Structure-feature relationship of nuclear receptors and fluorescent proteins

by

Mian Huang

B.S., South China Normal University, 2007
M.S., South China Normal University, 2011
M.S., University of Hawai‘i at Mānoa, 2017

AN ABSTRACT OF A DISSERTATION

submitted in partial fulfillment of the requirements for the degree

DOCTOR OF PHILOSOPHY

Department of Biochemistry and Molecular Biophysics
College of Arts and Sciences

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2022

Abstract

The protein structure-feature relationship describes how the three-dimensional structure of a protein impacts its feature. The structure-feature relationship of a fluorescent protein (FP) and two nuclear receptors (NRs) were explored in this work.

Residues noncovalently interacting with chromophores noticeably impact FP spectra features, but the effects of residues distant from chromophores have not been rigorously characterized. To understand these long-range effects, an FP called darkmRuby was derived from mRuby3 by mutating residues distant from the chromophore. darkmRuby shows dim and bell-shaped pH-dependent fluorescence different from mRuby3. darkmRuby was crystallized at pH 5.0, 8.0, and 9.0. The *in silico* analysis of the crystal models and site-directed mutagenesis of darkmRuby identify a long-range interaction responsible for its unusual features. Met94 and Phe96 regulate the conformation of His197, adjacent to the chromophore, over 15 Å, mediated by water molecules in a channel. The channel links the chromophore and His197 to external solvent, allowing water molecules to quench fluorescence and poor brightness. The solvent exposure also affects the His197 protonation state, resulting in pH sensitivity. This study provides the first detailed mechanism for long-range effects in FPs. This study extends the mechanistic understanding of FPs and empowers the rational design of new FP biosensors.

Human NR RORɣ is the master transcription factor of cytokine IL-17 in T helper 17 cells. Since the overexpression of IL-17 can lead to various autoimmune disorders, RORɣ has drawn attention as a potential drug target. In this project, we selected 15 compounds by computer methods for their potential for binding RORɣ and tested them by differential scanning fluorimetry (DSF). One compound, RG14-2, shows strong binding against the RORɣ ligand-binding domain (LBD). A reporter gene assay showed it inhibited RORɣ transcriptional activity

with an IC_{50} value of 1.5 μM. NMR experiments determined a moderate binding affinity of RG14-2 against the LBD with a K_d value of 5.7 μM. The co-crystal structure of the RG14-2-bound LBD demonstrates an atomic mechanism for how RG14-2 inhibits the RORɣ transcriptional function. RG14-2 is a promising lead compound for developing a novel class of RORɣ inverse agonists.

Breast cancer ranks as the second highest cause of death among cancers in women. About 70% of the cases are ERα-positive ($ERα^+$), making the NR ERα an ideal drug target for treating breast cancer. Unfortunately, almost half of the patients carrying $ERα^+$ breast cancer develop drug resistance caused by mutations in ERα. In this project, I targeted the most aggressive mutant, Y537S, and attempted to discover new ERα modulators defeating Y537S-mediated drug resistance. A virtual screening workflow was designed to select potential modulators from millions of small molecules. Five top-ranking compounds were chosen and tested by DSF. One compound, ERA1, binds tightly against both the wild-type ERα and the Y537S mutant, showing its potential as a lead compound to develop into a new generation of ERα modulators.

Structure-feature relationship of nuclear receptors and fluorescent proteins

by

Mian Huang

B.S., South China Normal University, 2007
M.S., South China Normal University, 2011
M.S., University of Hawai‘i at Mānoa, 2017

A DISSERTATION

submitted in partial fulfillment of the requirements for the degree

DOCTOR OF PHILOSOPHY

Department of Biochemistry and Molecular Biophysics
College of Arts and Sciences

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2022

Approved by:

Major Professor
Ho Leung Ng

Copyright

Abstract

The protein structure-feature relationship describes how the three-dimensional structure of a protein impacts its feature. The structure-feature relationship of a fluorescent protein (FP) and two nuclear receptors (NRs) were explored in this work.

Residues noncovalently interacting with chromophores noticeably impact FP spectra features, but the effects of residues distant from chromophores have not been rigorously characterized. To understand these long-range effects, an FP called darkmRuby was derived from mRuby3 by mutating residues distant from the chromophore. darkmRuby shows dim and bell-shaped pH-dependent fluorescence different from mRuby3. darkmRuby was crystallized at pH 5.0, 8.0, and 9.0. The *in silico* analysis of the crystal models and site-directed mutagenesis of darkmRuby identify a long-range interaction responsible for its unusual features. Met94 and Phe96 regulate the conformation of His197, adjacent to the chromophore, over 15 Å, mediated by water molecules in a channel. The channel links the chromophore and His197 to external solvent, allowing water molecules to quench fluorescence and poor brightness. The solvent exposure also affects the His197 protonation state, resulting in pH sensitivity. This study provides the first detailed mechanism for long-range effects in FPs. This study extends the mechanistic understanding of FPs and empowers the rational design of new FP biosensors.

Human NR RORɣ is the master transcription factor of cytokine IL-17 in T helper 17 cells. Since the overexpression of IL-17 can lead to various autoimmune disorders, RORɣ has drawn attention as a potential drug target. In this project, we selected 15 compounds by computer methods for their potential for binding RORɣ and tested them by differential scanning fluorimetry (DSF). One compound, RG14-2, shows strong binding against the RORɣ ligand-binding domain (LBD). A reporter gene assay showed it inhibited RORɣ transcriptional activity

with an IC_{50} value of 1.5 μM. NMR experiments determined a moderate binding affinity of RG14-2 against the LBD with a K_d value of 5.7 μM. The co-crystal structure of the RG14-2-bound LBD demonstrates an atomic mechanism for how RG14-2 inhibits the RORɣ transcriptional function. RG14-2 is a promising lead compound for developing a novel class of RORɣ inverse agonists.

Breast cancer ranks as the second highest cause of death among cancers in women. About 70% of the cases are ERα-positive ($ERα^+$), making the NR ERα an ideal drug target for treating breast cancer. Unfortunately, almost half of the patients carrying $ERα^+$ breast cancer develop drug resistance caused by mutations in ERα. In this project, I targeted the most aggressive mutant, Y537S, and attempted to discover new ERα modulators defeating Y537S-mediated drug resistance. A virtual screening workflow was designed to select potential modulators from millions of small molecules. Five top-ranking compounds were chosen and tested by DSF. One compound, ERA1, binds tightly against both the wild-type ERα and the Y537S mutant, showing its potential as a lead compound to develop into a new generation of ERα modulators.

Table of Contents

List of Figures

List of Tables

Acknowledgements

Thank Dr. Ho Leung Ng for believing in me all the time, encouraging me to overcome difficulties, and supporting me to try new ideas and techniques. Thank my previous and current lab colleagues, Samson, Ye, John, Bo, and Bowen. Thank the RORγ team members, Hannah, Shelby, Anna, and the ERα team members, Jacob and Alex.

I would like to thank my committee members, Dr. Brian Geisbrecht and Dr. Om Prakash for generously sharing knowledge, facilities, and advice with me, Dr. Zhilong Yang for your time and feedback.

Thank all faculties and graduate students at the Department of Biochemistry and Molecular Biophysics in Kansas State University for always being supportive. Especially, thank Nitin for the technical support in the NMR experiment.

Thank Dr. Jun Chu and Dr. Mengyin Deng for the professional support while collaborating on the darkmRuby project.

Thank all my friends and relatives around the world for their encouragement.

I would like to especially appreciate my parents, Zhenming Huang and Yuling Liang, for their unconditional love and support.

Chapter 1 - General Introduction

Structural biology is an interdisciplinary field for studying three-dimensional structures of macromolecules, especially proteins and nuclear acids, and the structure-feature relationship. It encompasses the principles and techniques of molecular biology, biochemistry, and biophysics.

Since protein tertiary and quaternary structures usually determine their characteristics and bioactivities, deciphering protein structures will enhance the understanding of life science topics to an atomic level. It is crucial for the research related to diseases and treatment. Many diseases are caused by inappropriate activation or overexpression of critical proteins such as IL-17-related diseases and ERα-positive breast cancer [1,2], and misfolded proteins and related aggregation such as Alzheimer's disease and Parkinson's disease Creutzfeldt-Jacob disease, mad cow disease, Type 2 diabetes mellitus [3–5]. Uncovering the structural grounds of the diseases allows pharmaceutical scientists to develop therapeutical strategies through structure-based drug design methods to improve drug efficacy dramatically. Also, 3D structures of some proteins displaying unique features, like fluorescent proteins and enzymes, can help identify critical amino acid residues or motifs in controlling the protein features and then manipulate the features in protein engineering.

Three biophysical techniques, X-ray crystallography, nuclear magnetic resonance (NMR) spectroscopy, and cryogenic electron microscopy (cryo-EM), are widely applied to solve 3D structures of proteins and protein-ligand complexes. All of them show advantages and disadvantages. X-ray crystallography can provide high-resolution crystal models of proteins in a broad range of molecular weight, but a lot of proteins are tough to crystallize, and the structures obtained could be distinct from the natural protein state in solution. NMR spectroscopy is

suitable for investigating protein structures and dynamics in solution, but it usually deals with relatively small proteins (< 40-50 kD), and the protein samples need to be isotope-labeled and prepared in high purity. cryo-EM can handle the macromolecule systems in large size and offer structures in the native state, but the models are solved in poor resolution (> 3.5 Å). Structural biologists often choose the technique based on the sample condition and the project purpose. In this work, I applied X-ray crystallography to successfully build crystal models of two different proteins and a protein-ligand complex. I also used NMR spectroscopy to measure the binding affinity of a small molecule against a protein.

Two families of proteins, fluorescent proteins and nuclear receptors, were intimately involved in the projects. They were briefly introduced as follows.

Fluorescent proteins

The first green fluorescent protein (FP), avGFP, was isolated from jellyfish *Aequorea Victoria* in 1962 [6], and its crystal structure was solved in 1996 (Figure 1.1) [7,8]. It is composed of an 11-stranded β-barrel and a chromophore, which is embedded in the core of the barrel and connects to the barrel with an α-helix. The chromophore is formed by autocyclization of three amino acids, Ser65, Tyr66, Gly67, under oxygen participation [9,10].

Ever since that, other naturally existing FPs have been discovered in over 100 species of organisms originated from 16 divisions, including bacteria, cyanobacteria, green algae, ascomycetes, bony fishes, coral anemones, crustaceans, eudicots, hydrozoans, lancelets, mat anemones, sea anemones, sea pens, soft corals, stony corals, and tube anemones [11]. Some of them can emit fluorescence in other colors, besides green fluorescence. Based on the visible color, FPs are commonly classified into cyan/blue FPs, GFPs, yellow FPs, orange FPs, red FPs,

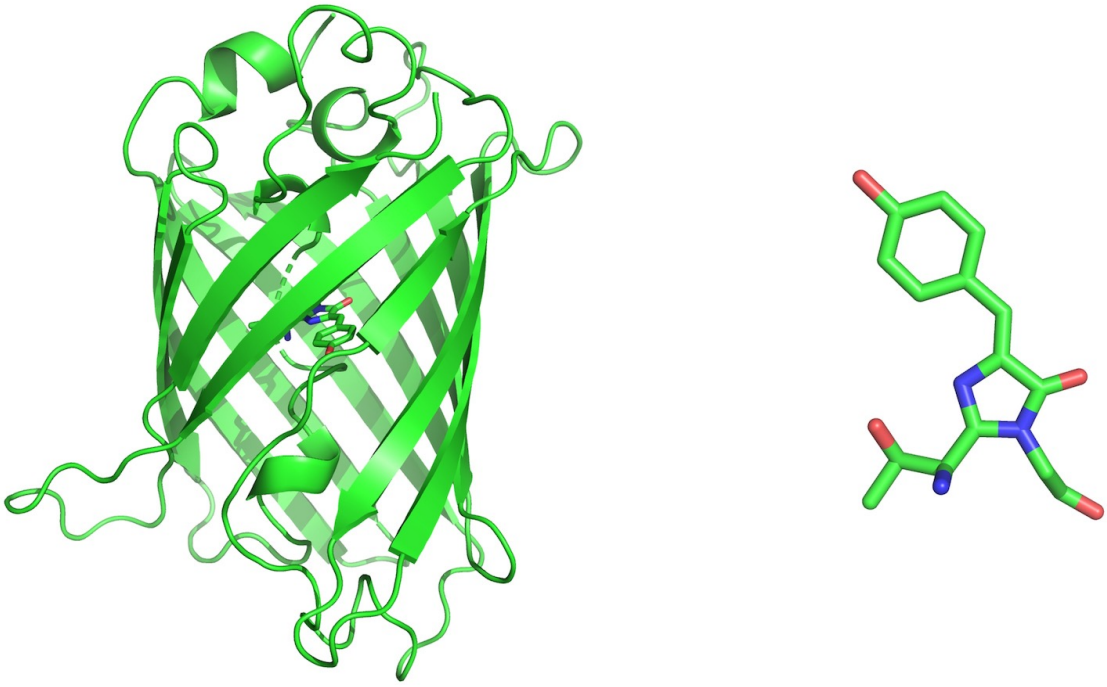


Figure 1.1 Crystal model of avGFP (left, 1EMA) and its chromophore (right).

and purple FPs. Most FPs share a protein architecture similar to avGFP and are designated as GFP-like FPs. However, the amino acid sequences of these GFP-like proteins are less conserved in the chromophore and barrel manners. For example, the chromophore in the yellow FP Zfp538 is constructed by Lys66, Tyr67, and Gly68 [12]; the first red FP DsRed forms its chromophore with Gln66, Tyr67, and Gly68 [13]; the red FP eqFP611, the ancestor of the FP I am interested here, cyclizes Met63, Tyr64, Gly65 to produce the chromophore [14]. It is believed that the chemical diversity of the chromophores, residues on the barrels, and their noncovalent interactions determine the features of the proteins, which include maturation, oligomerization, wavelengths of the maximal absorbance, excitation, and emission, brightness, pH stability, photostability, barrel rigidity, and so on. Hundreds of engineered FPs have been engineered from the natural

ones for decades through manipulating residue components. The visible spectrum has been well covered by both the naturally existing and engineered FPs.

In 1994, Chalfie et al. successfully expressed the avGFP in the first stage *Caenorhabditis elegans* larva without losing the green fluorescence or showing cytotoxic effect [15]. As a jumping-off point, it opened the door of a research field exploring FP applications. Until now, the FPs and FP-based biosensors have been widely applied in multiple aspects of life science studies, such as labeling from DNA, RNA, proteins to organelles and cells, tracking interactions between different biomolecules, sensing drug activities, and so on [16]. Driven by the application purposes, more and more FPs are derived from the naturally existing or previously engineered ones.

With an increasing number of FP crystal structures solved, scientists have analyzed the structural information and rationalized the potential roles of some critical residues in the protein features. It pushes the field to switch the techniques from random mutagenesis to structure-based FP design. Chapter 2 structurally studied a new dark FP engineered from a bright protein and proposed an atomic mechanism of a long-range regulation on the protein features, which has never been discussed before.

Nuclear receptors

Human nuclear receptors (NRs) are a superfamily of 48 transcription factors in the human genome. They all play vital roles in diverse human body development and metabolism processes. For example, estrogen receptor α (ERα) is responsible for maturation of male and female reproductive phenotypes, maintaining bone integrity, and central nervous development [17]; peroxisome proliferator-activated receptors (PPAR) participates in multiple aspects of energy homeostasis and metabolic function, including triglyceride level, insulin sensitization, and fatty

acid metabolism [18]; vitamin D receptor is involved in intestinal and renal transport of calcium and other minerals [19]; retinoic acid-related orphan receptor γ (RORγ) promotes T cell differentiation into the T helper 17 (Th17) subtype and regulates the IL-17 expression in Th17 cells [20,21]. In the meantime, the disorder of their transcriptional functions due to inappropriate activation or deactivation and overexpression of NRs can lead to multiple serious diseases, such as autoimmune diseases and cancers.

Except for several special ones, NRs are composed of five domains, an N-terminal domain (NTD), a DNA-binding domain (DBD), a hinge region (HR), a ligand-binding domain (LBD), and a C-terminal domain (CTD). In most NRs, NTDs exist as random loops and contain the activator Function-1 (AF-1) region to interact with cofactors promoter-specifically [22]. DBDs are highly conserved in NRs, containing two zinc fingers to clamp the DNA sequence sites [23]. HRs are loop linkers in the least sequence and size conservation; post-translational modifications related to signal transfer are found on HRs [24]. LBDs are globular protein domains with the orthostatic binding pockets located in the core; when a small-molecule ligand embeds in the pocket, the AF-2 region on the LBD surface is conformationally impacted and recruiting a co-activator or co-depressor, resulting in triggering or inhibiting the transcriptional function. CTDs with variable sequence lengths are found after LBDs, but their functions are poorly understood. The proteins can attach to the DNA sequences in monomer, homodimer, or heterodimer. Because of the generally flexible architectures of NTDs and HRs in the proteins, it is challenging to obtain the structures of a full-length NR. Only several 3D structures of heterodimers and homodimers are solved by X-ray crystallography, which are the PPARγ-RXR heterodimer (3DZU, 3DZY, 3E00) [25], the RXR-LXE heterodimer (4NQA) [26], the RARβ-RXRα heterodimer (5UAN) [27], and the HNF-4α homodimer (4IQR) [28].

Despite lacking full-length models for mapping signal transfer between domains, it is commonly accepted that the LBDs take the primary responsibility of interacting with ligands and controlling the transcriptional activity of NRs. Over 13% of FDA-approved drugs target NR LBDs. For instance, tamoxifen, toremifene, raloxifene, and fulvestrant bind against ERα; flutamide, bicalutamide, enzalutamide, and apalutamide are antagonists of androgen receptor; tretinoin is a RAR agonist [29].

In Chapter 3 and 4, I investigated the novel small-molecule modulators binding against the LBDs of RORγ and the ERα mutant by computational and experimental methods. Detailed introduction of the proteins was presented in the relative chapters.

Chapter 2 - A Long-Range Interaction Regulating Brightness and pH Stability of a Dark Fluorescent Protein

Abstract: In fluorescent proteins (FPs), residues noncovalently interacting with chromophores noticeably impact FP spectra features, but those distal from chromophores lack study on their roles in the features. To decipher the unknown, the collaborating lab derived an FP model darkmRuby from mRuby3 through mutating residues far from the chromophore. Compared with mRuby3, darkmRuby presents a dim and bell-shaped pH-dependent fluorescence. The crystal models of darkmRuby show a conformational change of the chromophore-surrounding residue His197 at different pHs, hinting that the mutated residues impact the protein through His197. B-factor analysis and site-directed mutagenesis guide me to a double mutant M94T/F96Y, which recovers the intense fluorescence brightness and pH resistance of the parent protein. The crystal model of the mutant confirms the structural regulation of Met94 and Phe96 on His197, evoking the existence of a long-range interaction in a 15-Å distance. Molecular dynamics simulations and channel prediction hint that the interaction is mediated by water molecules diffused in a channel connecting Met94 and Phe96 to His197. The channel exposes the chromophore and His197 to the solvent environment, allowing water molecules to quench the fluorescence emitted by the chromophore, resulting in weakened brightness. Moreover, it impacts the protonation state of His197, leading to pH sensitivity of the protein. To my best knowledge, this is the first long-range interaction reported in FPs. This study proves the role of the faraway residues as crucial as the close ones in modulating FP features. It will extend the understanding of FPs and offer fundamental knowledge of new strategies to pilot the rational design of new FP biosensors.

2.1 Introduction

Fluorescent protein (FP) was often developed using random mutagenesis and feature-based screening, resulting in mutated residues spread in the entire protein. With more and more FP crystal structures accessible, a list of residues is found playing critical roles in controlling protein characteristics. For example, the conformation of Arg67 determines the photobleaching and photoconverting features of mEos2 and Dendra2 [30]; the positively charged Arg92 is intimately involved in rapid autocyclization of the avGFP chromophore [31,32]; chemical properties of residues 143 and 158 strongly affect the *cis/trans* states of eqFP611 and mKate chromophores [33,34]; types and intensity of the interaction between chromophores and residue 197 can change the spectra features of FPs [35,36]. The structural information has fed back into the FP development and pushed the field further. Interestingly, all residues mentioned above posit closely (< 4 Å) to FP chromophores and directly contact them with hydrogen bonding, π-stacking, and/or hydrophobic effects. To ease the following discussion, I would like to assign this kind of residues as inner-shell residues (ISRs) which non-covalently interact with FP chromophores within 4 Å, and the rest as outer-shell residues (OSRs). Compared with ISRs, OSRs are undervalued and lack mechanistic study despite that random mutagenesis has brought them to light in many FP-engineering cases for ages. When ISRs intensely alter the chromophore performance due to close contact, it is also possible for OSRs to significantly influence chromophores through other mechanisms.

To explore this potential, building an FP model with drastic feature changes under OSR mutations only is necessary. Although bright FPs have been frequently developed and employed, dark FPs with high extinction coefficient (EC, ⩾40 mM-1cm-1) and low quantum yield (QY, ⩽ 0.05) are increasingly desired due to their superiority in diminishing spectrum contamination,

releasing FP selection scope, and conjugating more FPs in Förster resonance energy transfer (FRET) assays. Given the easy monitoring of FP brightness in every experimental step and practicality of a dark FP, the collaborating lab derived a dark FP model, darkmRuby, from bright mRuby3, belonging to the mRuby family they had studied in-depth before [37–39].

The protein darkmRuby was engineered *via* mutating 21 OSRs of mRuby3 (Figure 2.1a). It retains the maximum wavelengths of excitation and emission of mRuby3 at 556 nm and 595 nm, respectively (Figure 2.1b). At the same time, it features a low QY of 0.05 and bell-shaped pH dependence peaked at pH 6.5 (Figure 2.1c), apparently altered from mRuby3. This chapter focused on the structure-feature study of darkmRuby and deciphering the roles of OSRs in its unique features. Comprehensive analysis combining experimental and computational experiments uncovers a long-range interaction that the OSRs Met94 and Phe96 regulate the ISR His197 conformation in a 15-Å distance under the mediation of water molecules flowing in a channel. The channel exposes the chromophore and His197 to the solvent environment, facilitating water molecules to quench fluorescence emitted by the chromophore, leading to low QY, as well as impacting His197 protonation state, resulting in pH sensitivity. To my best knowledge, this is the first time reporting a long-range interaction in FPs and presenting an atom-level mechanism of OSRs modulating FP features. This study proves the roles of OSRs as vital as ISRs, and more importantly, extends the understanding of FPs to a new stage and offers fundamental knowledge to support the development of new tools for maneuvering FP features.

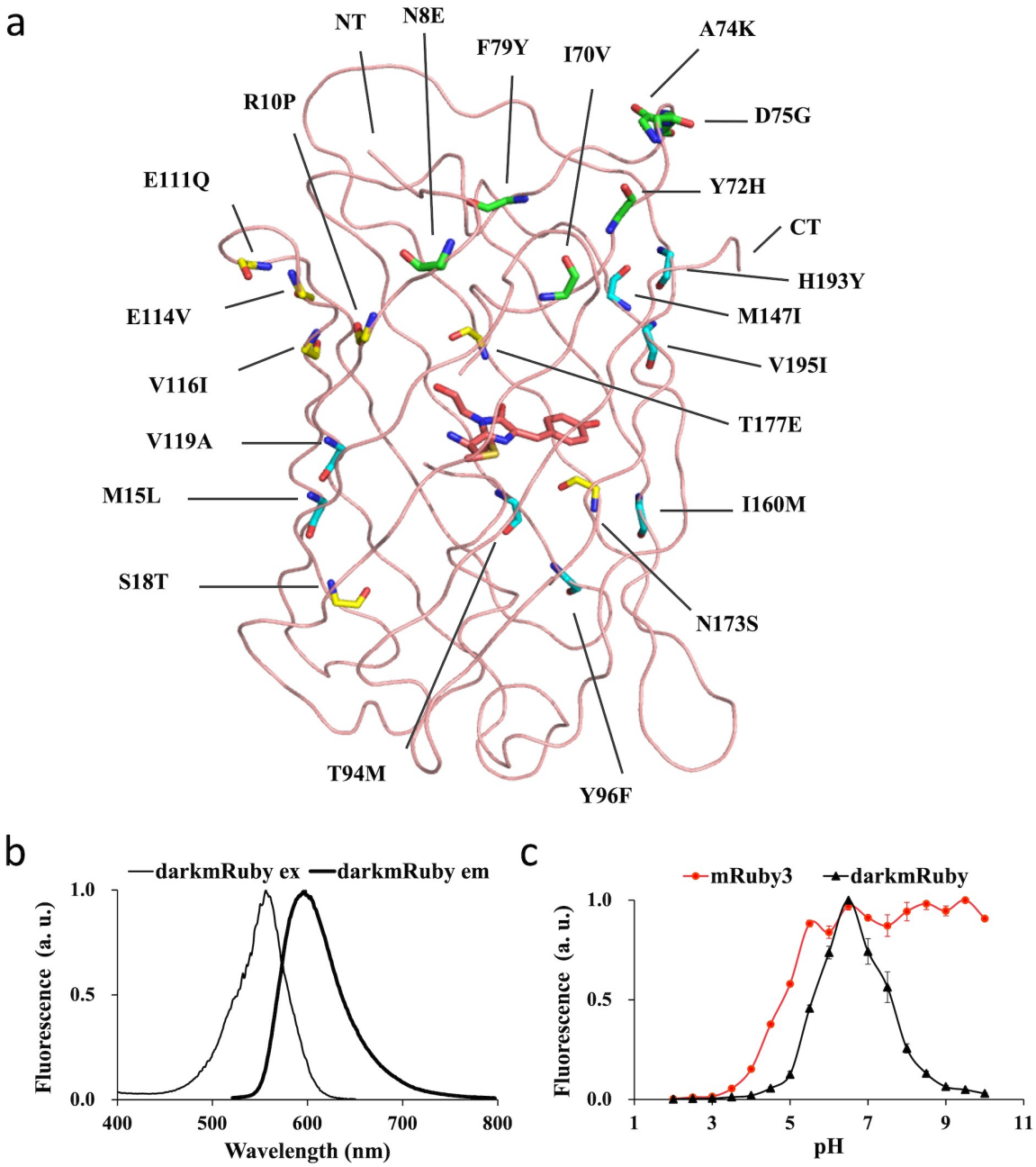


Figure 2.1 Designing darkmRuby.

(a) Crystal structure of mRuby1 (PDB: 3U0M) showing mutations between mRuby3 and darkmRuby. The β-barrel and chromophore are depicted with ribbons and sticks, respectively. (b) Normalized fluorescence spectra of darkmRuby. The inset image shows the color of purified darkmRuby under room lighting. (c) pH dependence of mRuby3 and darkmRuby fluorescence. Each curve is the mean of three independent experiments with the error bars showing the standard deviation.

2.2 Materials and Methods

2.2.1 Protein preparation

For crystallizing darkmRuby, the pNCS plasmid containing the nucleic acid sequence of darkmRuby was transformed into E. cloni 10G SOLOs chemically competent cells (Lucigen). The cells were grown in terrific broth media at 37 ℃ for 24 h, then at 18 ℃ for another 24 h. After being washed by lysis buffer (20 mM Tris pH 7.2, 5% glycerol, 0.5 mM TCEP, 0.5 mM PMSF), the cells were lysed by sonication and then centrifuged to remove cell fragments. The protein was purified by HisTrap HP column (GE Bioscience) with elution buffer (200 mM imidazole, 20 mM Tris pH 8.0) and Superdex 75 10/300 GL column (GE Bioscience) with equilibration buffer (20 mM Tris pH8.0).

After purification, the His tag of darkmRuby was removed by enterokinase to crystallize it in an acidic condition. To do that, I mixed the protein with the enzyme and incubated the mixture at 12 ℃ for approximately 48 hours, followed by isolating the untagged protein with HisTrap HP column.

The plasmid for expressing the mutant darkmRuby-TY was derived from the plasmid containing the wild-type gene by using Q5® Site-Directed Mutagenesis Kit (NEW ENGLAND Biolabs Inc.). The mutant protein was expressed and purified by following the protocol identical to above.

2.2.2 Protein crystallization and structure determination

Vapor diffusion by sitting drop method was applied to crystallize darkmRuby and its mutant. I screened the crystallization conditions through MCSG Crystallization Suite (Anatrace), including MCSG1, MCSG2, MCSG3, and MCSG4 sections with 384 conditions. Crystal

Gryphon (Art Robbins Instruments) dispensed 0.2 μL of protein solution and 0.2 μL of crystallization reagents on the micro-well surface of the 96-well crystallization plate. The reservoir was filled with 60 μL of relative crystallization reagents. The plates were incubated at 12 ℃ for about two weeks before the first piece of crystal was observed. If necessary, the crystallization condition was optimized by manipulating protein concentration, buffer pH, the volume ratio of protein solution to crystallization reagent, etc. The exact crystallization conditions for crystallizing darkmRuby and the darkmRuby-TY mutant were listed in Table 2.1.

Table 2.1 Crystallization conditions of darkmRuby and darkmRuby-TY.

Protein	His tag (Y/N)	Protein concentration	Protein buffer	Crystallization reagent	Vp : Vc *
darkmRuby	N	15 mg/mL	20 mM Tris pH 8.0, 50 mM NaCl, 0.5 mM TCEP	0.2 M lithium sulfate, 0.1 M sodium acetate: acetic acid pH 4.5, 30% (w/v) PEG8000	1:1
darkmRuby	Y	6.5 mg/mL	20 mM Tris pH8.0	1 M NaH_2PO_4 / K_2HPO_4	1:1
darkmRuby	Y	13 mg/mL	20 mM Tris pH8.0	0.1 M CHES: NaOH pH 9.5, 30% (w/v) PEG3000	3:1
darkmRuby-TY	Y	18 mg/mL	20 mM HEPES pH 7.4, 50 mM NaCl	0.1 M HEPES: NaOH, pH 7.5, 20 % (w/v) PEG 8000	1:1

* The volume ratio of protein solution to crystallization reagent.

X-ray diffraction data for the darkmRuby crystal at pH 5.0 and the darkmRuby-TY mutant model were respectively collected by SSRL beamlines 9-2 and 12-2 at SLAC National Accelerator Laboratory in Stanford University, while the ones at pH 8.0 and 9.0 were collected by the SIBYLS beamline at the Advanced Light Source at Lawrence Berkeley Lab in the University of California, Berkeley. All of them were integrated by X-ray Detector Software (XDS) [40]. Molecular replacement was conducted to build crystal structure models by Phaser [41] in the CCP4 suite. The crystal structures were refined by Coot 0.8.9-pre EL [42] manually and

Refmac5 [43] in the CCP4 suite computationally. Crystallography data collection and refinement statistics were shown in Table 2.2.

Table 2.2 X-ray diffraction data collection and model refinement statistics of darkmRuby and darkmRuby-TY.

Model	darkmRuby at pH 5.0	darkmRuby at pH 8.0	darkmRuby at pH 9.0	darkmRuby-TY
PDB ID	7RHA	7RHB	7RHC	7RHD
Data collection				
Space group	C2221	C121	C121	C121
Cell dimensions				
a/b/c (Å)	73.57/101.69/134.98	108.09/77.27/66.81	107.54/76.93/66.65	90.45/37.51/69.12
α/β/γ (°)	90.00/90.00/90.00	90.00/92.11/90.00	90.00/90.99/90.00	90.00/111.94/90.00
Wavelength (Å)	0.97946	0.97946	1.11583	0.97946
Resolution range (Å)	47.62-1.80 (1.90-1.80)	38.63-2.51 (2.61-2.51)	45.84-2.80 (2.95-2.80)	34.24-1.90 (2.00-1.90)
Total reflections	348672 (45429)	57481 (6446)	63424 (9338)	56770 (8067)
Unique reflections	47367 (6812)	18477 (2069)	13453 (1955)	16938 (2422)
Completeness (%)	100.00 (99.9)	98.0 (98.7)	99.4 (99.6)	98.3 (97.5)
I/σ (I)	10.4 (2.6)	6.1 (1.65)	9.1 (2.1)	7.8 (2.1)
Rmerge	0.112 (0.681)	0.087 (0.539)	0.116 (0.597)	0.084 (0.670)
CC 1/2	0.997 (0.842)	0.995(0.822)	0.994 (0.815)	0.996 (0.752)
Model refinement				
R_{work}/R_{free} (%)	17.6 / 21.1	22.7/26.6	19.7 / 27.1	19.1 / 24.8
Resolution range (Å)	47.58 – 1.80	38.63 – 2.51	47.58 – 1.80	32.32 – 1.90
Number of non-H atoms				
Protein	3726	3563	3600	1814
Water	246	69	40	69
RMSD values				
Bond lengths (Å)	0.009	0.007	0.008	0.008
Bond angles (°)	1.569	1.580	1.583	1.555
Ramachandran (%)				
Favored / outliers	96.54 / 0	92.21 / 0	95.47 / 0	97.37 / 0

2.2.3 B-factor normalization

B factor, also known as the Debye–Waller factor, temperature factor, or atomic displacement parameter, reflects the vibrational motion or disorder of an atom in a crystal structure. A large B factor value usually indicates an atom located in a flexible or unstable portion of the protein. For investigating the rigidity of the β-barrel, normalized B factor (B') of the atom *i* was calculated by the following equation: $B_i' = (B_i - \text{Mean}) / \text{SD}$, where B*i* represents the B factor of the atom *i* obtained from the refined structure model, SD represents standard deviation.

2.2.4 Channel predictions

The software CAVER is a tool for analyzing and visualizing tunnels and channels in protein structures, which is essential in drug design and molecular enzymology [44]. CAVER 3.0 PyMOL Plugin is the version incorporated in PyMOL for calculations. Here I applied it to predict available channels in darkmRuby and darkmRuby-TY crystal models. In the calculations, the minimum bottleneck radius of a channel was set to 0.6 Å; all amino acid residues and the chromophore were included, yet water and other solvent molecules were excluded. The channels were visualized by surface mode on PyMOL.

2.2.5 Molecular dynamics simulations

The darkmRuby crystal model at pH 8.0 was mutated with M94T and F96Y substitutions to build an *in silico* darkmRuby-TY model on PyMol. The built model was prepared by removing water molecules and solvent ions, then subjected to the molecular dynamics (MD) simulations under the AMBER14 force field in YASARA version 19.12.14.L.64 [45,46]. The

system was performed with explicit solvent in a cube box of 66.42-Å side length, containing 8674 water molecules. The default parameter settings were used in the MD run with the pressure at 1 bar, the temperature of 298 K, pH 7.4, and 1-fs time steps. In this project, the MD run was in the 160.10-ns length with snapshots taken at every 100-ps interval. The trajectory was analyzed by YASARA version 19.12.14.L.64.

2.3 Results and Discussion

2.3.1 Structural basis of low QY and pH sensitivity in darkmRuby

For atomically gaining insight into the OSR-feature relationship, I analyzed the X-ray crystal structures of darkmRuby at pH 5.0 (7RHA), 8.0 (7RHB), and 9.0 (7RHC) in resolutions of 1.80, 2.54, and 2.83 Å, respectively (Figure 2.2a). As expected, each chain in the crystals shows a typical GFP-like structure with an approximately co-planar chromophore. Since mRuby3 lacks a crystal model, I compared the darkmRuby crystal models with mRuby1 ones (Figure 2.3a). Like mRuby1, darkmRuby lodges a *trans* chromophore (PDB NRQ) at pH 8.0, 9.0, and mixed chromophores at pH 5.0 - about 40% in *cis* and 60% in *trans* (Figure 2.2b). The less-than-0.35-Å root-mean-square deviations (RMSDs) of Cα atoms between darkmRuby models suggest a generally stable protein scaffold at different pHs (Table 2.3). No significant difference is observed between the backbones of darkmRuby and mRuby1, indicating that the mutated OSRs may adjust darkmRuby in a more delicate manner. However, no contact of the OSRs with the chromophore hindered us from evaluating their impacts on the chromophore merely through eyeballing the crystal models.

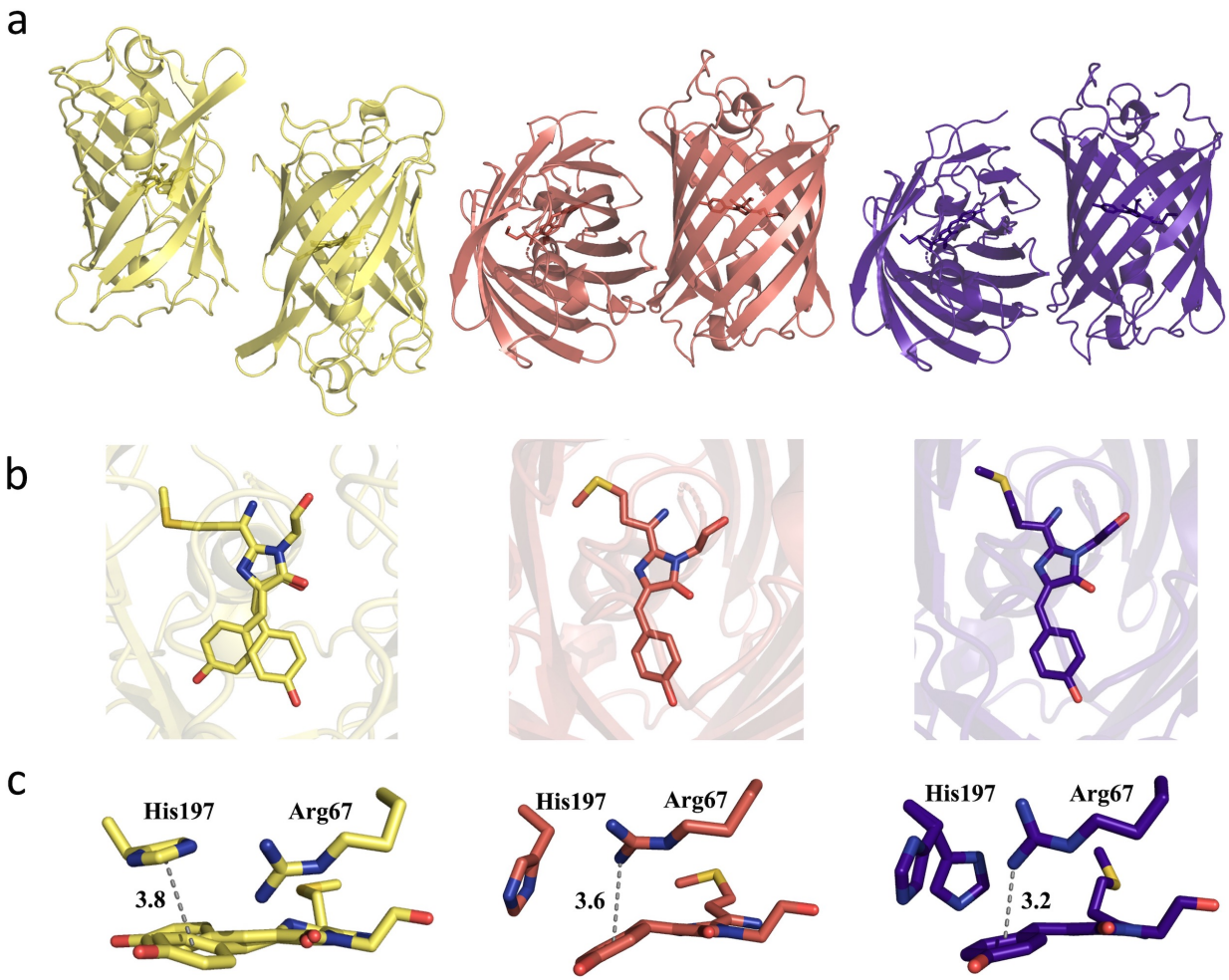


Figure 2.2 Crystal structures of darkmRuby at pH 5.0 (PDB 7RHA, yellow), 8.0 (PDB 7RHB, salmon) and 9.0 (PDB 7RHC, purple).

The chain B model in each crystal was presented. (a) Overview of the models. (b) Conformations of the chromophore in the models. (c) Arg67 and His197 display different conformations in the models. The π-stacking effects are shown with grey dash lines with bond lengths in angstroms labeled.

Table 2.3 The RMSD values (Å) of the backbone comparison among Chain B of darkmRuby crystal models.

Models	darkmRuby at pH 5.0	darkmRuby at pH 5.0	darkmRuby at pH 5.0
darkmRuby at pH 5.0	--	0.330	0.293
darkmRuby at pH 5.0	0.330	--	0.193
darkmRuby at pH 5.0	0.293	0.193	--
mRuby1 at pH 4.5	0.208	--	--
mRuby1 at pH 8.5	--	0.319	0.277

I expanded the model inspection to the ISRs to gather more clues. Most ISRs are conformationally comparable in darkmRuby and mRuby1 models, except Arg67 and His197. For darkmRuby, with pH increased, the His197 imidazole (I_{H197}) ring switches the pose from parallel to vertical against the chromophore phenolic (P_{NRQ}) ring, accompanying π-stacking broken; while Arg67 occupies the top space and interacts with the P_{NRQ} ring *via* cation-π (Figure 2.2c). In contrast, the I_{H197} ring in mRuby1 holds fast to the parallel pose at pH 4.5 and 8.5 (Figure 2.3a). Given that mRuby1 (QY = 0.35) is much brighter and better resistant to pH change, the pH-dependent movements of Arg67 and His197 could be an immediate cause for the low QY and pH sensitivity of darkmRuby. Moreover, the ISRs are conserved in the mRuby family, so their conformational changes must originate from the altered OSRs.

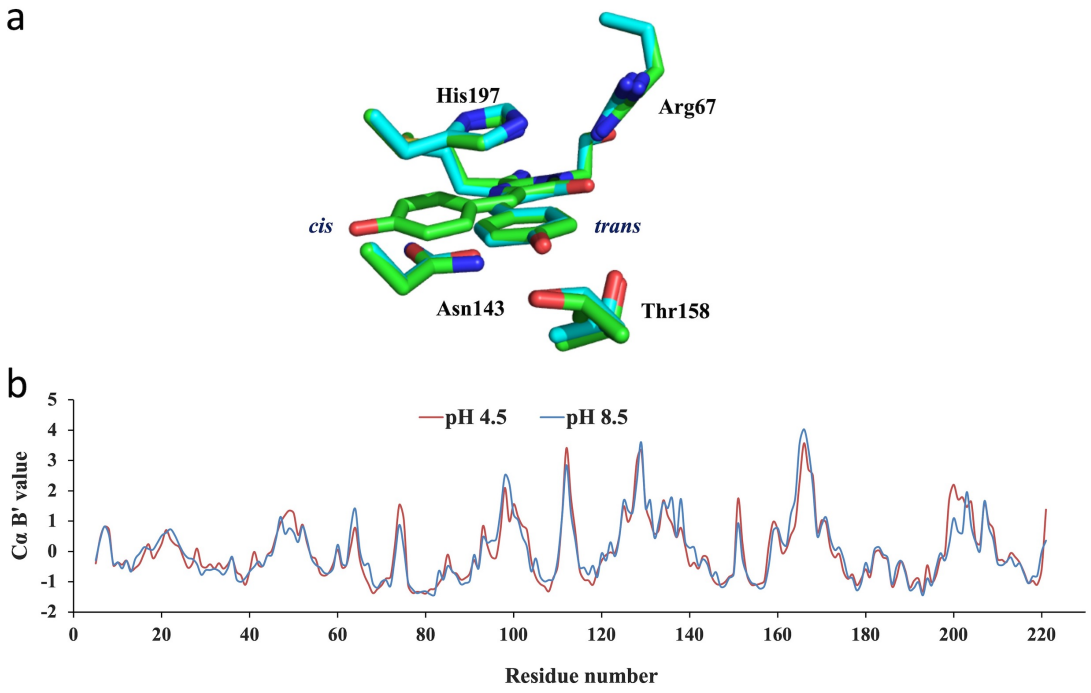


Figure 2.3 Crystal structures of mRuby1.

(a) His197 maintains the interactions with the chromophore of mRuby1 at pH 4.5 (3U0L, green) and 8.5 (3U0M, cyan). The other parts of the models are hidden for a clear view. (b) The Cα B' values of two mRuby1 crystal models are well aligned, indicating that the barrel conformation is resistant to pH change.

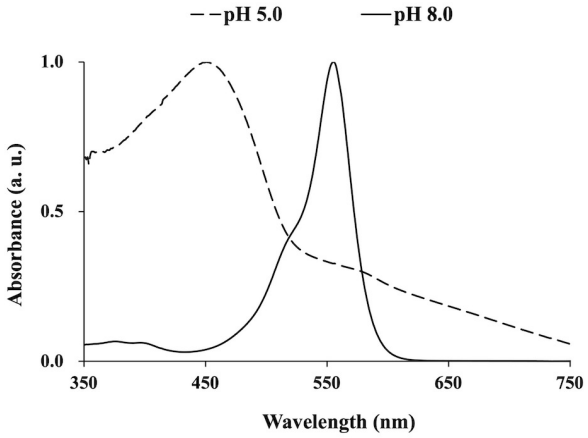


Figure 2.4 Absorption of darkmRuby at pH 5.0 and 8.0.

It is worth mentioning that the P_{NRQ} ring of darkmRuby is protonated in the acidic environment, according to the spectrophotometric scanning result (Figure 2.4). The protonation disfavors electron redistribution under photoexcitation and facilitates the cis-trans isomerization in the competition with fluorescence [47,48]. Thus, it is not surprising to observe low fluorescence intensity and an isomerized chromophore in darkmRuby at pH 5.0. In the rest of the chapter, I only focused on exploring the darkmRuby features in an alkaline environment unless specified.

2.3.2 A long-range interaction regulates His197 conformation in darkmRuby

Closer examination of the crystal models, mutagenesis, and X-ray crystallography were executed to figure out the key OSR mutations responsible for the conformational switches of Arg67 and His197. Rigid β-barrel structures of FPs have been proven essential for maintaining chromophore conformation and chemistry by preventing small molecules (such as water, gas) overflowing in the protein to cause photobleaching and fluorescence quenching, and so on [47,49,50]. I reasoned that one or more mutation sites might weaken the darkmRuby barrel rigidity, allowing small molecules to impact His197 and Arg67 directly or indirectly, leading to decreased QY, pH

stability. To simplify the study, I only discussed His197 from then on, as its less elastic side chain dominates the Arg67 pose in the His197-Arg67 interaction.

Here I attempted to explore the influence of the mutations on the barrel rigidity by B' factor, an excellent indicator of the Cα motion. As a control, the Cα B' values of two mRuby1 models were compared (Figure 2.3b), and it shows an almost perfect alignment, indicating a steady barrel resistant to pH change. In contrast, darkmRuby displays a tangled view of the model alignment (Figure 2.5a). The pH-dependently fluctuated B' values of several portions imply the flexible architectures sensitive to pH change. After excluding loop data due to the intrinsically disordered nature, I pinpointed four mutated residues, Met94, Phe96, Asp159, and His171, with B' values increasing strikingly at pH 8.0 and/or 9.0 (Figure 2.5b). In darkmRuby models, Met94 and Phe96 are conformationally identical among the models, but the improving B' values hint that the residues are actively affected by pH. Moreover, when the Thr94 and Tyr96 side chains construct a hydrogen-bonding network with water molecules in mRuby1, Met94 and Phe96 cannot fasten any water molecule in darkmRuby (Figure 2.5c). The failure of establishing the network may undermine the barrel rigidity, releasing more water molecules into the protein and quenching fluorescence. Asp159 and His171 situate in front of the P_{NRQ} ring with a hydrogen bond formed at pH 5.0 and broken at pH 8.0, 9.0. It could be a pH-sensitive, open-close gate controlling water diffusion and impacting the chromophore in darkmRuby (Figure 2.5d). The collaborating lab mutated these residues (M94T/F96Y, D159H/H171S) to test their roles in darkmRuby (Table 2.4). Two other residues 147 and 195 with trivial B' value shift in darkmRuby but different space occupation to mRuby1 were also selected for mutagenesis to briefly check the necessity of the B' value analysis (Figure 2.5e).

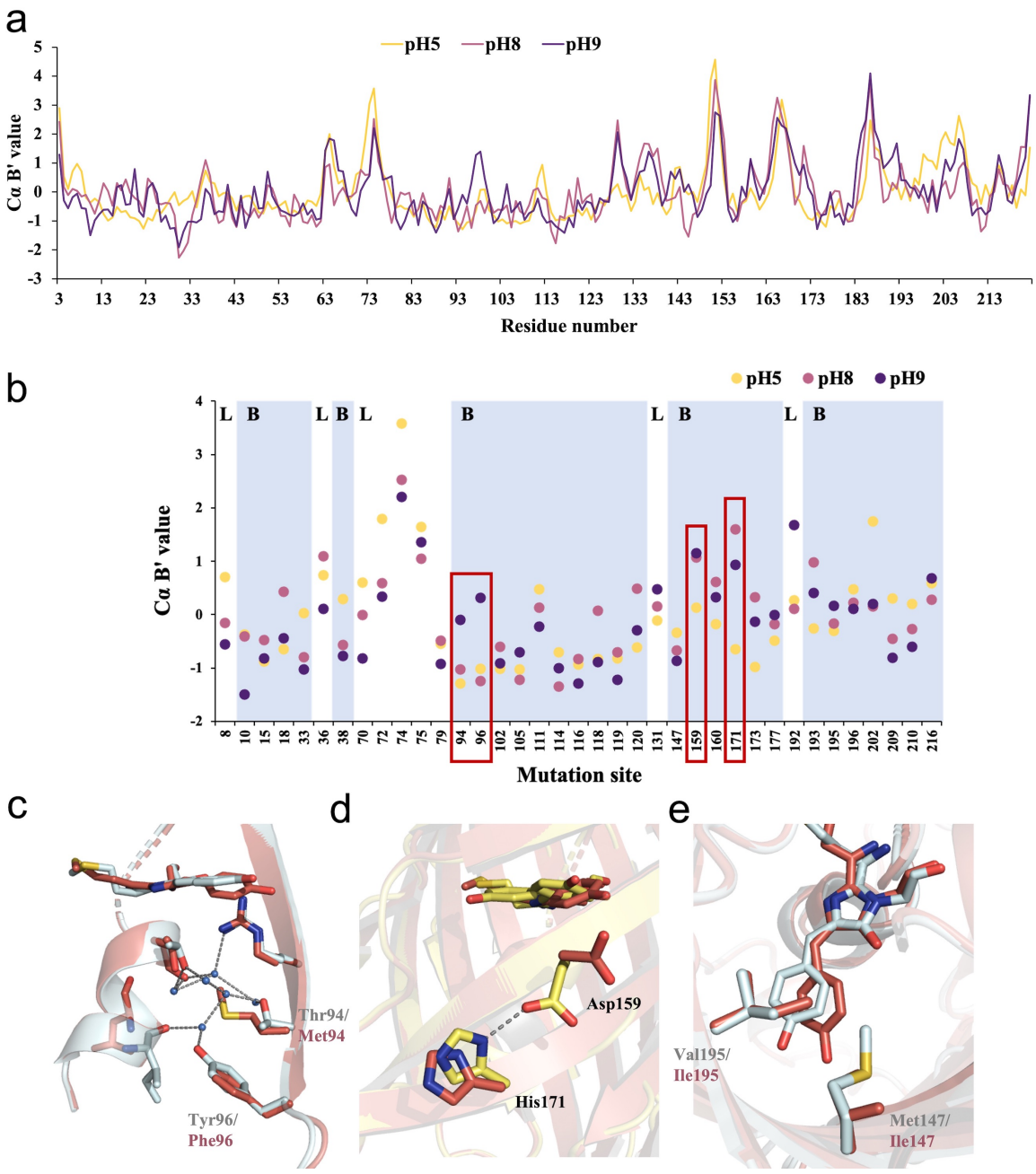


Figure 2.5 Structural analysis of darkmRuby.

(a) The fluctuated Cα B' values of the darkmRuby models show portions of the protein change with increasing pH. (b) The Cα B' values of the mutated residues in the three structures. The sites on the β strands (Bs) are shown with light blue backgrounds, while the sites on loops (Ls) are shown with blank backgrounds. The sites showing large B value shifts with pH on the β strands are marked with red rectangles. (c) A hydrogen-bonding network was formed by the residues 94, 96 and the surrounding residues, water molecules in mRuby1 at pH 8.5 (grey), yet it did not happen on darkmRuby at pH 8.0 (salmon). (d) His171 and Asp159 formed a hydrogen bond at pH 5.0 (yellow), but the bond was broken at pH 8.0 (salmon). (e) The space occupancy of the residues 147 and 195 was different in mRuby1 (grey) and darkmRuby (salmon).

Table 2.4 Interested mutation sites of darkmRuby.

Mutation sites	94	96	147	159	171	195
mRuby1	Thr	Tyr	Met	His	Ser	Val
mRuby2	Thr	Tyr	Met	His	Ser	Val
mRuby3	Thr	Tyr	Met	Asp	His	Val
darkmRuby	Met	Phe	Ile	Asp	His	Ile

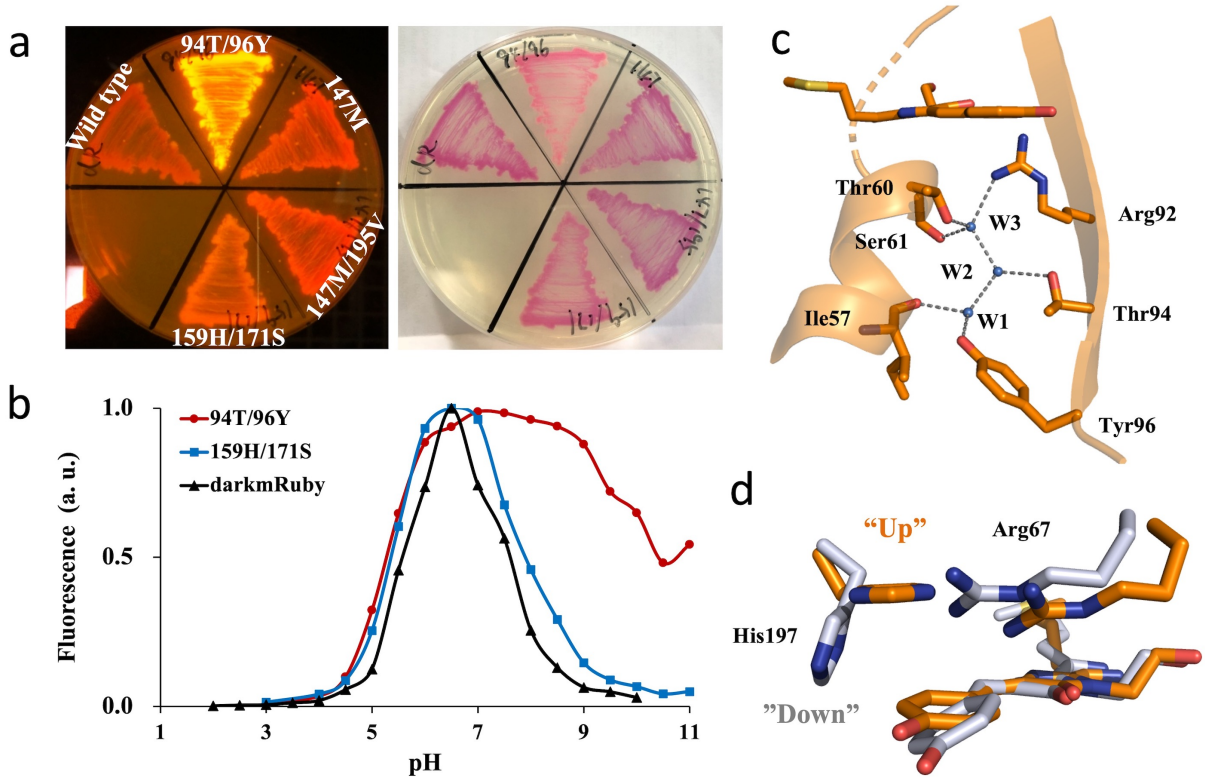


Figure 2.6 Site-directed mutagenesis on darkmRuby.

(a) Fluorescent brightness (left) and colors under room lighting (right) of darkmRuby mutants expressed in E. coli. (b) pH titration of the selected darkmRuby mutants. Each curve is the mean of three independent experiments. (c) A hydrogen-bonding network was formed by Thr94, Tyr96, surrounding residues and water molecules in darkmRuby-TY (7RHD, orange). (d) The I_{H197} ring reposited to the parallel top of the P_{NRQ} ring, and Arg67 was pushed backward in darkmRuby-TY. As a reference, darkmRuby at pH 8.0 was presented in the grey color.

Compared with the wild type, the mutants M94T/F96Y and D159H/H171S light up, yet the mutations I147M and I195V make only a marginal difference in brightness (Figure 2.6a). B' value shift seems to be a good indicator in mutation site selection. Additionally, the M94T/F96Y mutation enhances the alkaline resistance up to pH 9.5 (Figure 2.6b). The mutagenesis evidences the prime responsibility of Met94 and Phe96 in the low-QY and pH-sensitive features of darkmRuby. I crystallized the M94T/F96Y mutant, named darkmRuby-TY from here on, at pH 7.5 and gained a crystal model at 1.90 Å. A hydrogen-bonding network compatible with mRuby1 is formed by Thr94, Tyr96, three water molecules (W1, W2, W3) in darkmRuby-TY (Figure 2.6c), and other surrounding residues. More importantly, His197 parallelly π–stacks against the P_{NRQ} ring (Figure 2.6d). The mutant model structurally confirms the regulatory potential of the OSRs 94 and 96 on His197 conformation, acting as an "Up-Down" switch of the fluorescence. Interestingly, positions 94 and 96 are approximately 15-Å away from His197. To my best knowledge, this is the first long-range interaction reported in FPs.

2.3.3 The long-range interaction is mediated by water molecules in a channel

Fifteen Å is such a long distance that no single noncovalent contact could reach. By comparing darkmRuby and darkmRuby-TY models, I found noticeable pose transitions of residues in-between Thr94 and His197, including Asn143, Thr158, and Phe174 (Figure 2.7a). Accordingly, I rationalized that these residues were related to the long-range interaction. To verify the hypothesis, MD simulations were conducted against an *in silico* darkmRuby-TY model derived from the darkmRuby crystal model at pH 8.0. After trials under different combinations of simulation conditions, I found that only when W1, W2, W3 were included and His197 was protonated could the I_{H197} ring possibly convert to the “Up” mode in the simulation

system (Figure 2.7b), coinciding with the darkmRuby-TY crystal model. A series of transformations was captured in the run (Figure 2.8). The water molecules wobbled around Phe174, and Phe174 lifted towards the P_{NRQ} ring; then Asn143 turned the side chain from the protein surface into the barrel, followed by the I_{H197} ring flipping; after that, W1 went up along Phe174, Thr158 and escaped from the barrel through a gap between Asn143 and Thr158. I also noticed in other runs that W1 moved towards the other direction at the beginning, and Phe174 and protonated His197 remained the "Down" mode. The simulations plausibly reveal three main factors required to successfully reproduce the mutant crystal structure: involvement of the intermediate residues and water molecules, movement of the water molecules, and protonation of His197.

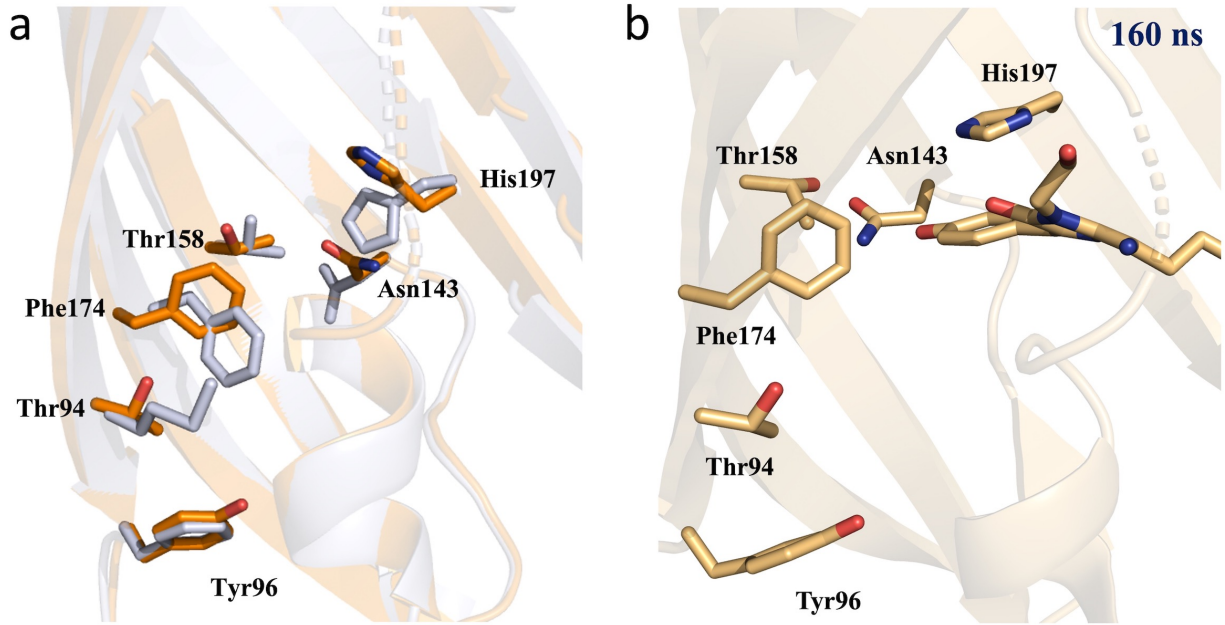


Figure 2.7 Conformational differences of Thr94 and Tyr96 in darkmRuby-TY.

(a) The residues between the residue 94 and His197 changed the poses after mutagenesis. The wild-type model at pH 8.0 was presented in grey color, and darkmRuby-TY in orange color. (b) The last frame of the 160-ns MD simulations against the darkmRuby-TY model computationally prepared by the darkmRuby model.

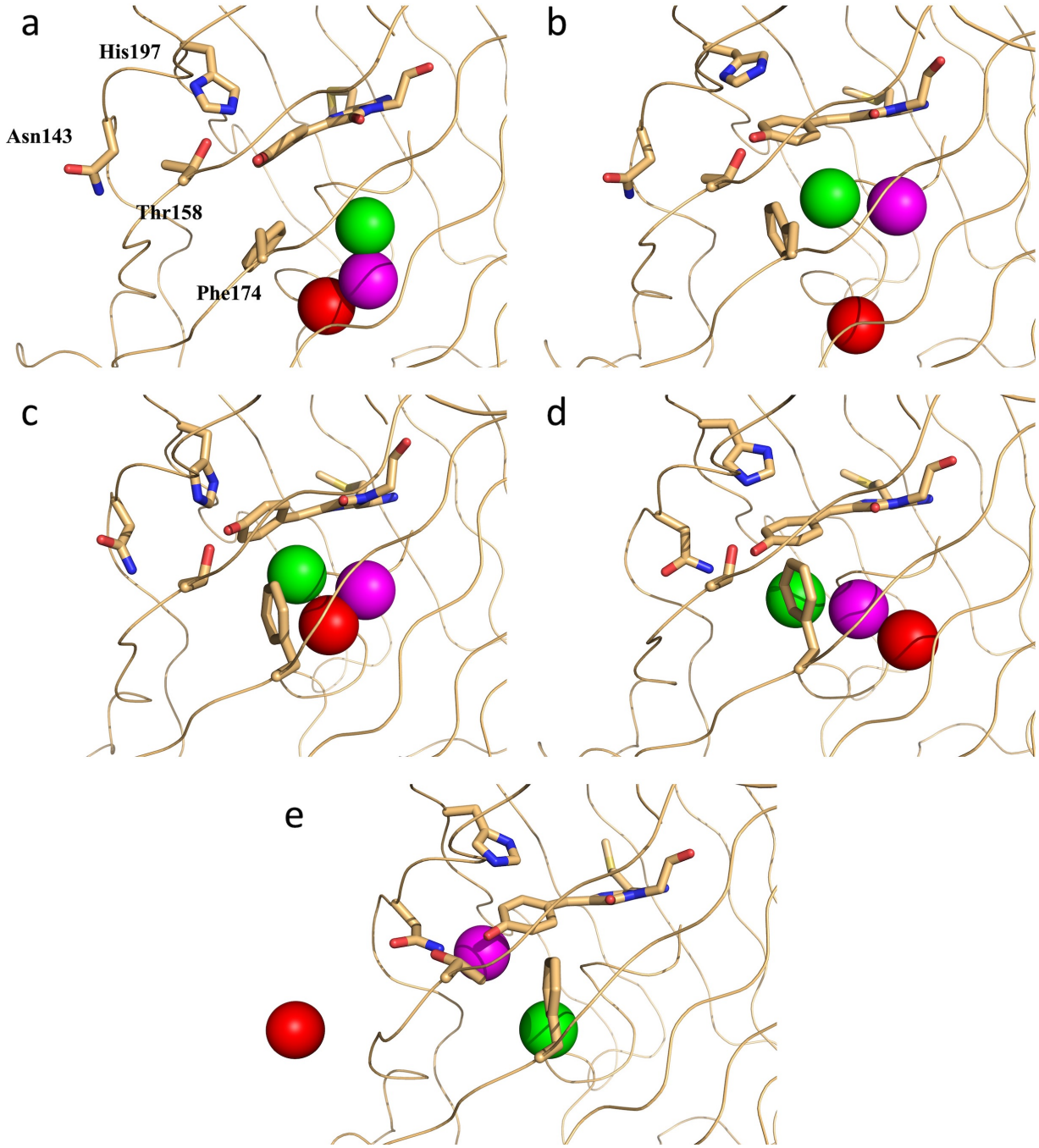


Figure 2.8 Snapshots of MD simulations at 0 ns (a), 2.2 ns (b), 2.5 ns (c), 4,1 ns (d), and 17.5 ns (e).

This simulation lasted for 160 ns with His197 protonated and water molecules W1 (red ball), W2 (magenta ball), W3 (green ball) included in the system. The protein model was shown in sand ribbon, and the chromophore and Asn143, Thr159, Phe174, His197 were in sand sticks.

For the in-between residues, Asn143 and Thr158 reside at crucial positions capable of impacting the chromophore performance, and extra caution is required in mutagenesis to prevent unwanted features created. The residue 174 is relatively mild, and the Phe174 conformation seems to associate with the His197 pose switch, according to the simulations. Hence, it was chosen to examine the mediating role of the intermediate residues by replacing it with hydrophilic, bulky residues (Tyr, His, Gln) in darkmRuby and darkmRuby-TY, respectively (Figure 2.9). The Tyr and Gln mutations canceled the light-up effort of the M94T/F96Y mutation in darkmRuby-TY, and the Gln mutation even modified the visible color of the protein. They implicate the vital mediating role of Phe174 and its hydrophobicity in the effect of Thr94 and Tyr96. However, the role of Phe174 seems vague in Met94 and Phe96 of darkmRuby. On the other hand, residue 174 locates in such an intermediate position that it can be defined as either an ISR or OSR, depending on the pose of the side chain. For example, Phe174 is an OSR in the darkmRuby model at pH 8.0 but an ISR in the darkmRuby-TY model. This feature complicates the result analysis, including the performance of the F174H mutation in the proteins. The brighter F174H mutants are possible consequences of His174 behaving like His197 in a position close enough to the P_{NRQ} ring. Site-directed mutagenesis can only provide limited information, and further investigation is needed to interpret the phenomenon better.

The MD results have shown that moving paths of the water molecules are essential to flip I_{H197}. It may be further scrutinized under the assistance of a computer program Caver. A hub of channels neighbor residues 94 and 96 in the crystal models (Figure 2.10a, b), presenting multiple potential water diffusion pathways. In darkmRuby-TY, it includes a channel perfectly overlapping with the moving path of W1, which stays far from the chromophore and His197 (Figure 2.10d). In darkmRuby, instead, a channel avoids the interested residues and passes

through the surfaces of the P_{NRQ} ring and the I_{H197} ring (Figure 2.10c). Obviously, chemical features and conformations of OSRs 94 and 96 have changed channel distribution in the proteins. It also intriguingly guides to a thought that His197 without contacting water molecules favors its "Up" mode pose. This finding can help interpret the low-QY feature of darkmRuby because water molecules flowing on the chromophore surface will quench the fluorescence [51,52].

The protonated state of His197 in darkmRuby-TY could be an outcome of no attack from water molecules. A cavity is formed around His197 to insulate it from water flow, presumably obliged by the parallel pose and its interactions with the surroundings (Figure 2.10d). The void benefits a proton boarding on the I_{H197} ring even in the moderately basic environment, thus darkmRuby-TY displays a pH-resistant capability until pH 9.5. In comparison, the channel in darkmRuby exposes the I_{H197} ring to the solvent environment. The solvent exposure can draw pKa of the His197 side chain close to the aqueous value 6.2; the I_{H197} ring becomes deprotonated and transforms to the "Down" mode in the neutral and basic solvent. Therefore, the darkmRuby fluorescence fades with pH increase after pH 6.5.

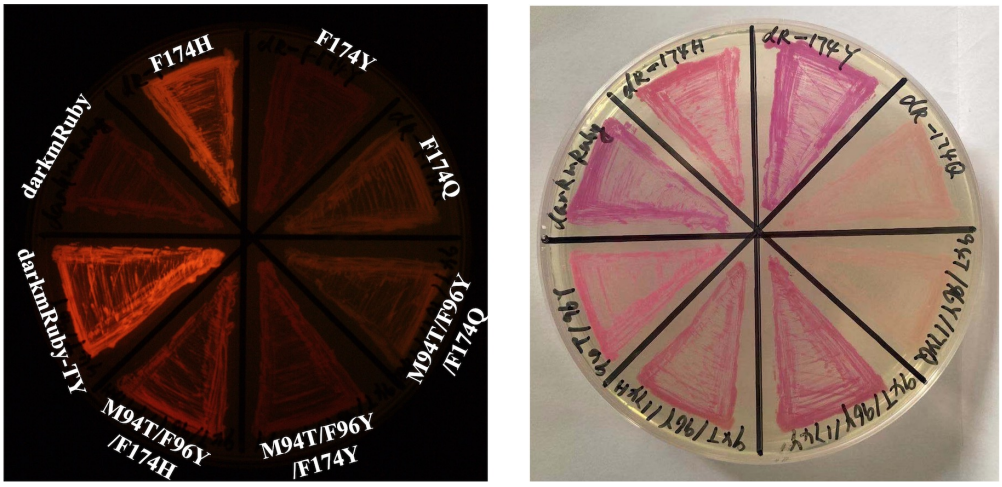


Figure 2.9 Site-directed mutagenesis of Phe174 in darkmRuby and darkmRuby-TY.
The mutants were expressed in bacteria and compared by fluorescent brightness (left) and color under room light (right). Each patch represents corresponded mutants, respectively.

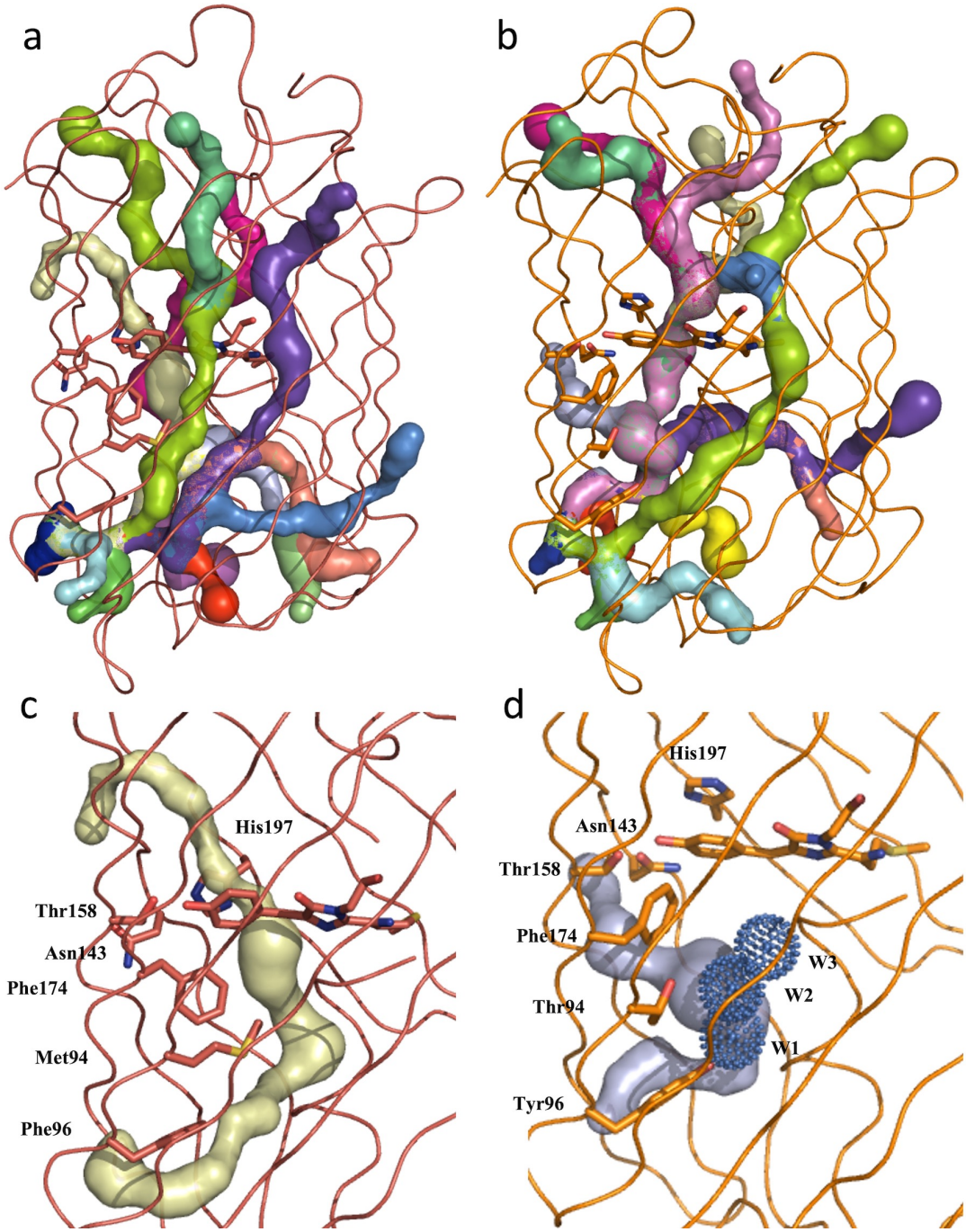


Figure 2.10 Channels predicted in darkmRuby and darkmRuby-TY.

(a) The channels near Met94 and Phe96 were shown in the darkmRuby model at pH 8.0. (b) The channels near Thr94 and Tyr96 were shown in the darkmRuby-TY model. (c) In the darkmRuby model at pH 8.0, a channel linked Phe96 to a gap between the P_{NRQ} ring and the I_{H197} ring without contacting Asn143, Thr158 or Phe174. (d) In the darkmRuby-TY model, a channel passed the vicinity of Thr94, Tyr96, and Phe174 and connected to the solvent environment *via* a gap between Asn143 and Thr158. It partially overlapped with the Van de Waals surface of W1 (shown in blue dots). Different channels were presented in different colors.

2.4 Conclusion

In summary, a dark FP, darkmRuby, was successfully derived from bright mRuby3 by mutating multiple OSRs. The protein displays a dim and pH-dependent fluorescence, an outcome of a long-range interaction between outer-shell Met94, Phe96 and inner-shell His197. I propose that Met94 and Phe96 regulate the His197 conformation under the mediation of water molecules flowing in a channel; the affected pose facilitates water molecules to quench the chromophore, leading to low QY of darkmRuby; in the meantime, the channel exposes His197 to the solvent and modulates its protonation state with pH change, causing pH sensitivity of darkmRuby. I also conclude that the bright mRuby members (mRuby1, mRuby2, mRuby3, darkmRuby-TY) rely on Thr94 and Tyr96 to protect the protonated His197 from water flow, keeping bright fluorescence and pH resistance.

No long-range interaction has been stated in FPs before this discovery. The OSR-feature relationship is excluded from the classic understanding of ISRs impacting the FP features. Water and oxygen diffusing in FP β barrels and directly interacting with chromophores have been explored on the maturation of mCherry and TurboGFP, and the photosensitizing feature of KillerRed [51,50,53]. However, this study tells a more complicated story about water molecules participating in an indirect influence of OSRs against the chromophore. I believe long-range interaction could be one of the critical feature-controlling elements widely existing in FPs but ignored previously. Moreover, in addition to water molecules, the mediators of long-range interactions can be other diffusible small molecules and adjacent residues. More long-range interactions may be recognized if more FPs are inspected with their structural data. Building a sophisticated research method for exploring FP long-range interactions will add a powerful tool to the box and pilot the rational design of FP biosensors into a new stage.

Chapter 3 - Inverse Agonism of a Novel Modulator Targeting Human RORγ

Abstract: As a transcription factor in T helper 17 cells, human retinoic acid receptor-related orphan receptor ɣ (RORɣ) plays a crucial role in controlling the expression of pro-inflammatory cytokine IL-17. Since the overexpression of IL-17 can lead to various autoimmune disorders, RORɣ has drawn the attention of scientists to its potential as a drug target. It is known that when an inverse agonist binds to the ligand-binding domain (LBD) of RORɣ, the transcriptional function of RORɣ is down-regulated, and IL-17 expression is suppressed. Accordingly, drugs targeting RORɣ LBD will offer a strategy to treat autoimmune diseases. This project used a computer program iDock to virtually screen a drug-like small molecule library against a RORɣ LBD crystal model. Depending on accessibility, 15 top-ranking compounds with unique chemical structures were purchased for experimental investigation. Their binding capability was preliminarily tested by differential scanning fluorimetry; a compound RG14-2 displays outstanding binding data against the purified LBD protein, while other compounds also show modest binding. RG14-2 was further explored by reporter gene assay and identified to inhibit the RORɣ transcriptional activity with IC_{50} values at 1.5 μM. The NMR experiments were conducted to determine the binding affinity of RG14-2 against the LBD, which results in the K_d value of 5.7 μM, indicating a moderate binding strength of RG14-2. To structurally study the interaction between RG14-2 and the LBD, a double mutant with C-terminal truncation was constructed and co-crystallized with RG14-2. A monomeric crystal model at 3.5 Å demonstrates that RG14-2 interrupts the agonistic conformation of the LBD, preventing the co-activator from being recruited on the protein surface and triggering the RORɣ transcriptional function. RG14-2 bears a chemical scaffold different from other known RORɣ inverse agonists, making it a

superior lead compound for developing a novel class of RORɣ inverse agonists. Compared with classic laboratory-based drug discovery, computer-aided drug discovery can dramatically improve efficiency and save costs in the initial drug discovery stage.

3.1 Introduction

As a member of the nuclear receptor (NR) superfamily in the human genome, retinoid-related orphan receptor γ (RORγ) contains two isoforms, the thymus-specific RORγt and the ubiquitous RORγ. The RORγt isoform is essential in promoting T cell differentiation into the T helper 17 (Th17) subtype and regulating IL-17A gene transcription in Th17 cells [21]. The RORγ isoform acts as a negative regulator of adipocyte differentiation to impact adipogenesis and insulin sensitivity [54]. It may also be involved in supporting the Th17 differentiation [20]. Overexpression of RORγ isoforms can lead to obesity-associated insulin resistance and multiple autoimmune disorders such as multiple sclerosis, rheumatoid arthritis, Crohn's disease, and psoriasis. The pathological functions make them intriguing drug targets. Besides, several other diseases are found associated with Th17 cells. For example, RORγt and IL-17 expression in increased hepatic Th17 cells are observed in the progression from one subtype nonalcoholic fatty liver disease to the other subtype nonalcoholic steatohepatitis [1,55]; Th17 cells and IL-17 play mixed pro-tumor roles in the development of colorectal cancer [56]. Given the regulatory effects of RORγ isoforms in Th17 differentiation and IL-17 transcription, discovering drugs targeting RORγ may provide alternative therapeutic strategies for these Th17-mediated diseases.

The RORγt isoform is different from the RORγ isoform by missing the first 21 residues. To simplify the description in the dissertation, the term "RORγ" is utilized to represent both isoforms and number the residues by following the ubiquitous isoform. Like other NRs, RORγ is

composed of an N-terminal domain (1–30 a.a.), a DNA-binding domain (31–95 a.a.), a hinge domain (96–265 a.a.), a ligand-binding domain (LBD, 266–505 a.a.) and a C-terminal domain (506–518 a.a.). The knowledge obtained from other NRs facilitates scientists to quickly recognize the orthosteric binding site located in the LBD. The engagement of pharmaceutical companies, including Merck, Takeda, and Biogen, boosts the development of novel RORγ-targeting drugs that hundreds of small-molecule inverse agonists (Appendix B) binding in the orthosteric site have been investigated. However, no compound has yet passed Phase II clinical trials, owing to off-target effect, toxicity, and poor therapeutic efficacy found in the stages of preclinical studies or clinical trials [57,58].

The orthosteric binding site of RORγ locates in the core of the globular LBD, embraced by α-helices H3, H5, H6, H7, and H11 with the critical H12 in the vicinity of one pocket exit (Figure 3.1). Since the putative endogenous agonists, cholesterol analogs (hydroxycholesterols [59], 4α-carboxy and 4β-methyl-zymosterol [60]), were found binding in this site; thousands of small molecules have been screened as potential ligands and the structure-activity relationships have been thoroughly studied. It is believed that a RORγ agonist binding in the orthosteric site facilitates contact of the coactivator with the LBD surface, initiating the transcriptional function. In contrast, a RORγ inverse agonist in the same site achieves inhibition by preventing the coactivator recruited to the LBD. Depending on the ligand type, H12 changes its conformation to behave as an on-and-off switch.

H12 was primarily considered a static helix distal from the orthosteric pocket in apo NR LBDs for a long time, based on artificial crystal-packing interactions in co-crystals of the RARγ LBD-ligand complexes [61]. This changed in 2003 when Kallenberger et al. proposed a more dynamic model based on tracking the H12 movement in the PPARγ LBD by fluorescence

anisotropy techniques [62]. They found that H12 is highly dynamic and disordered in the apo LBD, but agonist binding induces H12 to become a helix and lie on the pocket exit with a well-defined pose, facilitating recruitment of the coactivator to the LBD surface and consequently activating transcription.

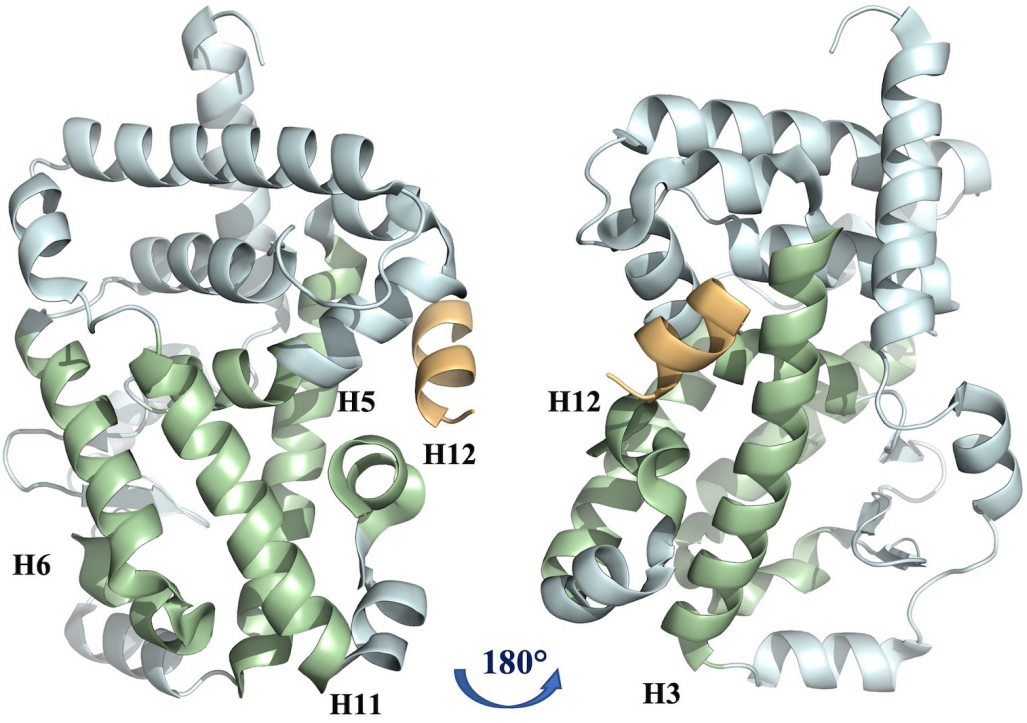


Figure 3.1 The globular structure of the apo RORγ LBD (5X8U) bound with a coactivator. H3, H5, H6, H7, H11 and H12 are presented in green color to indicate the orthosteric binding pocket, while other helixes are presented in light blue color, and the coactivator in light orange color.

This dynamic model also applies to the H12 behavior in the RORγ LBD. NMR results showed that the H12 chemical shifts in the unbound LBD were missing or shifted in backbone ^{15}N-TROSY and methyl-^{13}C-HSQC spectra, demonstrating disorder [63]. This was reflected in the difficulty of crystallizing the apo LBD in the absence of the coactivator [63,64]. HDX-MS results also demonstrated reduced dynamic properties of H12 when the RORγ LBD transitioned from

the apo state to the agonistic state, supporting a disorder-to-order transition [65]. However, the H12 dynamic properties in the apo and inverse-agonist states are similar, suggesting that H12 may not have a discrete conformation when an inverse agonist binds in the orthosteric site. In some co-crystals of RORγ LBD with inverse agonists, H12 could not be modeled due to missing electron density [66]; in many cases, H12 had to be truncated to cocrystallize the RORγ LBD with inverse agonists [67–69].

The term "agonist lock" is introduced to describe the atomic mechanism of LBD agonism only existing in the ROR subfamily [67]. In the agonist-bound LBD, a hydrogen bond formed between His479 and Tyr502 locks H12 in the correct position (Figure 3.2), facilitating coactivator recruitment. Meanwhile, Phe506 assists in stabilizing the lock by π-stacking with His479 [63]. It is now commonly accepted that RORγ orthosteric binders primarily regulate the transcriptional function by manipulating the stability of the lock. The molecular dynamics simulations further confirm it by showing lower fluctuations in interaction energy for the lock in the agonist-bound LBD and higher fluctuations for the lock in the unbound or inverse agonist-bound LBD [70]. In other words, the discovery of RORγ orthosteric inverse agonists is to discover drugs effectively breaking the agonist lock and disrupting the H12 conformation.

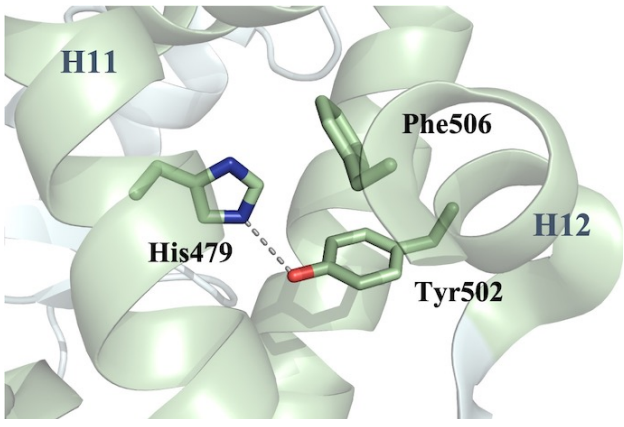


Figure 3.2 The agonist lock in the RORγ LBD (5X8U).

In contrast to the classical high-throughput screening method of screening chemicals for RORγ-targeting drugs, virtual screening (VS) accomplishes the searching process *in silico* by docking a library of drug-like small compounds into a structural model, scoring each compound by mathematic algorithms. It immensely saves time and cost in the initial screening stages of narrowing down the chemical pool. However, only two research groups have applied this method to seek RORγ orthosteric inverse agonists. In 2014, the Xu group conducted VS on a 220,000-compound database against the orthosteric pocket in an inverse-agonistic LBD crystal model using the Schrödinger Glide software, resulting in 115 hits [71]. In 2016, they did another VS to a five-times larger database against the same model using Glide and Phase simultaneously, resulting in 28 hits [72]. Most of the hits showed moderate or strong activity in biological and biochemical assays, but the models of the atomic structures for the best hits were only modeled. In 2020, Liu et al. created a homemade analog database of their interested compound and employed Glide for VS to find more potent compound [73]. These cases illustrate the feasibility of applying VS in the discovery of RORγ LBD-targeting modulators, although the powerful function and superiority of VS are currently underestimated.

In this project, my supervisor Dr. Ng has performed VS to screen the Zinc drug-like compound library against a crystal model of the RORγ LBD in inverse agonistic mode by a VS program named iDock [74]. Fifteen top-ranking hits were then handpicked basing on the chemical scaffolds distinctive to the known inverse agonists and the commercial accessibility.

In this chapter, I focused on the research accomplished in the laboratory to verify the bioactivity of the 15 hits. After the compounds arrived, their binding capacity and bioactivity were analyzed by differential scanning fluorimetry and reporter-gene assay. A compound RG14-2 presents excellent experimental results thus was further investigated by NMR experiments and

X-ray crystallography. The co-crystal model of the protein-ligand complex reveals a protein conformation with a broken agonist lock and a disordered H11, which results in inhibition of the RORɣ transcriptional function. The unique structure of RG14-2 makes it a superior lead compound for developing a novel class of RORɣ inverse agonists.

3.2 Materials and Methods

3.2.1 Protein preparation

The pMAL-C5e plasmid containing the gene sequence of the MBP-tagged RORγ LBD was ordered from GenScript. The protein mutants were developed from this plasmid, as mentioned in 3.2.2.

To express the protein and its mutants, I firstly transformed the plasmid into Rosetta2 (DE3) Singles competent cells (Novagen) by heat shock method. The cells were grown on a carbenicillin-pretreated LB media plate at 37 ℃ for at least 12 hours. Then a colony was picked and added to 20 mL of the liquid LB media with carbenicillin, followed by shaking the cell broth at 37 ℃ at 250 rpm for 16 hours. After that, the cell broth was poured in 1L of the TB media with carbenicillin and shaken at 37 ℃ at 250 rpm until the cell density reached the OD_{600} value of 0.6. A milliliter of 1M IPTG was added to the TB cell broth. The cells were incubated at 18 ℃ with a 200-rpm shake for another 22 hours.

The cells were harvested by centrifuging at 4000 rpm for 20 minutes. After discarding the supernatant, I resuspended the cells with Buffer 1 and spun down the cells again. The same operation was repeated until the supernatant became colorless. The recipe of buffer 1 is as follows: 20 mM Tris, 200 mM NaCl, 20% glycerol, 2 mM TCEP, pH 7.7. The cleaned cell paste was resuspended by Buffer 1 and lysed on ice by sonication with 20 seconds of pausing after

every 10 seconds of pulsing for a 4-minute pulsing. Then the cell fragment was spun down by ultracentrifuge at 40,000 rpm for 30 minutes. The supernatant was retained for protein purification in a workflow shown in Figure 3.3.

Figure 3.3 The workflow of protein purification.

To isolate the protein from others, the supernatant was loaded to a 5-mL MBPTrap HP prepacked column (GE Bioscience), which was subsequently rinsed by the elution buffer to release the target protein. The elution buffer was prepared by dissolving 10 mM of maltose in Buffer 1. The eluant was then concentrated by a 30-kD ultracentrifuge filter (Amicon) to around 1 mL and injected into a Superdex 200 10/300 GL column (GE Bioscience) equilibrated with Buffer 1 for size exclusion chromatography. The interested fractions were gathered, combined, concentrated, and stored at -80 ℃.

Depending on the next experiment, the MBP tag of the protein would be cleaved by enterokinase. The tagged protein was treated with the enzyme at 12 ℃ for 72 hours, then flowed through a 1-mL MBPTrap HP prepacked column (GE Bioscience) to remove MBP and undigested protein as the protocol mentioned above. The RORγ LBD carries several histidine residues on the protein surface, and it shows weak binding against the nickel column, thus I eluted the untagged protein by 50 mM of imidazole in a 1-mL HiTrap HP prepacked column (GE Bioscience). The eluant was buffer-exchanged with Buffer 1 for removing imidazole and preparing the next experiment or storage at -80℃.

3.2.2 Site-directed mutagenesis

I constructed different RORγ LBD mutants using Q5® Site-Directed Mutagenesis Kit (NEW ENGLAND Biolabs Inc.). The sequences of the forward and reverse primers containing mutation sites were generated by NEBaseChanger and ordered from Integrated DNA Technologies Inc. (Table 3.1). Then a PCR sample, composed of primers, Q5 master mix, template DNA, and nuclease-free water, was amplified in PCR for 25 cycles by following the instructions provided from NEBaseChanger. The resulted DNA was ligated and transformed into E. cloni 10G SOLOs chemically competent cells (Lucigen) for plasmid propagation. The cells were grown on an LB plate pretreated with carbenicillin overnight, and several colonies were picked for DNA sequencing. The plasmids with the correct sequence were kept at -20 ℃ for further experiments.

Table 3.1 Primers designed to create the RORγ LBD mutants.

Mutation	Forward primer	Reverse primer
$SRC2^{+}$	CTGCATCGCCTGCTGCAGGATAGC TAAGGATCCGAATTCCCTG	AATTTTATGTTTTTCGCCGCCGCCG GTGGAAAACAGTTCTTTATAC
$H12^{-}$	TAACACCCGATTGTTGTCCAAG	CAGATGTTGGAAGATCTGC
K469A/R473A	AAACTGGCATCGCTGTGTAGCCAA CAC	ACCTGCCGGCGGCAGTTTGGCCAG AAT

3.2.3 Reporter gene assay

I tested the bioactivity of a compound against RORγ by the Human RORγ Reporter Assay System (INDIGO Biosciences). The engineered human cells provided by the kit express the GAL4 DBD- RORγ LBD hybrid receptor. When the cells are treated by RORγ agonists, the receptor attaches to a specific DNA sequence and triggers the expression of luciferase, which catalyzes the substrate to luminate. In contrast, RORγ inverse agonists depress the light emission. Here I operated the experiments by following the instruction provided and treated the cells with the selected compounds in a series of concentrations at 37 ℃ for 22 hours, then added

the substrate to each treatment and quantified luminescence in a plate reader. Four aliquots were set for each treatment. The raw data was processed and plotted, and the IC_{50} value for each compound was measured by GraphPad.

3.2.4 Differential scanning fluorimetry

Differential scanning fluorimetry (DSF) is a biophysical method quantifying the stability of a protein or protein-ligand complex by measuring its melting temperature (T_m). The protein or complex is incubated with a hydrophobic fluorescent dye in this method. The fluorescence of the dye molecule is quenched by water molecules in the aqueous solution until it binds to the hydrophobic surface of the protein. Here I applied Vivoli's DSF recipe [75] with specific adjustments to preliminarily estimate the binding capacity of the VS hits.

A protein solution was prepared by mixing the protein stock with 3.7 µL of 0.5 M HEPES at pH 7.0, 28.0 µL of 1 M NaCl, and DI water to finalize the solution volume in 180 µL with protein concentration at 0.11 mg/mL. In each well of a 96-well qPCR reaction plate, 17 µL of the protein solution was mixed with 2 µL of the ligand solution and 1 µL of the 20x fluorescent dye, SyproOrange (Invitrogen ™). The plate was sealed by clear films and incubated on ice for 30 minutes to ensure sufficient protein-ligand interaction. After that, it was loaded to the Bio-Rad iCycler. I placed the excitation filter at 490 nm and the emission filter at 620 nm in the instrument and created a melting curve method on the connected computer program that the plate was equilibrated at 25 ℃ for 2 minutes, then gradually heated up to 90 ℃ in approximately 40 minutes. During the temperature ramp, the fluorescence intensity of the mixture in each well was detected in real time. When the run was finished, I collected the temperature-sensitive fluorescence data and uploaded it to an online server JTSA (http://paulsbond.co.uk/jtsa)

produced by Paul Bond in 2017. The curve of fluorescence reading versus temperature was fitted by Sigmoid-5 equation [76] to estimate the T_m of each sample. Generally, the higher the T_m value, the more stable the protein-ligand complex is.

3.2.5 Nuclear magnetic resonance (NMR) spectroscopy

The 2D ^{1}H-^{15}N HSQC NMR experiments were executed at 12 ℃ on a Bruker 600 MHz spectrometer equipped with a 5-mm cryogenic triple resonance probe. Complex points 2048 × 512 and spectral widths of 16 × 35 ppm were used to record all the spectra. High resolution 2D 1H-15N HSQC NMR spectra were collected using 150 μM of ^{15}N-labeled, no-tag RORγ LBD treated with 0-400 μM of RG14-2 dissolved in DMSO using 3 mm Wilmad 335 NMR sample tube. The collected NMR spectra were visualized and aligned by NMRFAM-Sparky [77]. The NMR titration plot and the K_d measurement were generated by python codes written based on the curve_fit and minimize_scalar modules in the scipy.optimize package (see Appendix C).

3.2.6 X-ray crystallography

The untagged protein at approximately 5 mg/mL was prepared in crystallization buffer (20 mM HEPES at pH 7.4, 50 mM NaCl, 5% glycerol, 2 mM TCEP). 50 mM of RG14-2 in DMSO was added to the protein solution in a 1:1 molarity ratio. The mixture was incubated on ice for at least 30 minutes and then spined at 12,000 rpm for 20 minutes to remove the precipitated protein and/or RG14-2. The 0.2 μL of protein solution and 0.2 μL of crystallization reagents were dispensed on the micro-well surface of a 96-well crystallization plate by Crystal Gryphon (Art Robbins Instruments). The reservoir next to each microwell was filled with 60 μL of relative crystallization reagents. Conditions in MCSG Crystallization Suite (MCSG1,

MCSG2) and NR-LBD™ + NR-LBD™ Extension HT-96 kit (Molecular Dimensions) were applied to screen suitable crystallization methods. The plates were incubated at 12 ℃ for 24 – 48 hours before the first piece of crystal was observed.

For an X-ray diffraction experiment, the crystals were picked by cryoloops, cryo-protected by 20% of ethylene glycol, and fast-frozen in liquid nitrogen. The diffraction data of the co-crystal was collected by SSRL beamline 12-1 at SLAC National Accelerator Laboratory at Stanford University. The data was integrated by X-ray Detector Software [40] and reduced by Pointless [78] and Aimless [79] in the CCP4 suite. Molecular replacement was conducted to build the crystal model by Phaser [41]. The crystal structure was preliminarily refined by Coot 0.8.9-pre EL [42] manually and Refmac5 [43] computationally.

3.3 Results and Discussion

3.3.1 Workflow of RORγ LBD preparation

The pMAL-C5e plasmid carrying the gene of the MBP-tagged RORγ LBD was transformed into Rosetta2 (DE3) Singles competent cells for overexpressing the protein under the IPTG induction. The cells were collected, lysed, and centrifuged to remove the cell fragments. The protein expression level in the supernatant was confirmed by Western Blot (Figure 3.4). The supernatant flowed through an MBPtrap HP prepacked column. The SDS-PAGE gel result shows that most of the proteins without the MBP tag were rinsed out from the column, and the MBP-tagged RORγ LBD was eluted by the elution buffer (Figure 3.4). To further remove the other proteins retained with RORγ LBD in the eluant, I concentrated the eluant and conducted size exclusion chromatography against it. As shown in Figure 3.5a, two significant peaks were presented on the UV spectrum. According to the SDS-PAGE gel result

(Figure 3.5b), RORγ LBD is the main ingredient in both peaks. The retention time of the left peak implies dimerization of the protein with other proteins attached to it. To ensure the homogeneity of the protein for the following experiments, I only collected the fractions belonging to the peak on the right side.

Given that the bulky MBP tag (42 kD) could influence the following NMR and co-crystallization experiments, it was removed from the protein by enterokinase. The MBP was mainly separated from the cleaved protein by MBPtrap HP prepacked column, but a traceable amount was flowed with the untagged protein, probably due to MBP misfolding. Taking advantage of several histidine residues clustering on the protein surface, I loaded the wash-through of the MBPtrap column in a HisTrap HP prepacked column. The untagged RORγ LBD was eluted from the column with a clean background (Figure 3.5c).

Different RORγ LBD constructs were produced in the experiments mentioned in the sections below, yet the protein preparation followed the same protocol described here.

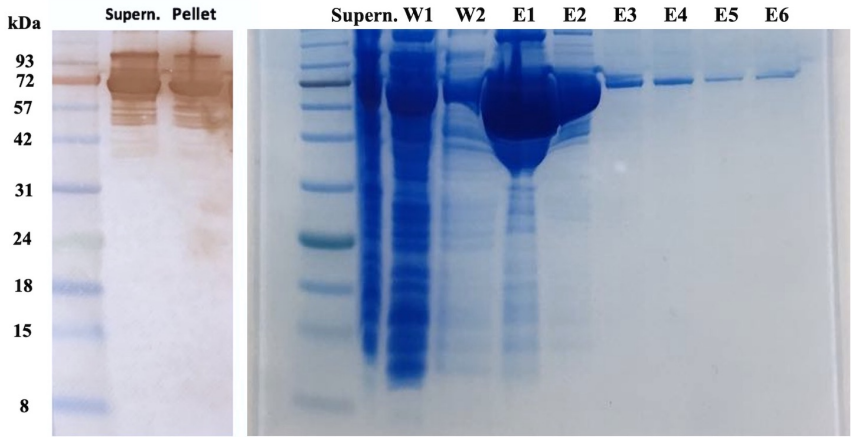


Figure 3.4 Expression and MBPtrap-column purification of the MBP-tagged RORγ LBD. Thick bands were shown at around 72 kDa of the prestained protein ladder on the western blot membrane (left) for both supernatant (supern.) and cell pellet. The SDS-PAGE gel (right) result showed most other proteins were separated in washing through fractions W1 and W2 after MBPtrap-column purification. E1, E2, E3, E4, E5, and E6 represent six fractions of the eluant collected from the column.

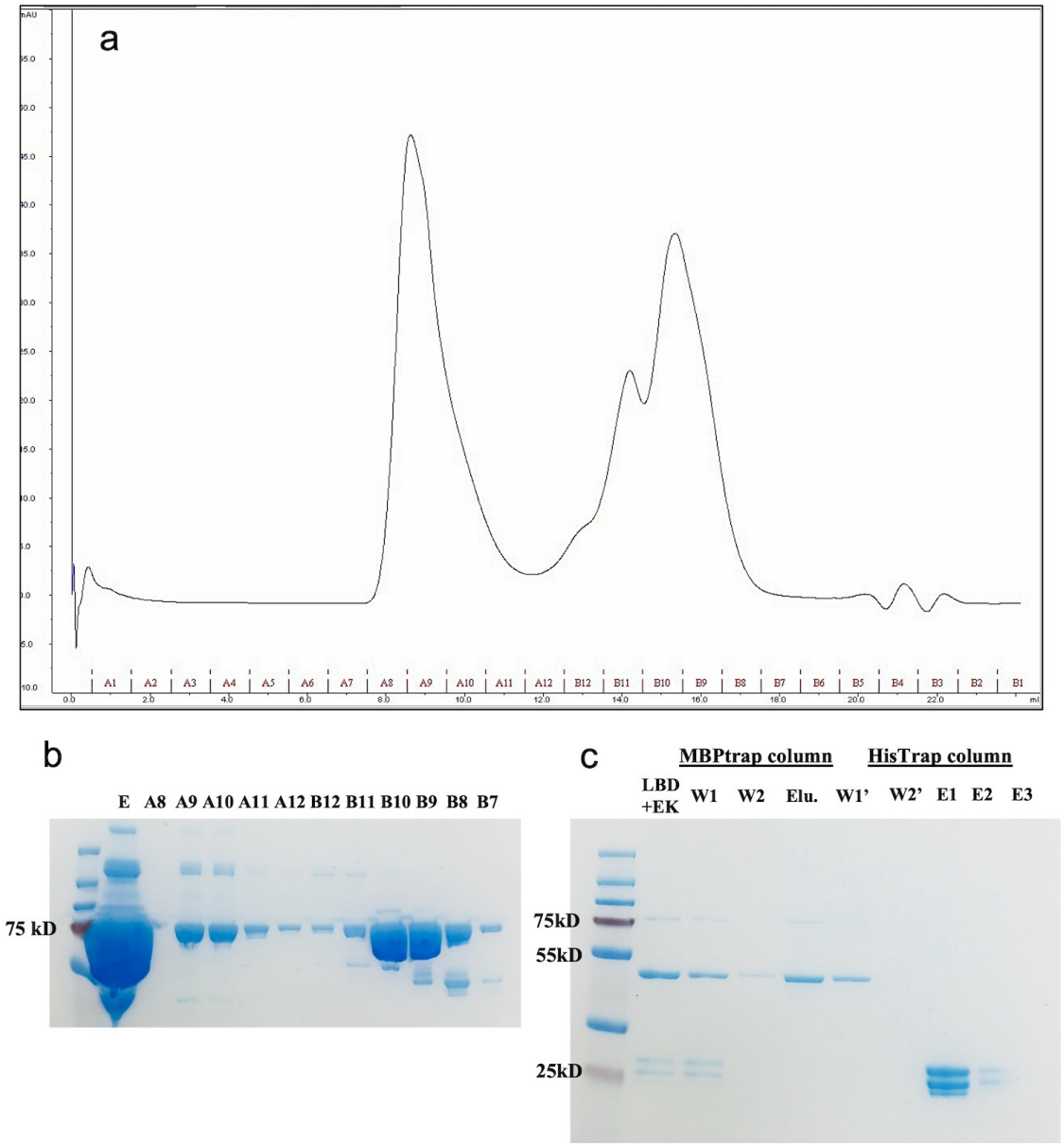


Figure 3.5 Size exclusion chromatography and MBP cleavage of the MBP-tagged RORγ LBD.

(a) The UV absorption spectrum of size exclusion chromatography showed two dominant peaks. (b) The fractions of two peaks were analyzed by SDS-PAGE gel. (c) After treated with enterokinase (EK), the protein was loaded to an MBPtrap column for removing most cleaved MBP tag, and then the washing through fraction W1 was loaded to a HisTrap column to further purify the protein.

3.3.2 Preliminary binding screening of the selected compounds

The DSF assay was applied to quickly detect the binding capability of the purchased VS hits against an MBP-tagged RORγ LBD variant linking the co-activator SRC2, designated as the $SRC2^{+}$ mutant. The short peptide SRC2 can be recruited on the surface of the LBD when H12 attaches to the protein in a correct pose. Since the T_m value measured in DSF is an intrinsic constant describing the thermal stability of a protein or protein-ligand complex, introducing the peptide into the protein could count the effect of the compounds on the protein-coactivator interaction in the measuring system. In addition, the no-tag apo RORγ LBD displays strong fluorescence background, which covers the signal of protein unfolding with increasing temperature in the assay (Figure 3.6a). In contrast, retaining the MBP tag in the protein diminishes the background signal and ensures that the protein unfolding process is profiled by fluorescence for fitting curves and measuring the T_m values on JTSA.

In the assay, the known RORγ inverse agonist, ursolic acid (UA) [80], was utilized as a positive control, and the apo protein was treated with DMSO as a negative control. As can be seen in Figure 3.6b, the compound treatments carry the T_m values between the UA and DMSO treatments, suggesting the potential binding capability of the compounds against the MBP-tagged $SRC2^{+}$ mutant. In particular, the compound RG14-2 shows a T_m value standing out from the others and slightly lower than UA, making it the best candidate for further study. It is worth noting that RG14, RG14-1, RG14-2, RG14-3 are chemically analogous (Figure 3.7). Their distinct T_m values imply that a slight difference in the compound structures could impact the interaction with the protein.

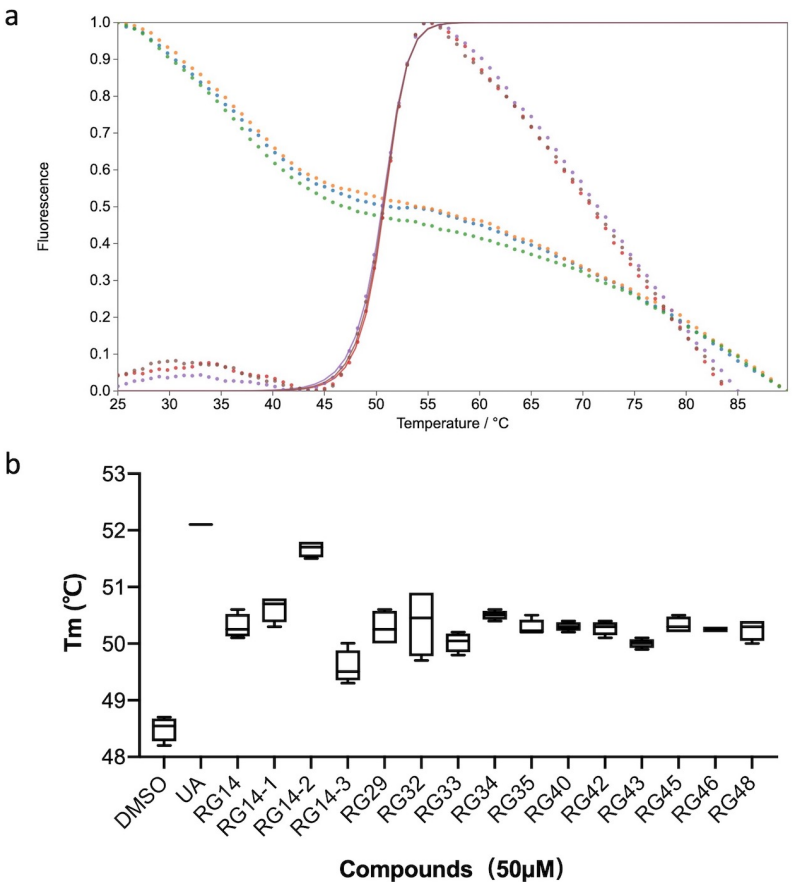


Figure 3.6 Binding capability of the selected compounds against the MBP-tagged RORγ LBD.

(a) The peaks of the no-tag RORγ LBD were covered by strong background signals (the decreasing dot lines). The peaks of the MBP-tagged protein were curve-fitted by JSTA (the increasing dot lines). The data was normalized. (b) Fifteen compounds were tested with the DMSO treatment as a negative control and ursolic acid (UA) as a positive control.

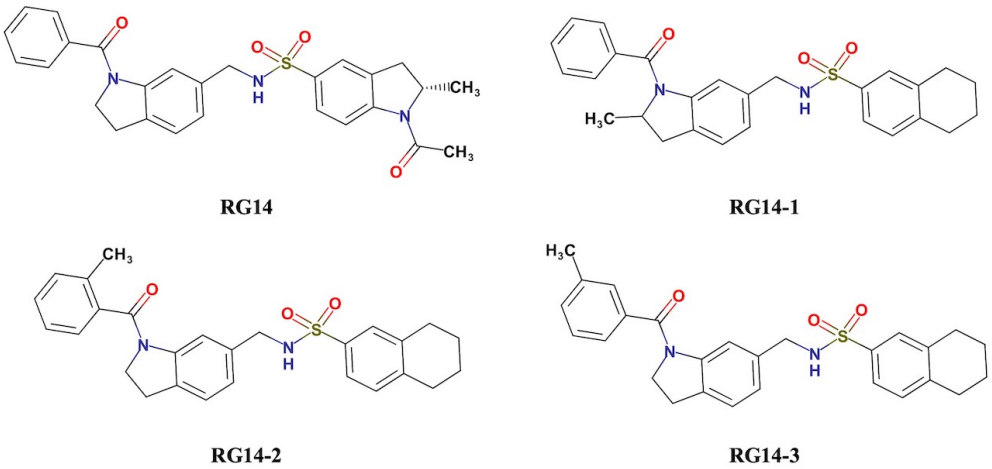


Figure 3.7 Chemical structures of RG14 and its analogs.

3.3.3 Inverse agonism of RG14 and RG14-2 on the RORγ transcriptional function

With the outstanding performance of RG14-2 in DSF and its scaffold unique from the known RORγ ligands, a cell-based reporter gene assay was conducted to investigate the bioactivity of this compound. RG14 was also included in the assay to compare the effect of the subtle structural difference (Figure 3.8). The positive control UA exhibits the most vigorous inverse agonistic activity against the RORγ transcriptional function in the smallest IC_{50} value of 0.11 μM, while RG14 and RG14-2 undermine the function in the IC_{50} values of 3.2 and 1.5 μM, separately. The data coincides with the DSF result mentioned above in numerical order and excludes the possibility of the compounds binding against the MBP tag rather than RORγ LBD. Despite the inhibitory efficacy being weaker than UA, RG14-2 has demonstrated the potential as a lead compound to be developed into the RORγ inverse agonists. Therefore, its interaction with the RORγ LBD was investigated in the following experiments.

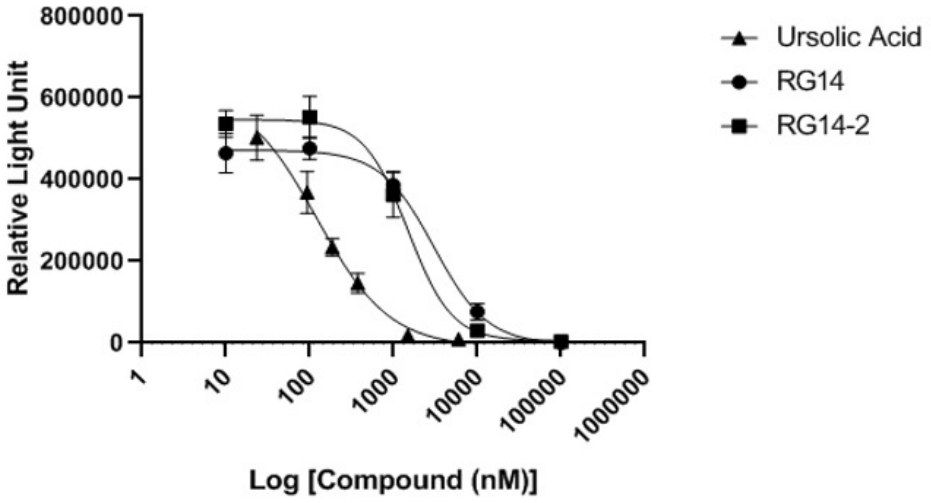


Figure 3.8 Inhibition of RG14 and RG14-2 on the RORγ transcriptional function. Reporter gene assay was conducted to test the compound bioactivity with ursolic acid as a positive control.

3.3.4 Binding affinity of RG14-2 against the RORγ LBD

The DSF assay is time- and sample-saving to preliminarily screen the binding capacity of plenty of compounds against a target protein, but it is difficult to precisely distinguish the signals in submicromolar difference. Thus, it is unsuitable for applying it to measure the dissociation constant (K_d) of the ligands with moderate or strong binding. To better estimate the K_d value of RG14-2 against the RORγ LBD, a 2D ^{1}H-^{15}N HSQC NMR experiment was conducted. Here ^{15}N-labeled RORγ LBD without tag was prepared as described in 3.3.1, except that M1 media with ^{15}N-labeled ammonium chloride substituted the TB media in the cell culture step. The protein at 150 μM was then treated with RG14-2 at a series of final concentrations (0, 0.5, 5.5, 25.0, 50.0, 125.0, 200.0, 400.0 μM) and scanned by NMR at every concentration. I overlapped the spectra to analyze the chemical shift of the signals (Figure 3.9a). A signal was chosen for calculating the K_d value of RG14-2, considering its sharp peak and its smooth movement with the concentration change (Figure 3.9b, c). RG14-2 features a K_d value of approximately 5.7 μM, elucidating a moderate binding affinity against the RORγ LBD.

3.3.5 Structural study of the interaction between RORγ LBD and RG14-2

I tried to co-crystalize the wild-type RORγ LBD with RG14-2 to explore the interaction structurally. Unfortunately, no crystal was grown after testing in over 300 crystallization conditions. I suspected that H12 in the bound RORγ LBD could be highly mobile in the solvent environment due to the inverse agonistic effect of RG14-2, preventing the complex molecules from forming ordered crystal lattices in the space. H12 is missing in many co-crystal models of RORγ LBD-inverse agonist complexes released on Protein Data Bank, implying the possible disruption in crystallization and the necessity of removing it.

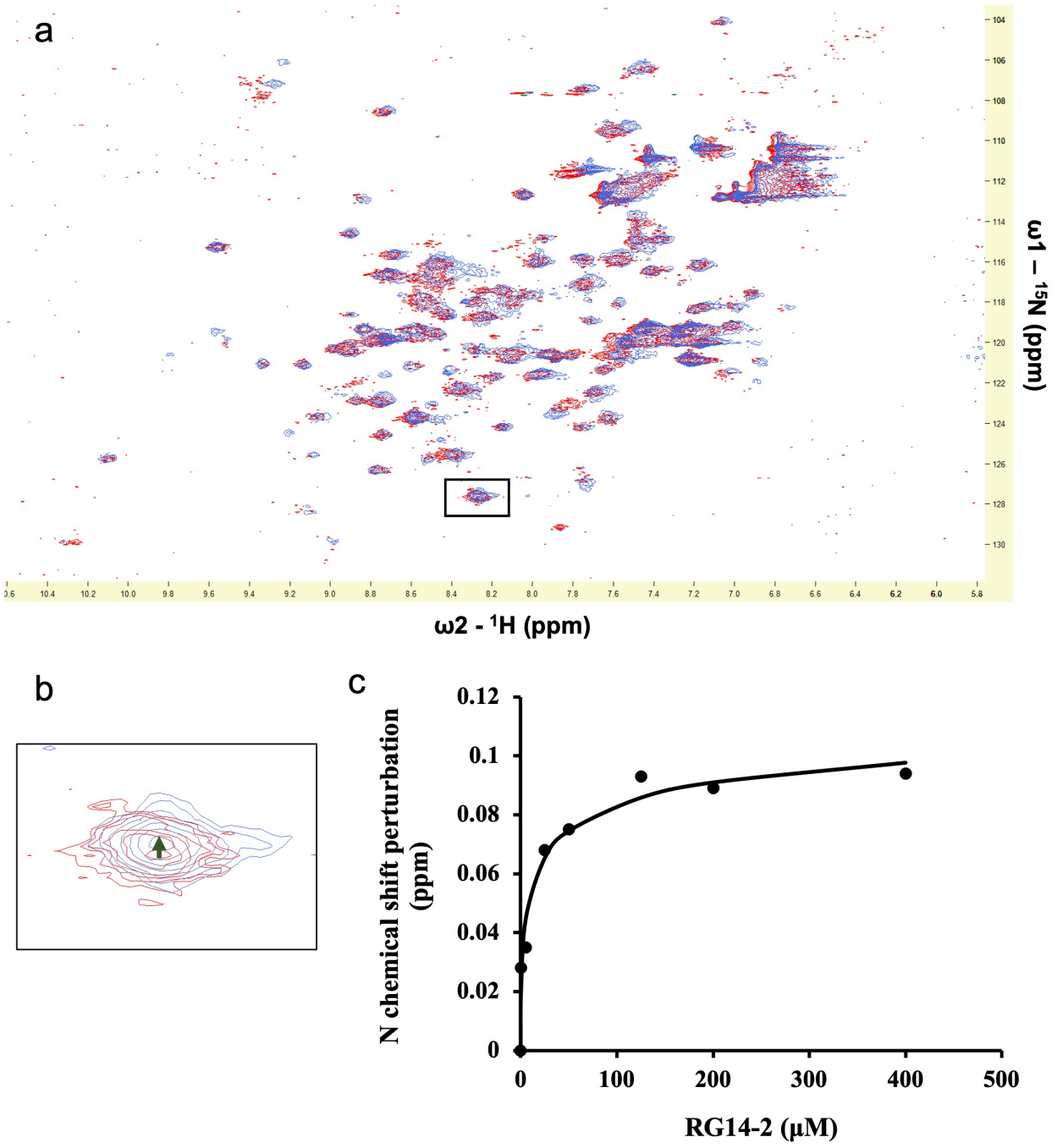


Figure 3.9 Binding affinity of RG14-2 against the RORγ LBD was measured by ^{1}H-^{15}N HSQC NMR experiment.

(a) Two NMR spectra of the RG14-2 treatment at 0 μM (red) and 125 μM (blue) were aligned. The selected signal for K_d measurement was highlighted by a square. (b) The amplified view of the selected signal showed the chemical shift with RG14-2 concentration. (c) The titration plot of chemical shift was built and curve-fitted to measure the K_d value.

Thus, I truncated the H12 and constructed an $H12^-$ mutant specifically for the co-crystallization experiment. However, co-crystals still could not grow.

Noguchi et al. stated that substituting the residues Lys469 and Arg473 with Alanine could overcome the co-crystallization problem they confronted in their studies [64,81]. It could be related to the high surface entropy of Lys469 and Arg473 side chains subverting protein crystallization. Accordingly, I mutated them on the $H12^-$ mutant and then treated the mutant protein ($H12^-$/K469A/R473A) with RG14-2. Clear, diamond-shape crystals were grown (Figure 3.10a). The crystals were picked and sent for an X-ray diffraction experiment. A piece of crystal was diffracted in the resolution of 3.5 Å. Since no other crystal displayed better resolution, I processed the diffraction data of this crystal and built an early 3D structural model of the protein-ligand complex for glancing at the interaction.

It is a monomeric protein model in a $P6_122$ space group with the R and R_{free} values at 0.25 and 0.32, respectively. The backbone of the protein model could be generally fitted into the 2Fo-Fc electron density map, except for several C-terminal residues and a loop moiety between Pro468 and Ala473 (Figure 3.10c), hinting a dynamic state of the loop in the environment. It also could be a reason for the difficulty in crystallizing the bound $H12^-$ mutant. A Y-shape blot was found in the empty core of the protein model in the 2Fo-Fc map, and the positive density of the Fo-Fc map indicates the actual signal of the blot (Figure 3.10b). RG14-2 could be fitted into the blot in a folded shape, but the poor resolution prevents me from further refining the posture of the compound.

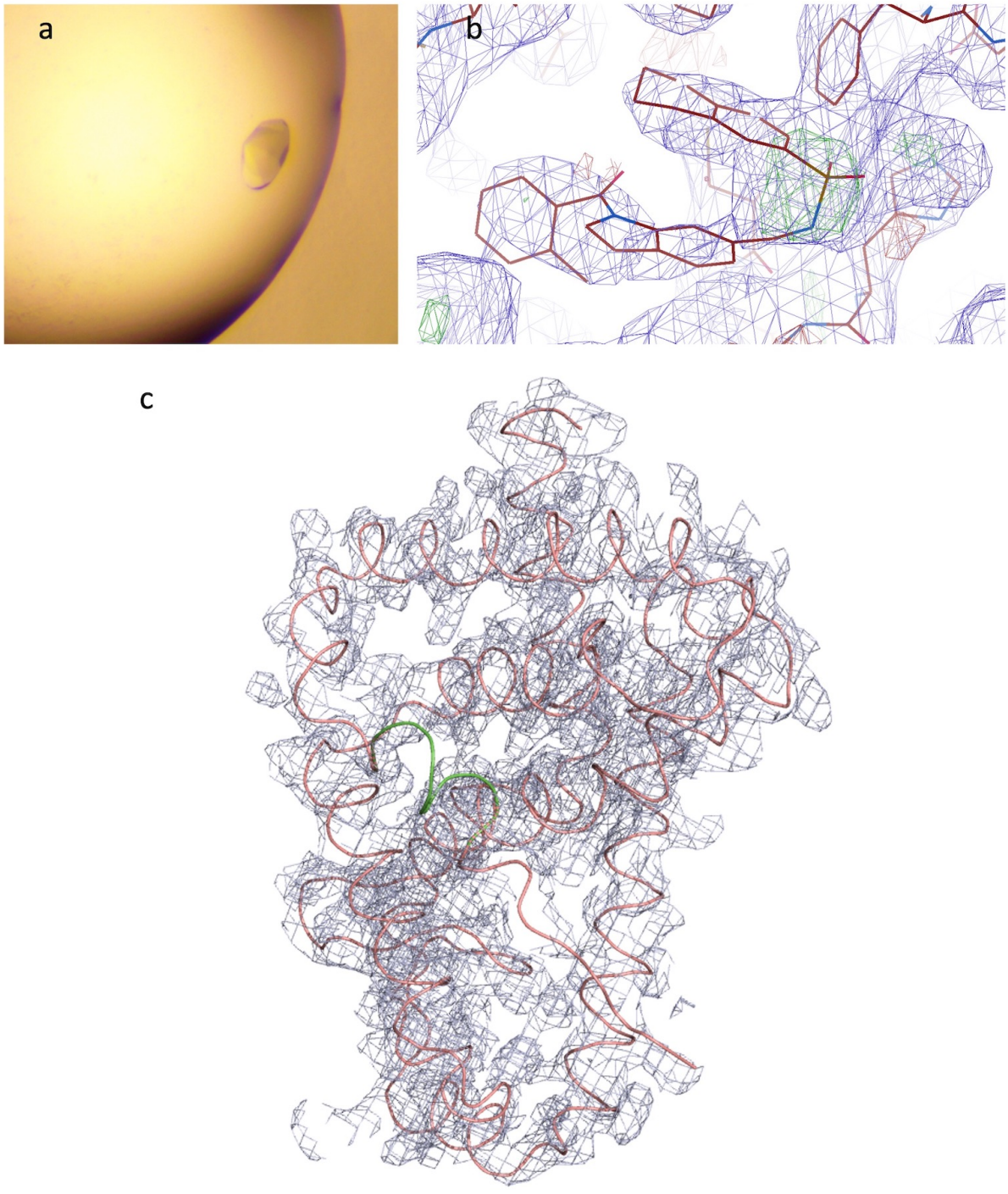


Figure 3.10 X-ray crystallography of the RORγ LBD H12^{-}/K469A/R473A mutant bound with RG14-2.

(a) A crystal of the complex was photographed under the microscope. (b) RG14-2 was laid in the blots with σ values of the 2Fo-Fc and positive Fo-Fc electron density maps at 0.7 Å (blue) and 3.0 Å (green). (c) The backbone of the entire protein crystal (salmon ribbon), except a loop portion between Pro468 and Ala473 (green ribbon), was well fitted in the 2Fo-Fc map (grey).

RG14-2 impacts the RORγ LBD conformation in several aspects. Firstly, compared with the models of the apo protein (5X8U) and the H12^{-}/K469A/R473A mutant bound with an inverse agonist (6A22), part of H11 transforms from a helix to a loop in the RG14-2-bound protein (Figure 3.11a, b). It is more common for an inverse agonist to break the helical conformations of H11' and H12, while H11 is affected by leaning the entire helix to the other side [82]. Only two cases have been reported that inverse agonists interrupt the H11 conformation [83,84]. This could result from the protrusion of the methylbenzene ring in RG14-2 pushing H11 away from the original position. Also, its hydrophobicity rebuilds the noncovalent interactions with the surrounding residues and disfavors the maintenance of the helical state. Secondly, the density-missing loop noted above shows a conformation different from 6A22 (Figure 3.11d), although they share identical residues in this portion, which includes two mutated residues Ala469 and Ala473. However, it can align with 5X8U (Figure 3.11c). More investigation is needed to determine its role and the effects of two mutation sites. Thirdly, the key residue His479 shifts downwards because of the H11 disordered and extent in the RG14-2-bound model (Figure 3.11e). Though H12 is truncated, it is reasonable to rationalize that this shift could destroy the agonist lock, release H12 to the space, and disfavor the co-activator recruitment. Compared with the one in 6A22, the His479 side chain flips to an opposite direction (Figure 3.11d), implying a different mode of RG14-2 in disrupting the agonist lock.

Restricted by the low resolution of the co-crystal model, the side chains of many key residues are unreliable in their poses, which impedes the detailed analysis of the protein-ligand interaction. A model with better resolution is required for digging more structural information.

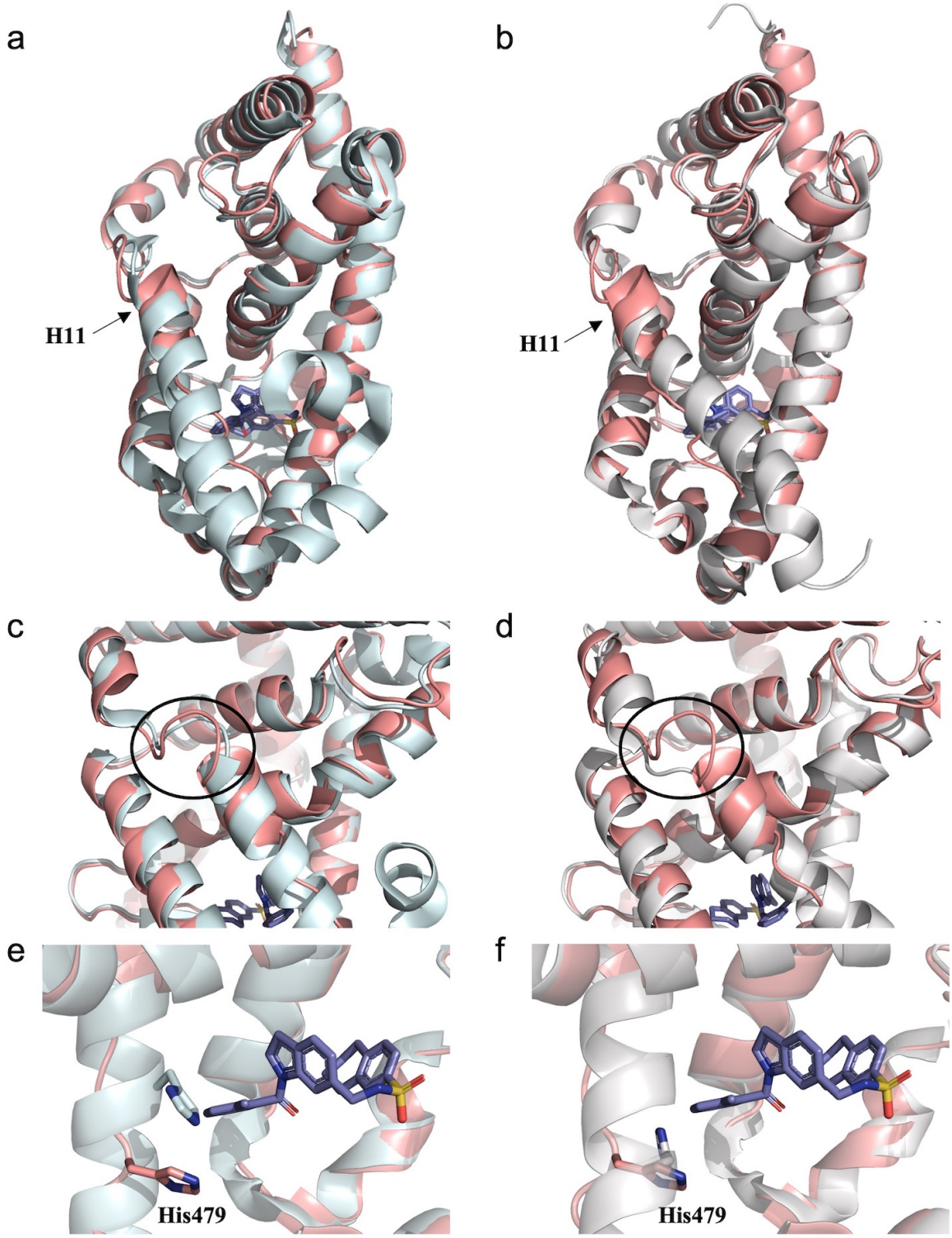


Figure 3.11 Crystal model alignment with the apo protein (5X8U, light blue) and the published bound mutant (6A22, light grey).

The built protein model was presented in salmon cartoon, and RG14-2 was in blue sticks. (a, b) The H11 was disordered. (c, d) The loop containing two mutated residues Ala469 and Ala473 were circled. (e, f) The His479 conformation differed from 5X8U and 6A22.

3.4 Conclusion

In this project, a small molecule RG14-2 is selected by computational experiments for potentially inhibiting the transcriptional function of a human nuclear receptor RORγ as an inverse agonist. With experimental verification, it inhibits the protein activity in the IC_{50} value of 1.5 μM and binds against the protein LBD in the K_d value of 5.7 μM. The co-crystal model of the RORγ LBD mutant bound with RG14-2 evidences that the inverse agonistic activity of RG14-2 is accomplished by breaking the agonist lock between H11 and H12, which disfavors the recruitment of a co-activator on the protein surface to trigger the protein function.

Despite hundreds of RORγ inverse agonists being explored, and some of them are pushed to clinical trials, no one has survived the Phase II trial or obtained approval from the FDA. RORγ is an excellent target for drugs treating autoimmune diseases that more efforts are needed to develop new drugs. RG14-2 was picked from a drug-like compound database containing millions of small molecules. It shows a chemical backbone distinct from all other known RORγ inverse agonists, making it valuable to further develop into a more potent compound. Moreover, its unique structure may lead to a different performance in higher-level biomedical experiments, such as animal experiments, worthy of in-depth investigation.

In contrast to the traditional pre-clinical drug discovery pathway, the computer-aided method saves a lot of time and cost. Virtual screening against millions of compounds can be finished in several weeks, yet it is infeasible to achieve by laboratory experiments in such a short time. After shortening the compound list to hundreds of compounds or an even smaller amount, experimental screening can be quickly finished and allow scientists to move to the next stage of drug development. The workflow of this project is an excellent example to show the superiority of the computer-aided drug discovery method, especially operated by an academic lab. For

example, it is less labor-dependent, given that part of the human work can be substituted by the computer; it is more affordable since the chemicals required in the experiments are narrowed down to a much smaller pool. It is easy to predict that computer-aided drug discovery will become the mainstream strategy in the recent future, and more advanced computer methods will be built to boost the study in the drug discovery field.

Chapter 4 - Computer-Aided Discovery of an ERα Modulator Overcoming the Y537S Mutation

Abstract: Breast cancer ranks the second biggest death-leading reason in women, and approximately 70% of the cases are detected ERα-positive ($ERα^+$), making ERα an ideal drug target for treating breast cancer. Unfortunately, almost half of the patients carrying $ERα^+$ breast cancer acquire drug resistance issues caused by mutations on ERα. In this project, I targeted the most aggressive mutant Y537S and attempted to discover a new generation of ERα modulators overcoming the Y537S-related drug resistance. Virtual screening was firstly performed on two small-molecule libraries, MolPort database and OTAVA SERMs-Like Library, against six crystal models of the ERα Y537S ligand-binding domain, respectively. The top-ranking hits were then redocked and rescored by Glidescores, RF-scores. The shortened hit list was further profiled by applying multiple filtering indexes, including molecular weight, lipophilicity, hydrogen-bonding numbers between the protein and the ligand, accessibility, chemical difference to the known drugs, and so on. Five compounds were chosen for purchase and tested by differential scanning fluorimetry. A compound ERA1 is found capable of binding against both the wild-type ERα and the mutant in a comparable manner, showing its potential as a lead compound for developing into new SERMs.

4.1 Introduction

With all cancers found in women, breast cancer occupies 25% of them and ranks the second biggest death-leading reason. Approximately 70% of the detected breast cancer cases are estrogen receptor α-positive ($ERα^+$), as ERα mediates proliferation and metastasis of the cancer cells through estrogen binding against it [2]. It makes ERα such an ideal drug target that several

selective ERα-response modulators (SERMs) have been developed to block the estrogen-ERα interaction by pharmaceutical companies. Several of them are approved by FDA, such as tamoxifen [85], toremifene [86], and raloxifene [87]. However, up to 50% of the patients holding $ERα^+$ breast cancer acquire drug resistance issues after about five years of the canonical treating strategy [88]. Intensive studies reveal that the resistance results from mutations found in the ERα gene [89–92].

The human nuclear receptor ERα is a transcription factor broadly expressed throughout the body. Similar to RORγ, the protein is constituted of five domains, an N-terminal domain (1 – 184 a.a.), a DNA-binding domain (DBD, 185 – 250 a.a.), a hinge-region (251 – 310 a.a.), a ligand-binding domain (LBD, 311 – 547 a.a.), and a C-terminal domain (548 – 595 a.a.). The structural study is mainly concentrated on the LBD because the orthosteric binding site posits in the core of the LBD. The LBD is formed by 12 helices. Although H12 is the most dynamic helix playing a crucial role in the co-activator recruitment, just like the RORγ LBD, no agonist lock is observed in the ERα LBD. When the endogenous agonist (estradiol, estrone, estriol, or estetrol, etc.) binds in the orthosteric binding pocket, H12 covers the pocket exit (Figure 4.1), allowing the co-activator with the "LXXLL" motif to interact with the activation function (AF)-2 cleft on the LBD surface, facilitating the DBD attachment at the DNA sequence and triggering the downstream transcription in the cells (e.g., the $ERα^+$ breast cancer cells). Most of the antagonistic SERMs compete with estrogens to bind against the orthosteric site for prohibiting the transcriptional function. It is accomplished by disrupting the H12 pose on the pocket exit and favoring it to occupy the AF-2 cleft with its "LYDLL" motif (536 - 540 a.a.), blocking the coactivator recruitment [93].

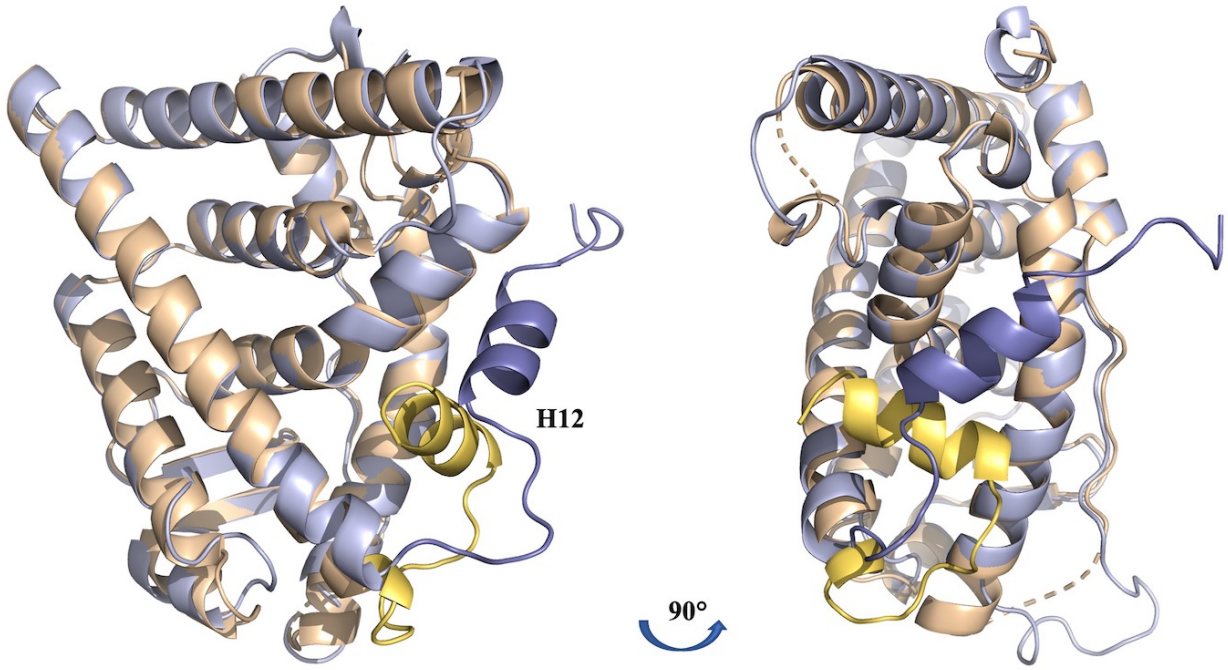


Figure 4.1 Alignment of the ERα LBD models bound with an agonist (1ERE, wheat) and an antagonist (3ERT, light blue).

To indicate the H12 dynamics, the sequence moiety from Lys529 to the C terminus was colored in bright yellow for 1ERE and slate for 3ERT.

Multiple mutations have been discovered on the ERα LBD, including E380Q, L536Q/R, Y537C/N/S, and D538G [88]. Mutations Y537N/S and D538G are the most predominant forms detected in the drug-resistant ERα^{+} breast cancer cells, while Y537S is the most aggressive type. According to the structural data, the Y537S mutation on H12 enables an additional hydrogen bond formed between Asp351 and Ser537 (Figure 4.2), which stabilizes the agonistic pose of H12 and constitutively promotes the coactivator recruitment, maintaining the agonistic state of the protein even without a ligand bound in the pocket [94,95]. This mutation impairs the binding affinity of 4-hydroxytamoxifen (the active metabolite of tamoxifen) against the LBD in an 8-fold decreasing manner [95]. Therefore, it is urgent to develop a new generation of SERMs for targeting both the wild type and the mutant to overcome the drug resistance. This topic started drawing attention in the last decade, whereas only limited reports have been released on discovering new

SERMs overcoming drug resistance caused by the Y537S mutation [96,97]. The process of drug discovery is too slow to meet the urgent need.

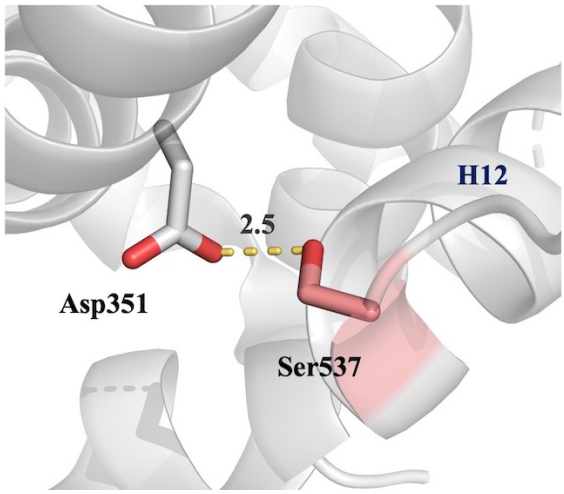


Figure 4.2 Ser537 forms a hydrogen bond with Asp351 in the ERα Y537S mutant (3UUD).

Here I attempted to discover new SERMs through virtually screening two different small-molecule libraries, MolPort database and OTAVA SERMs-Like Library, followed by picking the hits with multiple filtering indexes, including Glidescores, RF-scores, molecular weight, lipophilicity, hydrogen-bonding numbers between the protein and the ligand, accessibility, chemical difference to the known SERMs, and so on. Five selected compounds were purchased and tested by differential scanning fluorimetry. In particular, a compound ERA1 is capable of binding against both the wild-type ERα and the mutant in a comparable manner, showing its potential as a lead compound for further developing into new SERMs.

4.2 Materials and Methods

4.2.1 Pharmit

Pharmit (http://pharmit.csb.pitt.edu/index.php) is an open-source online server for high-throughput screening drug-like compounds on browsers [98]. It currently contains 1,156,210,113 conformations of 248,031,356 compounds from nine built-in compound libraries, which are CHEMBL25, ChemDiv, ChemSpace, MCULE, MCUL-ULTIMATE, MolPort, NCI Open Chemical Repository, LabNetwork, and ZINC. It detects the binding site based on the co-crystallized ligand in the crystal model and lists the noncovalent protein-ligand interactions. A VS run could be set up by specifying the pharmacophore noncovalent interactions, spatial arrangement, and hit chemical features (molecular weight, LogP value, hydrogen-bond donor and acceptor numbers, and so on) (Figure 4.3).

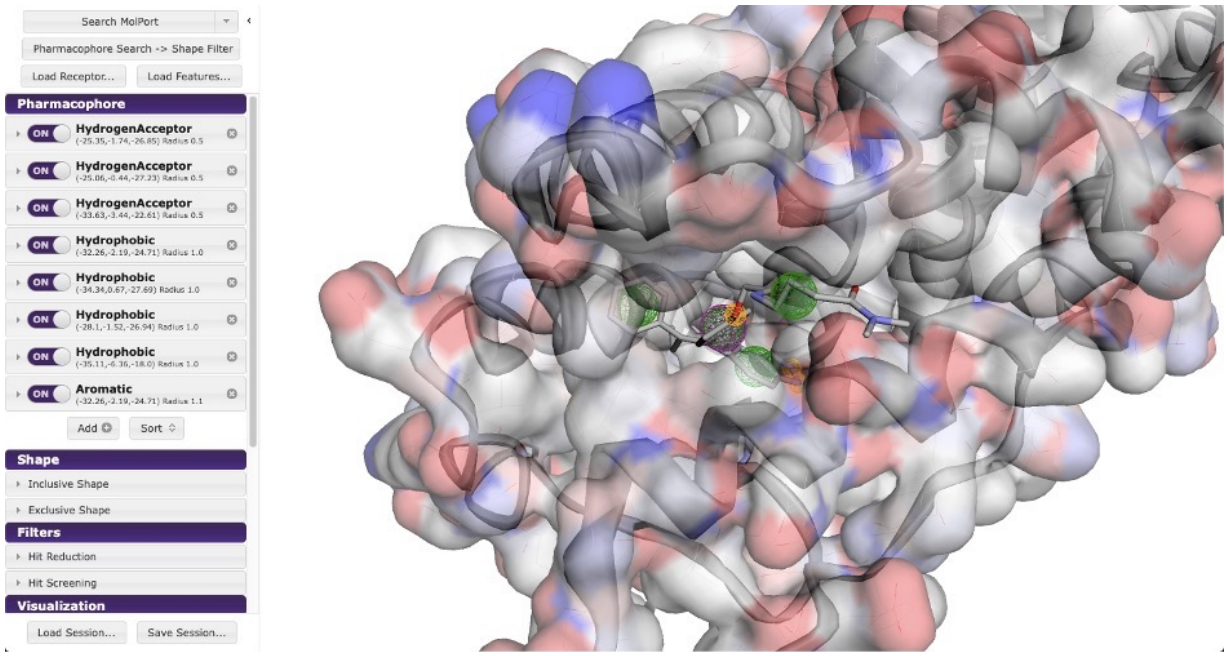


Figure 4.3 The interface of Pharmit online server.

The screening usually takes a few seconds to several minutes to finish. Since the docked poses of the hits are directly extracted from libraries, energy minimization is required to further recognize the reliable hits after the run. The built-in programs AutoDock Vina [99] scoring function and smina [100] accomplish it by scoring and ranking the hits with specialized algorithms. The hits could be downloaded for further analysis.

4.2.2 Glide

Glide is a program produced by Schrödinger LLC for high-throughput screening small compounds against the receptors with exporting optimized scoring functions [101,102]. The general workflow of VS on Glide is summarized as shown in Figure 4.4. The experiment was operated on Maestro, a graphical user interface of Schrödinger's computational programs, by following the instruction provided.

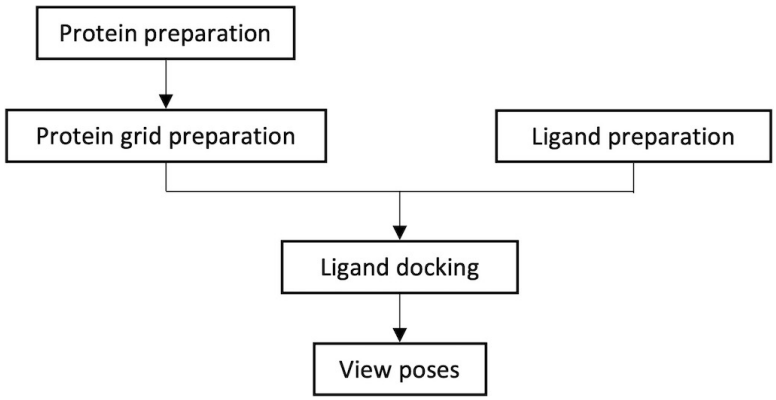


Figure 4.4 The workflow of docking and scoring by Glide.

4.2.3 RF-Score

RF-Score is a machine learning approach developed by Ballester et al. to predict protein-ligand binding affinity and score the hits after molecular docking [103]. Over a thousand protein-

ligand complexes trained the Random Forest (RF) model to construct RF-Score. A value is exported for each tested compound or pose against the receptor. The larger the value, the tighter the predicted binding is.

4.2.4 DataWarrior

DataWarrior is a data visualization and analysis program of small molecules, which supports multiple chemical features [104]. I applied it to organize the information of the VS hits obtained from Pharmit, including glide score, RF-score molecular weight, LogP, number of hydrogen donors and acceptors, vendor information, etc. The filtering function was frequently used to exclude the hits not meeting specific criteria.

Molecular similarity analysis was also applied on DataWarrior to cluster compounds with high similarity in chemical structure. DataWarrior contains a default descriptor FragFp, a binary fingerprint dictionary of 512 substructure fragments. In the analysis, every compound was described by FragFp. The more fragments shared in two molecules, the higher the similarity value shows.

4.2.5 OTAVA SERMs-Like Library

The company OTAVA Chemicals has produced a SERMs-like library, a collection of drug-like compounds with predicted selective ERα modulators activity. Two ligand-based approaches, pharmacophore screening and Bayesian modeling, have been employed to gather the compounds in the library. It contains 1140 commercially available compounds now. I inquired the SDF file of the library from OTAVA Chemicals for VS.

4.2.6 Protein preparation

The pET-14b plasmid containing the nucleic acid sequence of the His-tagged ERα Y537S mutant LBD was purchased from GenScript.

To express the protein, I firstly transformed the protein into Rosetta2 (DE3) Singles competent cells (Novagen) by heat shock method. The cells were grown on an LB media plate pretreated with carbenicillin at 37 ℃ overnight. Then a colony was picked and added to 20 mL of the liquid LB media with carbenicillin, shaken at 250 rpm at 37 ℃ for 16 hours. After that, the cell broth was poured in 1 L of the TB media with carbenicillin and shaken at 250 rpm at 37 ℃ until the cell density reached the OD_{600} value of 0.6. A milliliter of 1M IPTG was added to the TB cell broth. The cells were incubated at 18 ℃ with a 200-rpm shake for another 22 hours.

The cells were harvested by centrifuging at 4000 rpm for 20 minutes. After discarding the supernatant, I resuspended the cells with Buffer 1 and spun down the cells. The operation was repeated until the supernatant became colorless. The cleaned cells were lysed on ice by 4-minute-pulsing sonication with 20 seconds of pausing after every 10 seconds of pulsing. Then the cell fragment was spun down by ultracentrifuge at 40,000 rpm for 30 minutes. The supernatant was retained for protein purification.

The 1-mL HisTrap HP prepacked column (GE Bioscience) was utilized to isolate the protein from the supernatant. The elution buffer was prepared by dissolving 200 mM of imidazole in Buffer 1. The eluant was concentrated by a 10-kD ultracentrifuge filter (Amicon), followed by size exclusion using Superdex 75 10/300 GL column (GE Bioscience) equilibrated with Buffer 1. The interested fractions were gathered, combined, concentrated, and stored at -80 ℃ for further experiments.

4.2.7 Site-directed mutagenesis

The mutagenesis protocol for constructing a plasmid of the wild-type ERα LBD is the same as 3.2.2. The primers were designed in Table 4.1.

Table 4.1 Primers of site-directed mutagenesis.

Primers	Nucleic acid sequence
Forward primer	5'-TGACCTGCTGCTGGAAATGCTG-3'
Reverse primer	5'-TACAGCGGAACCACGTTTTTACATTTC-3'

4.2.8 Thermal shift assay

The protocol is the same as 3.2.4.

4.3 Results and Discussion

4.3.1 Docking receptor selection

Although over 100 crystal models of the ligand-bound ERα LBD Y537S mutant have been released on Protein Data Bank, only 11 of them display typical inverse agonistic conformation that H12 is displaced. Since I aim to develop novel inverse agonists targeting the mutant, choosing the inverse agonistic models as research objects at the beginning is critical. The PDB IDs are as follows: 6CHW, 6CHZ, 6OWC, 6PSJ, 6V87, 6V8T, 6VGH, 7VJ1, 6VPK, 6VVP, 7KCA. In the models 6CHW and 6OWC, the ligands covalently bind against the protein; since Pharmit is designed for searching ligands noncovalently interacting with target proteins, they are beyond the ability of Pharmit for VS. The models 6VGH, 7VJ1, and 6VPK show twinning fractions in their X-ray crystallography statistics, implying structural issues existing in the models. To ensure the reliability of the receptors, I excluded these five models from Pharmit VS. For the rest six models, their backbones are generally identical in the space (Figure 4.5a), but Ser479 is affected by ligands and moves dramatically (Figure 4.5b). Therefore, reckoning all

six models would take into account of the Ser479 flexibility in the VS process, probably promoting the success rate of active hits.

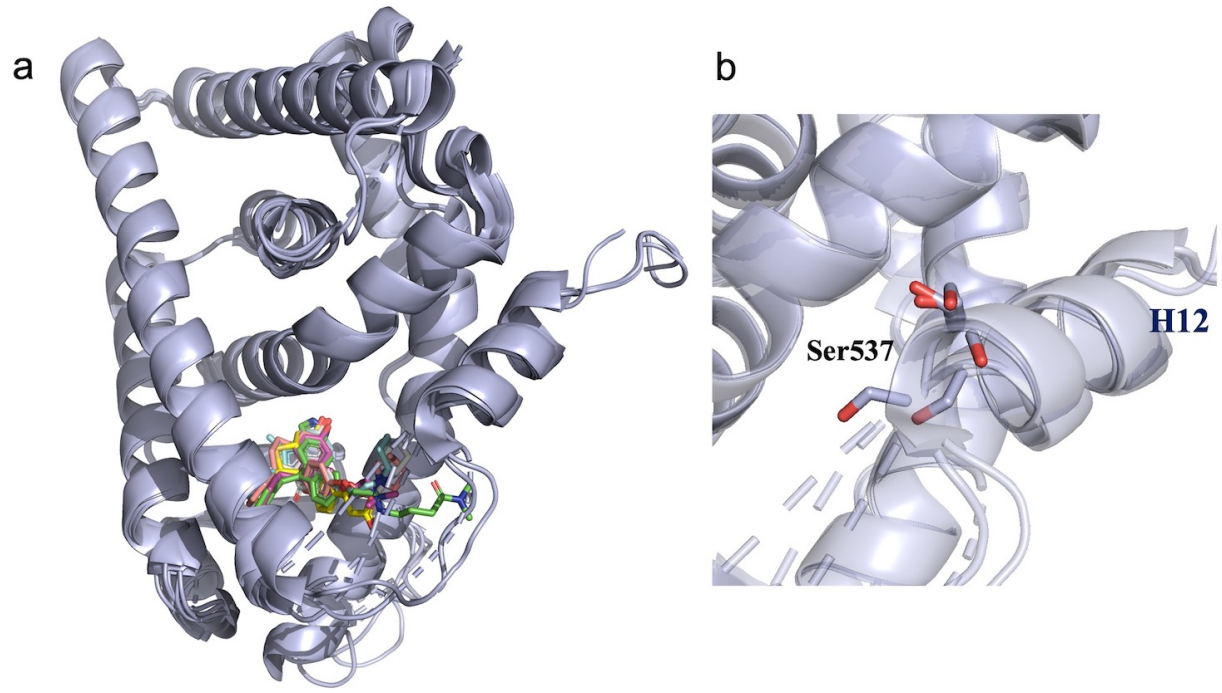


Figure 4.5 Alignment of six models of the ERα LBD Y537S mutant.
The crystal models were 6CHZ, 6PSJ, 6V87, 6V8T, 6VVP, 7KCA. (a) The protein models were shown in light blue cartoon, while the ligands were shown in sticks with different color. (b) Ser537 in the models were aligned in sticks.

4.3.2 Workflow of VS against the *MolPort database*

Considering the feasibility of the compounds for the following laboratory experiments, I selected the MolPort database for VS here, which contained 112, 939, 594 conformers of 7, 875, 286 molecules at the time of conducting the screening. The VS statistics are summarized in Figure 4.6.

Firstly, the six models were imported on Pharmit online server, respectively. For each model, four to eight VS runs were processed with different combinations of pharmacophore

interactions. Every run was set up by following the criteria listed in Table 4.2. It is worth noticing that when selecting the hit screening filters, I referred the Lipinski's rule of five [105] (Appendix D) with some reasonable adjustments, owing to the hydrophobicity of the binding pocket in the LBD. After energy minimization, the hits with any conformation scored lower than -9 were downloaded in every run. The hits from different runs were combined by models.

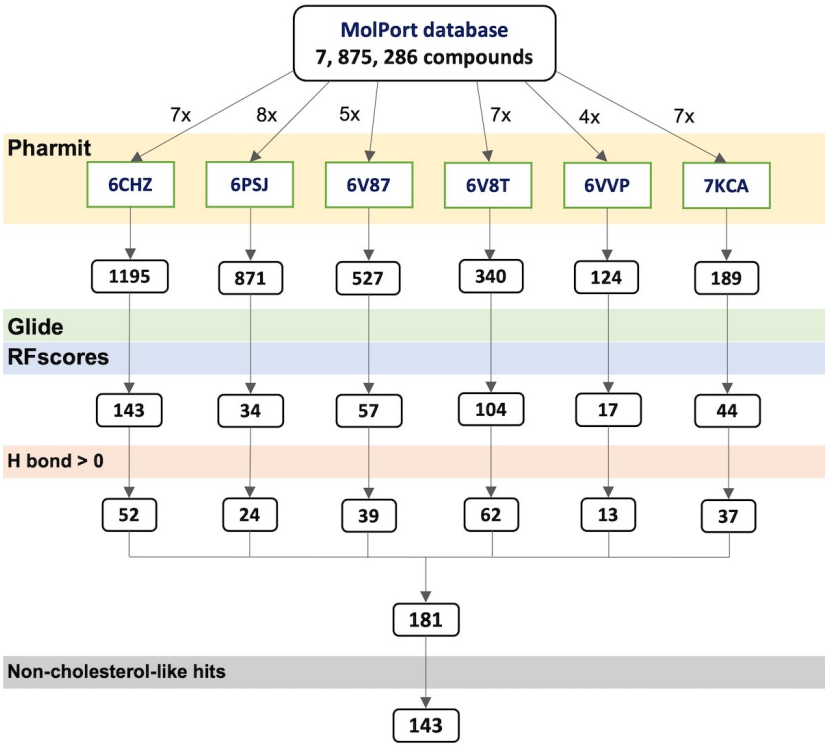


Figure 4.6 The statistics of VS against the MolPort database.

Table 4.2 The criteria of Pharmit VS and hit filtering.

Features	Criteria
Pharmacophore	1) Strong interactions, especially hydrogen bonding, are preferred in the selection; 2) Three to five interactions are selected in each run; 3) Selected interactions should be spread out in the binding pocket as possible.
Exclusive shape	The space occupied by the receptor model should be excluded during the run.
Hit screening filters	1) Molecular weight of a hit is between 200 and 600; 2) The LogP value of a hit is between 0 and 6; 3) Hydrogen-bond donor number of a hit is no more than 5; 4) Hydrogen-bond acceptor number of a hit is no more than 10.
Minimization	1) Maximum score was set to be – 9; 2) Single conformer was checked.

Employ two to three different docking methods and scoring functions to conduct VS is usually necessary to improve the prediction's reliability. Since Glide is renowned for its high accuracy in screening active compounds binding against proteins, I fed Glide with the mutant models and the relative Pharmit hits for redocking and rescoring in the second round of VS. Then the Glide output was further rescored by RF-score, a Random Forest (RF)-based program on predicting the binding affinity of a compound against a receptor. According to the instruction of Glide, hits with a Glide score lower than -10.0 usually indicate a potentially good binder, but for a binding pocket predominantly filled with hydrophobic residues, like ERα LBD, the hits between -8.0 and -10.0 could be still capable of a binding. Thus, I set a cutoff Glide score value of -8.0 and abandoned the compounds that scored higher than it. RF-score output the predicted binding value as low as 5.9. Although the hits with an RF-score higher than 6.0 are considered more valuable for further study, I found a number of hits showing a good Glide score but a poor RF-score. To balance the significance of both scoring functions and avoid missing potent hits, I combined the cutoff values in the decision criteria below:

(1) If the Glide score was smaller than -10.0, the hit was retained for further study;

(2) If the Glide score was between -10.0 and -8.0 and the RF-score was higher than 6.1, the hit was also retained for further study;

(3) Other hits not meeting the above were discarded.

After filtering, 17 to 143 hits were left in each group. As docking and scoring programs traditionally provide more optimistic prediction on hydrogen bonding than other noncovalent interactions, I included the number of hydrogen bonds formed as another filter here. The hits with zero hydrogen bond against the receptor were excluded because it implies a weak protein-

ligand interaction. In the last step, I merged six groups of the hits, and 181 unique hits were virtually gathered.

The 181 hits were imported to DataWarrior for chemical analysis and further filtering. I found that 38 out of them are cholesterol-like compounds. Though estradiol, an estrogen steroid hormone, has been proved tightly binding against the ERα LBD as an agonist [106], steroids or cholesterol-like compounds are usually not ideal drugs. They may non-specifically bind against multiple NRs, such as ROR subfamily members (RORα, RORβ, RORγ) and the steroid receptor subfamily members (ERs, progesterone receptors, androgen receptors, glucocorticoid receptors, and mineralocorticoid receptor). Therefore, I moved these compounds from the list.

The rest hits were analyzed by their Glide scores and RF scores (Figure 4.7a). Several hits locate on the top-left corner of the scattering plot, demonstrating their outstanding performance in scoring. Notably, two of them are endoxifen (red square) and afimoxifene (also named 4-hydroxytamoxifen, blue circle), two active metabolites of the authentic breast cancer drug tamoxifen [107]. Through comparing with nine known ligands of the Y537S mutants (Figure 4.7b), I find that most of these top-ranking hits are analogs of the knowns in structural similarity more than 60% (Figure 4.8). Another three analogs are also found in the good-Glide-score area (the bottom-left corner of the plot). The results illustrate the accuracy of the VS strategy applied here.

Besides the analogs of the known ligands, some compounds are clustered into groups by sharing over 80% of structural similarity, hinting the likelihood of their scaffolds being adapted in the binding pocket. In the light of the architecture distinction from the knowns, laboratory experiments are required to test their binding capacity and bioactivity, as they are potentially new families of lead compounds targeting the ERα Y537S mutant.

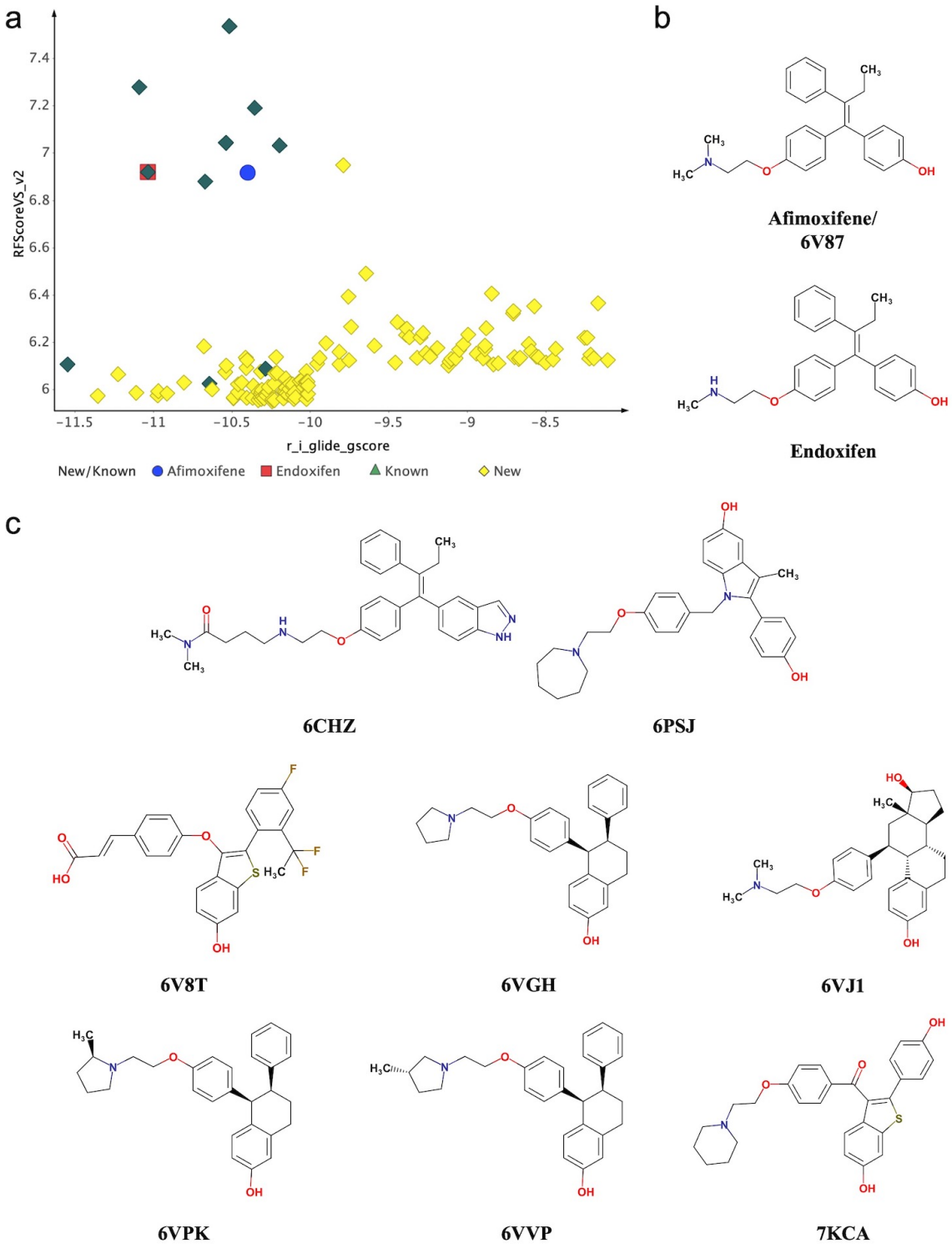


Figure 4.7 Top-ranking hits in the VS against the MolPort database.

(a) The hits were plotted by the scores obtained in Glide and RF-score. The deep green plots are the hits >60% similar to nine known ligands. (b) Chemical structures of afimoxifene and endoxifen. (c) Chemical structures of eight known ligands. The 6V87 ligand is identical to afimoxifene.

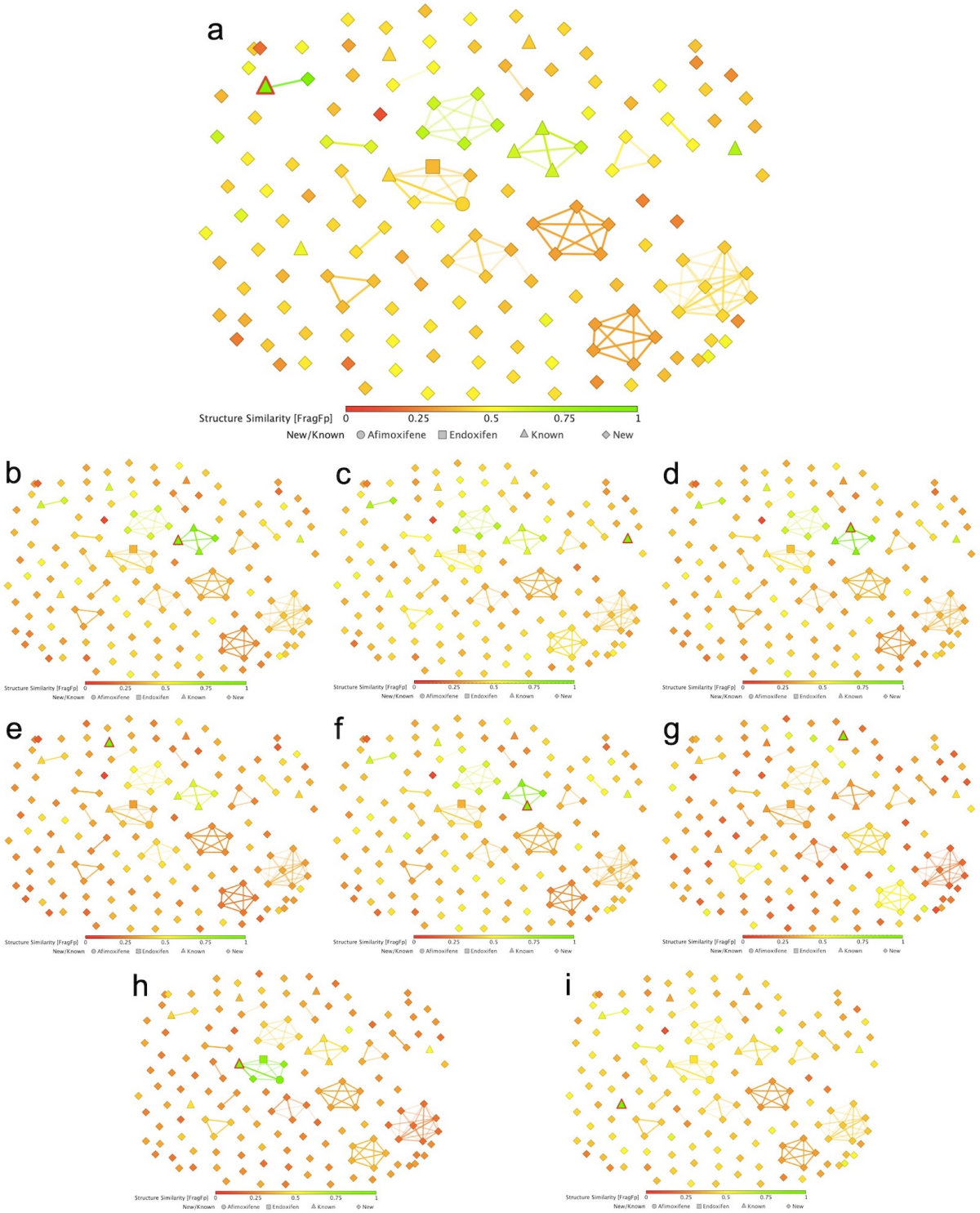


Figure 4.8 Structural similarity of the VS hits and the known ligands.

A mark represents a compound with fixed position in the comparisons. Two marks link together when their structures are >80% similar. In each comparison, the interested known ligand is labeled as a green triangle with red edge; other compounds change color to indicate the similarity to the interested one. All nine known ligands were tested in an order of 6CHZ (a), 6PSJ (b), 6V87 (c), 6V8T (d), 6VGH (e), 6VJ1 (f), 6VPK (g), 6VVP (h), 7KCA (i).

4.3.3 Redocking, rescoring, and ranking OTAVA compounds

Given that the ERα Y537S mutant is of one-residue difference from the wild type, it is highly possible that an ERα modulator could also bind against the mutant, despite the binding affinity may be hindered due to the Y537S mutation. In contrast to searching new compounds targeting the ERα Y537S mutant from giant libraries, it is timesaving to start with the existing ERα-targeting drug-like compound database. The selective ER modulators (SERM)-like library built by OTAVA Chemicals is one of them.

Since the 1140 compounds in the SERM-like library had been virtually screened in advance, here I redocked and rescored them by Glide and RF-score and filtered the hits in a way similar to 4.3.2 (Figure 4.9). OTAVA Chemicals has excluded cholesterol-like compounds from the library ahead, so no action was needed here. Totally 65 unique hits were eventually obtained in the screening.

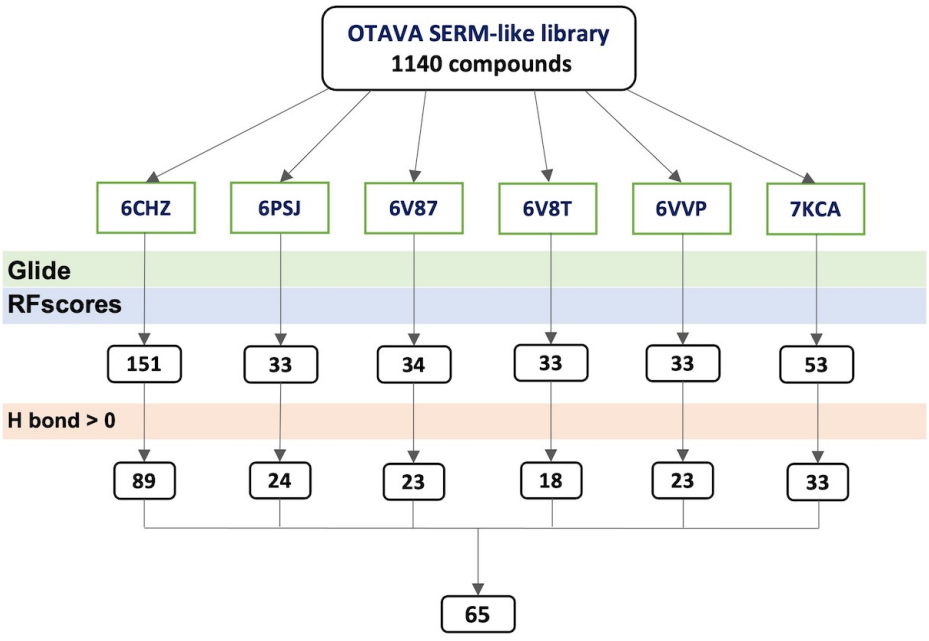


Figure 4.9 The statistics of VS against the OTAVA SERM-like library.

In contrast to the ones from MolPort, the hits from the SERM-like library are less similar (< 60%) to the knowns, implying better chances to obtain new families of lead compounds. However, the scoring functions spread the hits on the plot with a different pattern that no hit posits in the top-left corner, meaning no perfect candidate (Figure 3.10). They scored well in only one of the functions. Additionally, only a group of hits could form a big cluster on the similarity chart (Figure 3.11). The extreme diversity becomes troublesome in evaluating the VS efficacy. To gain more information, five compounds were picked for laboratory experiments based on scores, structural uniqueness, and clustering. They were labeled as ERA1, ERA2, ERA3, ERA4, and ERA5 (Table 4.3).

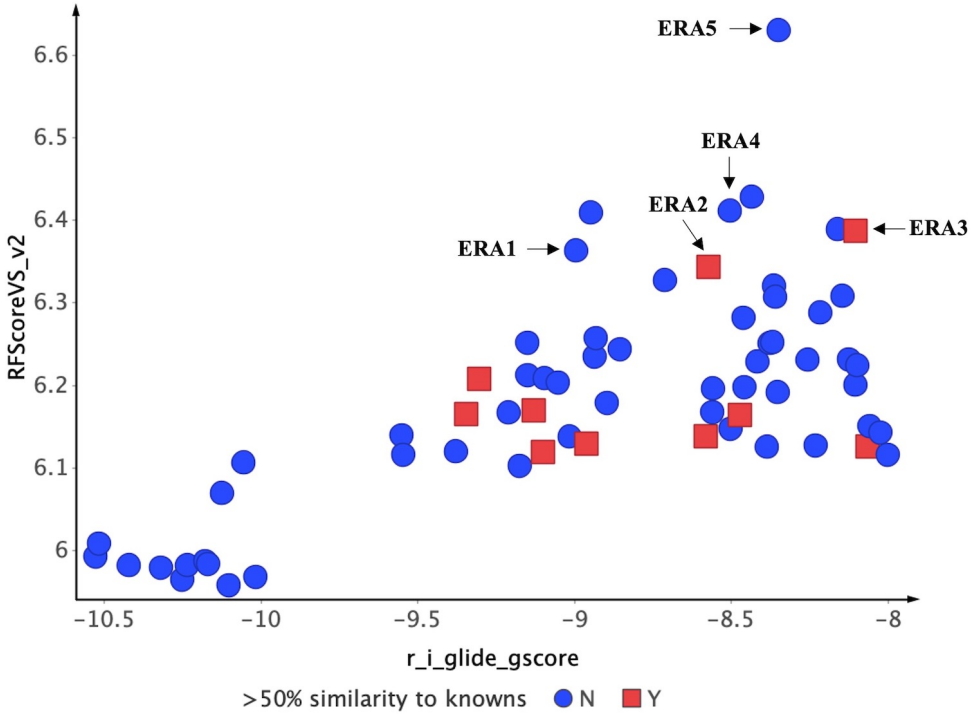


Figure 4.10 Top-ranking hits in the VS against the OTAVA SERM-like library. The hits were plotted by the scores obtained in Glide and RF-score.

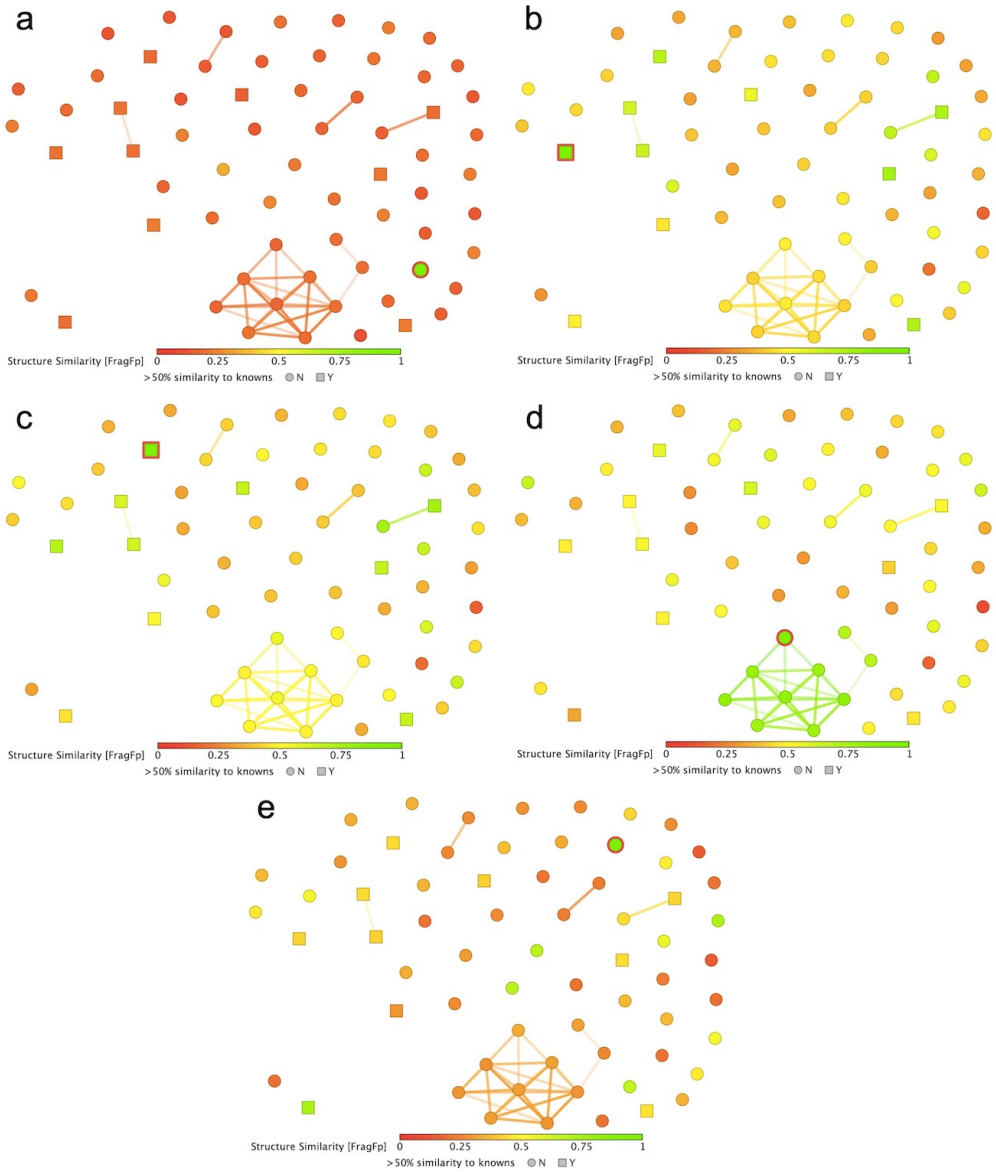


Figure 4.11 Structural similarity of five selected hits with other VS hits.

The selective hits are ERA1 (a), ERA2 (b), ERA3 (c), ERA4 (d), ERA5 (e). In each comparison, the interested hit is labeled as a green mark with red edge; other compounds change color to indicate the similarity to the interested one.

Table 4.3 Chemical information of the purchased OTAVA compounds.

Label	ERA1	ERA2	ERA3	ERA4	ERA5
M.W.	254.33	393.49	349.43	417.42	330.34
LogP	4.3	4.9	4.5	3.4	3.9
HDN	2	1	1	1	1
HAN	2	5	4	5	4
Glide score	-9.28	-8.57	-8.10	-8.50	-8.35
RF-score	6.42	6.34	6.39	6.41	6.63

4.3.4 Preparation of the ERα LBD and its Y537S mutant

The pET-14b plasmid containing the ERα LBD Y537S gene was transformed in Rosetta2 cells, and the protein was overexpressed through IPTG induction. The protein expression was analyzed by western blot. As shown on the membrane (Figure 4.12a), although more protein stayed in the cell pellet section, the supernatant of the lysed cell broth still contained a reasonable amount of the protein, which was enough for the laboratory experiments. Thus, I kept the supernatant for protein purification. The supernatant flowed through the HisTrap column, and the protein was eluted by increased concentrations of imidazole. The protein purity of each eluant was checked by SDS-PAGE gel (Figure 4.12b). Since the eluants with imidazole concentration lower than 100 mM contained a noticeable portion of other proteins, I only collected the eluants at 100 mM and 200 mM of imidazole for size exclusion chromatography. Two eluants were combined, concentrated, and injected in FPLC. The retention volume of the prominent peak was approximately 11 mL, consistent with the column feature on a monomeric 26-kD protein. The SDS-PAGE gel result confirmed the protein (Figure 4.12c). Although the trivial amount of miscellaneous proteins could not be removed entirely through SE purification, the LBD mutant was pure enough for DSF assays.

Given that I am searching for a compound overcoming drug resistance caused by Y537S mutation, it is necessary to include the wild-type ERα LBD as a control group and treating group in the research. The plasmid carrying the wild-type gene was derived from the Y537S mutant and transformed in Rosetta2 cell lines, and the wild-type protein was expressed and purified by procedures similar to previous. The SDS-PAGE gel results indicate the purity of the mutant protein with purification steps (Figure 4.13).

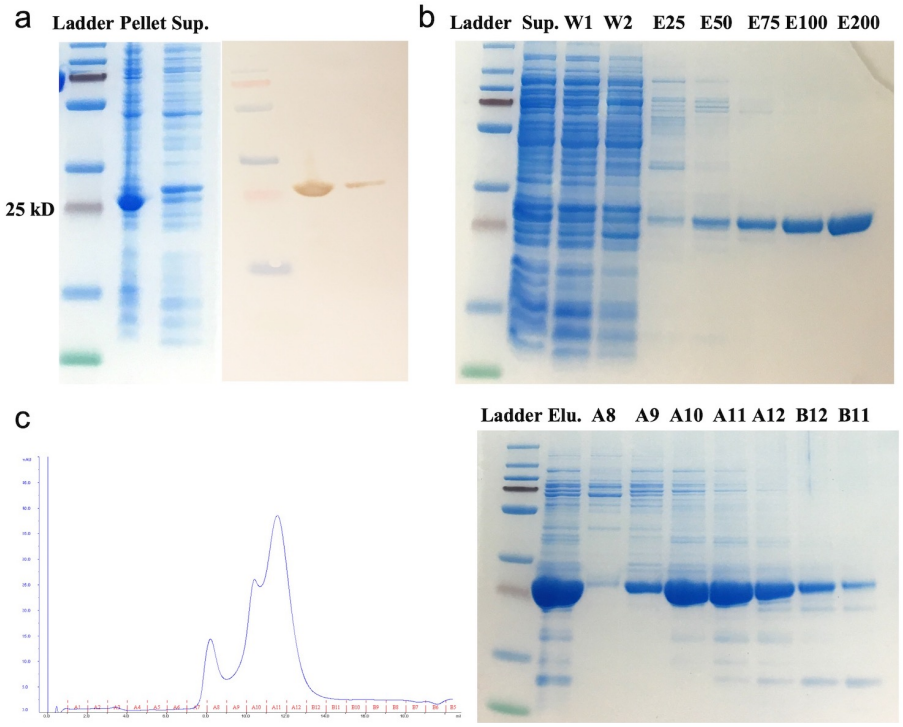


Figure 4.12 Expression and purification of the ERα LBD Y537S mutant.

(a) SDS-PAGE (left) and western blot (right) results. (b) SDS-Page was run after HisTrap column purification. W1 and W2 represent two fractions of washing through. E25, E50, E75, E100, and E200 represent the eluants released by 25, 50, 75, 100, and 200 mM of imidazole. (c) UV spectrum of size exclusion chromatography (left) and SDS-Page of the peak fractions (right).

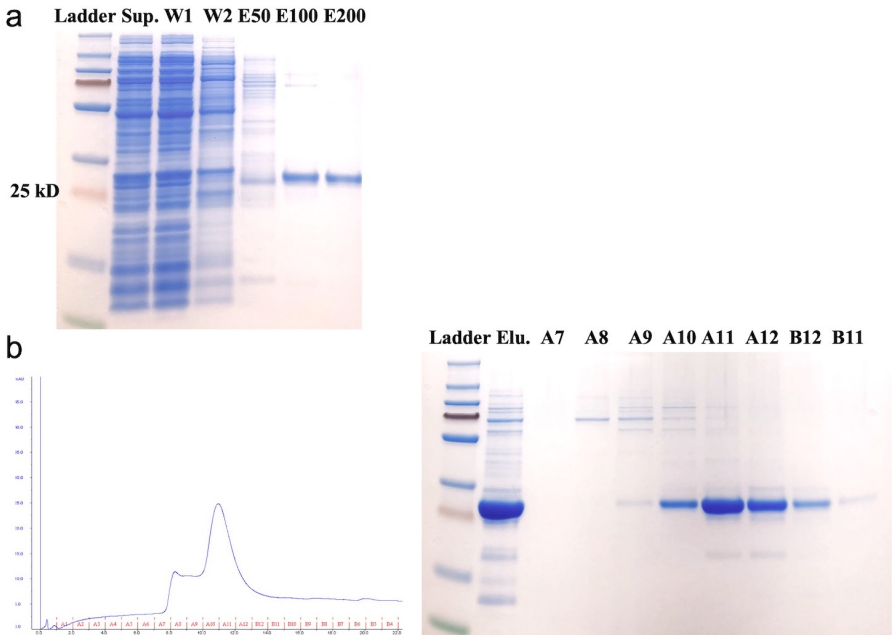


Figure 4.13 Expression and purification of the ERα LBD.

(a) SDS-Page was run after HisTrap column purification. (b) UV spectrum of size exclusion chromatography (left) and SDS-Page of the peak fractions (right).

4.3.5 Binding capacity of selected compounds against ERα and the Y537S mutant

The selected compounds were purchased from OTAVA Chemicals. To quickly analyze their binding capability against the wild type and the mutant in the lab, I conducted a Differential scanning fluorimetry (DSF) assay. In the assay, the purified ERα and the Y537S mutant proteins were treated with a high dose of compounds, 50 μM, to maximize the binding effect. The DMSO-treated group was considered a negative control, as the compound stock solutions were prepared in 100% DMSO. Endoxifen (EndX) [107,108], an active metabolite of tamoxifen, was featured as a positive control here. Because of its poor solubility in DMSO, I prepared the EndX stock solution in 100% methanol, and a methanol-treated group was also introduced in the assay. As shown in Figure 4.14a, EndX shows excellent binding on both proteins, though methanol seems harmful to the protein stability, compared with the same amount of DMSO. For the groups treated with the OTAVA compounds, ERA1 noticeably stabilizes two proteins, indicating its binding capacity of ERA1 against both proteins, while the other four compounds are probably of no avail to the proteins. Another DSF assay was performed for ERA1 in a series of concentrations. As shown in Figure 4.14b, the mutant stability is enhanced with the increasing concentration of ERA1. It confirms the result of the previous assay and implies a binding affinity of ERA1 in a moderate-to-strong range. Interestingly, ERA1 performs slightly better on the Y537S mutant than the wild type in both assay runs. Theoretically, the structure of the apo mutant is more thermally stable because the mutated residue Ser517 forms an extra hydrogen bond with Asp351, which could contribute to the difference between the two data set. Howbeit, the T_m data of the apo proteins cannot prove the hypothesis. In the meantime, given the weakness of DSF also discussed in 3.3.4, it is still too early to conclude that ERA1 is more potent on the mutant than the wild type. Further investigation with advanced methods on ERA1 is necessary to

explain the phenomenon. Moreover, ERA1 carries a unique chemical structure in low molecular weight (Figure 4.14c), favoring addition of more functional groups on the skeleton to enhance the potency. It is an excellent lead compound for developing into new SERMs.

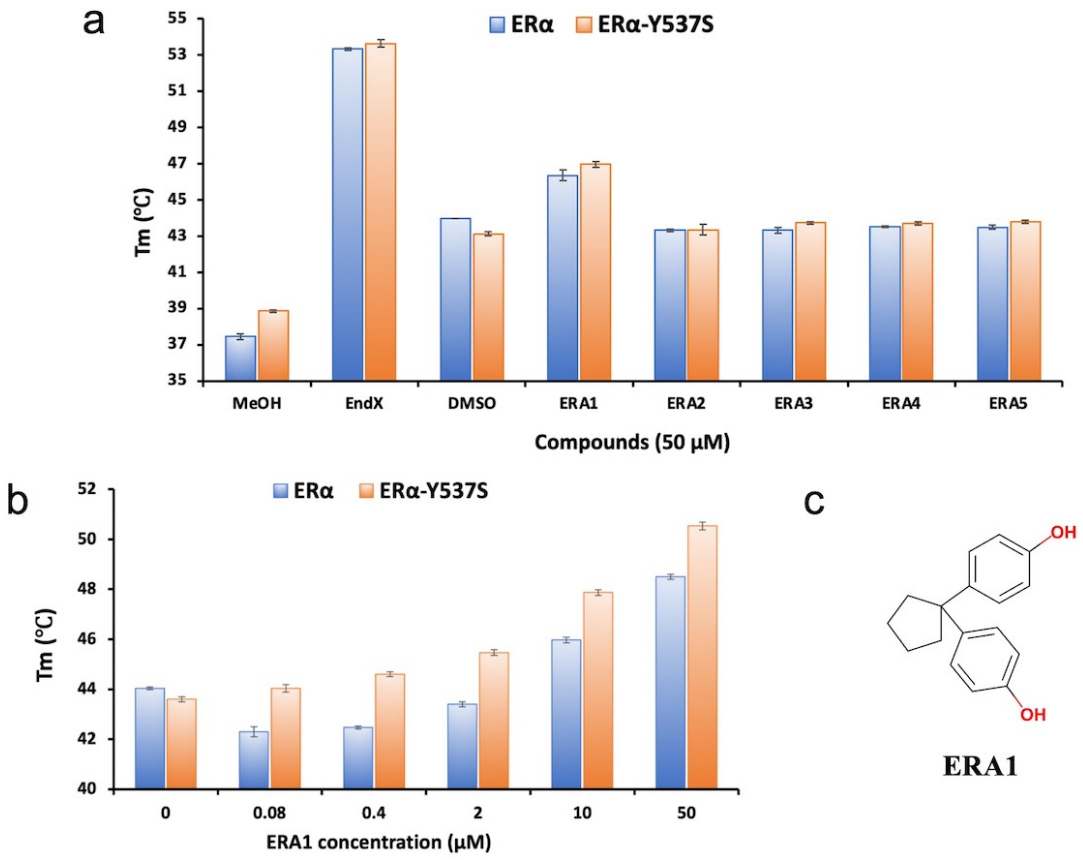


Figure 4.14 Binding capability of the selected hits against the ERα LBD wild type and Y537S mutant.

(a) ERA1 shows promising binding against both proteins. EndX in methanol (MeOH) was applied as a positive control in the test. (b) The binding of ERA1 against the proteins was tested in different concentrations. (c) Chemical structure of ERA1.

4.4 Conclusion

In summary, a workflow of VS was designed to search for a new generation of SERMs that can defeat the drug-resistant issue observed during the treatment of $ER\alpha^+$ breast cancer. Millions of small molecules were docked against crystal models of the ERα Y537S LBD, respectively. A list of the top-ranking hits was then redocked, rescored, and profiled for further narrowing down the pool of the potential SERMs. Five compounds that performed well *in silico* were chosen and tested by differential scanning fluorimetry. A compound ERA1 comparably binds against both the wild-type and the mutant of ERα LBD, showing its potential as a lead compound for developing into new SERMs.

In this project, the computational experiments were accomplished in a week, which saves much time in the pre-clinical drug discovery stage, compared with the traditional laboratory-based method. In view of the output, quite a number of the top-ranking hits picked from the MolPort database are known SERMs and their analogs, elucidating the efficacy of the workflow employed here. Also, as one out of five tested compounds, ERA1 is preliminarily determined as a binder of both the wild-type and the mutant protein, demonstrating a 20%-efficacy rate of the *in silico* prediction. This project is another good example to show the superiority of computer-aided drug discovery methods accessible by an academic lab.

Chapter 5 - Future Directions

In this chapter, I briefly highlighted further works that can be expanded based on the achievements stated in the previous chapters.

5.1 A long-range interaction regulating brightness and pH stability of a dark fluorescent protein

The darkmRuby excites at 556 nm maximally, locating on the 540-570 nm region of the spectrum without dark FPs reported before. With its monomericity and fast maturation, darkmRuby unveils its aptitude as a protein tag and organelle marker in cell-based assays and an excellent receptor in FRET assays. The pH-sensitive feature may make it further stand out from the crowds in the pH-dependent assays. Additionally, darkmRuby is the first dark FP thoroughly studied in the atomic manner, which will provide an indispensable guide when searching for its unique use. In short, darkmRuby deserves more exploration for its value in application.

To soundly prove the significance of long-range interactions in regulating FP features, more FP models need to be inspected in-depth. The choice of FP samples should cover multiple dimensions, such as the FP color, brightness, chromophore structures, the *cis/trans* isoforms of the chromophores, amino acid sequence similarity, and so on. Protein Data Bank and FPbase [11] are excellent resources for searching FP crystal structures and spectra features.

5.2 Inverse agonism of a novel modulator targeting human RORγ

The experimental results show the potential of RG14-2 as a lead compound to develop into RORγ invest agonists. However, the binding mode of the compound is unclear because the

poor quality of the co-crystal model prevents me from inspecting the atomic interactions. To solve this issue, reproducing co-crystals with high resolution is necessary.

It is intriguing that RG14-3, an analog of RG14-2, presents noticeably weaker binding against the RORγ LBD while they are just different in the position of a methyl group on the benzene ring. It implies the importance of the spatial arrangement of a ligand in the pocket to maximize the drug efficacy. More RG14-2 analogs could be developed with accommodating the pocket interior space to improve the potency.

5.3 Computer-aided discovery of an ERα modulator overcoming the Y537S mutation

The compound ERA1 is able to bind against both the ERα wild type and the Y537S mutant. However, DSF cannot distinguish antagonists from agonists. Cell-based assays are needed to determine its bioactivity. Also, co-crystallization of the LBDs bound with ERA1 will provide structural information of the interaction, benefiting the analog development. Furthermore, the small size of ERA1 leaves a noticeable void in the pocket, allowing more functional groups to be introduced to ERA1. Therefore, it is possible to develop more potent ERα modulators based on ERA1.

References

(1) Tang, Y.; Bian, Z.; Zhao, L.; Liu, Y.; Liang, S.; Wang, Q.; Han, X.; Peng, Y.; Chen, X.; Shen, L.; Qiu, D.; Li, Z.; Ma, X. Interleukin-17 Exacerbates Hepatic Steatosis and Inflammation in Non-Alcoholic Fatty Liver Disease. *Clin. Exp. Immunol.* **2011**, *166* (2), 281–290. https://doi.org/10.1111/j.1365-2249.2011.04471.x.

(2) Sommer, S.; Fuqua, S. A. W. Estrogen Receptor and Breast Cancer. *Semin. Cancer Biol.* **2001**, *11* (5), 339–352. https://doi.org/10.1006/scbi.2001.0389.

(3) Chaudhuri, T. K.; Paul, S. Protein-Misfolding Diseases and Chaperone-Based Therapeutic Approaches. *FEBS J.* **2006**, *273* (7), 1331–1349. https://doi.org/10.1111/j.1742-4658.2006.05181.x.

(4) M. Ashraf, G.; H. Greig, N.; A. Khan, T.; Hassan, I.; Tabrez, S.; Shakil, S.; A. Sheikh, I.; K. Zaidi, S.; Akram, M.; R. Jabir, N.; K. Firoz, C.; Naeem, A.; M. Alhazza, I.; A. Damanhouri, G.; A. Kamal, M. Protein Misfolding and Aggregation in Alzheimer's Disease and Type 2 Diabetes Mellitus. *CNS Neurol. Disord. - Drug Targets- CNS Neurol. Disord.* **2014**, *13* (7), 1280–1293.

(5) Shamsi, T. N.; Athar, T.; Parveen, R.; Fatima, S. A Review on Protein Misfolding, Aggregation and Strategies to Prevent Related Ailments. *Int. J. Biol. Macromol.* **2017**, *105*, 993–1000. https://doi.org/10.1016/j.ijbiomac.2017.07.116.

(6) Shimomura, O.; Johnson, F. H.; Saiga, Y. Extraction, Purification and Properties of Aequorin, a Bioluminescent Protein from the Luminous Hydromedusan, Aequorea. *J. Cell. Comp. Physiol.* **1962**, *59* (3), 223–239. https://doi.org/10.1002/jcp.1030590302.

(7) Ormö, M.; Cubitt, A. B.; Kallio, K.; Gross, L. A.; Tsien, R. Y.; Remington, S. J. Crystal Structure of the Aequorea Victoria Green Fluorescent Protein. *Science* **1996**. https://doi.org/10.1126/science.273.5280.1392.

(8) Yang, F.; Moss, L. G.; Phillips, G. N. The Molecular Structure of Green Fluorescent Protein. *Nat. Biotechnol.* **1996**, *14* (10), 1246–1251. https://doi.org/10.1038/nbt1096-1246.

(9) Cody, C. W.; Prasher, D. C.; Westler, W. M.; Prendergast, F. G.; Ward, W. W. Chemical Structure of the Hexapeptide Chromophore of the Aequorea Green-Fluorescent Protein. *Biochemistry* **1993**, *32* (5), 1212–1218. https://doi.org/10.1021/bi00056a003.

(10) Heim, R.; Prasher, D. C.; Tsien, R. Y. Wavelength Mutations and Posttranslational Autoxidation of Green Fluorescent Protein. *Proc. Natl. Acad. Sci.* **1994**, *91* (26), 12501–12504. https://doi.org/10.1073/pnas.91.26.12501.

(11) Lambert, T. J. FPbase: A Community-Editable Fluorescent Protein Database. *Nat. Methods* **2019**, *16* (4), 277–278. https://doi.org/10.1038/s41592-019-0352-8.

(12) Remington, S. J.; Wachter, R. M.; Yarbrough, D. K.; Branchaud, B.; Anderson, D. C.; Kallio, K.; Lukyanov, K. A. ZFP538, a Yellow-Fluorescent Protein from Zoanthus, Contains a Novel Three-Ring Chromophore,. *Biochemistry* **2005**, *44* (1), 202–212. https://doi.org/10.1021/bi048383r.

(13) Gross, L. A.; Baird, G. S.; Hoffman, R. C.; Baldridge, K. K.; Tsien, R. Y. The Structure of the Chromophore within DsRed, a Red Fluorescent Protein from Coral. *Proc. Natl. Acad. Sci.* **2000**, *97* (22), 11990–11995. https://doi.org/10.1073/pnas.97.22.11990.

(14) Petersen, J.; Wilmann, P. G.; Beddoe, T.; Oakley, A. J.; Devenish, R. J.; Prescott, M.; Rossjohn, J. The 2.0-Å Crystal Structure of EqFP611, a Far Red Fluorescent Protein from the Sea Anemone Entacmaea Quadricolor *. *J. Biol. Chem.* **2003**, *278* (45), 44626–44631. https://doi.org/10.1074/jbc.M307896200.

(15) Chalfie, M.; Tu, Y.; Euskirchen, G.; Ward, W. W.; Prasher, D. C. Green Fluorescent Protein as a Marker for Gene Expression. *Science* **1994**, *263* (5148), 802–805.
(16) Chudakov, D. M.; Matz, M. V.; Lukyanov, S.; Lukyanov, K. A. Fluorescent Proteins and Their Applications in Imaging Living Cells and Tissues. *Physiol. Rev.* **2010**, *90* (3), 1103–1163. https://doi.org/10.1152/physrev.00038.2009.
(17) Bondesson, M.; Hao, R.; Lin, C.-Y.; Williams, C.; Gustafsson, J.-Å. Estrogen Receptor Signaling during Vertebrate Development. *Biochim. Biophys. Acta BBA - Gene Regul. Mech.* **2015**, *1849* (2), 142–151. https://doi.org/10.1016/j.bbagrm.2014.06.005.
(18) Tyagi, S.; Gupta, P.; Saini, A. S.; Kaushal, C.; Sharma, S. The Peroxisome Proliferator-Activated Receptor: A Family of Nuclear Receptors Role in Various Diseases. *J. Adv. Pharm. Technol. Res.* **2011**, *2* (4), 236–240. https://doi.org/10.4103/2231-4040.90879.
(19) Fleet, J. C.; Schoch, R. D. Molecular Mechanisms for Regulation of Intestinal Calcium Absorption by Vitamin D and Other Factors. *Crit. Rev. Clin. Lab. Sci.* **2010**, *47* (4), 181–195. https://doi.org/10.3109/10408363.2010.536429.
(20) Ruan, Q.; Kameswaran, V.; Zhang, Y.; Zheng, S.; Sun, J.; Wang, J.; DeVirgiliis, J.; Liou, H.-C.; Beg, A. A.; Chen, Y. H. The Th17 Immune Response Is Controlled by the Rel–RORγ–RORγT Transcriptional Axis. *J. Exp. Med.* **2011**, *208* (11), 2321–2333. https://doi.org/10.1084/jem.20110462.
(21) Ivanov, I. I.; McKenzie, B. S.; Zhou, L.; Tadokoro, C. E.; Lepelley, A.; Lafaille, J. J.; Cua, D. J.; Littman, D. R. The Orphan Nuclear Receptor RORγt Directs the Differentiation Program of Proinflammatory IL-17+ T Helper Cells. *Cell* **2006**, *126* (6), 1121–1133. https://doi.org/10.1016/j.cell.2006.07.035.
(22) Kumar, R.; Thompson, E. B. Transactivation Functions of the N-Terminal Domains of Nuclear Hormone Receptors: Protein Folding and Coactivator Interactions. *Mol. Endocrinol.* **2003**, *17* (1), 1–10. https://doi.org/10.1210/me.2002-0258.
(23) Schwabe, J. W. R.; Chapman, L.; Finch, J. T.; Rhodes, D. The Crystal Structure of the Estrogen Receptor DNA-Binding Domain Bound to DNA: How Receptors Discriminate between Their Response Elements. *Cell* **1993**, *75* (3), 567–578. https://doi.org/10.1016/0092-8674(93)90390-C.
(24) Anbalagan, M.; Huderson, B.; Murphy, L.; Rowan, B. G. Post-Translational Modifications of Nuclear Receptors and Human Disease. *Nucl. Recept. Signal.* **2012**, *10* (1), nrs.10001. https://doi.org/10.1621/nrs.10001.
(25) Chandra, V.; Huang, P.; Hamuro, Y.; Raghuram, S.; Wang, Y.; Burris, T. P.; Rastinejad, F. Structure of the Intact PPAR-γ–RXR-α Nuclear Receptor Complex on DNA. *Nature* **2008**, *456* (7220), 350–356. https://doi.org/10.1038/nature07413.
(26) Lou, X.; Toresson, G.; Benod, C.; Suh, J. H.; Philips, K. J.; Webb, P.; Gustafsson, J.-A. Structure of the Retinoid X Receptor α–Liver X Receptor β (RXRα–LXRβ) Heterodimer on DNA. *Nat. Struct. Mol. Biol.* **2014**, *21* (3), 277–281. https://doi.org/10.1038/nsmb.2778.
(27) Chandra, V.; Wu, D.; Li, S.; Potluri, N.; Kim, Y.; Rastinejad, F. The Quaternary Architecture of RARβ–RXRα Heterodimer Facilitates Domain–Domain Signal Transmission. *Nat. Commun.* **2017**, *8* (1), 868. https://doi.org/10.1038/s41467-017-00981-y.
(28) Chandra, V.; Huang, P.; Potluri, N.; Wu, D.; Kim, Y.; Rastinejad, F. Multidomain Integration in the Structure of the HNF-4α Nuclear Receptor Complex. *Nature* **2013**, *495* (7441), 394–398. https://doi.org/10.1038/nature11966.

(29) Zhao, L.; Zhou, S.; Gustafsson, J.-Å. Nuclear Receptors: Recent Drug Discovery for Cancer Therapies. *Endocr. Rev.* **2019**, *40* (5), 1207–1249. https://doi.org/10.1210/er.2018-00222.
(30) Berardozzi, R.; Adam, V.; Martins, A.; Bourgeois, D. Arginine 66 Controls Dark-State Formation in Green-to-Red Photoconvertible Fluorescent Proteins. *J. Am. Chem. Soc.* **2016**, *138* (2), 558–565. https://doi.org/10.1021/jacs.5b09923.
(31) Sniegowski, J. A.; Lappe, J. W.; Patel, H. N.; Huffman, H. A.; Wachter, R. M. Base Catalysis of Chromophore Formation in Arg96 and Glu222 Variants of Green Fluorescent Protein*. *J. Biol. Chem.* **2005**, *280* (28), 26248–26255. https://doi.org/10.1074/jbc.M412327200.
(32) Wood, T. I.; Barondeau, D. P.; Hitomi, C.; Kassmann, C. J.; Tainer, J. A.; Getzoff, E. D. Defining the Role of Arginine 96 in Green Fluorescent Protein Fluorophore Biosynthesis,. *Biochemistry* **2005**, *44* (49), 16211–16220. https://doi.org/10.1021/bi051388j.
(33) Kredel, S.; Nienhaus, K.; Oswald, F.; Wolff, M.; Ivanchenko, S.; Cymer, F.; Jeromin, A.; Michel, F. J.; Spindler, K.-D.; Heilker, R.; Nienhaus, G. U.; Wiedenmann, J. Optimized and Far-Red-Emitting Variants of Fluorescent Protein EqFP611. *Chem. Biol.* **2008**, *15* (3), 224–233. https://doi.org/10.1016/j.chembiol.2008.02.008.
(34) Pletnev, S.; Shcherbo, D.; Chudakov, D. M.; Pletneva, N.; Merzlyak, E. M.; Wlodawer, A.; Dauter, Z.; Pletnev, V. A Crystallographic Study of Bright Far-Red Fluorescent Protein MKate Reveals PH-Induced Cis-Trans Isomerization of the Chromophore ∗. *J. Biol. Chem.* **2008**, *283* (43), 28980–28987. https://doi.org/10.1074/jbc.M800599200.
(35) Lin, M. Z.; McKeown, M. R.; Ng, H.-L.; Aguilera, T. A.; Shaner, N. C.; Campbell, R. E.; Adams, S. R.; Gross, L. A.; Ma, W.; Alber, T.; Tsien, R. Y. Autofluorescent Proteins with Excitation in the Optical Window for Intravital Imaging in Mammals. *Chem. Biol.* **2009**, *16* (11), 1169–1179. https://doi.org/10.1016/j.chembiol.2009.10.009.
(36) Muslinkina, L.; Pletnev, V. Z.; Pletneva, N. V.; Ruchkin, D. A.; Kolesov, D. V.; Bogdanov, A. M.; Kost, L. A.; Rakitina, T. V.; Agapova, Y. K.; Shemyakina, I. I.; Chudakov, D. M.; Pletnev, S. Two Independent Routes of Post-Translational Chemistry in Fluorescent Protein FusionRed. *Int. J. Biol. Macromol.* **2020**, *155*, 551–559. https://doi.org/10.1016/j.ijbiomac.2020.03.244.
(37) Lam, A. J.; St-Pierre, F.; Gong, Y.; Marshall, J. D.; Cranfill, P. J.; Baird, M. A.; McKeown, M. R.; Wiedenmann, J.; Davidson, M. W.; Schnitzer, M. J.; Tsien, R. Y.; Lin, M. Z. Improving FRET Dynamic Range with Bright Green and Red Fluorescent Proteins. *Nat. Methods* **2012**, *9* (10), 1005–1012. https://doi.org/10.1038/nmeth.2171.
(38) Bajar, B. T.; Wang, E. S.; Lam, A. J.; Kim, B. B.; Jacobs, C. L.; Howe, E. S.; Davidson, M. W.; Lin, M. Z.; Chu, J. Improving Brightness and Photostability of Green and Red Fluorescent Proteins for Live Cell Imaging and FRET Reporting. *Sci. Rep.* **2016**, *6* (1), 20889. https://doi.org/10.1038/srep20889.
(39) Xu, Y.; Deng, M.; Zhang, S.; Yang, J.; Peng, L.; Chu, J.; Zou, P. Imaging Neuronal Activity with Fast and Sensitive Red-Shifted Electrochromic FRET Indicators. *ACS Chem. Neurosci.* **2019**, *10* (12), 4768–4775. https://doi.org/10.1021/acschemneuro.9b00501.
(40) Kabsch, W. XDS. *Acta Crystallogr. D Biol. Crystallogr.* **2010**, *66* (2), 125–132. https://doi.org/10.1107/S0907444909047337.
(41) McCoy, A. J.; Grosse-Kunstleve, R. W.; Adams, P. D.; Winn, M. D.; Storoni, L. C.; Read, R. J. Phaser Crystallographic Software. *J. Appl. Crystallogr.* **2007**, *40* (4), 658–674. https://doi.org/10.1107/S0021889807021206.

(42) Emsley, P.; Cowtan, K. Coot: Model-Building Tools for Molecular Graphics. *Acta Crystallogr. D Biol. Crystallogr.* **2004**, *60* (12), 2126–2132. https://doi.org/10.1107/S0907444904019158.
(43) Murshudov, G. N.; Skubák, P.; Lebedev, A. A.; Pannu, N. S.; Steiner, R. A.; Nicholls, R. A.; Winn, M. D.; Long, F.; Vagin, A. A. REFMAC5 for the Refinement of Macromolecular Crystal Structures. *Acta Crystallogr. D Biol. Crystallogr.* **2011**, *67* (4), 355–367. https://doi.org/10.1107/S0907444911001314.
(44) Pavelka, A.; Sebestova, E.; Kozlikova, B.; Brezovsky, J.; Sochor, J.; Damborsky, J. CAVER: Algorithms for Analyzing Dynamics of Tunnels in Macromolecules. *IEEE/ACM Trans. Comput. Biol. Bioinform.* **2016**, *13* (3), 505–517. https://doi.org/10.1109/TCBB.2015.2459680.
(45) Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C. Ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from Ff99SB. *J. Chem. Theory Comput.* **2015**, *11* (8), 3696–3713. https://doi.org/10.1021/acs.jctc.5b00255.
(46) Krieger, E.; Vriend, G. New Ways to Boost Molecular Dynamics Simulations. *J. Comput. Chem.* **2015**, *36* (13), 996–1007. https://doi.org/10.1002/jcc.23899.
(47) Park, J. W.; Rhee, Y. M. Electric Field Keeps Chromophore Planar and Produces High Yield Fluorescence in Green Fluorescent Protein. *J. Am. Chem. Soc.* **2016**, *138* (41), 13619–13629. https://doi.org/10.1021/jacs.6b06833.
(48) Romei, M. G.; Lin, C.-Y.; Mathews, I. I.; Boxer, S. G. Electrostatic Control of Photoisomerization Pathways in Proteins. *Science* **2020**. https://doi.org/10.1126/science.aax1898.
(49) Danny Hsu, S.-T.; Blaser, G.; E. Jackson, S. The Folding, Stability and Conformational Dynamics of β-Barrel Fluorescent Proteins. *Chem. Soc. Rev.* **2009**, *38* (10), 2951–2965. https://doi.org/10.1039/B908170B.
(50) Chapagain, P. P.; Regmi, C. K.; Castillo, W. Fluorescent Protein Barrel Fluctuations and Oxygen Diffusion Pathways in MCherry. *J. Chem. Phys.* **2011**, *135* (23), 235101. https://doi.org/10.1063/1.3660197.
(51) Li, B.; Shahid, R.; Peshkepija, P.; Zimmer, M. Water Diffusion in and out of the β-Barrel of GFP and the Fast Maturing Fluorescent Protein, TurboGFP. *Chem. Phys.* **2012**, *392* (1), 143–148. https://doi.org/10.1016/j.chemphys.2011.11.001.
(52) Dobretsov, G. E.; Syrejschikova, T. I.; Smolina, N. V. On Mechanisms of Fluorescence Quenching by Water. *Biophysics* **2014**, *59* (2), 183–188. https://doi.org/10.1134/S0006350914020079.
(53) Roy, A.; Carpentier, P.; Bourgeois, D.; Field, M. Diffusion Pathways of Oxygen Species in the Phototoxic Fluorescent Protein KillerRed. *Photochem. Photobiol. Sci.* **2010**, *9* (10), 1342–1350. https://doi.org/10.1039/C0PP00141D.
(54) Adipogenesis and Insulin Sensitivity in Obesity Are Regulated by Retinoid-Related Orphan Receptor Gamma. *EMBO Mol. Med.* **2011**, *3* (11), 637–651. https://doi.org/10.1002/emmm.201100172.
(55) Rau, M.; Schilling, A.-K.; Meertens, J.; Hering, I.; Weiss, J.; Jurowich, C.; Kudlich, T.; Hermanns, H. M.; Bantel, H.; Beyersdorf, N.; Geier, A. Progression from Nonalcoholic Fatty Liver to Nonalcoholic Steatohepatitis Is Marked by a Higher Frequency of Th17 Cells in the Liver and an Increased Th17/Resting Regulatory T Cell Ratio in Peripheral Blood

and in the Liver. *J. Immunol.* **2016**, *196* (1), 97–105. https://doi.org/10.4049/jimmunol.1501175.
(56) Cui, G. TH9, TH17, and TH22 Cell Subsets and Their Main Cytokine Products in the Pathogenesis of Colorectal Cancer. *Front. Oncol.* **2019**, *9*.
(57) 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.; Wong, H. Reduction in Lipophilicity Improved the Solubility, Plasma–Protein Binding, and Permeability of Tertiary Sulfonamide RORc Inverse Agonists. *Bioorg. Med. Chem. Lett.* **2014**, *24* (16), 3891–3897. https://doi.org/10.1016/j.bmcl.2014.06.048.
(58) Sun, N.; Guo, H.; Wang, Y. 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.* **2019**, *29* (9), 663–674. https://doi.org/10.1080/13543776.2019.1655541.
(59) Santori, F. R.; Huang, P.; van de Pavert, S. A.; Douglass, E. F.; Leaver, D. J.; Haubrich, B. A.; Keber, R.; Lorbek, G.; Konijn, T.; Rosales, B. N.; Rozman, D.; Horvat, S.; Rahier, A.; Mebius, R. E.; Rastinejad, F.; Nes, W. D.; Littman, D. R. Identification of Natural RORγ Ligands That Regulate the Development of Lymphoid Cells. *Cell Metab.* **2015**, *21* (2), 286–298. https://doi.org/10.1016/j.cmet.2015.01.004.
(60) Jin, L.; Martynowski, D.; Zheng, S.; Wada, T.; Xie, W.; Li, Y. Structural Basis for Hydroxycholesterols as Natural Ligands of Orphan Nuclear Receptor RORγ. *Mol. Endocrinol.* **2010**, *24* (5), 923–929. https://doi.org/10.1210/me.2009-0507.
(61) Renaud, J.-P.; Rochel, N.; Ruff, M.; Vivat, V.; Chambon, P.; Gronemeyer, H.; Moras, D. Crystal Structure of the RAR-γ Ligand-Binding Domain Bound to All-Trans Retinoic Acid. *Nature* **1995**, *378* (6558), 681–689. https://doi.org/10.1038/378681a0.
(62) Kallenberger, B. C.; Love, J. D.; Chatterjee, V. K. K.; Schwabe, J. W. R. A Dynamic Mechanism of Nuclear Receptor Activation and Its Perturbation in a Human Disease. *Nat. Struct. Biol.* **2003**, *10* (2), 136–140. https://doi.org/10.1038/nsb892.
(63) Li, X.; Anderson, M.; Collin, D.; Muegge, I.; Wan, J.; Brennan, D.; Kugler, S.; Terenzio, D.; Kennedy, C.; Lin, S.; Labadia, M. E.; Cook, B.; Hughes, R.; Farrow, N. A. Structural Studies Unravel the Active Conformation of Apo RORγt Nuclear Receptor and a Common Inverse Agonism of Two Diverse Classes of RORγt Inhibitors. *J. Biol. Chem.* **2017**, *292* (28), 11618–11630. https://doi.org/10.1074/jbc.M117.789024.
(64) Noguchi, M.; Nomura, A.; Murase, K.; Doi, S.; Yamaguchi, K.; Hirata, K.; Shiozaki, M.; Hirashima, S.; Kotoku, M.; Yamaguchi, T.; Katsuda, Y.; Steensma, R.; Li, X.; Tao, H.; Tse, B.; Fenn, M.; Babine, R.; Bradley, E.; Crowe, P.; Thacher, S.; Adachi, T.; Kamada, M. Ternary Complex of Human RORγ Ligand-Binding Domain, Inverse Agonist and SMRT Peptide Shows a Unique Mechanism of Corepressor Recruitment. *Genes Cells* **2017**, *22* (6), 535–551. https://doi.org/10.1111/gtc.12494.
(65) Strutzenberg, T. S.; Garcia-Ordonez, R. D.; Novick, S. J.; Park, H.; Chang, M. R.; Doebellin, C.; He, Y.; Patouret, R.; Kamenecka, T. M.; Griffin, P. R. HDX-MS Reveals Structural Determinants for RORγ Hyperactivation by Synthetic Agonists. *eLife* **2019**, *8*, e47172. https://doi.org/10.7554/eLife.47172.
(66) Olsson, R. I.; Xue, Y.; von Berg, S.; Aagaard, A.; McPheat, J.; Hansson, E. L.; Bernström, J.; Hansson, P.; Jirholt, J.; Grindebacke, H.; Leffler, A.; Chen, R.; Xiong, Y.; Ge, H.; Hansson, T. G.; Narjes, F. Benzoxazepines Achieve Potent Suppression of IL-17 Release in

Human T-Helper 17 (TH17) Cells through an Induced-Fit Binding Mode to the Nuclear Receptor RORγ. *ChemMedChem* **2016**, *11* (2), 207–216. https://doi.org/10.1002/cmdc.201500432.
(67) Chao, J.; Enyedy, I.; Van Vloten, K.; Marcotte, D.; Guertin, K.; Hutchings, R.; Powell, N.; Jones, H.; Bohnert, T.; Peng, C.-C.; Silvian, L.; Hong, V. S.; Little, K.; Banerjee, D.; Peng, L.; Taveras, A.; Viney, J. L.; Fontenot, J. Discovery of Biaryl Carboxylamides as Potent RORγ Inverse Agonists. *Bioorg. Med. Chem. Lett.* **2015**, *25* (15), 2991–2997. https://doi.org/10.1016/j.bmcl.2015.05.026.
(68) Marcotte, D. J.; Liu, Y.; Little, K.; Jones, J. H.; Powell, N. A.; Wildes, C. P.; Silvian, L. F.; Chodaparambil, J. V. Structural Determinant for Inducing RORgamma Specific Inverse Agonism Triggered by a Synthetic Benzoxazinone Ligand. *BMC Struct. Biol.* **2016**, *16* (1), 7. https://doi.org/10.1186/s12900-016-0059-3.
(69) Sasaki, Y.; Odan, M.; Yamamoto, S.; Kida, S.; Ueyama, A.; Shimizu, M.; Haruna, T.; Watanabe, A.; Okuno, T. Discovery of a Potent Orally Bioavailable Retinoic Acid Receptor-Related Orphan Receptor-Gamma-t (RORγt) Inhibitor, S18-000003. *Bioorg. Med. Chem. Lett.* **2018**, *28* (22), 3549–3553. https://doi.org/10.1016/j.bmcl.2018.09.032.
(70) Atomistic simulations shed new light on the activation mechanisms of RORγ and classify it as Type III nuclear hormone receptor regarding ligand-binding paths | Scientific Reports https://www.nature.com/articles/s41598-019-52319-x (accessed 2022 -03 -06).
(71) Zhang, Y.; Xue, X.; Jin, X.; Song, Y.; Li, J.; Luo, X.; Song, M.; Yan, W.; Song, H.; Xu, Y. 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.* **2014**, *78*, 431–441. https://doi.org/10.1016/j.ejmech.2014.03.065.
(72) Song, Y.; Xue, X.; Wu, X.; Wang, R.; Xing, Y.; Yan, W.; Zhou, Y.; Qian, C.-N.; Zhang, Y.; Xu, Y. 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.* **2016**, *116*, 13–26. https://doi.org/10.1016/j.ejmech.2016.03.052.
(73) Liu, Q.; Batt, D. G.; Weigelt, C. A.; Yip, S.; Wu, D.-R.; Ruzanov, M.; Sack, J. S.; Wang, J.; Yarde, M.; Li, S.; Shuster, D. J.; Xie, J. H.; Sherry, T.; Obermeier, M. T.; Fura, A.; Stefanski, K.; Cornelius, G.; Khandelwal, P.; Tino, J. A.; Macor, J. E.; Salter-Cid, L.; Denton, R.; Zhao, Q.; Dhar, T. G. M. Novel Tricyclic Pyroglutamide Derivatives as Potent RORγt Inverse Agonists Identified Using a Virtual Screening Approach. *ACS Med. Chem. Lett.* **2020**, *11* (12), 2510–2518. https://doi.org/10.1021/acsmedchemlett.0c00496.
(74) Li, H.; Leung, K.-S.; Wong, M.-H. Idock: A Multithreaded Virtual Screening Tool for Flexible Ligand Docking. In *2012 IEEE Symposium on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB)*; 2012; pp 77–84. https://doi.org/10.1109/CIBCB.2012.6217214.
(75) Vivoli, M.; Novak, H. R.; Littlechild, J. A.; Harmer, N. J. Determination of Protein-Ligand Interactions Using Differential Scanning Fluorimetry. *J. Vis. Exp. JoVE* **2014**, No. 91, 51809. https://doi.org/10.3791/51809.
(76) Schulz, M. N.; Landström, J.; Hubbard, R. E. MTSA—A Matlab Program to Fit Thermal Shift Data. *Anal. Biochem.* **2013**, *433* (1), 43–47. https://doi.org/10.1016/j.ab.2012.10.020.
(77) Lee, W.; Tonelli, M.; Markley, J. L. NMRFAM-SPARKY: Enhanced Software for Biomolecular NMR Spectroscopy. *Bioinformatics* **2015**, *31* (8), 1325–1327. https://doi.org/10.1093/bioinformatics/btu830.

(78) Evans, P. R. An Introduction to Data Reduction: Space-Group Determination, Scaling and Intensity Statistics. *Acta Crystallogr. D Biol. Crystallogr.* **2011**, *67* (Pt 4), 282–292. https://doi.org/10.1107/S090744491003982X.
(79) Evans, P. R.; Murshudov, G. N. How Good Are My Data and What Is the Resolution? *Acta Crystallogr. D Biol. Crystallogr.* **2013**, *69* (Pt 7), 1204–1214. https://doi.org/10.1107/S0907444913000061.
(80) Xu, T.; Wang, X.; Zhong, B.; Nurieva, R. I.; Ding, S.; Dong, C. Ursolic Acid Suppresses Interleukin-17 (IL-17) Production by Selectively Antagonizing the Function of RORgamma t Protein. *J. Biol. Chem.* **2011**, *286* (26), 22707–22710. https://doi.org/10.1074/jbc.C111.250407.
(81) Noguchi, M.; Nomura, A.; Doi, S.; Yamaguchi, K.; Hirata, K.; Shiozaki, M.; Maeda, K.; Hirashima, S.; Kotoku, M.; Yamaguchi, T.; Katsuda, Y.; Crowe, P.; Tao, H.; Thacher, S.; Adachi, T. Ternary Crystal Structure of Human RORγ Ligand-Binding-Domain, an Inhibitor and Corepressor Peptide Provides a New Insight into Corepressor Interaction. *Sci. Rep.* **2018**, *8* (1), 17374. https://doi.org/10.1038/s41598-018-35783-9.
(82) Huang, M.; Bolin, S.; Miller, H.; Ng, H. L. RORγ Structural Plasticity and Druggability. *Int. J. Mol. Sci.* **2020**, *21* (15), 5329. https://doi.org/10.3390/ijms21155329.
(83) Kallen, J.; Izaac, A.; Be, C.; Arista, L.; Orain, D.; Kaupmann, K.; Guntermann, C.; Hoegenauer, K.; Hintermann, S. Structural States of RORγt: X-Ray Elucidation of Molecular Mechanisms and Binding Interactions for Natural and Synthetic Compounds. *ChemMedChem* **2017**, *12* (13), 1014–1021. https://doi.org/10.1002/cmdc.201700278.
(84) Narjes, F.; Xue, Y.; von Berg, S.; Malmberg, J.; Llinas, A.; Olsson, R. I.; Jirholt, J.; Grindebacke, H.; Leffler, A.; Hossain, N.; Lepistö, M.; Thunberg, L.; Leek, H.; Aagaard, A.; McPheat, J.; Hansson, E. L.; Bäck, E.; Tångefjord, S.; Chen, R.; Xiong, Y.; Hongbin, G.; Hansson, T. G. Potent and Orally Bioavailable Inverse Agonists of RORγt Resulting from Structure-Based Design. *J. Med. Chem.* **2018**, *61* (17), 7796–7813. https://doi.org/10.1021/acs.jmedchem.8b00783.
(85) Jordan, V. C. Fourteenth Gaddum Memorial Lecture. A Current View of Tamoxifen for the Treatment and Prevention of Breast Cancer. *Br. J. Pharmacol.* **1993**, *110* (2), 507–517.
(86) Miller, W. R.; Ingle, J. N. *Endocrine Therapy in Breast Cancer*; CRC Press, 2002.
(87) Fabian, C. J.; Kimler, B. F. Selective Estrogen-Receptor Modulators for Primary Prevention of Breast Cancer. *J. Clin. Oncol.* **2005**, *23* (8), 1644–1655. https://doi.org/10.1200/JCO.2005.11.005.
(88) Pavlin, M.; Gelsomino, L.; Barone, I.; Spinello, A.; Catalano, S.; Andò, S.; Magistrato, A. Structural, Thermodynamic, and Kinetic Traits of Antiestrogen-Compounds Selectively Targeting the Y537S Mutant Estrogen Receptor α Transcriptional Activity in Breast Cancer Cell Lines. *Front. Chem.* **2019**, *7*.
(89) Toy, W.; Shen, Y.; Won, H.; Green, B.; Sakr, R. A.; Will, M.; Li, Z.; Gala, K.; Fanning, S.; King, T. A.; Hudis, C.; Chen, D.; Taran, T.; Hortobagyi, G.; Greene, G.; Berger, M.; Baselga, J.; Chandarlapaty, S. ESR1 Ligand-Binding Domain Mutations in Hormone-Resistant Breast Cancer. *Nat. Genet.* **2013**, *45* (12), 1439–1445. https://doi.org/10.1038/ng.2822.
(90) Jeselsohn, R.; Yelensky, R.; Buchwalter, G.; Frampton, G.; Meric-Bernstam, F.; Gonzalez-Angulo, A. M.; Ferrer-Lozano, J.; Perez-Fidalgo, J. A.; Cristofanilli, M.; Gómez, H.; Arteaga, C. L.; Giltnane, J.; Balko, J. M.; Cronin, M. T.; Jarosz, M.; Sun, J.; Hawryluk, M.; Lipson, D.; Otto, G.; Ross, J. S.; Dvir, A.; Soussan-Gutman, L.; Wolf, I.; Rubinek, T.;

Gilmore, L.; Schnitt, S.; Come, S. E.; Pusztai, L.; Stephens, P.; Brown, M.; Miller, V. A. Emergence of Constitutively Active Estrogen Receptor-α Mutations in Pretreated Advanced Estrogen Receptor–Positive Breast Cancer. *Clin. Cancer Res.* **2014**, *20* (7), 1757–1767. https://doi.org/10.1158/1078-0432.CCR-13-2332.

(91) Merenbakh-Lamin, K.; Ben-Baruch, N.; Yeheskel, A.; Dvir, A.; Soussan-Gutman, L.; Jeselsohn, R.; Yelensky, R.; Brown, M.; Miller, V. A.; Sarid, D.; Rizel, S.; Klein, B.; Rubinek, T.; Wolf, I. D538G Mutation in Estrogen Receptor-α: A Novel Mechanism for Acquired Endocrine Resistance in Breast Cancer. *Cancer Res.* **2013**, *73* (23), 6856–6864. https://doi.org/10.1158/0008-5472.CAN-13-1197.

(92) Robinson, D. R.; Wu, Y.-M.; Vats, P.; Su, F.; Lonigro, R. J.; Cao, X.; Kalyana-Sundaram, S.; Wang, R.; Ning, Y.; Hodges, L.; Gursky, A.; Siddiqui, J.; Tomlins, S. A.; Roychowdhury, S.; Pienta, K. J.; Kim, S. Y.; Roberts, J. S.; Rae, J. M.; Van Poznak, C. H.; Hayes, D. F.; Chugh, R.; Kunju, L. P.; Talpaz, M.; Schott, A. F.; Chinnaiyan, A. M. Activating ESR1 Mutations in Hormone-Resistant Metastatic Breast Cancer. *Nat. Genet.* **2013**, *45* (12), 1446–1451. https://doi.org/10.1038/ng.2823.

(93) Shiau, A. K.; Barstad, D.; Loria, P. M.; Cheng, L.; Kushner, P. J.; Agard, D. A.; Greene, G. L. The Structural Basis of Estrogen Receptor/Coactivator Recognition and the Antagonism of This Interaction by Tamoxifen. *Cell* **1998**, *95* (7), 927–937. https://doi.org/10.1016/S0092-8674(00)81717-1.

(94) Nettles, K. W.; Bruning, J. B.; Gil, G.; Nowak, J.; Sharma, S. K.; Hahm, J. B.; Kulp, K.; Hochberg, R. B.; Zhou, H.; Katzenellenbogen, J. A.; Katzenellenbogen, B. S.; Kim, Y.; Joachimiak, A.; Greene, G. L. NFκB Selectivity of Estrogen Receptor Ligands Revealed by Comparative Crystallographic Analyses. *Nat. Chem. Biol.* **2008**, *4* (4), 241–247. https://doi.org/10.1038/nchembio.76.

(95) Fanning, S. W.; Mayne, C. G.; Dharmarajan, V.; Carlson, K. E.; Martin, T. A.; Novick, S. J.; Toy, W.; Green, B.; Panchamukhi, S.; Katzenellenbogen, B. S.; Tajkhorshid, E.; Griffin, P. R.; Shen, Y.; Chandarlapaty, S.; Katzenellenbogen, J. A.; Greene, G. L. Estrogen Receptor Alpha Somatic Mutations Y537S and D538G Confer Breast Cancer Endocrine Resistance by Stabilizing the Activating Function-2 Binding Conformation. *eLife* **2016**, *5*, e12792. https://doi.org/10.7554/eLife.12792.

(96) Puyang, X.; Furman, C.; Zheng, G. Z.; Wu, Z. J.; Banka, D.; Aithal, K.; Agoulnik, S.; Bolduc, D. M.; Buonamici, S.; Caleb, B.; Das, S.; Eckley, S.; Fekkes, P.; Hao, M.-H.; Hart, A.; Houtman, R.; Irwin, S.; Joshi, J. J.; Karr, C.; Kim, A.; Kumar, N.; Kumar, P.; Kuznetsov, G.; Lai, W. G.; Larsen, N.; Mackenzie, C.; Martin, L.-A.; Melchers, D.; Moriarty, A.; Nguyen, T.-V.; Norris, J.; O'Shea, M.; Pancholi, S.; Prajapati, S.; Rajagopalan, S.; Reynolds, D. J.; Rimkunas, V.; Rioux, N.; Ribas, R.; Siu, A.; Sivakumar, S.; Subramanian, V.; Thomas, M.; Vaillancourt, F. H.; Wang, J.; Wardell, S.; Wick, M. J.; Yao, S.; Yu, L.; Warmuth, M.; Smith, P. G.; Zhu, P.; Korpal, M. Discovery of Selective Estrogen Receptor Covalent Antagonists for the Treatment of ERαWT and ERαMUT Breast Cancer. *Cancer Discov.* **2018**, *8* (9), 1176–1193. https://doi.org/10.1158/2159-8290.CD-17-1229.

(97) Shylaja, R.; Loganathan, C.; Kabilan, S.; Vijayakumar, T.; Meganathan, C. Synthesis and Evaluation of the Antagonistic Activity of 3-Acetyl-2H-Benzo[g]Chromen-2-One against Mutant Y537S Estrogen Receptor Alpha via E-Pharmacophore Modeling, Molecular Docking, Molecular Dynamics, and in-Vitro Cytotoxicity Studies. *J. Mol. Struct.* **2021**, *1224*, 129289. https://doi.org/10.1016/j.molstruc.2020.129289.

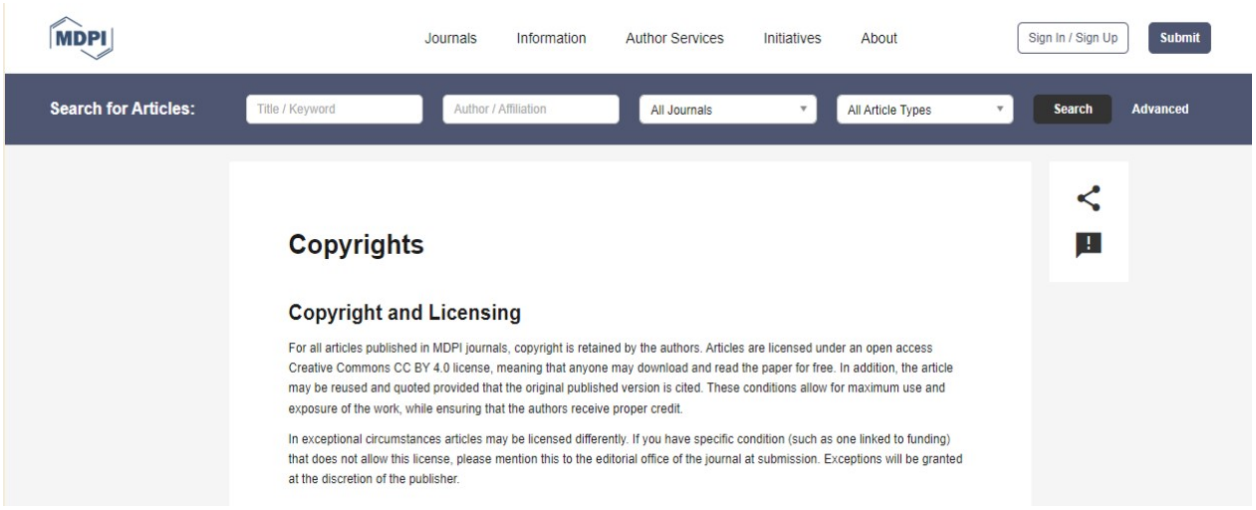
(98) Sunseri, J.; Koes, D. R. Pharmit: Interactive Exploration of Chemical Space. *Nucleic Acids Res.* **2016**, *44* (W1), W442–W448. https://doi.org/10.1093/nar/gkw287.
(99) Trott, O.; Olson, A. J. AutoDock Vina: Improving the Speed and Accuracy of Docking with a New Scoring Function, Efficient Optimization, and Multithreading. *J. Comput. Chem.* **2010**, *31* (2), 455–461. https://doi.org/10.1002/jcc.21334.
(100) Koes, D. R.; Baumgartner, M. P.; Camacho, C. J. Lessons Learned in Empirical Scoring with Smina from the CSAR 2011 Benchmarking Exercise. *J. Chem. Inf. Model.* **2013**, *53* (8), 1893–1904. https://doi.org/10.1021/ci300604z.
(101) Friesner, R. A.; Banks, J. L.; Murphy, R. B.; Halgren, T. A.; Klicic, J. J.; Mainz, D. T.; Repasky, M. P.; Knoll, E. H.; Shelley, M.; Perry, J. K.; Shaw, D. E.; Francis, P.; Shenkin, P. S. Glide: A New Approach for Rapid, Accurate Docking and Scoring. 1. Method and Assessment of Docking Accuracy. *J. Med. Chem.* **2004**, *47* (7), 1739–1749. https://doi.org/10.1021/jm0306430.
(102) Halgren, T. A.; Murphy, R. B.; Friesner, R. A.; Beard, H. S.; Frye, L. L.; Pollard, W. T.; Banks, J. L. Glide: A New Approach for Rapid, Accurate Docking and Scoring. 2. Enrichment Factors in Database Screening. *J. Med. Chem.* **2004**, *47* (7), 1750–1759. https://doi.org/10.1021/jm030644s.
(103) Ballester, P. J.; Mitchell, J. B. O. A Machine Learning Approach to Predicting Protein–Ligand Binding Affinity with Applications to Molecular Docking. *Bioinformatics* **2010**, *26* (9), 1169–1175. https://doi.org/10.1093/bioinformatics/btq112.
(104) Sander, T.; Freyss, J.; von Korff, M.; Rufener, C. DataWarrior: An Open-Source Program For Chemistry Aware Data Visualization And Analysis. *J. Chem. Inf. Model.* **2015**, *55* (2), 460–473. https://doi.org/10.1021/ci500588j.
(105) Lipinski, C. A.; Lombardo, F.; Dominy, B. W.; Feeney, P. J. Experimental and Computational Approaches to Estimate Solubility and Permeability in Drug Discovery and Development Settings1PII of Original Article: S0169-409X(96)00423-1. The Article Was Originally Published in Advanced Drug Delivery Reviews 23 (1997) 3–25.1. *Adv. Drug Deliv. Rev.* **2001**, *46* (1), 3–26. https://doi.org/10.1016/S0169-409X(00)00129-0.
(106) Anstead, G. M.; Carlson, K. E.; Katzenellenbogen, J. A. The Estradiol Pharmacophore: Ligand Structure-Estrogen Receptor Binding Affinity Relationships and a Model for the Receptor Binding Site. *Steroids* **1997**, *62* (3), 268–303. https://doi.org/10.1016/S0039-128X(96)00242-5.
(107) Maximov, P. Y.; Abderrahman, B.; Fanning, S. W.; Sengupta, S.; Fan, P.; Curpan, R. F.; Rincon, D. M. Q.; Greenland, J. A.; Rajan, S. S.; Greene, G. L.; Jordan, V. C. Endoxifen, 4-Hydroxytamoxifen and an Estrogenic Derivative Modulate Estrogen Receptor Complex Mediated Apoptosis in Breast Cancer. *Mol. Pharmacol.* **2018**, *94* (2), 812–822. https://doi.org/10.1124/mol.117.111385.
(108) Wu, X.; Hawse, J. R.; Subramaniam, M.; Goetz, M. P.; Ingle, J. N.; Spelsberg, T. C. The Tamoxifen Metabolite, Endoxifen, Is a Potent Antiestrogen That Targets Estrogen Receptor α for Degradation in Breast Cancer Cells. *Cancer Res.* **2009**, *69* (5), 1722–1727. https://doi.org/10.1158/0008-5472.CAN-08-3933.

Appendix A - Copyright permission

Appendix A acknowledges illustrations reprinted for use in this work.

Part of the article * was reused in 3.1 Introduction, at Page 30 – 35, under an open access Creative Commons CC BY 4.0 license, according to the MDPI website:

https://www.mdpi.com/authors/rights



* Huang, M.; Bolin, S.; Miller, H.; Ng, H. L. RORγ Structural Plasticity and Druggability. *Int. J. Mol. Sci.* **2020**, *21* (15), 5329. https://doi.org/10.3390/ijms21155329.

Appendix B - Definitions of agonist, inverse agonist, and antagonist

The following definitions are cited from:

Solt L.A., Burris T.P. Action of RORs and their ligands in (path)physiology. *Trends Endrocrinol. Metab.* **2012**, *23*, 619-627. http://dx.doi.org/10.1016/j.tem.2012.05.012

Agonist

An agonist binds to the nuclear receptor ligand-binding domain (LBD) and induces a conformational change, leading to increased recruitment of coactivators.

Inverse agonist

An inverse agonist binds to the LBD and induces a conformational change of the protein to suppress the cofactor recruitment, resulting in inhibition of the basal constitutive transcriptional activity of the protein.

Antagonist

An antagonist binds to the LBD and prevents an agonist from binding and activating the protein, but it does not elicit a response from the protein.

Appendix C - Python codes for K_d measurement

```
#!/usr/bin/env python
# coding: utf-8

import matplotlib as mpl
import matplotlib.pyplot as plt
import numpy as np
from pylab import cm
from scipy.optimize import curve_fit
from scipy.optimize import minimize_scalar

def power_law(x, a, b):          #power law fitting
    return a*np.power(x, b)

xcoords = np.array([0, 0.5, 5.5, 25, 50,125, 200, 400])         #ligand concentrations
ycoords = np.array([0, 0.028, 0.035, 0.068, 0.075, 0.093, 0.089, 0.094])
                                                    #experimental changes of chemical shifts

pars, cov = curve_fit(f=power_law, xdata=xcoords, ydata=ycoords, p0=[0, 0], bounds=(-np.inf,
np.inf))

stdevs = np.sqrt(np.diag(cov))          # calculate the parameters in the power-law equation
print(pars)
print(stdevs)

plt.plot(xcoords, power_law(xcoords, *pars), linestyle='--', linewidth=2, color='black')
Rsq_sum = lambda kd: sum((ycoords - 0.047*np.power((xcoords/kd), 0.18377823))**2)
                                                    # sum of R*R related to Kd

Kd = minimize_scalar(Rsq_sum, method='brent').x    # Kd value when sum of R*R is minimized
print(Kd)
```

Appendix D - Lipinski's rule of five

Lipinski's rule of five is as follows:

Number of hydrogen-bond donor is no more than 5;

Number of hydrogen-bond acceptor is no more than 10;

Molecular weight is less than 500 daltons;

LogP is no more than 5.