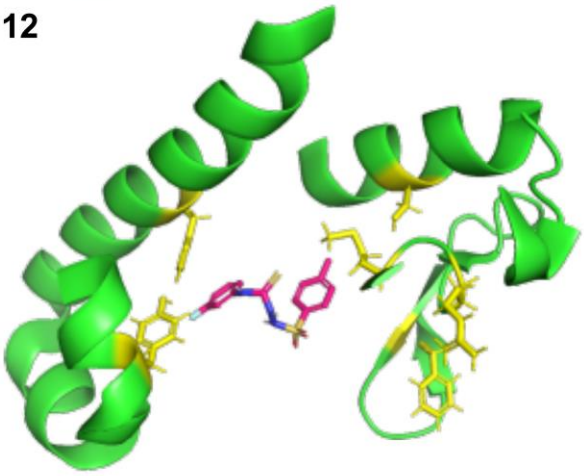
10



11



12



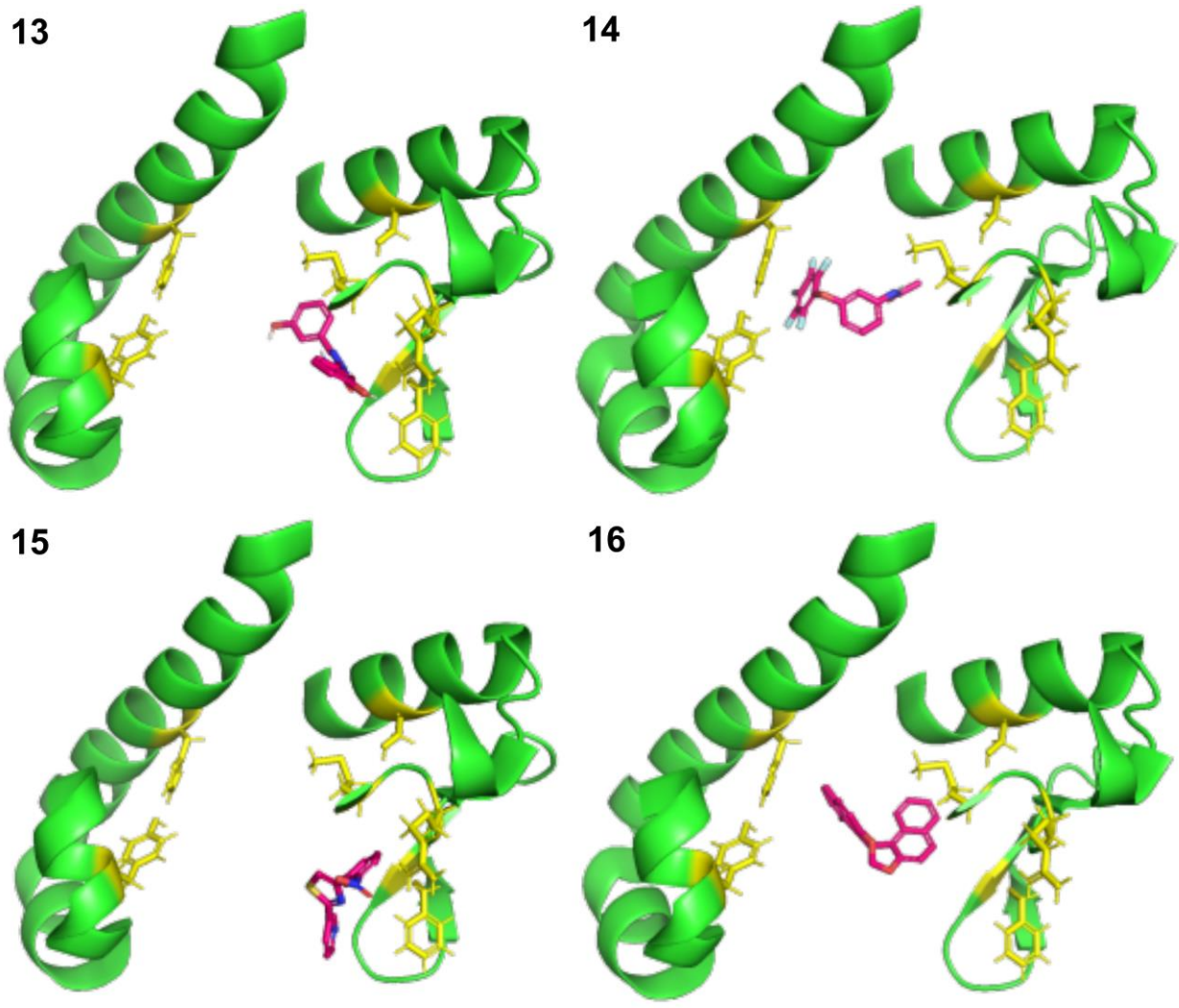


Figure 3.3. Fragments Docked to the Ligand Binding Domain of RORγ (PDB Id: 6XAE).

The Pymol diagrams show the 3-dimensional docking predictions of the 16 fragments to the ligand binding domain of RORγ. The protein is presented in ribbons with green color, and the residues interacting with the ligand (magenta) are represented by yellow sticks.

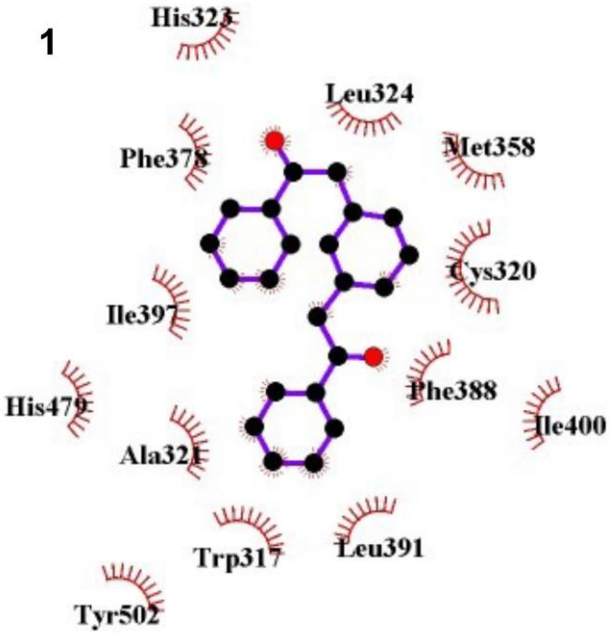
Table 3.1. Chemical Names and Molecular Details of the Top Fragments

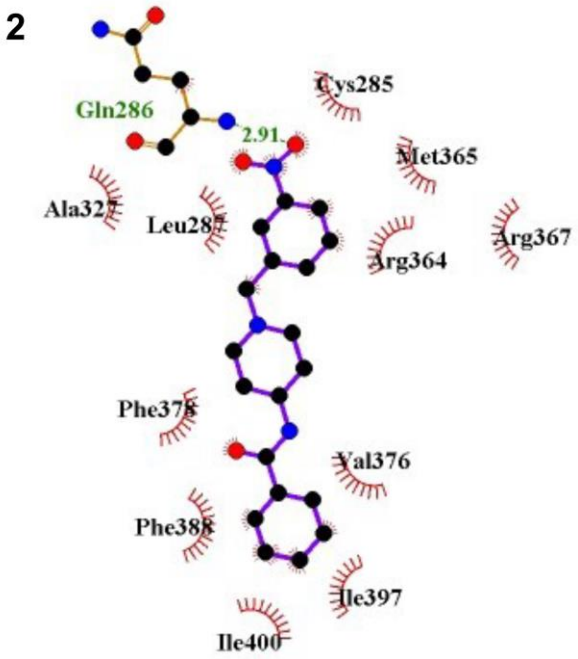
Fragment	MolPort/Chemspace Catalog Numbers	M.W. (Da)	PSA* (A^2)	log P
1	MolPort-002-927-269	366.41	34.14	4.76
2	MolPort-002-915-866	339.40	46.38	-2.46
3	MolPort-002-912-863	343.45	73.39	2.25
4	MolPort-046-591-928	331.39	98.26	2.85
5	MolPort-002-926-423	337.32	51.81	4.34
6	MolPort-002-911-222	336.40	50.70	4.69
7	MolPort-002-916-755	318.41	46.01	5.79
8	MolPort-002-910-914	307.31	92.93	3.12
9	MolPort-002-923-558	313.35	58.64	3.30
10	MolPort-000-466-748	302.82	24.39	4.51
11	MolPort-000-659-671	316.31	52.60	4.59
12	MolPort-002-912-339	339.40	70.23	3.44
13	MolPort-002-909-532	298.30	106.31	2.31
14	MolPort-003-882-167	317.22	38.33	3.42
15	CSC026570857	238.31	68.92	3.41
16	MolPort-001-918-059	298.34	18.46	5.22

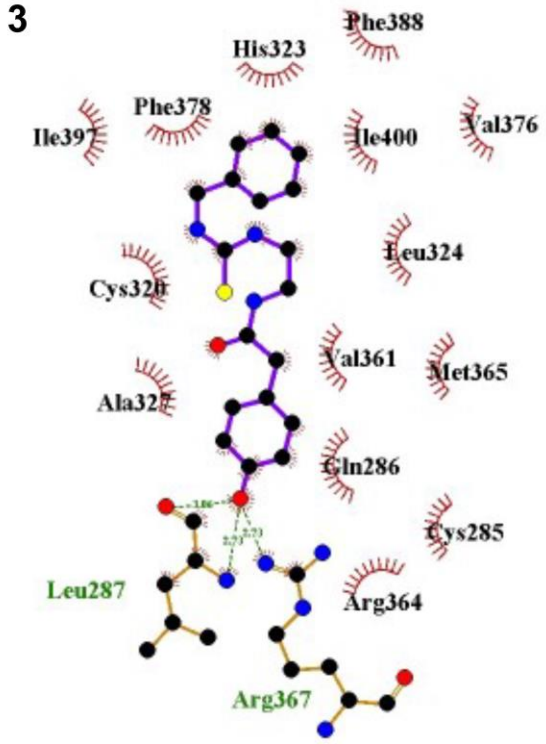
*** Polar Surface Area**

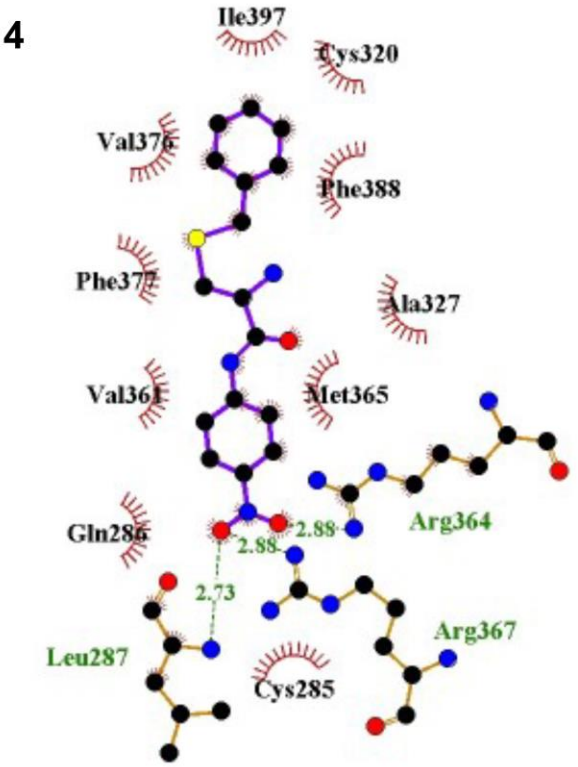
Table 3.2. Docking Scores of the Top Fragments

Fragment	Smina Score (kcal/mol)	K_d (mol/L)	Ligand Efficiency - (kcal/ (mol* atom))	RF-Score
1	-10.3	$2.81*10^{-8}$	0.468	6.34
2	-9.72	$7.45*10^{-8}$	0.405	6.23
3	-9.70	$7.76*10^{-8}$	0.422	6.36
4	-9.36	$1.38*10^{-7}$	0.425	6.30
5	-9.30	$1.51*10^{-7}$	0.404	6.40
6	-9.31	$1.50*10^{-7}$	0.358	6.36
7	-9.30	$1.52*10^{-7}$	0.423	6.31
8	-9.20	$1.83*10^{-7}$	0.399	6.28
9	-9.10	$2.16*10^{-7}$	0.413	6.34
10	-9.22	$1.73*10^{-7}$	0.439	6.36
11	-9.15	$1.96*10^{-7}$	0.381	6.26
12	-9.14	$1.98*10^{-7}$	0.416	6.25
13	-9.19	$1.84*10^{-7}$	0.417	6.47
14	-9.13	$2.01*10^{-7}$	0.415	6.33
15	-9.04	$2.35*10^{-7}$	0.452	6.51
16	-9.84	$6.05*10^{-8}$	0.410	6.23

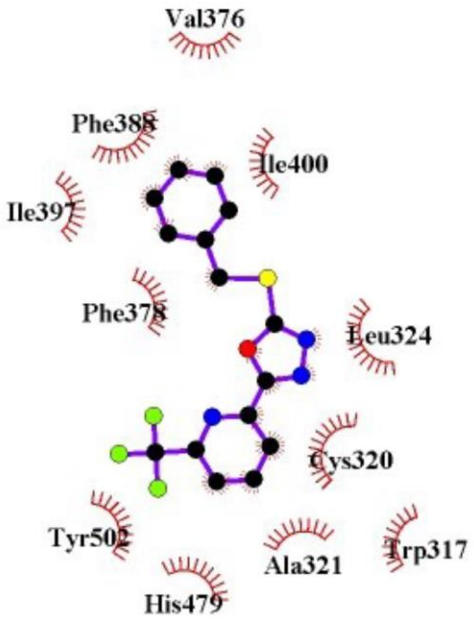

1
His323
Leu324
Phe378
Met358
Cys320
Ile397
His479
Phe388
Ile400
Ala321
Trp317
Leu391
Tyr502


2
Gln286
Cys285
2.91
Met365
Ala327
Leu287
Arg364
Arg367
Phe378
Val376
Phe388
Ile397
Ile400

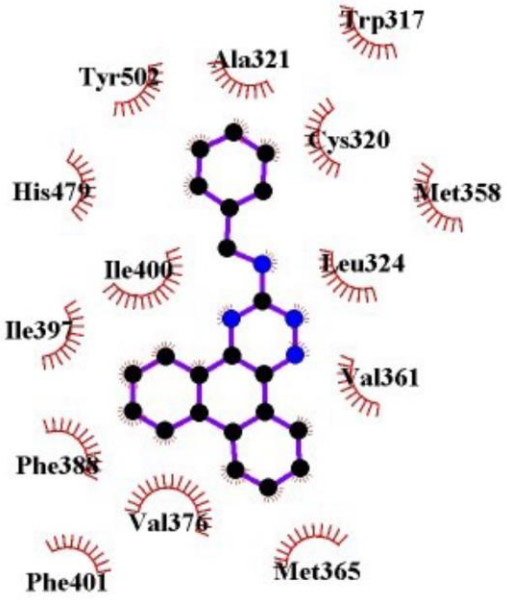

3
Phe388
His323
Phe378
Ile397
Ile400
Val376
Leu324
Cys320
Val361
Met365
Ala327
Gln286
Cys285
Leu287
Arg364
Arg367


4
Ile397
Cys320
Val376
Phe388
Phe377
Ala327
Val361
Met365
Arg364
Gln286
2.88
2.88
2.73
Arg367
Leu287
Cys285

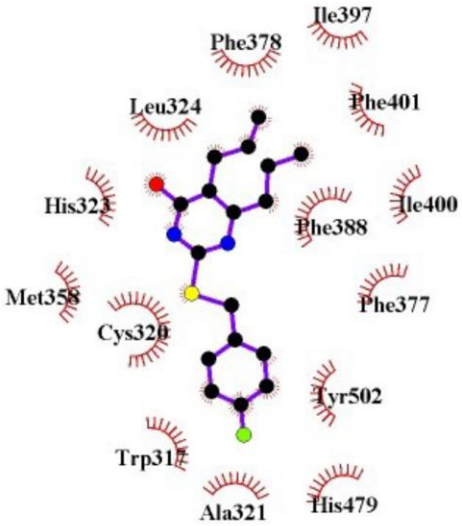
5


Val376
Phe388
Ile400
Ile397
Phe378
Leu324
Cys320
Tyr502
Ala321
Trp317
His479

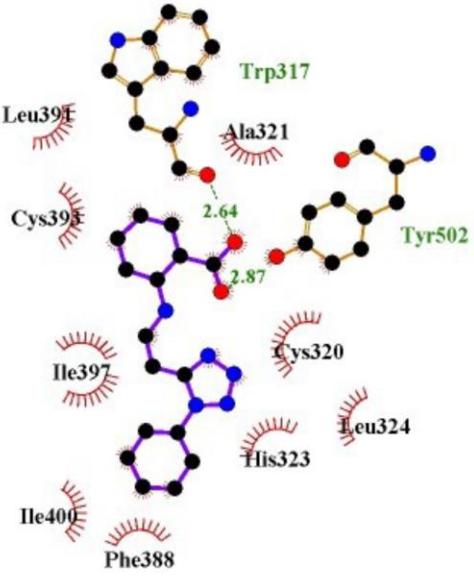
6


Trp317
Ala321
Tyr502
Cys320
His479
Met358
Ile400
Leu324
Ile397
Val361
Phe388
Val376
Met365
Phe401

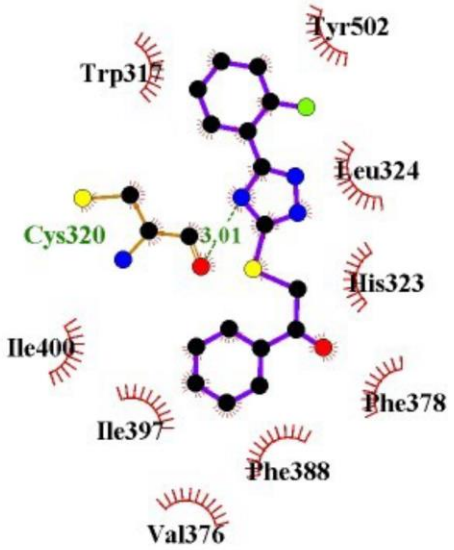
7


Ile397
Phe378
Leu324
Phe401
His323
Ile400
Phe388
Met358
Phe377
Cys320
Tyr502
Trp317
Ala321
His479

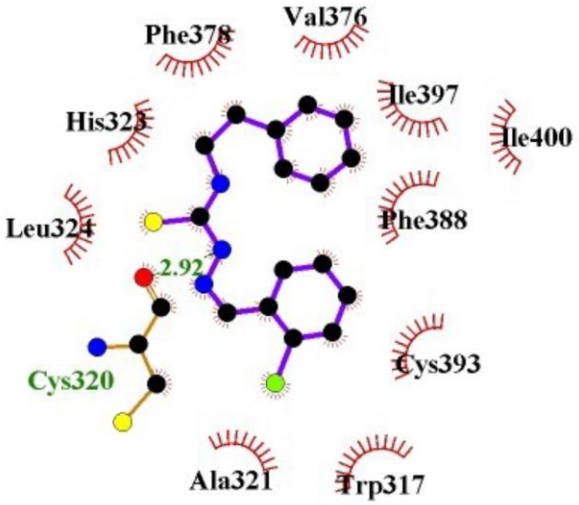
8


Trp317
Leu391
Ala321
2.64
Cys393
Tyr502
2.87
Cys320
Ile397
Leu324
His323
Ile400
Phe388

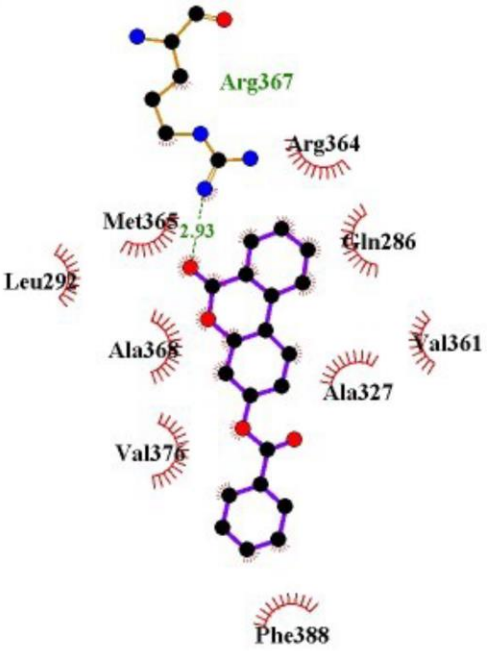
9


Tyr502
Trp317
Leu324
Cys320
3.01
His323
Ile400
Phe378
Ile397
Phe388
Val376

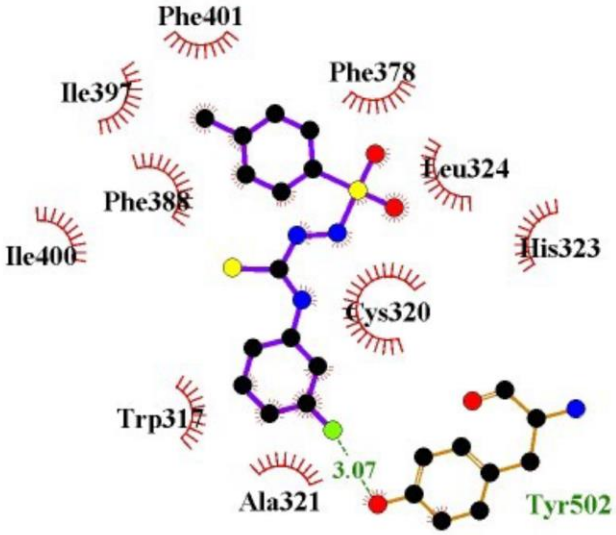
10


Val376
Phe378
Ile397
His323
Ile400
Leu324
Phe388
2.92
Cys393
Cys320
Ala321
Trp317

11


Arg367
Arg364
Met365
2.93
Gln286
Leu292
Ala368
Val361
Ala327
Val376
Phe388

12


Phe401
Ile397
Phe378
Leu324
Phe388
His323
Ile400
Cys320
Trp317
3.07
Ala321
Tyr502

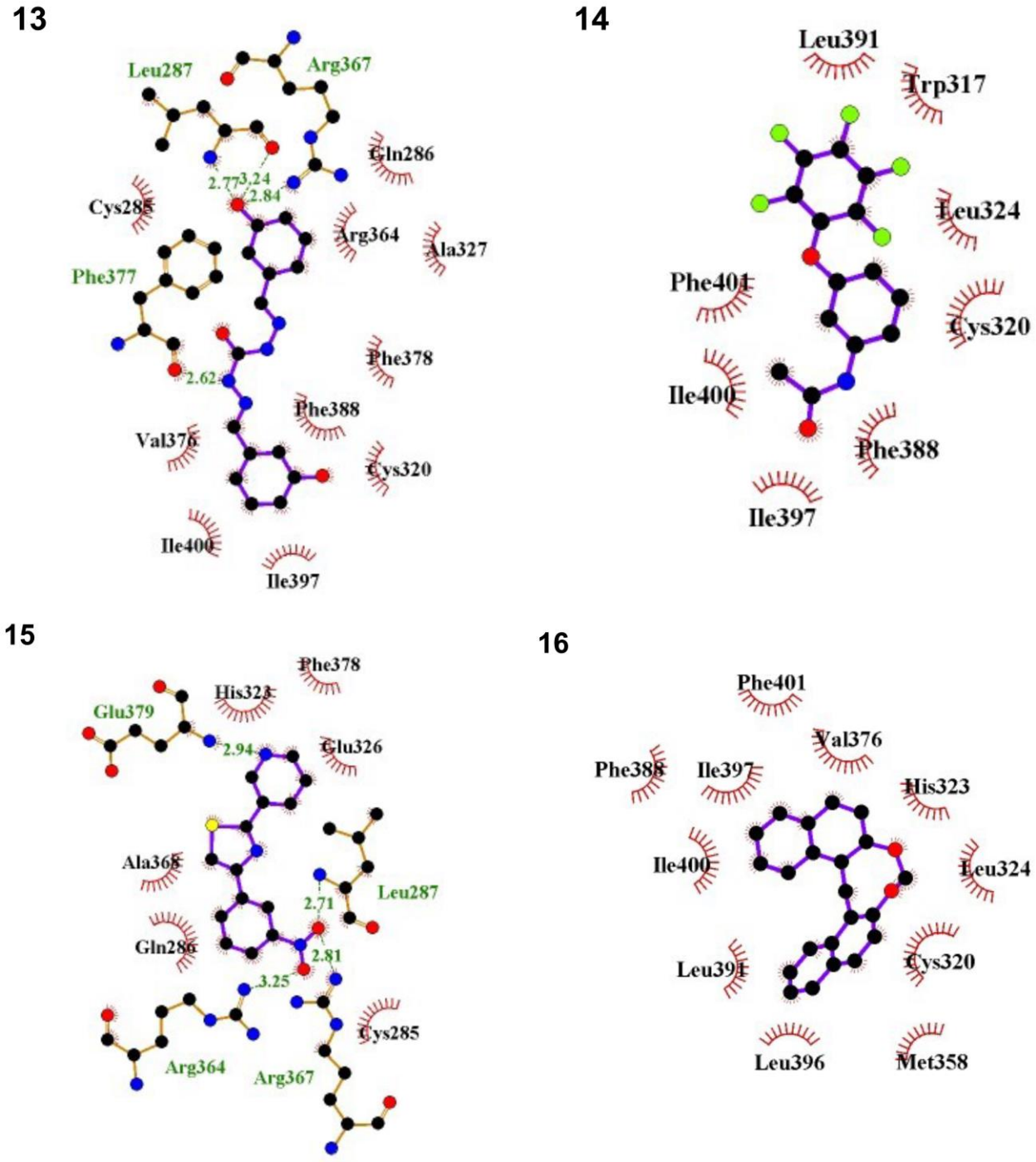


Figure 3.4. LigPlot Diagrams of the Fragments.

The two-dimensional interaction diagrams of the protein-fragment interactions. The fragments are presented in sticks with purple bonds, and other residues hydrogen-bonding against the

ligand are with green bonds. The hydrogen bond is presented in green dash lines, and the hydrophobic interactions are presented in red scattered lines.

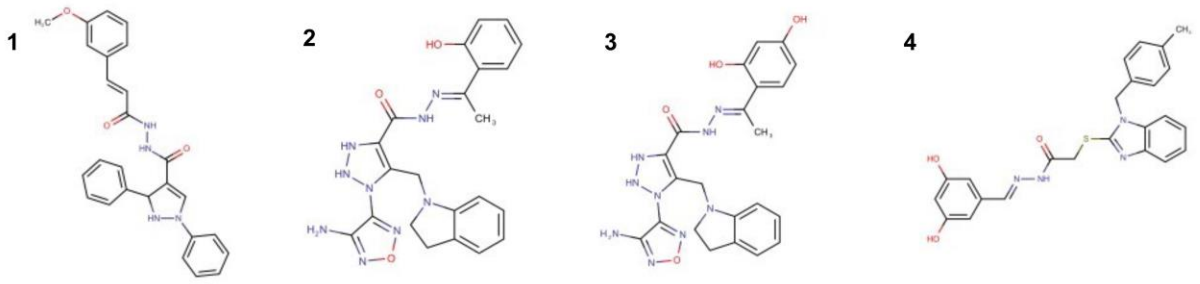


Figure 3.5. Compounds Identified with Pharmit.

Table 3.3. Compounds Identified with Pharmit

Compound	1	2	3	4
MolPort Catalog Number	MolPort-005-837-777	MolPort-002-585-480	MolPort-000-782-354	MolPort-000-445-187
Smina Score	-11.0	-10.2	-9.85	-11.1
RF-Score	7.03	6.73	6.72	6.64

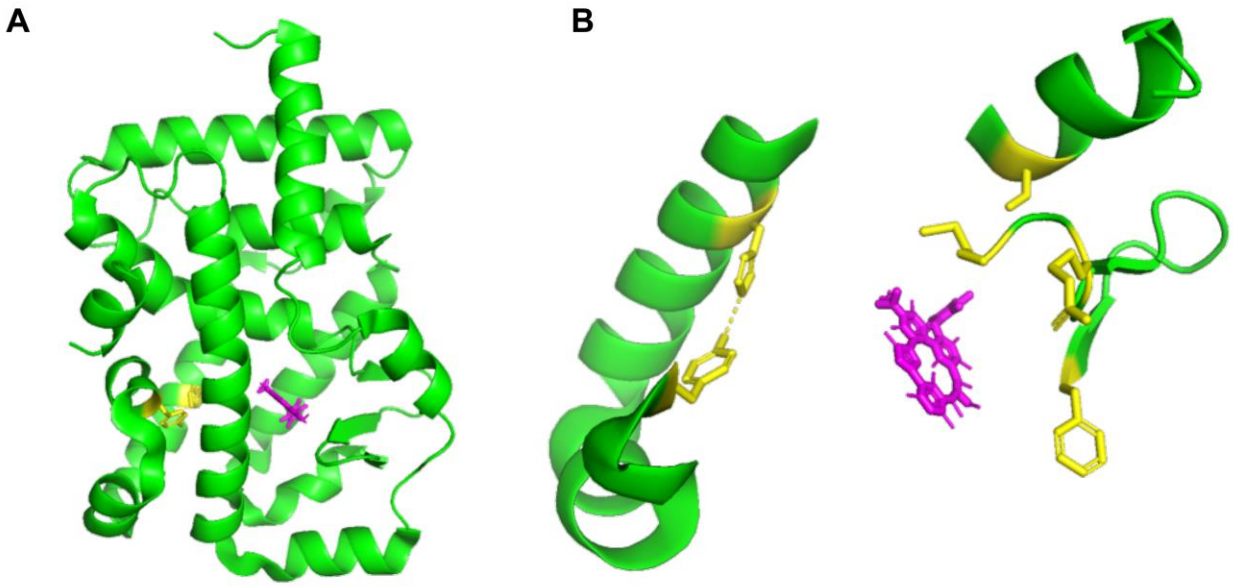


Figure 3.6. Fragment Linking.

Fragments 14 and 15 were linked together theoretically with two bonds and docked to the ligand binding domain of RORγ (PDB Id: 6XAE) using the YASARA platform. The Pymol diagrams show the best three-dimensional docking prediction. The protein is presented in ribbons with green color, and the critical residues are represented by yellow sticks. The ligand is shown in magenta.

Chapter 4 - Discussion and Conclusion

In this research, I compared different popular docking methods and did virtual screening using the MolPort and Chemspace databases. From all the comparisons of the ligand docking methods, 16 fragments were predicted to bind the active site of the ligand binding domain and interact with the same residues as the agonist compound found in the crystal structure. After docking, the energies and conformations of binding were compared among the fragments. Some fragments shared only hydrophobic interactions with the protein, while others shared both hydrogen bonds and hydrophobic interactions. Some fragments bind to the area of the binding pocket that is closest to the His479-Tyr502 hydrogen bond, while others are localized in the area that is closest to residues such as Arg367. Some fragments expand into both regions. Each fragment has different features that could dictate its degree of potency and whether it may behave as an agonist or antagonist. It is interesting how fragment 1 has the best smina score out of all the fragments (largest number with a negative sign), but a random forest score that is in the middle. Fragment 15 has the worst smina score, but the best (highest positive number) random forest score. This could be due to the difference in how the two scoring functions calculate docking scores. Smina measures the relative change in Gibbs free energy (ΔG), while the random forest score is determined by taking the negative logarithm of the K_d.

To interpret more fully the results, fragments 14 and 15 were linked together theoretically (Figure 3.6). These fragments each fit independently in the binding site. They were linked together theoretically with two bonds to fix their orientation with respect to each other. This combined fragment was docked to the cleaned ligand binding domain using the YASARA platform. The best docking pose indicates that this combined fragment binds with a predicted ΔG value of -14.32 kcal/mol. This ΔG value is better than those of the fragments before they were

linked together. The YASARA platform was also used to determine what residues this linked fragment interacts. This linked fragment shares only hydrophobic interactions with the residues Phe388, Phe378, His323, Ala327, Val361, Met365, Ile400, and Val376. It binds in the portion of the binding pocket that is closest to residues Arg367, F377, and Met365. It seems far away from His479 and Tyr502 and is unlikely to interrupt the agonist conformation of the ligand binding domain.

The compound hits identified here can be used as starting points for the development of new chemical series that are distinct from published compounds. These results offer experimental evidence as to what residues in the ligand binding pocket can be used as potential lead binding sites. They can also provide a basis for structure-based drug design to introduce polar interactions to optimize RORγ lead compounds. Potential polar interactions can be directed to engage Arg367 (from site 3), Phe377 (from multiple sites), and Ser404 and Met365 (site 1). The carbonyl of Phe377 can accept hydrogen bond interactions from different directions, that is, from either hotspot site 1, 3, or 4. These distinct hotspots as well as possible simultaneous binding of multiple fragments in the pocket provide inspiration in ligand design for hit-to-lead chemistry and lead optimization beyond fragment-based lead generation. It is evident that the fragment hits were able to probe the entire space of the ligand binding pocket and interactions with the corresponding residues in the different hotspots. The next step of this project would be to use the obtained compounds in Figure 3.5 to continue to look for compounds that have less similarity to known RORγ bioactive compounds. It would also be interesting to test some of those compounds experimentally to see if they have any bioactivity on RORγ. Increasingly, computational tools are providing models that enable data from one regime to inform the design of experiments in another, ultimately resulting in more efficient drug discovery.

References

1. Jetten, A. M. (2009) Retinoid-related orphan receptors (RORs): critical roles in development, immunity, circadian rhythm, and cellular metabolism. *Nucl. Recept. Signal.* 10.1621/nrs.07003

2. Huang, M., Bolin, S., Miller, H., and Ng, H. L. (2020) Rorγ structural plasticity and druggability. *Int. J. Mol. Sci.* **21**, 1–17

3. Jetten, A. M., and Joo, J. H. (2006) Retinoid-related orphan receptors (RORs): Roles in cellular differentiation and development. *Adv. Dev. Biol.* **16**, 313–355

4. Ivanov, I. I., McKenzie, B. S., Zhou, L., Tadokoro, C. E., Lepelley, A., Lafaille, J. J., Cua, D. J., and Littman, D. R. (2006) The Orphan Nuclear Receptor RORγt Directs the Differentiation Program of Proinflammatory IL-17+ T Helper Cells. *Cell*. **126**, 1121–1133

5. Ivanov, I. I., Zhou, L., and Littman, D. R. (2007) Transcriptional regulation of Th17 cell differentiation. *Semin. Immunol.* **19**, 409–417

6. Dong, C. (2008) TH17 cells in development: An updated view of their molecular identity and genetic programming. *Nat. Rev. Immunol.* **8**, 337–348

7. Leppkes, M., Becker, C., Ivanov, I. I., Hirth, S., Wirtz, S., Neufert, C., Pouly, S., Murphy, A. J., Valenzuela, D. M., Yancopoulos, G. D., Becher, B., Littman, D. R., and Neurath, M. F. (2009) RORγ-Expressing Th17 Cells Induce Murine Chronic Intestinal Inflammation via Redundant Effects of IL-17A and IL-17F. *Gastroenterology*. **136**, 257–267

8. Fan, J., Lv, Z., Yang, G., Liao, T. ting, Xu, J., Wu, F., Huang, Q., Guo, M., Hu, G., Zhou, M., Duan, L., Liu, S., and Jin, Y. (2018) Retinoic acid receptor-related orphan receptors: Critical roles in tumorigenesis. *Front. Immunol.* 10.3389/fimmu.2018.01187

9. Hu, X., Liu, X., Moisan, J., Wang, Y., Lesch, C. A., Spooner, C., Morgan, R. W., Zawidzka, E. M., Mertz, D., Bousley, D., Majchrzak, K., Kryczek, I., Taylor, C., Van Huis, C., Skalitzky, D., Hurd, A., Aicher, T. D., Toogood, P. L., Glick, G. D., Paulos, C. M., Zou, W., and Carter, L. L. (2016) Synthetic RORγ agonists regulate multiple pathways to enhance antitumor immunity. *Oncoimmunology*. **5**, 1–13

10. Hu, X., Majchrzak, K., Liu, X., Wyatt, M. M., Spooner, C. J., Moisan, J., Zou, W., Carter, L. L., and Paulos, C. M. (2018) In vitro priming of adoptively transferred T Cells with a RORγ agonist confers durablememory and stemness in vivo. *Cancer Res.* **78**, 3888–3898

11. Cui, G. (2019) TH9, TH17, and TH22 Cell Subsets and Their Main Cytokine Products in the Pathogenesis of Colorectal Cancer. *Front. Oncol.* **9**, 1–12

12. Qiu, R., and Wang, Y. (2018) Retinoic Acid Receptor-Related Orphan Receptor γt (RORγt) Agonists as Potential Small Molecule Therapeutics for Cancer Immunotherapy. *J. Med. Chem.* **61**, 5794–5804

13. Fauber, B. P., René, O., De Leon Boenig, G., Burton, B., Deng, Y., Eidenschenk, C., Everett, C., Gobbi, A., Hymowitz, S. G., Johnson, A. R., La, H., Liimatta, M., Lockey, P., Norman, M., Ouyang, W., Wang, W., and Wong, H. (2014) Reduction in lipophilicity improved the solubility, plasma-protein binding, and permeability of tertiary sulfonamide RORc inverse agonists. *Bioorganic Med. Chem. Lett.* **24**, 3891–3897

14. Sun, N., Guo, H., and Wang, Y. (2019) Retinoic acid receptor-related orphan receptor gamma-t (RORγt) inhibitors in clinical development for the treatment of autoimmune diseases: a patent review (2016-present). *Expert Opin. Ther. Pat.* **29**, 663–674

15. Pandya, V. B., Kumar, S., Sachchidanand, Sharma, R., and Desai, R. C. (2018) Combating Autoimmune Diseases with Retinoic Acid Receptor-Related Orphan Receptor-γ (RORγ or RORc) Inhibitors: Hits and Misses. *J. Med. Chem.* **61**, 10976–10995

16. Zhang, Y., Xue, X., Jin, X., Song, Y., Li, J., Luo, X., Song, M., Yan, W., Song, H., and Xu, Y. (2014) Discovery of 2-oxo-1,2-dihydrobenzo[cd]indole-6-sulfonamide derivatives as new RORγ inhibitors using virtual screening, synthesis and biological evaluation. *Eur. J. Med. Chem.* **78**, 431–441

17. Song, Y., Xue, X., Wu, X., Wang, R., Xing, Y., Yan, W., Zhou, Y., Qian, C. N., Zhang, Y., and Xu, Y. (2016) Identification of N -phenyl-2-(N -phenylphenylsulfonamido)acetamides as new RORγ inverse agonists: Virtual screening, structure-based optimization, and biological evaluation. *Eur. J. Med. Chem.* **116**, 13–26

18. Harigua-Souiai, E., Cortes-Ciriano, I., Desdouits, N., Malliavin, T. E., Guizani, I., Nilges, M., Blondel, A., and Bouvier, G. (2015) Identification of binding sites and favorable ligand binding moieties by virtual screening and self-organizing map analysis. *BMC Bioinformatics*. **16**, 11–15

19. Koes, D. R., Baumgartner, M. P., and Camacho, C. J. (2013) Lessons learned in empirical scoring with smina from the CSAR 2011 benchmarking exercise. *J. Chem. Inf. Model.* **53**, 1893–1904

20. Dastmalchi, S., Hamzeh-Mivehroud, M., and Sokouti, B. (2016) *Methods and Algorthms for Molecular Docking-Based Drug Design and Discovery*, Medical Information Scientific Reference

21. Ho, T. K. (1998) The random subspace method for constructing decision forests. *IEEE Trans. Pattern Anal. Mach. Intell.* **20**, 832–844

22. Ballester, P. J., and Mitchell, J. B. O. (2010) A machine learning approach to predicting protein-ligand binding affinity with applications to molecular docking. **26**, 1169–1175

23. Wójcikowski, M., Ballester, P. J., and Siedlecki, P. (2017) Performance of machine-learning scoring functions in structure-based virtual screening. *Sci. Rep.* **7**, 1–10

24. Sunseri, J., and Koes, D. R. (2016) Pharmit: interactive exploration of chemical space. *Nucleic Acids Res.* **44**, W442–W448

25. Koes, D. R. (2016) Pharmacophore Modeling: Methods and Applications. *Methods Pharmacol. Toxicol.*

26. Dror, O., Shulman-peleg, A., Nussinov, R., and Wolfson, H. J. (2004) Predicting Molecular Interactions in silico: I. A Guide to Pharmacophore Identification and its Applications to Drug Design. *Curr. Med. Chem.* **11**, 71–90

27. Flower, D. R. (1998) On the properties of bit string-based measures of chemical similarity. *J. Chem. Inf. Comput. Sci.* **38**, 379–386

28. Joseph-McCarthy, D., Campbell, A. J., Kern, G., and Moustakas, D. (2014) Fragment-based lead discovery and design. *J. Chem. Inf. Model.* **54**, 693–704

29. Xue, Y., Guo, H., and Hillertz, P. (2016) Fragment Screening of RORγt Using Cocktail Crystallography: Identification of Simultaneous Binding of Multiple Fragments. *ChemMedChem*. **11**, 1881–1885

30. Harikrishnan, L. S., Gill, P., Kamau, M. G., Qin, L. Y., Ruan, Z., O'Malley, D., Huynh, T., Stachura, S., Cavallaro, C. L., Lu, Z., J.-W. Duan, J., Weigelt, C. A., Sack, J. S., Ruzanov, M., Khan, J., Gururajan, M., Wong, J. J., Huang, Y., Yarde, M., Li, Z., Chen, C., Sun, H., Borowski, V., Murtaza, A., and Fink, B. E. (2020) Substituted benzyloxytricyclic compounds as retinoic acid-related orphan receptor gamma t (RORγt) agonists. *Bioorganic Med. Chem. Lett.* **30**, 127204

31. Olsson RI, Xue Y, von Berg S, Aagaard A, McPheat J, Hansson EL, Bernström J, Hansson P, Jirholt J, Grindebacke H, Leffler A, Chen R, Xiong Y, Ge H, Hansson TG, N. F. (2016) Benzoxazepines Achieve Potent Suppression of IL-17 Release in Human T-Helper 17 (TH 17) Cells through an Induced-Fit Binding Mode to the Nuclear Receptor RORγ. *ChemMedChem*. **11**, 207–216

32. René, O., Fauber, B. P., Boenig, G. D. L., Burton, B., Eidenschenk, C., Everett, C., Gobbi, A., Hymowitz, S. G., Johnson, A. R., Kiefer, J. R., Liimatta, M., Lockey, P., Norman, M., Ouyang, W., Wallweber, H. A., and Wong, H. (2015) Minor structural change to tertiary sulfonamide RORC ligands led to opposite mechanisms of action. *ACS Med. Chem. Lett.* **6**, 276–281

33. Erlanson, D. A., Fesik, S. W., Hubbard, R. E., Jahnke, W., and Jhoti, H. (2016) Twenty years on: The impact of fragments on drug discovery. *Nat. Rev. Drug Discov.* **15**, 605–619

34. Giordanetto, F., Jin, C., Willmore, L., Feher, M., and Shaw, D. E. (2019) Fragment Hits: What do They Look Like and How do They Bind? *J. Med. Chem.* **62**, 3381–3394

Appendix A - Copyright Permissions

Jetten, A. M. (2009) Retinoid-related orphan receptors (RORs): critical roles in development, immunity, circadian rhythm, and cellular metabolism. *Nucl. Recept. Signal.* 10.1621/nrs.07003

doi: 10.1621/nrs.07003

Ballester, P. J., and Mitchell, J. B. O. (2010) A machine learning approach to predicting protein-ligand binding affinity with applications to molecular docking. **26**, 1169–1175

https://s100.copyright.com/CustomerAdmin/PLF.jsp?ref=266138ee-2293-4c29-8866-c14efc5d0a10

Sunseri, J., and Koes, D. R. (2016) Pharmit: interactive exploration of chemical space. *Nucleic Acids Res.* **44**, W442–W448

Creative Commons

https://s100.copyright.com/AppDispatchServlet?publisherName=OUP&publication=1362-4962&title=Pharmit%3A%20interactive%20exploration%20of%20chemical%20space&publicationDate=2016-04-19&volumeNum=44&issueNum=W1&author=Sunseri%2C%20Jocelyn%3B%20Koes%2C%20David%20Ryan&startPage=W442&endPage=W448&contentId=10.1093%2Fnar%2Fgkw287&oa=CC%20BY&copyright=Oxford%20University%20Press&orderBeanReset=True

Flower, D. R. (1998) On the properties of bit string-based measures of chemical similarity. *J. Chem. Inf. Comput. Sci.* **38**, 379–386

PERMISSION/LICENSE IS GRANTED FOR YOUR ORDER AT NO CHARGE

This type of permission/license, instead of the standard Terms & Conditions, is sent to you because no fee is being charged for your order. Please note the following:

- Permission is granted for your request in both print and electronic formats, and translations.
- If figures and/or tables were requested, they may be adapted or used in part.
- "Reprinted (adapted) with permission from (Flower, D. R. (1998) On the properties of bit string-based measures of chemical similarity. *J. Chem. Inf. Comput. Sci.* **38**, 379–386). Copyright (1998) American Chemical Society."

- One-time permission is granted only for the use specified in your request. No additional uses are granted (such as derivative works or other editions). For any other uses, please submit a new request.
If credit is given to another source for the material you requested, permission must be obtained from that source.