Protein-ligand interactions of druggable protein targets

by

Samson Aeloa Souza

B.S., Chapman University, 2007
B.S., University of Hawai‘i at Mānoa, 2013

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

2020

Abstract

A druggable protein target is one in which an exogenous ligand will induce the desired response. In this work, small molecule interactions of three druggable protein targets will be detailed. The first of these is a bacterial enzyme involved in the synthesis of cofactor biotin, which is an essential cofactor exploited across all life domains. It is necessary for fatty acid biosynthesis, gluconeogenesis, and amino acid metabolism. Mammals lack the biosynthetic machinery to produce it and must acquire it in the diet. Meanwhile, bacteria such as *E. coli*, and *M. tuberculosis* can synthesize it endogenously. As such, enzymes involved in biotin synthesis are attractive targets in antimicrobial development. Diaminopelargonic acid synthase (BioA) catalyzes the second step in the conserved pathway from starting compounds pimeloyl-CoA and L-alanine. Unlike other bacteria, *Bacillus subtilis* requires L-lysine as a substrate for transamination of 7-keto-8-aminopelargonic acid (KAPA) to its diamino-product, 7,8-diaminopelargonic acid (DAPA), by BioA. I present kinetic work that suggests a donation of lysine ε-amino group to KAPA. I follow this with the crystal structure of PLP-conjugated lysine as an external aldimine (LLP). The adduct is stabilized by electrostatic interactions between the carboxylate and R410, and pi-cation interactions between the former lys α-amine and two aromatic side chains in the pocket.

In the latter segment of this work, I survey ligand interactions of two membrane proteins directly involved in estrogen signaling. The first of these two proteins, G-protein coupled estrogen receptor (GPER), is localized in the endoplasmic reticulum. This research, which was the first to demonstrate *in vitro* ligand binding with recombinant protein, focuses on steps to produce functional GPER for structural and binding assays. GPER is a potential non-nuclear strategy for breast cancer therapy since 10 – 20 % of diagnoses are estrogen receptor negative.

The second estrogen-related protein I will explore is the cytochrome P450 enzyme aromatase (Cyp19). It catalyzes the last biosynthetic step in the production of endogenous estrogens in mammals. To this end, it is a current target in the treatment of hormone-related illnesses and diseases such as endometriosis, ovarian cancer, and breast cancer. Current aromatase inhibitors (AIs), for instance, tamoxifen, are potent, yet they often lead to debilitating side effects. Eventual relapse creates a need for novel breast cancer therapeutics that improve patient outcome. Virtual screening of a library of millions of compounds is often employed to initially uncover drug candidates. I provide activity data of these top hit candidates against a putative Cyp19 allosteric site. Two lead compounds, AR11 and AR13, exhibit potent, anti-aromatase activity comparable to active tamoxifen metabolite, endoxifen. Inhibitory mechanisms of these compounds and the journey to find a promising construct for cocrystallization will be explored. This insight will aid in the search to unearth a novel class of allosteric aromatase inhibitors with diverse toxicity profiles.

Protein-ligand interactions of druggable protein targets

by

Samson Aeloa Souza

B.S., Chapman University, 2007
B.S., University of Hawai‘i at Mānoa, 2013

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

2020

Approved by:

Major Professor
Ho Leung Ng

Copyright

Abstract

A druggable protein target is one in which an exogenous ligand will induce the desired response. In this work, small molecule interactions of three druggable protein targets will be detailed. The first of these is a bacterial enzyme involved in the synthesis of cofactor biotin, which is an essential cofactor exploited across all life domains. It is necessary for fatty acid biosynthesis, gluconeogenesis, and amino acid metabolism. Mammals lack the biosynthetic machinery to produce it and must acquire it in the diet. Meanwhile, bacteria such as *E. coli*, and *M. tuberculosis* can synthesize it endogenously. As such, enzymes involved in biotin synthesis are attractive targets in antimicrobial development. Diaminopelargonic acid synthase (BioA) catalyzes the second step in the conserved pathway from starting compounds pimeloyl-CoA and L-alanine. Unlike other bacteria, *Bacillus subtilis* requires L-lysine as a substrate for transamination of 7-keto-8-aminopelargonic acid (KAPA) to its diamino-product, 7,8-diaminopelargonic acid (DAPA), by BioA. I present kinetic work that suggests a donation of lysine ε-amino group to KAPA. I follow this with the crystal structure of PLP-conjugated lysine as an external aldimine (LLP). The adduct is stabilized by electrostatic interactions between the carboxylate and R410, and pi-cation interactions between the former lys α-amine and two aromatic side chains in the pocket.

In the latter segment of this work, I survey ligand interactions of two membrane proteins directly involved in estrogen signaling. The first of these two proteins, G-protein coupled estrogen receptor (GPER), is localized in the endoplasmic reticulum. This research, which was the first to demonstrate *in vitro* ligand binding with recombinant protein, focuses on steps to produce functional GPER for structural and binding assays. GPER is a potential non-nuclear strategy for breast cancer therapy since 10 – 20 % of diagnoses are estrogen receptor negative.

The second estrogen-related protein I will explore is the cytochrome P450 enzyme aromatase (Cyp19). It catalyzes the last biosynthetic step in the production of endogenous estrogens in mammals. To this end, it is a current target in the treatment of hormone-related illnesses and diseases such as endometriosis, ovarian cancer, and breast cancer. Current aromatase inhibitors (AIs), for instance, tamoxifen, are potent, yet they often lead to debilitating side effects. Eventual relapse creates a need for novel breast cancer therapeutics that improve patient outcome. Virtual screening of a library of millions of compounds is often employed to initially uncover drug candidates. I provide activity data of these top hit candidates against a putative Cyp19 allosteric site. Two lead compounds, AR11 and AR13, exhibit potent, anti-aromatase activity comparable to active tamoxifen metabolite, endoxifen. Inhibitory mechanisms of these compounds and the journey to find a promising construct for cocrystallization will be explored. This insight will aid in the search to unearth a novel class of allosteric aromatase inhibitors with diverse toxicity profiles.

Table of Contents

List of Figures

List of Tables

Acknowledgements

Mahalo-

Dr. Ng for supporting me, believing in me, and being receptive to my ideas. Thank you Dr. Dane Kurohara, the aromatase crew- Shalin, Michelle, Brendan, Bryant, Raul. Dr. Jessy Lu, and fellow lab colleagues- Mian, John, Ye, and Anthony.

Kstate grad students, and my colleague and friend Dr. Nicole Green for lending your kokua as I made my transition as part of the K-State family. I'd like to thank my Committee members, Dr. Michal Zolkiewski and Dr. Brian Geisbrecht- for your guidance, Dr. Ping Li and Dr. John Schlup- for your time and feedback.

Mcnair program- Maile Goo, Kuaʻana and Native Hawaiian Student Services- Aunty Kuʻumealoha and Uncle Kinohi. Thank you Dr. Hemscheidt, Dr. Bingham, Dr. Williams, and Dr. Jarrett for your contributions towards my earlier work as a graduate student. I'd also like to acknowledge Hawaiʻi colleagues Hope, Josh, and Julie for their support and friendship.

My childhood friends and neighbors in my hometown of Wailupe Valley, Kamehameha Schools, and my life-long friends Liz, Chablis, Josh, D.J., and Jackie.

My brothers Kamalu, Miki, J.P., and Mikala who continuously support me outside of academia- Mahalo. I'd like to especially acknowledge my Mom and Dad, Nancy Ann Malulani and Michael Wayne for their unconditional love and support. *Me ke aloha pumehana kakou.*

Dedication

This work is dedicated to my Aunty Cathy Mahealani Brede Ogata, Aunty Lehua Opanaʻewa DeRamos Kalaʻi, Grandma Ethel Lee Souza, Grandma Marcella Kaopua Brede, and my mom Nancy Ann Malulani Brede Souza. These women are my puʻuhonua- my foundation and place of refuge. They are moʻo wahine- listening to their stories inspired me to champion this path of knowledge.

Preface

This work will be described in three major parts. The first will provide the support that lysine donates its ε-amino group to furnish PMP for transamination in *B. subtilis* diaminopelargonic acid synthase (BioA). I introduce preliminary work completed by the members of the Jarrett lab that shows ^{15}N incorporation in dethiobiotin when isotopically-labeled lysine is fed into the enzyme-coupled bioF/bioA/bioD pathway from substrates pimeloyl-coA and L-alanine. I then explore the active site of the solved crystal structure and propose a mechanism by which PMP forms.

In the final two parts, I investigate steroidal targets G protein-coupled estrogen receptor (GPER), and human aromatase (CYP19). In GPER, we sought a viable recombinant expression system to produce milligram quantities of protein for structural studies. Preliminary work explored expression levels in *E. coli*, while I focused my attention to heterologous expression in yeast, and production in our lab's optimized cell-free expression system. I provide evidence that we can recombinantly produce functional GPER.

Following my discussion of GPER, I will detail steps to uncover human aromatase inhibition by predicted, allosteric small molecules. The first part of this chapter will investigate *in silico* high throughput screening for potential allosteric inhibitors, and provide activity data to show that small molecule-binding of one lead inhibitor occurs at a distinct location from the active site. The second part will outline steps to show this structurally. I describe and characterize functional constructs for subsequent crystallization screens.

Chapter 1 - Introduction

At the most fundamental level, we need to look at the molecular structure to understand how things work. Crystal structures provide the blueprint for us to solve real-world problems, and in drug design, this is no exception. However, this is complicated if the dysregulated protein under study is not druggable- which I will leave for my final remarks. This work is focused on exploring the ligand interactions with three different druggable protein targets. Comprehensively, they are from two different species belonging to two different life domains. They also include soluble and membrane-associated proteins, enzymes and receptors, and enzyme-coupling and independently-functioning proteins. They represent unique structures and functions, yet all together, they are druggable. By this, I am referring to proteins with features that can be exploited to attain the desired response.

The expansive contributions of T-cell receptor-mimetics, peptides, immunotherapies, and natural products in drug discovery is beyond the scope of this work. I limit our definition of drugs to small molecules that loosely align with Lipinski's rule of 5. In 1997, Lipinski, Lombardo, and Feeney set guidelines that predict drug-likeness based on properties that influence absorption and solubility. These include LogP < 5, mass < 500 Da, < 5 H-bond donors, and < 10 H-bond acceptors.[1] Since then, understanding mechanisms that contribute to drug disposition has led to important discoveries. For instance, annotation of several hundreds of membrane transporter proteins belonging to ATP-binding cassette and solute carrier protein superfamilies led to our growing knowledge of their critical role in drug flux.[2] As such, these rules are taken loosely, since Lipinski and colleagues observed drugs that were FDA-approved over 20 years ago. Currently, with more FDA-approved drugs, a reassessment of these rules determined that current drug-like

compounds exceeded 500 Da, and averaged a logP of 5.[3] This ultimately brings to light the influence the FDA has on determining the course of drug discovery.

Enzymes in biotin biosynthesis are targets for antimicrobial development

Biotin, also known as vitamin B7 and vitamin H,[4] is a coenzyme involved in one-carbon transformations of various biosynthetic pathways.[5] These transformations include carboxylation/decarboxylation reactions[6] in amino acid metabolism, lipogenesis, and gluconeogenesis. To this end, biotin is an essential cofactor for proper cell function and maintenance in organisms across all domains of life.[5] In mammals, biotin deficiency may cause ketoacidosis, seizures, and teratogenic effects including mental retardation, physical malformations, and hypotonia.[7] Biotin deficiency may be seen in pregnant women, and malnourished individuals.[7] Mammals lack the biosynthetic enzymes to produce biotin, therefore biotin must be acquired in the diet. Plants and bacteria, such as *Mycobacterium tuberculosis* (Mtb), *Bacillus subtilis*, and *Escherichia coli* produce the cofactor endogenously by a pathway that is well characterized. This process will be detailed later.

DAPA synthase (BioA) is a target in antimicrobial development

BioA is an enzyme in a conserved pathway to produce biotin. Mammals lack these enzymes to produce biotin and must administer it in the diet, or absorb biotin from the microbiota. As such, these enzymes are attractive targets in antimicrobial drug development. In *Mycobacterium marinum*, a biotin auxotroph failed to grow on Sauton minimal media, while growth was restored upon addition of biotin in a dose-dependent manner.[8] In bacteria, such as Mtb, biotin biosynthetic enzymes can be targeted to inhibit its synthesis, thereby attenuating

lipogenesis in chronic infections. During the late stages of persistence, bacilli tubercles upregulate enzymes involved in fatty acid metabolism and utilize the glyoxylate cycle for carbon and energy.[9]

In *Mycobacterium smegmatis*, transposon mutagenesis at the BioA gene stunted stationary phase growth, despite growth in biotin-supplemented, nutrient-rich media. This suggests that *de novo* synthesis of biotin is necessary for growth due to the cell's inability to absorb biotin in the stationary phase.[10] In Mtb, a Δ*bioA* mutant *in vitro* was not viable when it was transferred from biotin-supplemented media to biotin-free media. However, viability could be rescued upon the addition of DTB. In murine models, Δ*bioA* mutants failed to establish acute infection, and Mtb survival was diminished in chronic infections when *bioA* was silenced. This aside, BioA activity needs to be inhibited by greater than 90 % in mice to induce a substantial Mtb growth defect.[10] As expected, enzyme vulnerability is likely to pose a challenge in drug development.

GPER-targeted therapy can modulate diverse signaling pathways in tissues throughout the body

G protein-coupled estrogen receptor (GPER) intimate role in modulating cell signaling pathways, and its wide distribution in tissues throughout the body makes it a novel target for understanding human health and disease. Chakrabarti and Davidge (2012) report that GPER activation attenuates up-regulation of adhesion proteins via TNFα disruption in the endothelium preventing the binding of leukocytes for inflammatory response.[9] Coupled with NO activation for vasodilation, GPER is beneficial in that it can prevent reperfusion injury, autoimmune disorders such as multiple sclerosis, and stroke. Several studies also showed myocardial and neuroprotective effects, albeit a sex-dependent role in glucose-uptake and bone growth was exhibited.[11] As such, GPER dysregulation has complex and deleterious physiological effects often resulting in disease.

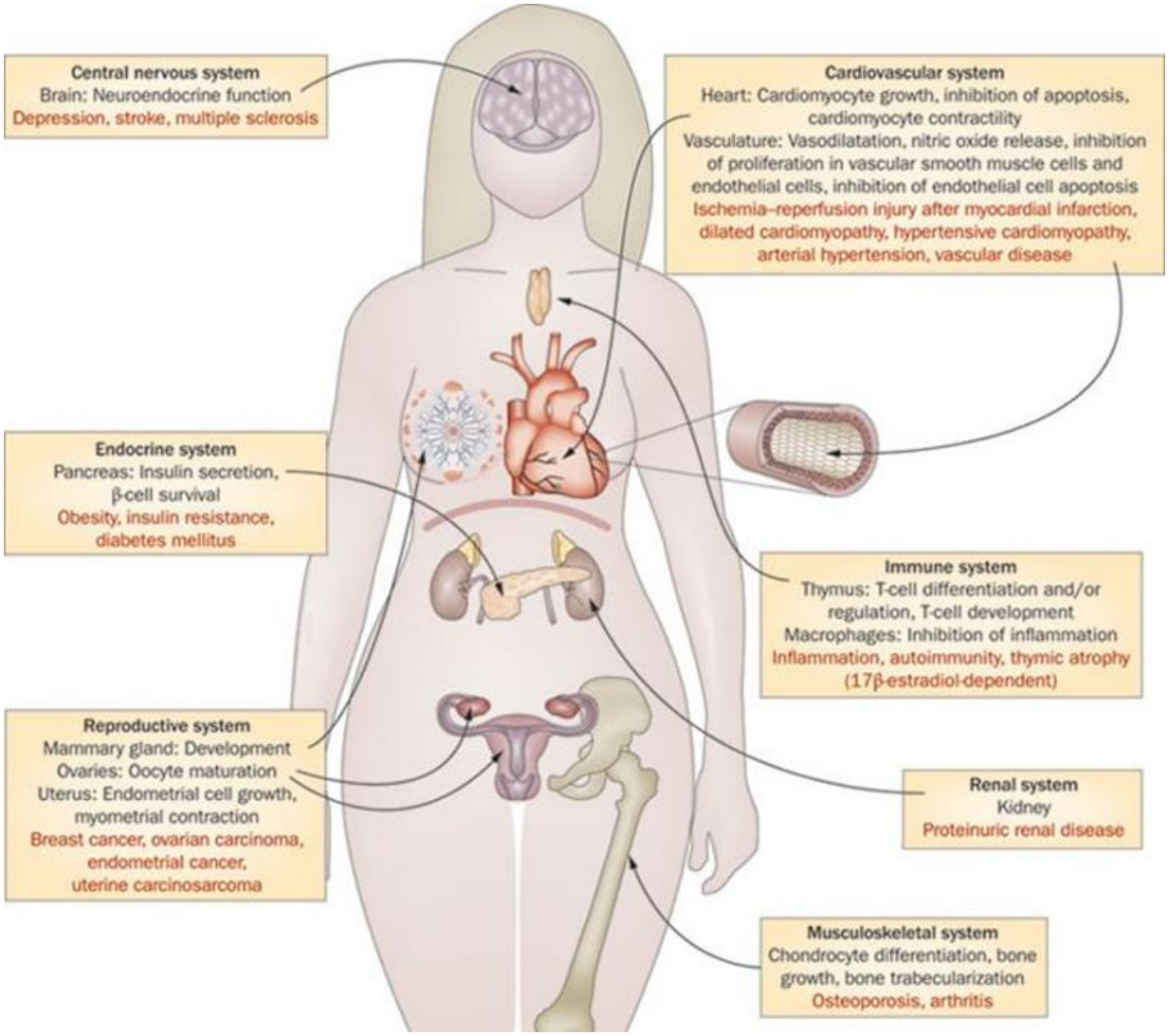


Figure 1.1 GPER distribution throughout tissues in the human body.

GPER dysregulation is experimentally or clinically shown to play a role in the development or progression of conditions/diseases highlighted in red. Reused with permission from Springer Nature.[12]

In the past decade, the focus has been placed on GPER's function in breast cancer and tumorigenesis. The preferred treatment for breast cancer in women is the administration of tamoxifen or aromatase inhibitors. However, these treatments are non-responsive in 15-20% of

women who have triple-negative breast cancer (TNBC). These women lack cancer cells expressing receptor-targets ERα, progesterone receptor, and human EGF-2 (HER2). Girgert and colleagues (2012) report increased Src kinase activity, HER2 activation, and *c-fos* expression in two TNBC cell lines upon incubation with 17 β-estradiol and tamoxifen.[13] In this study, a direct relationship between GPER expression and cell proliferation was inferred by the introduction of siRNA. Furthermore, knock-down of GPER expression resulted in the elimination of Src, EGFR, and *c-fos* signaling. This implies that GPER is upregulated with current treatments, and plays a role in TNBC cell proliferation, providing evidence for a novel target for women with breast cancer. In 2014, the first *in vivo* study of breast cancer and its correlation with GPER was conducted. Fewer metastases and smaller tumor size in GPER knock-out mice were reported relative to wild type PyMT mice [transgenic mouse model of mammary tumorigenesis (PyMT)]. Metastasis is the current indicator for morbidity and mortality of women with breast cancer, and tumor size is a measure of cancer aggressiveness.[14]

Allosteric inhibition of aromatase may provide a novel strategy for endocrine therapy

Aromatase (Cyp19) is a cytochrome P450 that catalyzes the three-step conversion of androgens to estrogens. As a mammalian P450 enzyme, it requires 3 NADPH in an enzyme-coupled reaction with cytochrome P450 reductase (CPR) for a single turnover. The first two reactions have been well-characterized by the classic hydroxylation pathways P450s are known. The result of these consecutive reactions is the formation of a gem diol at the C19 methyl.[15] However, the mechanism of the last step for the aromatization of the steroid A ring, where it owes its name to, remained debatable. It was not until 2014 that Guengerich and Yoshimoto produced

compelling evidence that FeO^{3+} was the active species that catalyzed the final step rather than an $FeOO^{-}$ specie.[16] The length of time taken to characterize this step is surprising since Cyp19 has been extensively studied, and used for endocrine therapy in women with estrogen receptor α (ERα) positive breast cancer for over 40 years.

The inception of the first-generation aromatase inhibitor (AI) aminoglutethimide occurred after it was withdrawn due to unintended effects on adrenal function in 1966- originally administered as an anticonvulsant.[17] Since then, clinical success in the treatment of hormone-dependent breast cancer has spurred interest in the development of the third generation inhibitors used today. These include reversible nonsteroidal azole inhibitors anastrazole and letrozole, and irreversible steroidal inhibitor exemestane. Since 2009, there have been 11 structures of human Cyp19 solved by Ghosh and colleagues.[18–20] Three of these structures contain steroidal inhibitors with different C6 aliphatic substitutions- one of these include exemestane (pdb ID 3S7S). Intriguingly, the of the set of structures deposited in 2018, one of them crystallized with polyethylene glycol (PEG) at the proximal heme site (pdb ID 5JKV). PEG is a common precipitant used in crystallization, and it appears that this was unintentional. Additionally, PEG was reported to decrease enzyme activity.[20] This demonstrated that Cyp19 could be regulated at an allosteric site, and likely due to altering Cyp19-CPR interactions.

Chapter 2 - Structure of BioA with Amino-donor Lysine

In *B. subtilis*, BioA selects for lysine as an amino source. In this work, I present preliminary kinetic data (provided by Joseph Jarrett Lab, U. of Hawai'i, Honolulu, HI), and the crystal structure of BioA as an external aldimine. It suggests that lysine donates its ε-amino group to PLP in the first step of transamination. I will then explore the evolutionary consequences for substrate divergence in *B. subtilis* BioA and implications for antimicrobial drug development. A portion of this chapter will be submitted for publication.

Introduction

The biotin biosynthetic pathway

Biotin is formed in two phases- the first phase results in the production of a pimelate thioester. During the second phase, the bicyclic heterocycle is formed in four separate conserved reactions. However, the pimelate thioester needed for the first enzymatic reaction in this phase can vary across different species. In *E. coli*, pimeloyl-ACP acts as the substrate for BioF,[21] the first enzyme in the second phase, while Mtb utilizes pimeloyl-CoA.[22] In *B. subtilis*, the productive turnover to 7-keto-8-aminopelargonic acid (KAPA) requires pimeloyl-CoA both *in vitro* and *in vivo*.[23] The second reaction involves the PLP-dependent enzyme BioA. It is a two-step reaction that requires a substrate amino donor that converts enzyme-bound PLP to a noncovalently-bound PMP. KAPA may subsequently enter the active site for transamination to produce 7,8-diaminopelargonic acid (DAPA) furnishing the enzyme-bound PLP. In the third reaction, BioD carboxylates DAPA to form a ureido ring in the presence of ATP and magnesium to yield dethiobiotin (DTB). In the final reaction, BioB catalyzes DTB to biotin by an S-

adenosylmethionine (SAM)-radical mechanism.[22] Figure 2.1 is a schematic of the second, conserved phase of the biotin biosynthetic pathway.

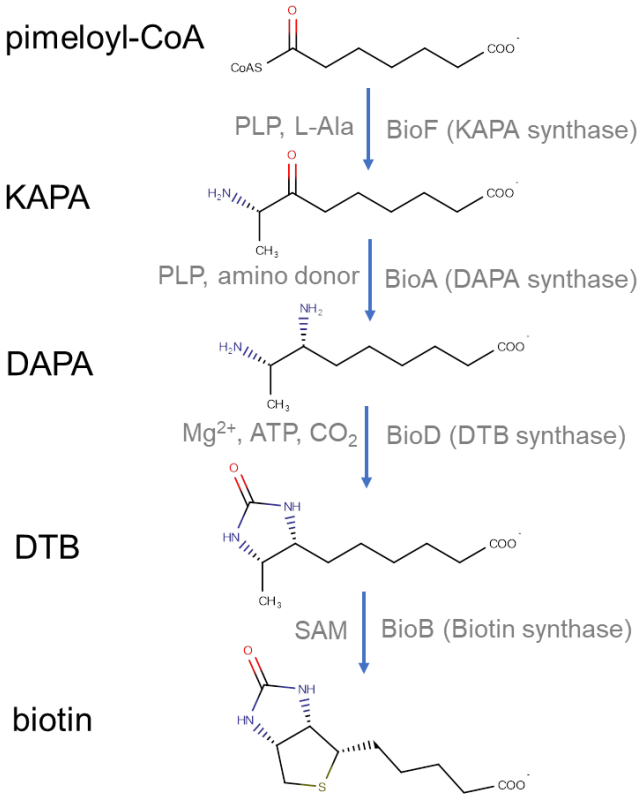


Figure 2.1 Schematic of the last four enzymatic steps of the biotin biosynthetic pathway. The last four biosynthetic steps in biotin synthesis involve four unique enzymes with a conserved function across species. Cofactors/substrates, and the enzyme responsible for each process are listed to the left, and right of each progress arrow, respectively.

SAM is classically known to methylate substrates in biological processes. BioA is unique in that it is the only transaminase to utilize S-adenosylmethionine (SAM) as an amino donor, and it is the only PLP enzyme to use it as a substrate.[24] Interestingly, the BioA gene is the only cistron of the biotin operon transcribed leftward. In 1974, Stoner and Eisenberg proposed that gene duplication and transposition of the BioF gene could explain this phenomenon, and this begs the question whether this correlates with a change in SAM's function. In *B. subtilis*, and only seen in

Bacillus spp., BioA uses substrate L-lysine rather than SAM as an amino source. Although *B. subtilis* is nonpathogenic, in a collaborative effort, we looked to understand this anomaly by characterizing it kinetically and structurally to seek answers on why this occurs.

Materials and Methods

Purified BioA was received from Dr. Joseph Jarrett lab from the University of Hawai‘i at Mānoa. Protocol for BioA preparation and mutant kinetic experiments in its entirety is available on the University of Hawai‘i dissertation repository on work by Dr. Julia Cramer. Preliminary ^{15}N labeled assays described in the results section were performed by Jennifer Morris- former member of the Jarrett lab.

Crystallization and data collection

BioA was co-crystallized with cofactor PLP, and substrate lysine by a sitting-drop vapor diffusion method. In brief, 10 mg/mL BioA prepared in 1 mM PLP, 10 mM L-lysine, and 50 mM HEPES (pH 7.5) was vapor diffused against reservoir solution at a 2:1 protein:reservoir ratio. Mother liquor contained 0.2 M NaOAC, 0.1 M Tris HCl (pH 8.5), and 30 % (w/v) PEG 4000. Though we do not report it here, high resolution data of apoenzyme was collected to 1.8 Å in 0.2 M NaCl, 0.1 M Tris-HCl (pH 8.5), and 40 % (w/v) PEG 3350 at a 2:1 protein:reservoir ratio. Needles, and microcrystals were apparent after 1 week of growth at 18˚C. After 8 weeks, crystals were submerged in LVCO, and flash frozen for data collection at Advanced Light Source, Berkeley Lab. Data processing, and molecular replacement was completed in XSCALE (XDS suite),[25] and PHASER (CCP4 interface 2.2.1)[26] with 3DOD (PDB ID), respectively.

Model building and refinement

Model building and restrained refinement routines were performed in COOT 0.8.6.1,[27] and RefMac5 (CCP4 suite 7.0.019).[28–32] PLP-Lys (LLP) ligand was regularized in J ligand,[33] then fit in the BioA active site electron density.

All graphics were generated in Chimera[34] software except where it is stated otherwise.

Results

Preliminary kinetics work show DTB product with ^{15}N incorporation from ($^{15}N\varepsilon$)-L-lysine

After a 15 hr enzyme-coupled BioF/A/D reaction, dethiobiotin was separated from other reaction components by HPLC with a C18 column (Waters Atlantis, 2.1 X 150 mm, 5 μM), and detected at 210 nm. Buffer A (distilled water, 0.1 % H_3PO_4) was used to equilibrate the column before applying a linear gradient for DTB collection. DTB retention time was 17.7 min with a 30 min linear gradient 2 – 30 % buffer B (80 % ACN, 0.1 % H_3PO_4). Reconstituted samples were auto injected onto an Agilent 6410 LC/MS for the detection of DTB products. Test substrates included L-lysine, ($^{15}N\alpha$)-L-lysine, and ($^{15}N\varepsilon$)-L-lysine. ^{15}N- labeled DTB was detected by a monoisotopic m/z at 216.1 when ($^{15}N\varepsilon$)-L-lysine was used as an amino donor in the BioF/A/D reaction. The other test substrates had an M+1 peak at 215.1. The catalytic constants reported in Table 2.1 were determined at the time this preliminary study was performed.

Table 2.1 Catalytic constants of BioA orthologs.

Specie	*Substrate*	K_m *(mM)*	K_{cat} *(min^{-1})*	K_{cat}/K_m *(M^{-1}s^{-1})*
B. subtilis	L- lysine	2.45 2-25*	0.09	0.59
E. coli	S- adenosyl methionine	0.17 0.2*	0.63 17*	62 87 – 1400*
M. tuberculosis	S- adenosyl methionine	1.17*	0.31-1*	4 – 21*

*Values published separately by Dey et al., Stoner et al., and Arsdell et al.

BioA crystallization and metrics

BioA crystallized as a triclinic lattice type in the P1 space group. There is a dimer per asymmetric unit. Chain A had sufficient electron density to model the PLP cofactor in its external aldimine form. There are 12 outliers corresponding to 1.45 % of the modeled residues. Four of these correspond to the active site lysine and methionine in both chains that exist as outliers in high-resolution structures of BioA. The remaining outliers correspond to glycine and serine residues, and the latter resides at the protein surface. The published 2.2 Å resolved structure of *B.subtilis* BioA with KAPA bound had 11 outliers reported at 1.23 % of the modeled residues. The official preliminary validation report is provided in Appdendix C.

Table 2.2 BioA data collection and refinement statistics as an external aldimine.

Metric	
Data Collection	
Space group	P1
Cell dimensions	
a, b, c (Å)	58.30, 60.54, 62.72
α, β, γ (deg)	96.1, 106.0, 99.2
Resolution (Å)	59.51 – 2.59 (2.65 – 2.59)
R_{pim}	0.134 (0.705)
I/σI	7.8 (2.1)
Completeness	91.8
No. lattices	1
Wavelength	0.9795
Refinement	
Resolution (Å)	2.59
No. reflections	25151
R_{work}/R_{free}	0.232/0.340
No. atoms	13204
B factor (Å^2)	46.0
rmsd from ideal	
bond length (Å)	0.012
bond angle (deg)	1.623
chiral volume (Å^3)	0.087

*Parenthesized values for the highest-resolution bin.

Substrate LLP at the active site

In its internal aldimine form, PLP binds Lys280. Catalysis by PLP requires a ~27˚ ring tilt that translocates the aldimine carbon and nitrogen 2.5 and 3.0 Å, individually to accommodate binding of the substrate lysine (Figure 2.2). In the active site, the lysyl carboxylate forms a salt bridge with Arg410 at 2.7 Å, while the α-amino substituent forms pi-cation interactions with Tyr146, and Phe17 (Figure 2.3).

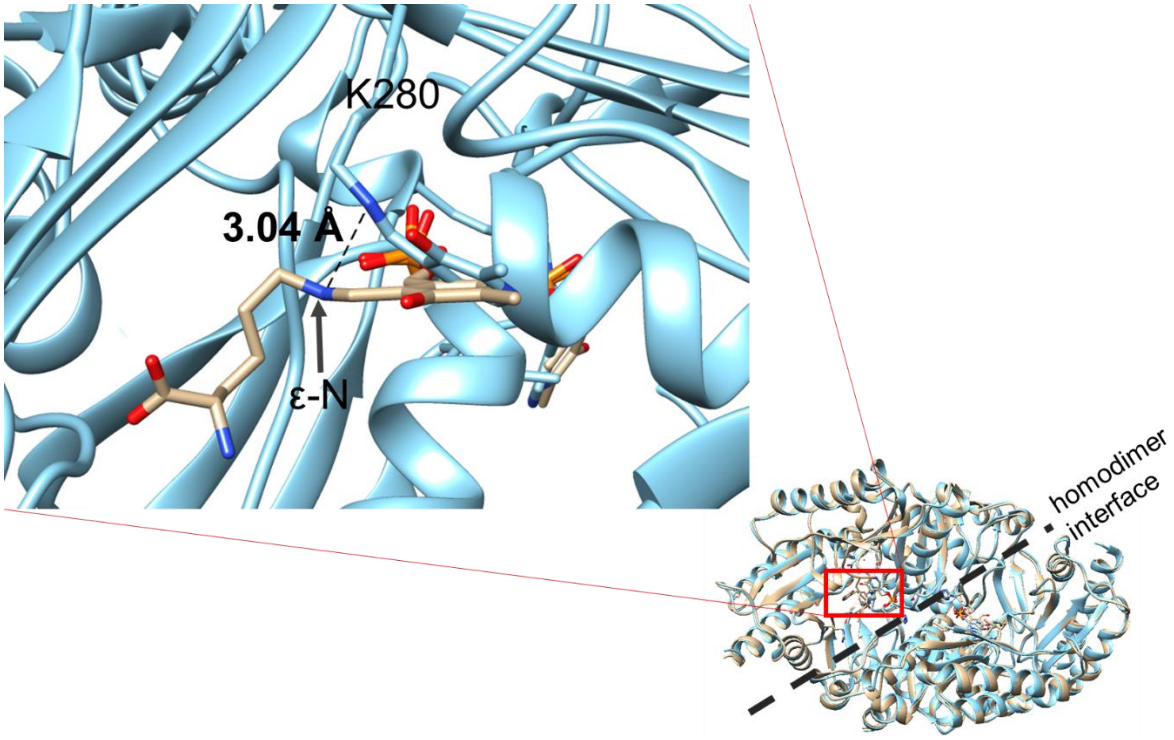


Figure 2.2 Superposition of holoprotein in the absence (pdb ID 3DOD, cyan) and presence (pdb ID 6WNN, beige) of lysine.

A ~27˚ tilt of the pyridoxal group results in a 3.04 Å distance between the ε-nitrogen of the internal aldimine and external aldimine structures.

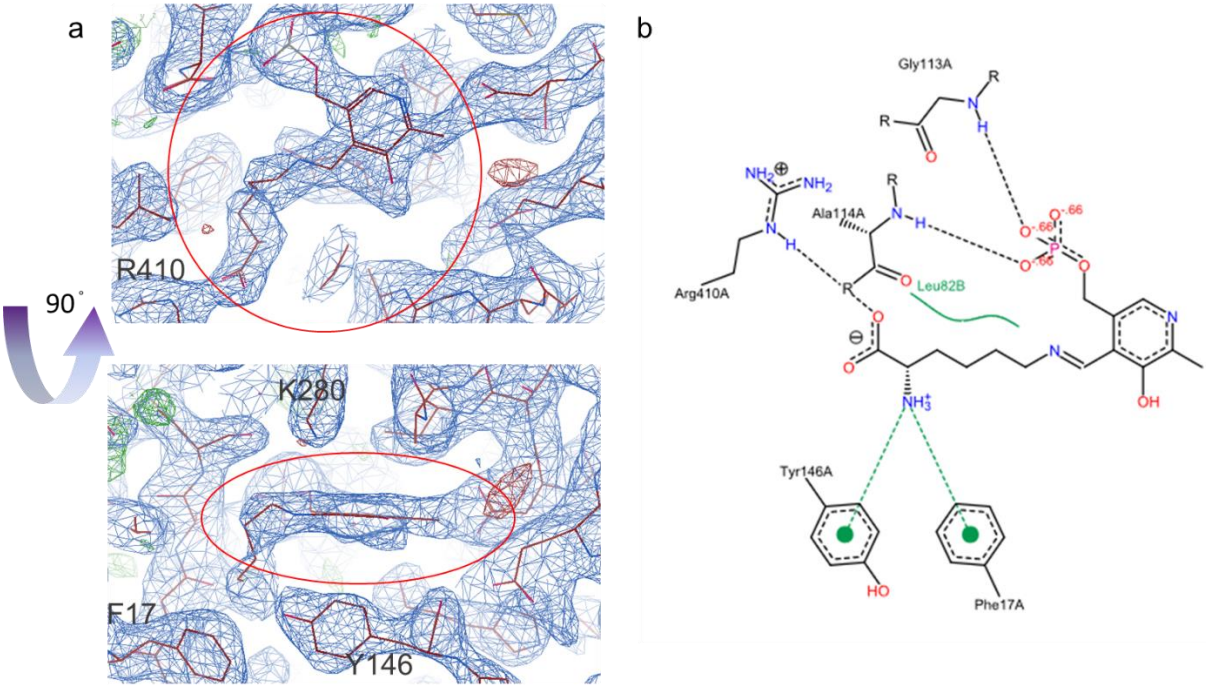


Figure 2.3 PLP-lysine in the BioA active site, pdb ID 6WNN.

a) 3-D depiction of cofactor PLP-lysine with $2F_O$-F_C map at 0.7 σ depicts planar electron density at the pyridoxal aromatic moiety through the ε-amine group. b) 2-D poseview shows the electrostatic interactions between the phosphate group with the backbone, and the charged α-groups of the ligand with the side chains of R410, Y146, and F17. L82 of the opposing B chain complements the ligand alkyl chain.

Discussion

Substrate divergence from homologs

In *B. subtilis*, L-lysine has a K_m close to its endogenous levels, and a turnover rate that is 100-fold less than what is observed in *E. coli*. Both of these factors contribute to a lower catalytic efficiency than Mtb and at least three orders of magnitude lower than *E. coli*. Non-physiological substrates and enzymes that are poorly evolved will contribute to suboptimal efficiencies. However, *B. subtilis* may have evolved to require lower levels of biotin or an efficient means to acquire biotin from its environment. Further, the ability to use lysine may be less energetically costly than synthesizing SAM.

In the active site, *B. subtilis* differs from *E. coli* and Mtb in that residues F17, V53, and L82 correspond to Tyr, Trp, and Gly. Although, beyond the scope of this chapter, it should be noted that single and combinatorial mutants were assayed for activity and binding by the collaborating lab. F17Y had no resounding effect on BioA activity as the structure I report here would demonstrate. Further, a triple mutant maintained 70 % of the native activity (unpublished results). Also, to note, BioA WT and mutants were unable to produce DTB in the presence of SAM, although a change in PLP absorption was observed for the L82G mutant. Consequently, substrate preference is not due to interactions of side chains in the immediate vicinity. Rather, we rationalize that substrate divergence is a consequence of mutations that change transaminase dynamism. Ultimately, this can alter substrate access or the size and shape of the binding pocket.

Transaminase regioselectivity for ω-amine

Transaminases that have a ping pong-bi-bi mechanism belong to subgroup II, fold type I such as E. coli GABA-aminotransferase, and human ornithine aminotransferase (OAT). Amino transferases in subgroup I are no exception as they show a similar mechanism for substrate specificity.[35] These include lysine-ε-aminotransferases in Mtb (UniProtKB- P9WQ77, pdb ID 2CJD), and *Streptomyces clavuligerus* (UniProtKB- Q01767). Both of these enzymes are believed to undergo a switch-type mechanism that selects for the 6-carbon substrate, lysine, in the first reaction. Like homolog OAT, PLP stabilizes a Glu-Arg interaction that is lost upon PMP formation.[36] Weakening of the salt bridge deshields arginine, which then becomes available to bind the 5-carbon substrate, α-ketoglutarate, at the δ-carboxylate for reductive amination at the α-carbon. The switch mechanism is corroborated by kinetic studies on mutants, and the crystal

structures of Mtb LAT in its internal aldimine, external aldimine, PMP, and α-ketoglutarate-bound arrangements.[37]

Here, the switch model is not viable in *B. subitlis* BioA for two reasons- the first being that Glu-Arg is not conserved, but rather correspond to alanine residues at both positions. The second reason this model is disfavored is that the substrate lysine has the same binding mode as KAPA- the substrate in the second reaction. Both form a salt bridge with R410, corresponding to L414 in homologous Mtb LAT. KAPA is a 9-carbon substrate, and the difference between the α-carboxy group and the site of conjugation with the pyridoxal group is seven carbons. This is a 1-carbon extension from substrate lysine. The methyl and amino groups at the KAPA C-8 position may mediate substrate binding mode simply due to spatial constraints. An extension of a ligand chain out-of-plane (as represented in Figure 2.4) may occupy a space that causes a realignment of the backbone or side chain rotamers that interrupt the dimer interface.

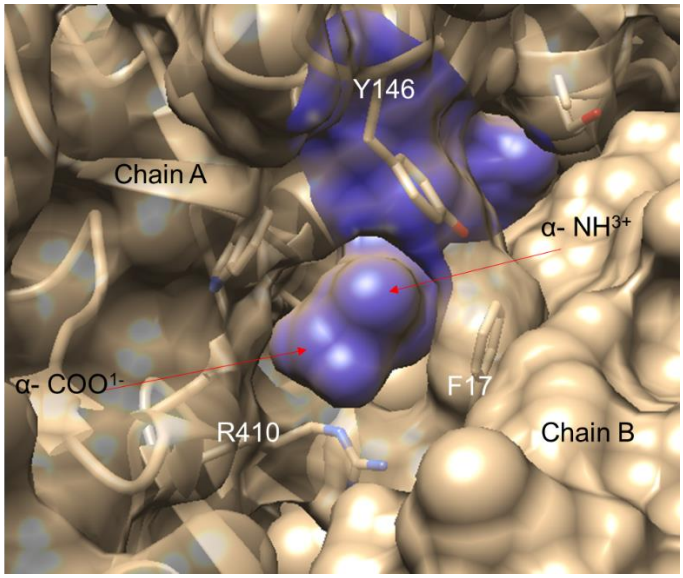


Figure 2.4 Surface representation of the lysyl α-terminus of the LLP ligand (blue) in the active site.

The transparent surface represents chain A of the homodimer.

Proposed mechanism

In the proposed mechanism for the first half of the transimination reaction, the schiff base is formed at the ε-amine of substrate lysine. Lys280 acts as a catalytic base to remove an ε-proton to form the transient quinonoid. Aromaticity is restored upon electrophilic addition at the methine ring substituent. This form is observed by a blue shift in absorption in the absence of KAPA. In the crystal structure reported here, there is no residue within reasonable proximity for a subsequent S_N2 attack at the aldimine. Therefore, the reaction is presumed to occur by a catalytic water.

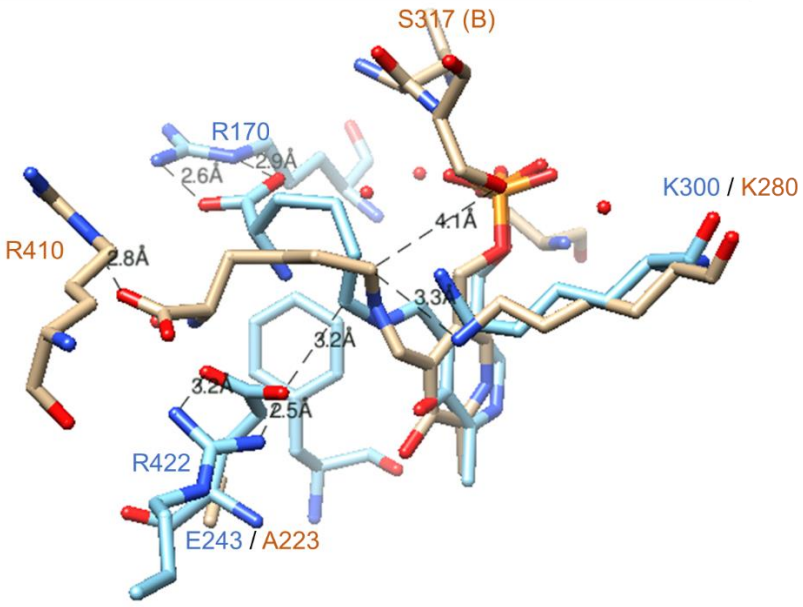


Figure 2.5 Overlay of LLP in *B. subtilis* BioA and Mtb LAT (pdb ID 2CJD) in the active site reveals no residue in proximity that can act as a catalytic base.

The lysine α-carboxylate forms a salt bridge that corresponds to R410 and R170 in BioA (beige) and LAT (cyan). The salt bridge between LAT E243 and R422 mediates the substrate preference for lysine in the first half of the transaminase glutamate switch mechanism. The nearest charged group to the ε-amine is K300 and E243. Proximal residues with polar side chains in BioA are K280 and S317 that are 3.3 Å and 4.1 Å away, respectively. BioA has an alanine at the position of the conserved E243. Waters are represented by red spheres in the LAT crystal structure.

In a $2F_0$-F_c map, there is positive density in the proximity of the electrophilic carbon that correlates well with an overlay of a KAPA-bound structure with waters modeled. However, at 2.7 Å resolution, this is clearly an interpretation that requires a higher resolution structure. The release of allysine product yields the PMP-bound enzyme equipped to complete a single turnover upon KAPA binding.

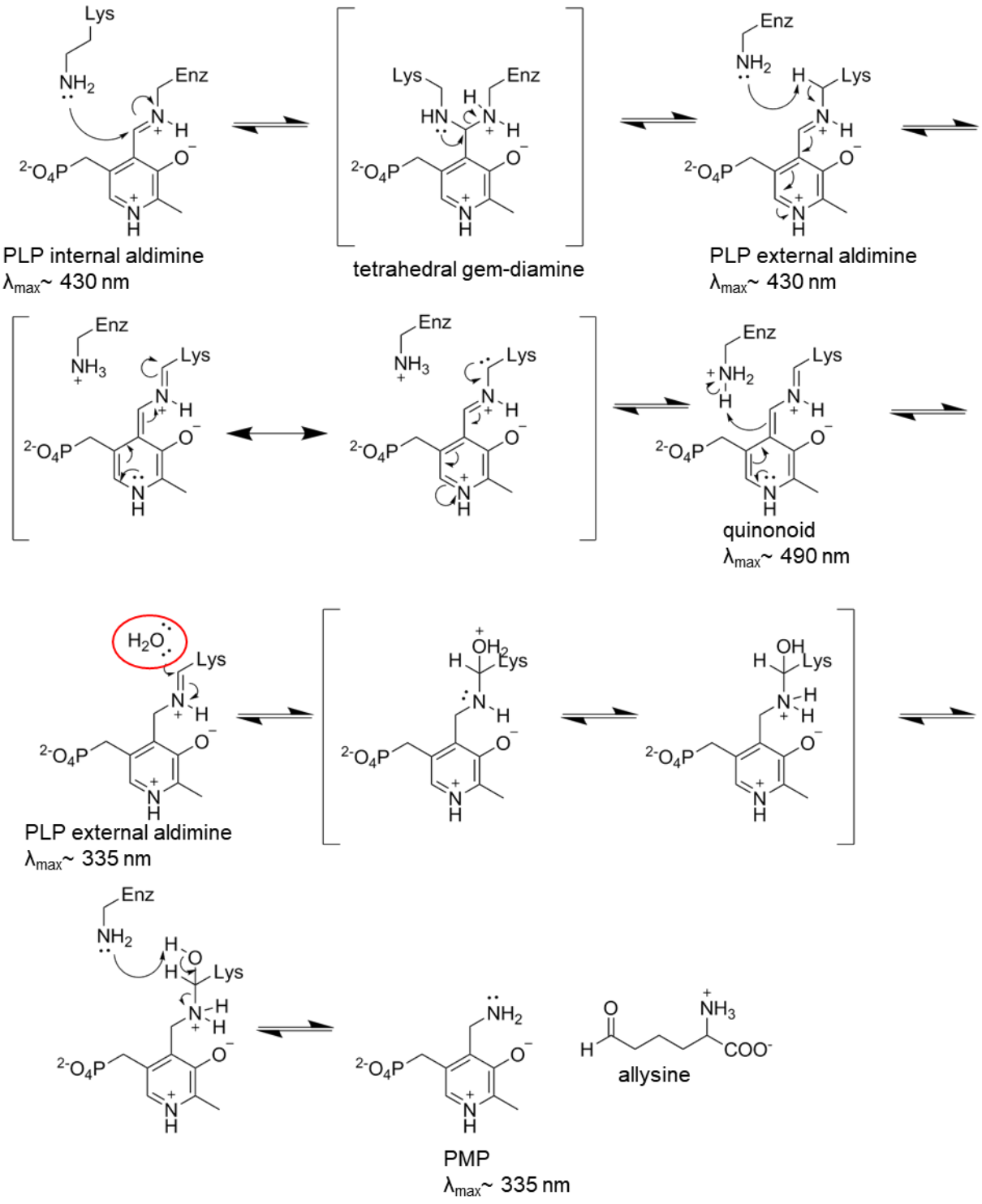


Figure 2.6 Proposed mechanism for the first half of the transimination reaction to furnish PMP.

An S_N2 attack at the aldimine by a catalytic water (circled in red) molecule results in allysine formation and PMP-bound enzyme. λ_{max} of the cofactor intermediate are given below each structure.

As a final thought, the ability to bind PLP has no bearing on whether a productive turnover event will occur. To demonstrate this, I briefly recount work performed by Dr. Cramer from our collaborating lab that is essential to this argument. Single and combinatorial mutations of F17Y, V53W, and L82G had binding affinities, absorption properties, and activities that were unpredictable. For example, lysine exhibited a 1.5-fold higher affinity for an L82G mutant, yet 91% of its activity was loss. The addition of a F17Y mutation decreased BioA's affinity for lysine by greater than 10^3-fold, while activity was rescued with the triple mutant, despite K_d > 100 mM. We believe that active site perturbations may change the torsion angle of the proton at the external aldimine carbon under nucleophilic attack, thereby altering its ability to form PMP (Figure 2.6). As described by Dunathan, H.C., the breaking of a bond periplanar to the p orbitals will result in a carbanion delocalized with the conjugated pi system.[38] Therefore, a low substrate binding affinity should not be diagnostic for low levels of activity. Although the spatial organization of the lysine chain affecting the level of activity according to Dunathan's hypothesis seems definitive, this is clearly speculation without a crystal structure of these BioA mutants. Nevertheless, this project is an example that random mutations in nature may offset others that result in a loss-of-function. In a broader context, this supports the dogma that evolution occurs such that mutations that are beneficial for an organism are selected for.

Chapter 3 - Production of functional GPER for structural studies

A portion of this chapter has been published in Current Protocols in Protein Science.

Samson A. Souza[1], Dane T. Kurohara[2], Chester Dabalos[3], Ho Leung Ng[1]

[1]*Department of Biochemistry and Molecular Biophysics, Kansas State University, Manhattan, KS*

[2]*Department of Biology, University of Hawai‘i at Mānoa, Honolulu, HI*

[3]*Department of Biology, University of Hawai‘i at Mānoa, Honolulu, HI*

In this study, I chronicle our efforts to produce the highest levels of recombinant GPER by *S. cerevisiae*, and a cell-free expression system. In preliminary work, our lab optimized the expression of enhanced green fluorescent protein (eGFP) by a cell-free protein synthesis (CFPS) method with an *E. coli* S12 extract prepared *de novo*. Here I demonstrate that monomeric GPER is most productively expressed and solubilized by precipitation-based cell-free protein synthesis. Furthermore, I provide qualitative evidence that it is produced in a functional state by an ESI-MS method for the detection of bound ligands tamoxifen and G-1 agonist.

Introduction

Endogenous estrogens play a crucial role in regulating cell proliferation, differentiation, apoptosis, and metabolism. They regulate genomic signaling by binding nuclear estrogen receptors (nERs). Nuclear ERs, ERα and ERβ, are classically known to oligomerize at estrogen response elements (EREs)- modulating transcription levels. These receptors include a conserved DNA-binding domain, a C-terminal ligand-binding domain, and an N-terminal domain with a

variable transcription activation 1 (AF1) subdomain. Upon binding the appropriate DNA sequence, transactivation is induced. Proper folding of AF1 directly or indirectly interacts, via cofactors or transcription factors, with transcription activation 2 (AF2) in the ligand-binding domain.[39] However, earlier findings infer a non-genomic role of estrogen that has yet to be fully understood.

Early work suggested that estrogens induced cAMP synthesis and calcium flux in the late 1970s.[40] It was much later in 2002 that expression levels of a 7 transmembrane orphan receptor, GPR30, were correlated with estrogen-mediated Erk1/2 activation and cAMP synthesis.[41] Later, ligand-binding studies revealed that an endogenous estrogen, 17 β-estradiol, bound GPR30 at concentrations 1000-fold less than that of 17 α-estradiol. Its preference for 17 β-estradiol was due to the configuration of both hydroxyl substituents and their positioning in the ligand-binding pocket.[39] GPR30 became appropriately known as G protein-coupled estrogen receptor (GPER). 17 β-estradiol acts as an agonist to ERα/β and GPER, binding at a K_d of 3nm. Tamoxifen, a selective estrogen receptor modulator (SERM), is an ERα/β antagonist and GPER agonist. The first highly selective GPER-agonist, G1, binds at a K_d of 10 μM.[42]

Non-genomic Signaling

Functional studies reveal that GPER is involved in the MAPK- Erk1/2 cascade via transactivation of epidermal growth factor receptor (EGFR), a tyrosine kinase. Upon GPER activation, $G_{\beta\gamma}$ stimulates heparin bound-EGFR (HB-EGF), which binds EGFR, causing dimerization and autophosphorylation at tyrosine residues. Activator proteins recognizing phosphotyrosyl residues activate Ras, which recruits Raf1, a serine/threonine kinase. The phosphorylation cascade leads to activated-Mek-1 that binds substrates Erk1/2.

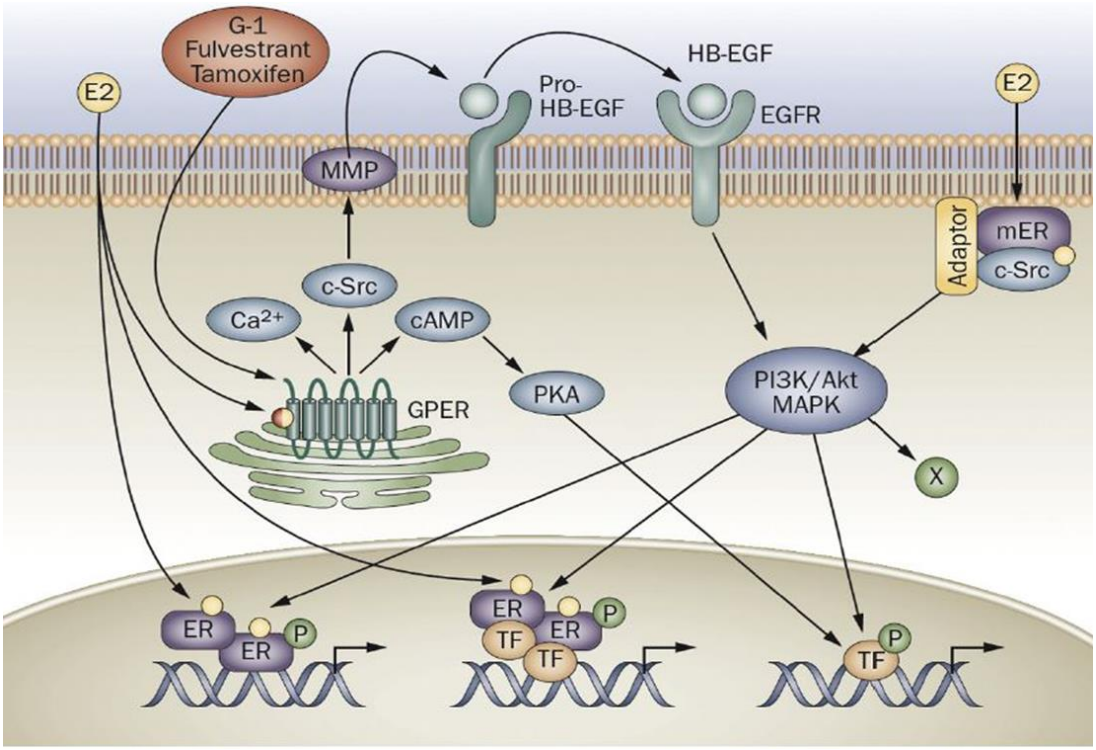


Figure 3.1 Estrogen-related cell signaling pathways.

GPER agonism upregulates cytosolic cAMP, Ca^{2+}, and Src levels. A major downstream effect includes the recruitment of transcription factors at EREs. Reused with permission from Springer Nature.[12]

GPER binding is also known to cause Src activation, which may directly phosphorylate EGFR. However, GPER-binding does not only cause cell proliferation via the MAPK-Erk1/2 cascade, but it initiates inhibitory effects as a means for regulation. GPER-binding triggers adenylyl cyclase (AC) activity by an unknown mechanism. This causes the production of secondary-messenger cAMP, resulting in attenuated Erk1/2 signaling via inhibition of Raf1-dependent protein kinase A (PKA).[41] Studies also suggest GPERs role in calcium mobilization via phospholipase C (PLC) activation with subsequent inositol triphosphate (IP_3) production. This requires transactivation of EGFR, a pathway distinct from ERα-Ca^{2+} mobilization.[43] Regulatory

function of GPER is tissue-type dependent, therefore different cell types have varied expression levels of proteins that respond differently to secondary messengers downstream of GPER, such as cAMP. Intuitively, GPER provides further insight into the complexity endogenous estrogens plays beyond their classical role of mediating nuclear receptor function.

Materials and Methods

Yeast expression

Primers for ligation independent cloning were designed to make N- and C-terminal truncations at codons encoding residues 40, and 353 respectively. A reverse primer complementary to the 3' end of GFP was also designed for cloning a GPER-GFP fusion construct. Primer flanks were made to be complementary to insertion at *Bst XI* restriction site of the pSGP36 expression vector. This vector is compatible for cloning in *E. coli*, containing a C-terminal His_{10} site, and selectable markers- *Ura3*, and ampicillin. Briefly, the vector was linearized by a standard digest protocol with *BstXI* (Fermentas) enzyme. PCR-amplified insert and vector were gel-purified with a QIAquick gel extraction kit (Qiagen). The vector and insert were mixed with T4 DNA polymerase (NEB) at 25 ̊C for 10 minutes, then cooled on ice to facilitate vector-insert annealing. DH5α cells were transformed for plasmid replication and then plated. Colony PCR was performed to confirm the presence of the DNA construct.

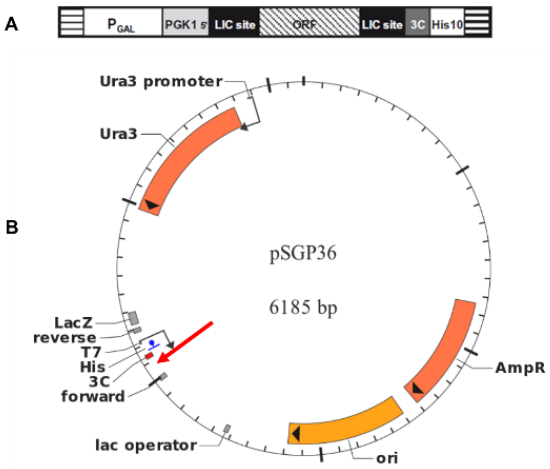


Figure 3.2 Dual system expression vector pSGP36.

A. DNA insertion region. B. pSGP36 plasmid. *BstXI* restriction site (indicated by the red arrow) is doubly digested for linearizing the vector before ligation-independent cloning.

DSY-6 yeast strain was chemically transformed with EZ-Transformation II kit (Zymo Research) for subsequent protein expression. DSY-6 cells are uracil-deficient, requiring orotidine 5' decarboxylase to grow on media lacking uracil, and contain multiple protease-encoding gene deletions. Yeast cultures were grown for 2-4 days at 30 ̊C on SD URA dropout agar-medium with 2% glucose, for selection of transformants. Colonies were picked and precultured in SD URA broth with 2% glucose between 12-22 hours on a shaker at 30 ̊C. Cultures were placed in YPD enriched media in a shaker at 30 ̊C and induced at an OD_{600}~ 0.6- 1.2. Cells were harvested at 4000 x g for 5-10 minutes. Cells were resuspended in buffer containing 50 mM Tris-HCl (pH 7.6), 1-5 mM EDTA, 10% glycerol (0.6 M sorbitol for 1 L cultures), protease inhibitor.[44] Cells were lysed with a glass bead-beater for 7 minutes in 30-second intervals. Unlysed cells and debris were pelleted by centrifugation for 5 seconds at 22,000 x g, 4 ̊C. Pellets were resuspended and subjected to the second round of lysis. Membrane-containing supernatant was pelleted by centrifugation for 60-120 minutes at 20,000 x g, 4 ̊C. Initial trials required preparation of a buffer with SDS and bromophenol blue (SB) to observe expression by in-gel fluorescence. Detergent solubilization

steps were done for 60 minutes on an end-over-end rotator at 4 ̊C. His-tagged protein was batch-purified by incubating in a Ni-NTA resin and subjected to slow agitation on an EOE rotator. The slurry was washed with 25 mM imidazole and transferred to a 0.22 µm spin-column for elution with 250 mM imidazole. Samples were characterized by SDS-PAGE/in-gel fluorescence (for fusion protein), and subsequently transferred to a PVDF membrane for detection by Ni-HRP and TMB substrate.

Cell-free protein synthesis construct design, protein production, and protein purification

pIVEX2.3d expression vector was digested with *NcoI* and *SmaI* restriction enzymes (Fermentas) DNA construct insertion. Primers were designed to include GPER1 residues 40 to 353 with a C-terminal His_6-tag and appropriate 5' *NcoI* and 3' *SmaI* flanking sites. E. cloni *DH5α* cells (Lucigen) were transformed for cloning and plated on carbenicillin-selective media. Transformants with DNA-insert were confirmed by colony PCR. BL21 (DE3) Rosetta-gami 2 cells (Novagen) were used to prepare an S12 cellular extract as outlined by Kim and colleagues.[8] Aliquots of the prepared S12 extract were stored at -80 ºC for use in the cell-free reaction mix.

A mastermix was prepared to aid the S12 translational machinery. Components of the mixture were pipetted in the order outlined in Table 3.1. The amino acid mix contained all naturally occurring amino acids, except Tyr, Cys, Pro, Gln, and Ser, due to differences in their concentration requirements. Each cell-free reaction was prepared for a 75 µL reaction volume. The appropriate amount of Millipore water was pipetted into each reaction vessel with the mastermix. S12 extract was pre-warmed at 30 ºC and pipetted to achieve 40 % (v/v). T7 RNA polymerase (prepared in-house), and tRNA solution was pipetted to a final concentration of 43 µg/mL and 0.16 mg/mL, respectively. Reactions were initiated by introducing ~ 1 µg of DNA per

75 µL of reaction. Reactions were incubated overnight at 30 ºC, 200 rpm in flat-bottom microcentrifuge vessels.

Table 3.1 Pipetting scheme for the preparation of the cell-free 2X mastermix.

Reagent	*Stock (mM)*	*Final Concentration (mM)*	*Volume (µL) 250 µL 2X*	*Volume (µL) 400 µL 2X*	*Volume (µL) 500 µL 2X*	*Order*
Bis-tris methane	2000	52.5	13.1	21	26.3	1
$Mg(OAc)_2$	800	12.5	7.8	12.5	15.6	3
Potassium glutamate	4500	212	23.6	37.7	47.1	4
Amino acid mix	50	3	30	48	60	5
Tyrosine	50	1	10	16	20	6
Cysteine	50	1	10	16	20	7
Glutamine	50	4	40	64	80	8
Serine	50	2	20	32	40	9
$NH_4(OAc)$	8000	27.4	1.7	2.7	3.4	10
NTP mix	50	0.85	8.5	13.6	17	11
ATP	120	1.2	5	8	10	12
cAMP	100	0.64	3.2	5.1	6.4	13
Creatine phosphate	2000	61.3	15.3	24.5	30.7	14
DTT	1000	1.7	0.9	1.4	1.7	15
CK	20 mg/mL	125	3.1	5	6.3	16
Folinic acid	17	68	2	3.2	4	17
Total (all except water)	-	-	**194.2**	**310.7**	**388.4**	-
Type 1 water	-	-	55.8	89.3	111.6	2

In the D-CF method, detergent was added to the cell-free reaction mix prior to induction. The detergents assayed by this method were n-dodecyl-β-D-maltoside (DDM), Brij 35, Brij 58, and Brij 78. After the overnight incubation, reaction mixtures were centrifuged at 16,000 x g for 20 minutes to pellet insoluble debris. In the P-CF method, the reactions were pelleted at 5,000 x

g for 5 minutes. The protein-containing pellet was washed 3 times with 75 µL P-CF buffer (15 mM NaH_2PO_4 (pH 6.8), 1 mM DTT, 0.5 mM PMSF). After the final wash step, the pellet was resuspended in loading buffer for analysis, or solubilized with P-CF buffer and 2% LMPG [1-myristoyl-2-hydroxy-s*n*-glycero-3-phospho-(1'-*rac*-glycerol)] for 1-hour at room temperature on an end-over-end (EOE) rotator. Subsequently, the solution was centrifuged to pellet insoluble debris, and the supernatant was retained for purification.

Characterization

Samples were loaded onto SDS-PAGE gels for western blot analysis. Ni-HRP (Thermo Scientific) probed for levels of His_6-tagged protein. Rabbit anti-human GPR30 polyclonal antibodies with goat anti-rabbit IgG secondary antibody with conjugated HRP confirmed the presence of Ni-purified GPER by CFPS for our ligand-binding study. 3,3',5,5'-tetramethylbenzidine (TMB) was used as a substrate for colorimetric detection.

Proteolysis of unfolded protein

The fraction of folded protein during the course of the CFPS was determined by initiating the reaction at different times. The samples were pelleted and subjected to 3 wash routines with buffer A (20 mM HEPES, 10 mM $CaCl_2$, and 1 mM DTT). Samples were solubilized with buffer B (buffer A, and 2% LMPG), and incubated with 0.05 mg/mL of thermolysin (TL) at 37 ºC. Samples were quenched with 20 mM EDTA and loaded onto an SDS-PAGE gel for western blot analysis with 1º and 2º antibodies referenced in the materials and methods section. Relative band intensities were analyzed with unScan-it-gel software. Experimental controls included a 70 ºC heat denaturation step for 18-hour reactions in the absence and presence of TL.

Assessing 17-β estradiol-GPER interaction by an HPLC based method

To measure estradiol-GPER ligand binding affinity, we employed an HPLC-based method. In short, free ligand and bound ligand would be separated by a 10 kDa and 50 kDa cut-off spin filter. The flow-through and retentate were applied toa BDS Hypersil C18 column (30 X 3 mm, 3 µM) to separate reaction components that include an internal standard. At 0.7 mL/min, and an isocratic elution with mobile phase (50 % 10 mM KH_2PO_4, 40 % ACN, 10 % MeOH), estradiol retention time was 0.83 min. The limit of detection and quantification were 7.5 ng and 20 ng on-column, respectively. Linearity was maintained between 20 – 10,0000 ng.

Ligand-binding study with GPER agonists

Ni-NTA resin was equilibrated with 10 mM imidazole in either P-CF buffer + detergent, or buffer B. Cell lysates were incubated with equivolume Ni-NTA slurry at 4 ºC for 1 hour on an EOE rotator. Samples were transferred to 0.22 µm mini spin columns for purification. Columns were washed with buffer at 25 mM imidazole and eluted with 200 mM imidazole. 0.05 mg/mL of ligand was incubated with elutes on ice for 5 minutes and at 25 ºC for 1 hour on an EOE rotator. Ligands used in this study included tamoxifen (4HT) (Tocris Bioscience), and G-1 GPER agonist (Azano Biotech). A size-exclusion chromatography (SEC) method was employed to separate free and protein-bound ligand. Briefly, a mini spin-column was loaded with 150 mg of prepared Bio-Rad P6 resin (P6), and primed with P-CF buffer + detergent. 20 µL of the sample was applied to the top of the resin bed, and the spin columns were centrifuged for 2 minutes at 1,000 x g. A 4.6 mm x 150 mm Phenomenex C_{18} column (5 µm particle size) in tandem with an Agilent 6410 QQQ MS detected bound ligand at 280 nm. Components in each sample were separated by a 15-minute

elution profile at 0.7 mL/min. A gradient elution profile was employed for the first 8 minutes at 10 – 100% ACN (H_2O, 0.1% HCO_2H/ ACN). Experimental controls included sample loading of the free ligand in the buffer (positive control), and SEC elutes of free ligand and free protein (negative controls).

Results

Expression of GPER-GFP fusion in yeast

The presence of chaperone-additive DMSO or His after induction was reported to increase whole cell-expression levels of transmembrane proteins in yeast by up to 30 percent.[44] 2.5% DMSO in growth medium during induction improved GPER expression levels. In all trials, uninduced cells produced fusion protein, suggesting Gal 1 promoter leakiness. A blot analysis with a histidine (His)-tag probe detected protein in the solubilized lysate at approximately 61.5 kDa and 70 kDa. Bands larger than 61.5 could be glycosylated protein, resulting in 8 – 10 kDa gel shifts.

We solubilized GPER with 1% DDM. However, cell membrane disruption with a bead beater was product limiting. Further, during nickel purification, the fusion protein was degraded, apparent by SDS-PAGE and fluorescence of sample fractions. A negative control, in the absence of the open reading frame (ORF), was used to reference background fluorescence and endogenous contaminants in the system. It yielded no fluorescence or degraded bands of GFP. Figure 3.3 illustrates fusion protein loss with each successive wash, and nearly no fluorescence in fraction elutes.

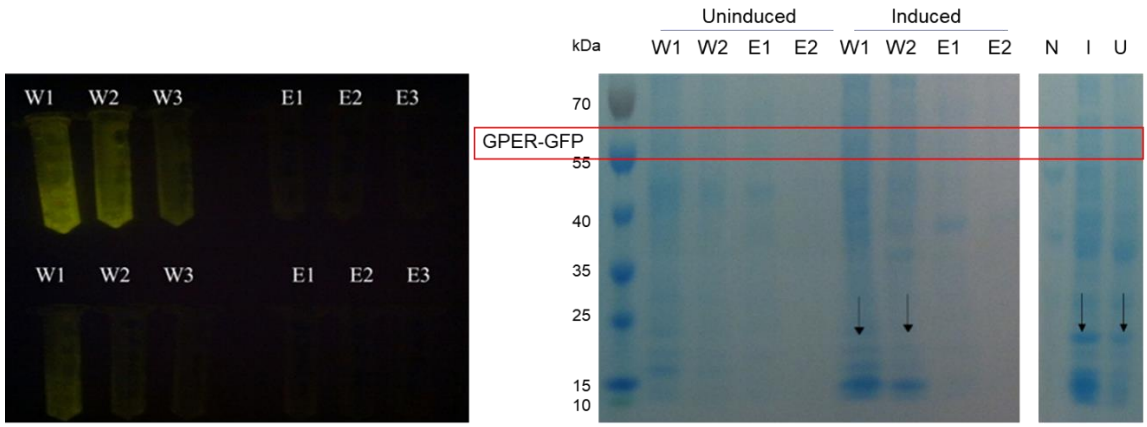


Figure 3.3 GPER expression in yeast is observed by GFP fluorescence of wash and elution fractions during Ni-purification and an SDS-PAGE gel.

Fluorescence in the top and bottom row (left) corresponds to induced and uninduced cultures. Consecutive 1 mL wash (W) and elution (E) fractions are numbered. SDS-PAGE shows proteolysis with a faint band above the 55 kDa marker that corresponds to the 61.5 kDa GPER-GFP band in the appended western blot. The 50 kDa band may be proteolyzed protein. Yeast expression of the plasmid without the ORF, and the supernatant of the induced (I) and uninduced (U) cultures prior to Ni-purification are provided (far right). Black arrows indicate GFP and its functional degraded products.

Expression and solubilization by CFPS

CFPS by the D-CF method did not yield detectable levels of GPER by western blot analysis. However, GPER expression was upregulated in the absence of detergent in the reaction mixture. Solubilization with SDS buffer produced a band at an apparent mass of 27 kDa. Membrane proteins have varied ion concentration requirements for cell-free protein synthesis, and earlier work demonstrated that GPER expression levels were optimal at 212 mM potassium, and 18 mM magnesium (Figure 3.4).

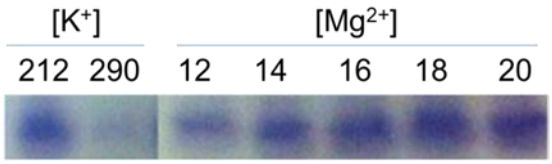


Figure 3.4 Blot of GPER expression levels at a varied concentration (mM) of potassium and magnesium.

Cell-free reaction mixture with 212 mM potassium glutamate and 18 mM magnesium acetate yielded the highest levels of protein expression by band intensity.

Solubilization with 2 % LMPG produced a band intensity for monomeric GPER comparable to solubilization with 1 % SDS.

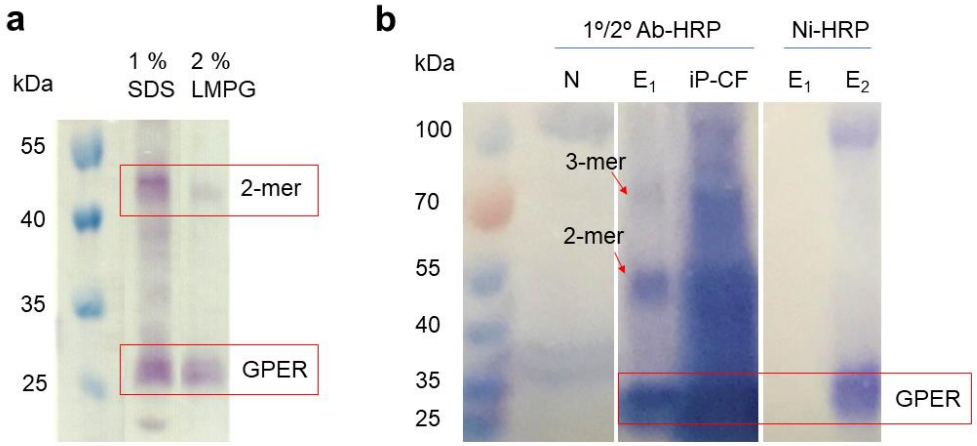


Figure 3.5 Detergent solubilized GPER detected by western blot analysis.

a) Detection by Ni-HRP. GPER solubilized by P-CF mode with detergents SDS, and LMPG yields monomeric and dimeric forms at apparent masses 27 kDa and 54 kDa, respectively. b) Ni-HRP detection indicates Ni-purified GPER elutes in the second fraction, while a GPER antibody detects monomeric, and oligomeric (red arrows) forms in the first elution, and iP-CF fraction (N- P-CF of reaction mixture without DNA template, iP-CF- insoluble P-CF fraction, E_1- first Ni-purified elution, E_2- second Ni-purified elution).

Proteolysis of unfolded protein

An 18-hour cell-free reaction time yields the highest expression levels, although nearly 70 % is in an unfolded state evidenced by TL proteolysis. A 6-hour reaction time yielded an intense band after TL treatment, and nearly all solubilized protein existed in a folded state.

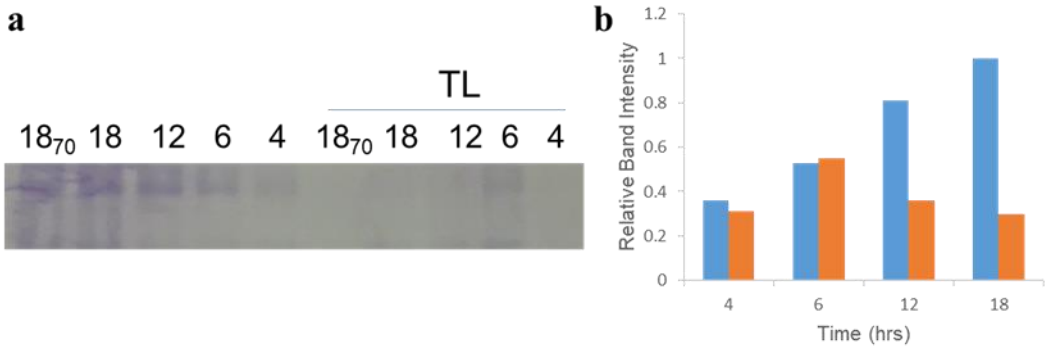


Figure 3.6 GPER band intensity before and after thermolysin treatment at varied cell-free reaction time.

a) Western blot and b) relative band intensities normalized to an 18- hour reaction time. Control samples subjected to a 70 ºC heat denaturation step are indicated by the subscripts. The absence of a band from TL treatment after exposure to 70 ºC indicates that there are no detectable levels of GPER in its native state.

Ligand-binding study

Tamoxifen was incubated with the receptor in the P-CF buffer for the initial binding study. It eluted from the RP-LC column with a retention time of 8.3 minutes at m/z 372.30 (see Appdx C) Incubation of receptor-ligand in HEPES buffer resulted in detection at 8.8 min. GPER-bound G1 agonist eluted from the column at 11.2 minutes with expected mass to charge ratios at 412 and 414 at a 1:1 peak intensity expected of ^{79}Br and ^{81}Br natural abundance. Positive and negative controls are described in the methods section. Extracted ion chromatograms are appended (App. B).

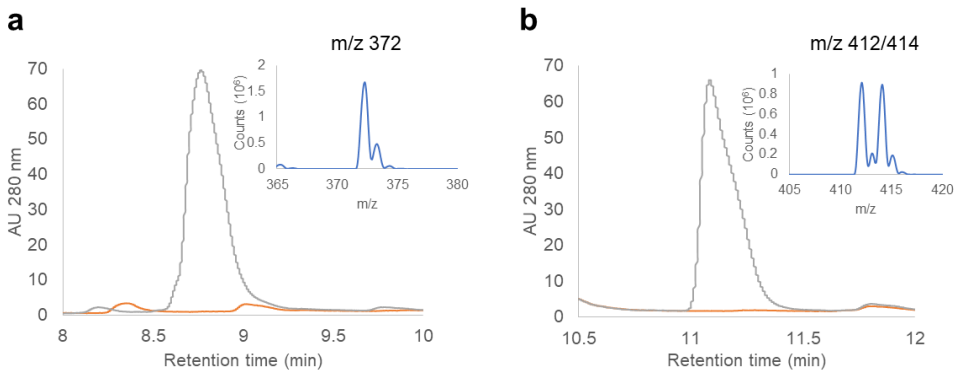


Figure 3.7 Absorbance overlay spectra of negative control (yellow), and bound ligand (grey) at 280 nm with insets of ligand m/z.

GPER-bound tamoxifen (a) and G1 agonist (b) elutes at 8.8 min and 11.2 min, respectively.

Discussion

Yeast expression was successful in that 40-353-GFP-His_{10} was produced since GFP was introduced as a C-terminal fusion. However, I will highlight the direction our lab was moving towards, and unable to complete in working with *S. cerevisiae* DSY-6 cells expression system. First, we sought to show functionality by including a hemagglutinin tag for us to view protein localization in the cell. We also planned to push our sample through a column with estradiol conjugated resin. The second opportunity we planned to move forward with yeast expression was to ramp up expression levels by induction with an ADH2 promoter. Although this has yet to be assayed, it appears that the Gal 1 promoter may be repressed by minute levels of glucose in the media.

The highest expression levels of GPER was achieved by producing it by a cell-free protein synthesis method. This involved minimizing components of the cell-free process that would

hinder translation. The build-up of inorganic phosphates from the energy source sequesters magnesium ions that aid in translation. Excess potassium and the presence of detergents during the reaction too hindered GPER production. We also found that GPER synthesis is unproductive beyond an 18-hour reaction time. In one experiment, a reaction time that extended beyond 6 hours resulted in a greater fraction of the proteolyzed receptor upon exposure to TL. TL selects for hydrophobic residues that are buried in the protein core, and proteolysis in its presence can be an indicator of unfolded protein. Although we have not assayed GPER expression levels beyond an 18-hour reaction time, protein denaturation would come as a trade-off to producing any appreciable increase in expression level.

The second portion of this project was to provide support for the recombinant expression of GPER in a functional state. In our earlier work, we evidenced that the endogenous ligand, 17-β estradiol, had a proclivity to partition into detergent micelles needed for solubilization. In a ligand-binding assay with an Estradiol EIA Kit (Cayman Chemical), estradiol was detected after overnight incubation with Ni-purified protein (App. B). Two negative controls were assayed in parallel. One negative control involved incubation of estradiol in a cell-free reaction in the absence of plasmid, and the other control was estradiol in the solubilization buffer. The latter control resulted in a positive read for estradiol binding. This is the result of ligand sequestration in LMPG micelles confirmed by the HPLC-based detection method highlighted in the materials and methods section. Negative controls (ligand and solubilization buffer) contained ligand in the retentate of spin columns with a 10 kDa and 50 kDa MWCO (App B). LMPG micellar size is approximately 44 kDa,[46] Therefore, estradiol is not a viable ligand to measure binding affinity when separating free and bound ligand by size. Although I do not present this data in the results section, the chromatogram is supplemented. Ligands 4HT and G1 were used to observe GPER functionality

because both ligands are well-documented, commercially available, and had the added benefit of easy detection by an LC-ESI-MS method. As we highlighted earlier, G1-agonist and 4HT exhibited signature m/z peaks that aligned with sample injections of ligand alone. Although we were unable to demonstrate binding of the endogenous ligand, if GPER is produced by this method on a preparative scale, and the appropriate controls are in place to correct for ligand sequestration, 17-β estradiol can be derivatized with 4-(dimethylamino)benzoyl chloride (DMABC) in excess.

This study suggests that GPER can be functionally produced by a CFPS method. However, quantifying ligand binding affinity was difficult due to the complexities that are intrinsic to the recombinant expression of GPCRs. They are unstable with a relatively short half-life *in vitro*, they are produced at low levels, and they require an amphipathic environment to maintain solubility. Earlier in this study, we attempted to show functionality by thermal and chemical denaturation assays. They were simplistic approaches that were quantified by measuring ligand-induced stabilization and densitometry. These included methods of Pulse Proteolysis and Fast Parallel Proteolysis described by David P. Minde[47] and Yu-Ran Na,[48], respectively. These methods make use of TL to cleave exposed residues, in the same way described earlier. and intact protein as detected by densitometry. Unfortunately, any noticeable ligand-induced stabilization of a detergent-solubilized GPCR would require homogenous sample preparations and instrumentation with high sensitivity. Moving forward, a viable method to quantify the binding affinity of steroidal ligands with GPER (and other lipophilic ligands with membrane proteins) - and currently used for high-throughput screening- is saturation transfer difference (STD) NMR.[49] In short, protons are first irradiated in a region -1 to 0 ppm. Spin diffusion will cause bound ligand to have resonances. The difference of an on and off-resonance would return a ligand signature if binding occurred.

The main limitations, though, include maintaining a stable protein for more than hours at a time and utilizing ligands with a moderate binding affinity ($K_d < 10^{-8}$ M).

We thank the Hawaii Community Foundation, the Victoria S. and Bradley L. Geist Foundation, the University of Hawaii Undergraduate Research Opportunities Program, and the National Science Foundation CAREER Award 1833181 for funding this research.

Chapter 4 - Modeling CYP19-P450 reductase interactions for use as a tool to modulate aromatase activity

Introduction

Bacillus megaterium P450-reductase fusion protein (P450BM3) is the paradigm for Cytochrome P450 (P450s) - CytochromeP450 reductase (CPR) interactions since its crystal structure determination in 1999 by Thomas Poulos. Much like human CPR, P450BM3 has an FMN/FAD reductase domain that interacts with the heme-binding domain in a way that facilitates electron transfer. The alternative P450 redox partner is a ferredoxin that – in mammals – is only found in the mitochondria.[50] The crystal structure of P450cam-putidaredoxin reductase together with mutagenesis studies indicate that arginine/lysine residues in the proximal heme site form important ion pairs that can be correlated to human P450-CPR coupling.[51,52] Currently, these interactions are relevant since there is a mounting interest in modulating P450s at its coupling interface. Currently, our understanding is limited to models and kinetic studies that have inherent drawbacks. Although P450s carry a very conserved fold, they have lower than a 30 % sequence identity across families. Further, a growing body of evidence suggests that P450s can form ternary complexes as a hetero- or homodimer.[53] CPR-P450 coupling is also complicated by the enzyme's ability to bind cytochrome b5 for completing a single turnover of molecular oxygen. There is no definitive answer on what interactions are necessary for P450 and cytochrome P450 reductase (CPR) coupling, except that it is likely driven by the negative potential in the FMN domain affinity for the positively charged proximal domain. From this, I draw upon past work, and our current work to assess the interactions between aromatase (Cyp19) and CPR.

The active site of cytochrome P450 enzymes is synonymous with the distal heme site since the chemistry occurs at the axial site opposite the residue that coordinates the iron. Conversely, the solvent exposed pocket on the opposing heme-face is defined as the proximal heme site. In Cyp19, it was proposed that K440 may build a positive electrostatic potential at the proximal heme site of CYP19 required for forming higher-order structures.[20] It was also suggested that electron shuttling in CYP19 occurs through the proximal heme site by way of a backdoor, solvent access tunnel. In a K440Q mutant, the enzyme formed lower-order oligomers and abolished activity.[19] This was an indication that the proximal heme site was involved in forming higher-order structures, and that a K440 mutant either compromises heme integrity and/or agitates CPR coupling.

In this work, the coupling interface of CPR and Cyp19 is predicted by an automated docking server. These interactions and the interactions of select P450s- that are major players in drug metabolism- are then modeled from a prediction-driven method. I will then illustrate that this interface can be targeted for allosteric regulation by utilizing endoxifen as a model molecule. It is an active metabolite of a clinically-used breast cancer therapeutic, tamoxifen, and it inhibits aromatase non-competitively.[54]

Materials and Methods

CYP19 (pdb ID 4KQ8), CYP2C19 (pdb ID 4GQS), CYP2D6 (pdb ID 2F9Q), and CYP3A4 (pdb ID 1TQN), truncated FMN domain of CPR (pdb ID 3QE2), and human-yeast chimeric CPR (pdb ID 3FJO) were selected for this study. Conserved surfaces were calculated by Bayes' theorem in ConSurf since maximum likelihood calculations yielded unreliable conservation scores. The maximum sequence identity cutoff among homologs was set at 30%, and the minimum cutoff was set at 10 %. DoGSiteScorer webserver was used to assess the druggability of all predicted Cyp19

binding pockets. Like other structure-based druggability prediction programs, it generates geometric and physicochemical descriptors that return potential binding pockets with corresponding drug scores. Other programs that perform similar tasks include DLID, MAP_{POD}, Fpocket, and SiteMap. Drug scores returned from these programs correlate well because their scoring functions are weighted on hydrophobic contributions within each predicted pocket.[55] Generally, a pocket that is more enclosed, strongly hydrophobic, and weakly hydrophilic in nature is druggable.[56] DoGSiteScorer distinguishes itself in that it is available as a webserver to new protein targets.[57] It develops a support vector machine (SVM) model that projects the druggability scores of binding pockets and subpockets with an 88 % accuracy.[57]

In this study, it should be noted we assume that the P450 acceptor protein couples to CPR as a heterodimer. We do not account for the possibility that there are functional higher-order heterooligomers. Prior studies show that P450 protein complexes form a dynamic equilibrium to modulate activity and substrate specificity.[18]

Automated docking

Initial docking of Cyp19 and human-yeast chimeric reductase was performed with ClusPro[58–60]. ClusPro was selected for its easy-to-use interface, only requiring two pdb files as input for the protein-receptor and protein-ligand. Furthermore, it clusters low energy complexes according to preset weighting factors represented as four different models. These models represent balanced, electrostatic, hydrophobic, and electrostatic + Van der Waals options described by the algorithm

$$E = w_1 E_{rep} + w_2 E_{att} + w_3 E_{elec} + w_4 E_{DARS},$$

where E_{rep}, E_{att}, E_{elec}, E_{DARS} represent repulsive, attractive, electrostatic, and desolvation contributions.[58] Weighting coefficients are denoted w_n of each term.

Prediction-driven docking

The Haddock webserver was used for a data-driven approach to model the protein-protein interactions between cytochrome P450 reductase and the P450 enzymes under study. The CPORT (Consensus Prediction Of interface Residues in Transient complexes) algorithm[61] was used to predict active and passive interfacial residues as part of the parameters in Haddock to return high scoring clusters. CPORT combines predictions from WHISCY, PIER, ProMate, cons-PPISP, SPPIDER, and PINUP. It proves to match or outperform the interface predictions from each stand-alone server.[61] In this study, residues that were solvent-exposed and surrounded by active residues were selected as passive residues (Table 1) for restraints input (referred to as ambiguous interaction restraints or AIRs) in Haddock. The Haddock program is available as a user-friendly webserver that performs at acceptable levels according to CAPRI (Critical Assessment of Prediction of Interactions) evaluations. CAPRI is a communitywide effort to assess protein-protein docking algorithms from blind predictions.[62] In addition to its easy-to-use interface, Haddock uses AIRs that allow for structural flexibility of the protein side chains and backbone.[63] In this study, the Easy Interface was used with automated semi-flexible segment definitions. The CPR chain B (pdb ID 3QE2) was used for the initial docking studies against Cyp19 in Haddock.

The CPR hinge region is intrinsically dynamic, causing the C-and N-terminal domains to adopt an open conformation.[64,65] Therefore, the P450s explored in this study were docked against the crystal structures of the human CPR FMN domain, and a human-yeast chimeric enzyme (pdb ID 3FJ0) in an open conformation. A sequence alignment of the yeast FMN domain showed a

high sequence similarity and a high conservation of secondary elements and tertiary fold with human CPR. Kinetic assays show that the chimeric protein maintained its ability to reduce cytochrome c, and human P450s.[66] Coordinates for the water molecules and the heme group were removed from all structures prior to docking in Haddock. The scoring function included desolvation, electrostatic, Van der Waals, and restraints energies of 200 surveyed samples. Similar models are then clustered into groups. A lower mean value equates to a higher scoring cluster.

Table 4.1 Residues selected for prediction-driven docking in Haddock.

Cytochrome P450	*Active residues selected for prediction-driven docking*	*Passive residues selected for prediction-driven docking*
CPR (B chain)	181, 193, 194, 196	144, 145, 160, 161, 163, 164, 177, 178, 179, 180, 182, 183, 184, 187, 188, 197
CPR (B chain FMN domain)	87, 89, 90, 91, 142, 143, 144, 145, 146, 147, 148, 150, 153, 156, 177, 178, 179, 180, 181, 182, 183, 184, 186, 189, 190	92, 93, 95, 96, 99, 114, 116, 117, 118, 119, 154, 157, 158, 160, 161, 193, 194, 197, 206, 208, 210, 211, 212, 213, 218
CPR (3FJ0, Chimera)	117, 119, 120, 153, 157, 158	118, 122, 123, 131, 154, 155, 156, 159, 160, 162, 165, 166, 184, 186, 187, 188
CYP19A1	149, 150, 151, 153, 154, 424, 426, 429, 430, 432, 433, 437, 440, 441	101, 104, 105, 108, 109, 145, 146, 147, 155, 156, 157, 158, 159, 202, 276, 281, 361, 364, 422, 423, 425, 434, 438
CYP2C19 (A chain)	125, 128, 129, 131, 132, 134, 135, 425, 435	121, 122, 133, 137, 138, 139, 140, 422, 423, 424, 427, 430, 431, 433, 434, 436, 437, 438, 439, 441
CYP2D6 (A chain)	436, 445, 446, 447, 449, 450	140, 427, 429, 433, 434, 435, 437, 438, 441, 442, 443, 444
CYP3A4	138, 139, 140, 434, 435, 442, 443, 444, 445, 446, 448, 449	130, 134, 135, 141, 143, 144, 351, 357, 361, 432, 436, 441, 447

P450-CPR binding affinity prediction

PRODIGY prediction web server was used to generate the Gibbs free energy and binding affinity of all protein-protein complexes. Values are generated from a model with optimized weighting factors of terms that include interfacial contacts and non-interacting surface (NIS)[67] contacts.[68,69] Apolar-polar, apolar-charged, and apolar-apolar inclusion as contributors to the interaction energy increased PRODIGY prediction accuracy with 1.89 kcal/mol RMSE of both flexible and rigid complexes.[68]

Docking endoxifen

Docking for E- and Z- endoxifen against substrate-free Cyp19 (pdb ID 3S79) was performed in YASARA-Vina software package with AMBER15IPQ forcefield. The lowest energy models from 25 docking runs were assessed for the best-fit binding mode.

Molecular graphics visualization

All molecular graphics were generated with Chimera software.[34] Figure 4.2 and electrostatic potential maps with APBS plugin were created in PyMOL.[70]

Results and Discussion

The proximal heme site has a positive electrostatic potential conserved among P450s. This is believed to drive reductase coupling, especially since class II members across the P450 superfamily have less than a 40 percent sequence identity.[71] A subject query with Cyp19 against 150 homologs yielded four conserved facial domains identified in Figure 4.1 as Site 1 - 4.

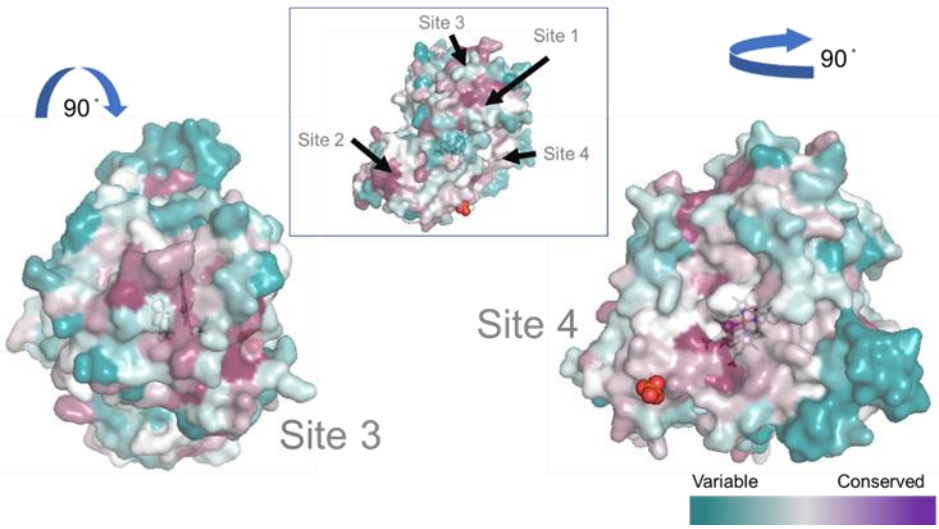


Figure 4.1 Surface representation of variable and conserved regions of aromatase (pdb ID 4KQ8) as predicted in ConSurf web server.

A color gradient from teal to magenta represents residues among P450 enzymes that are variable to conserved. Variable regions are solvent-exposed, while conserved residues are at the enzyme core. There are four conserved facial sites (sites 1-4) identified. Site 3 at the top of the protein relative to the structure boxed in blue, and site 4 is the proximal heme site.

Sites 2 and 4 are predicted by CPORT to contain residues that actively participate in protein-protein interactions. Site 2 corresponds to a portion of the structure that is membrane-embedded and spatially near the N-terminal transmembrane segment. Atoms from the facial N-terminal residues of the A' helix and β1-2 strands of the major sheet make up this surface. Site 4 is the proximal heme site containing basic and aromatic residues. It is ranked second to the active site in its druggability score as projected by DoGSiteScorer. Site 5 was ranked third in its druggability score, albeit there was a difference of 0.01 with site 4.

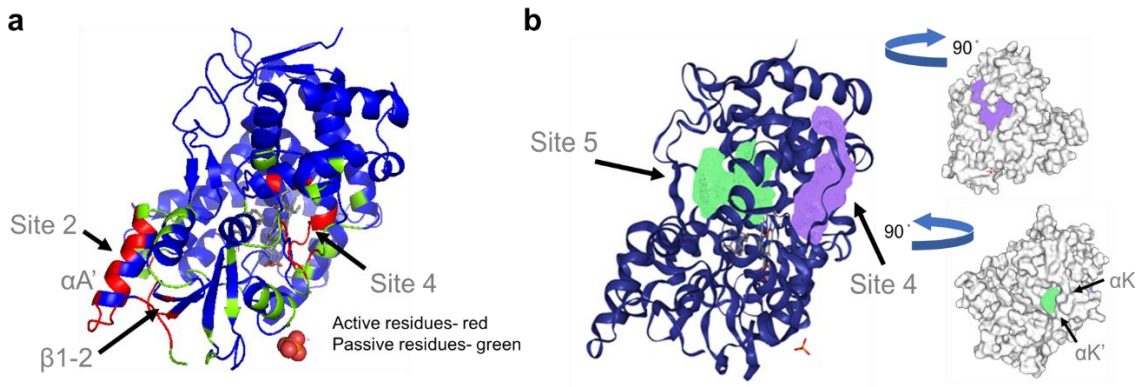


Figure 4.2 Ribbon diagrams of Cyp19 (pdb ID 4KQ8) that indicate sites projected to be protein interfaces (a), and druggable pockets (b).

a) Residues predicted to, directly and indirectly, participate in protein-protein interactions are highlighted in red and green, respectively. Site 2 is the N-terminal, membrane-associated region that includes A' helix, and β1- 2 strands of the four-stranded major β sheet.

The simple score reflects the disparity between the geometry of sites 4 and 5. Site 4 has a larger volume, greater surface area, and more solvent exposure- the three terms that make up the simple score function (Table 4.2). This infers that site 5 has physicochemical properties that offset these geometric terms to enhance the druggability score. Conceivably, pocket accessibility is likely to be restricted, and contingent on the plasticity and dynamism of helices K and K' that make up the pocket enclosure.

Table 4.2 Drug score and geometric terms of the three top-ranked druggable pockets.

Pocket	*Drug score*	*Simple score*	*Volume (Å^3)*	*Surface area (Å^2)*
Active site	0.86	0.53	836.16	801.03
Site 4	0.78	0.38	516.67	782.82
Site 5	0.77	0.15	393.02	328.69

The top 10 clusters returned from each of the four default models in ClusPro were initially observed to find the interactions that are most important for enzyme coupling without bias. Clusters considered viable were those that were oriented such that the FMN cofactor of CPR lay

at the interface of the proximal heme site. The balanced and hydrophobic models did not produce any feasible options, while the electrostatic model yielded 1 viable cluster. The model with weighted terms that favored electrostatic and Van der Waals contributions produced 9 viable clusters (App. B). However, three of these clusters were not likely to represent productive coupling because the distance between FMN and heme was too lengthy. The threshold for physiologically-sound electron transfer rates in redox proteins is 14 - 15 Å.[72] Through-bond tunneling of the P450-reductase archetype, P450BM3, was refuted since the theoretical rate constant is 0.02 s^{-1}.[72] As such, practical clusters were those with closer inter-cofactor distances. From this, Van der Waals and electrostatic interactions likely determine whether a productive enzyme-coupled event will occur.

In prediction-driven docking for the Cyp19-FMN domain, cluster 2 yielded the top Haddock score (Figure 4.3). When Cyp19 was docked against the chimeric reductase, the cluster size dramatically dropped to 6 (cluster 7). Further, the interface RMSD disparity from that of the other clusters was much more evident than the Cyp19 docking results against the FMN domain. This could be an indication of more reliable predictions. Alongside this, more stringent restraints (lesser active residues as input) would be expected to show a single cluster incongruent with the remaining clusters. Figure 4.3 also presents the fraction of common contacts (FCC) to illustrate population variance, as reported by Haddock. Plots for the remaining P450-CPR docking results are not reported here. Instead, the energies of the top three clusters each are supplemented (App. B). All models with the highest Haddock scores were assessed in this section.

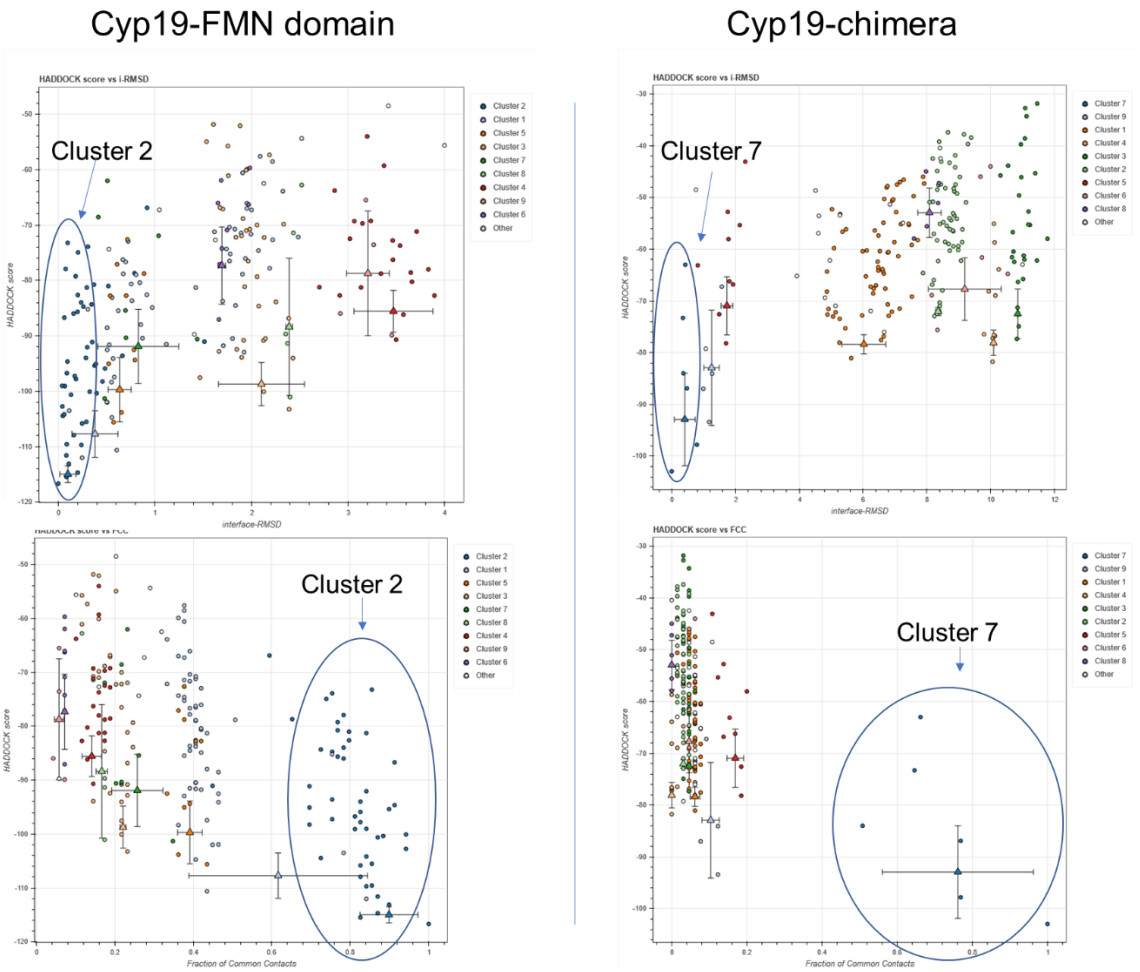


Figure 4.3 Plots of Haddock scores as a function of interface RMSD (top panels) and fraction of common contacts (bottom panels) for Cyp19-FMN domain and Cyp19-chimera. Clusters 2 and 7 returned the highest Haddock scores for Cyp19 docking against the FMN domain and chimeric CPR, respectively.[73]

An overlay of FMN domain and chimeric complexes did not yield the same binding mode, although both models were plausible for two reasons- 1) the cofactors were within the proximity of a 15 Å threshold to the heme, and 2) the N-terminus appeared to be oriented appropriately for the unmodeled transmembrane segment. An overlay of the two models with that of the crystal structure of *B. megaterium* P450BM3 fusion shows a stark contrast in binding mode. In fact, the crystal structure N5 atom of the cofactor was 22.7 Å from the iron. However, a rotation of the BM3 FMN domain towards the conserved residues in our Cyp19-FMN model moves the flavin

methyl groups 13.5 Å. This drives the N5 atom 9 Å closer to the heme center and within the physiological range for electron tunneling.

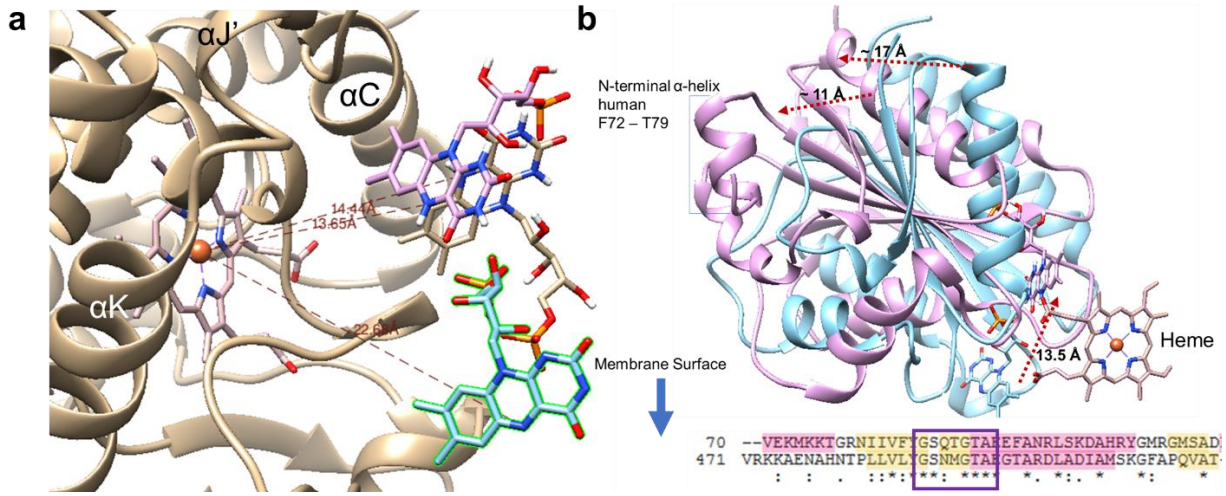


Figure 4.4 3-D overlay of coordinates from the crystal structure of P450BM3 (cyan), and models from prediction-driven docking of Cyp19-chimeric CPR (beige cofactor FMN), and Cyp19-FMN domain (magenta).

a) The ribbon structure of Cyp19 with distances of FMN N5 atom to the heme iron center at the proximal heme site-FMN domain interface. Ribbon diagrams of the FMN domains are removed. Correspondingly, distances for P450BM3, chimeric, and FMN-truncated structures are 22.7, 14.4, and 13.7 Å. b) Rotation of the P450BM3 FMN domain in the direction of the dotted arrows moves the N-terminal α-helix in the direction of the membrane surface where L22 – F42 is anchored moves the N5 atom of FMN 13.5 Å. An amino acid sequence alignment of the N-terminal region of human (top sequence) and *B. megaterium* (bottom sequence) shows a conserved secondary structure (α-helices- pink, β-strands- tan). A semi-conserved nucleotide-binding region is boxed in magenta.

Table 4.3 Physicochemical descriptions of P450-CPR complexes.

Complex	*Interfacial polar/charged interactions*	*ΔG (kcal/mol)*	*K_d (M)*	*K_d (M) experimental*	*FMN – Heme distance (Å)*
P450BM3 (1BVY)	17	-6.9	8.6×10^{-6}		22.68
Cyp19-FMN	25	-10.5	1.9×10^{-8}	9.5×10^{-9} ‡	13.65
Cyp19-3FJO	17	-10.2	3.2×10^{-8}		14.44
Cyp2C19-FMN	49	-12.2	1.1×10^{-9}	0.5×10^{-8} †	15.55
Cyp2C19-3FJO	25	-9.9	5.6×10^{-8}		14.64
Cyp2D6-FMN	31	-9.0	2.6×10^{-7}	2.0×10^{-8} †	15.94
Cyp2D6-3FJO	23	-11.3	4.8×10^{-9}		15.12
Cyp3A4-FMN	-	-	-	2.0×10^{-8} †	14.88
Cyp3A4-3FJO	-	-	-		17.62

† Experimentally determined and published by Shimada et al. with rat CPR and human P450.[74]
‡ K_m value determined by an estrone-based ELISA by Lo et al. with rat CPR and human P450.[19]

The reputed key driver for electron transfer between type II P450s and CPR is through the coupling of attractive electrostatic networks. After NADPH binds, a hydride is donated to FAD^+. Single-electron transfers are funneled from $FADH_2$ to FMN. The difference between electronic states is thought to mediate a change between an open and closed conformation between the FAD and FMN domains. In a closed conformation by the best model returned from prediction-driven docking, the closest proximity between the FMN N5 and the heme iron was 34.9 Å. After adopting an open conformation, the basic residues at the proximal heme site, and the acidic residues at the FMN site interact. Further, we are convinced that interfacial interactions are not conserved between two different productive coupling events of the same P450 with CPR. This can be rationalized by the low sequence identity among P450s at the protein's surface. It seems that CPR mediates whether coupling will occur, and that electron transfer is facilitated by the packing density between the proteins (proteins as dielectrics) and the proximity between the cofactors.

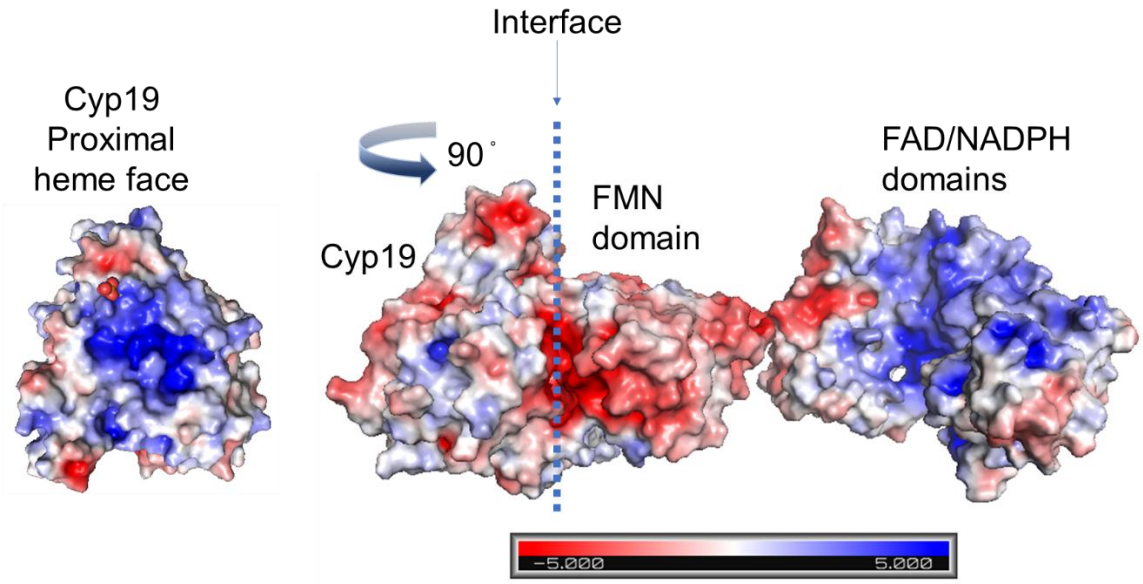


Figure 4.5 Electrostatic potential surface map of a proposed end-on binding mode of the Cyp19-CPR complex.

Cyp19 (pdb ID 4KQ8) and chimeric CPR (pdb ID 3FJO) interact at the Cyp19 proximal heme site when CPR is in an open conformation.

The lowest energy structures from 25 docking runs of E- and Z- endoxifen were evaluated, and the most probable binding mode for each is depicted in Figure 4.6. To note, in the absence of a simulation cell, the three highest-scoring E- endoxifen structures interacted with the proximal heme site. The mean binding energies from three runs as calculated by YASARA were -197.733 and -165.168 kcal/mol for E- and Z- endoxifen. There is currently no experimental data to distinguish the potency of each isomer since a racemate was used in prior work by Lu et al. Although it is clear that endoxifen and N-desmethyl tamoxifen (endoxifen without the para hydroxy substitution at the phenyl group) act non competitively with IC_{50} values 6.1 and 20.7 μM.[54] In any event, the hydroxyl group favorably contributes to a higher level of potency. This can be rationalized by the Z-isomer hydrogen bond with N421. In the E-isomer, water displacement will cause Van der Waals forces and long-range electrostatic contributions to weight heavier, ultimately contributing ~20 kcal/mol to the predicted binding energy.

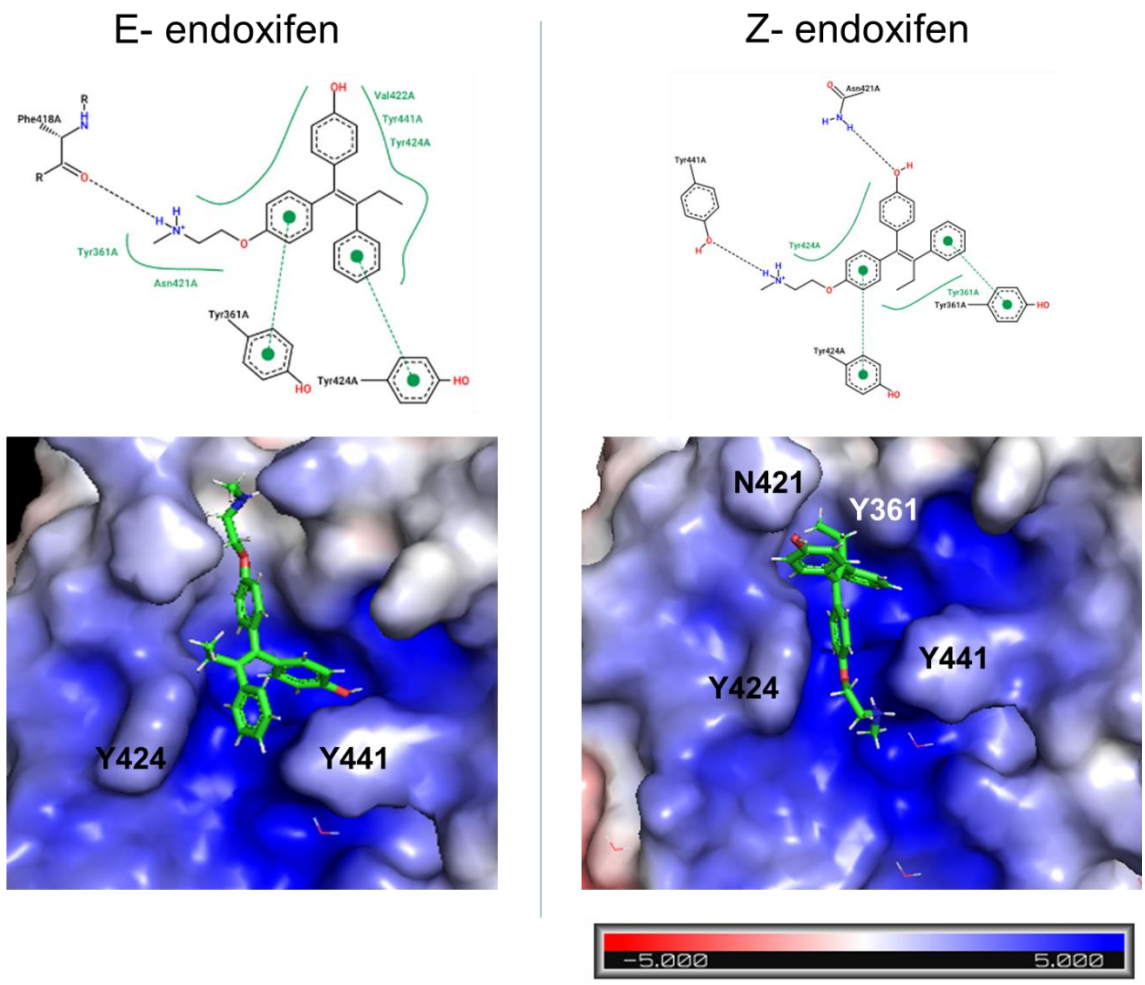


Figure 4.6 Putative binding modes for E- and Z- endoxifen depicted as a 2-D PoseView and 3-D representation.

Spatial complementarity and attractive forces contribute to endoxifen potency at the proximal heme. Ligand interaction sterically hinders and shields FMN interaction from the electropositive surface of the binding pocket.

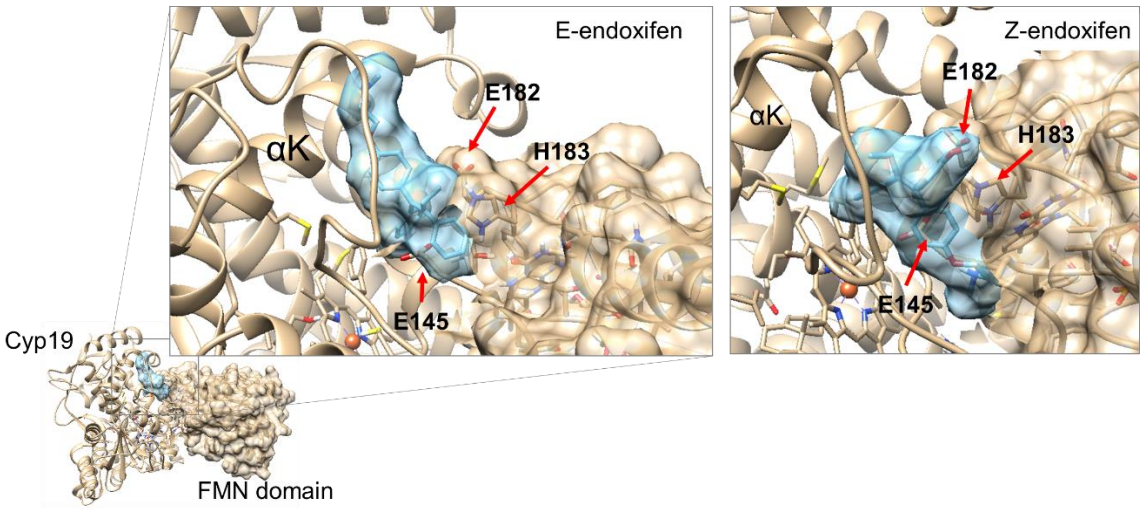


Figure 4.7 Superposition of E- and Z- endoxifen in proposed binding modes at the junction of the Cyp19-FMN complex.

Transparent surfaces of endoxifen (cyan) and FMN domain (beige) demonstrate a spatial conflict between E- endoxifen and residues E182 / H183. E/Z- endoxifen occupies a space reserved for the E145 side chain.

In this study, the proximal heme site is a high scoring putative binding pocket for nonsteroidal aromatase inhibitor endoxifen. In a prior study by Sgrignani and colleagues, three allosteric sites were predicted to bind endoxifen. However, the proximal heme site was ruled out as a candidate. It yielded the lowest interaction energy, and its center of mass exceeded a 4 Å distance during MD simulations in 8 ns.[75] This is unsurprising since the proximal heme site is solvent-exposed, having a low surface to volume ratio. In assessing the top candidate allosteric site proposed by Sgrignani, DoGSiteScorer ranked it fourth in druggability, with a 0.62 drug score, 333 $Å^3$ volume, and a 599 $Å^2$ surface area. These values indicate an enclosed allosteric pocket, that may be druggable since it purportedly occupies a dynamic space. Sgrignani concluded that allosteric interaction hampers G-H' loop mobility, subsequently occluding a substrate access channel.[75] However, endoxifen is a known noncompetitive inhibitor with a K_i equivalent to its K_i' value. Therefore, androstenedione is expected to have the same K_m value in the presence of

endoxifen. In our experience, the presence of either compound did not affect the ability of the other to bind the enzyme. Although the binding affinity and K_m are not equivalent, limiting solvent substrate access by increasing rigidity would affect the on-rate such that the K_d is increased. From this, we understand that the proximal heme site is a viable pocket to bind endoxifen. Cyp19 regulation by electron transfer disruption correlates well will our docking study. In the absence of a crystal structure, this is only conjecture and warrants further kinetic studies with active Cyp19 mutants.

CYP19 may serve as a model P450 to study the druggability of the proximal heme site of other CYPs that have extended meander loops. The loop region between helices K" and L contains the meander region of 21 residues long. Roughly 30 - 40 % of P450s have a loop 14 or 15 residues long- all of which are class II P450 enzymes. This disparity offers a niche to selectively target CYP19, as less than 5 percent of P450s have a meander loop as lengthy.[50] Convincingly, disruption of the electrostatic network in the proximal cavity would disproportionately perturb CYP19 heme stability relative to the vast majority of known P450s. P450 inhibitory activity of potential off-targets would likely be comparable to that of CYP19 if a small molecule can occupy the niche space. In our experience, a K440A mutant renders the enzyme inactive with a near-complete loss of the heme group. This is due to the K440 side chain amino group forming a heme-stabilizing electrostatic interaction with the CO backbone of G431 in the loop. This interaction maintains loop rigidity, and consequently, iron coordination to the C437 thiolate group. These results support prior work in which a K440 mutation destabilizes the heme. A substitution to glutamine results in a 3-fold loss of protein yield, a near 88 % reduction in specific activity, and a 10-fold loss in catalytic efficiency. It was inferred that heme stability was compromised due to the high A_{280}/A_{393} ratio from absorption scans.[19]

The Cyp19-P450 reductase interface is a promising target to modulate aromatase activity. Small molecules that bind the proximal heme site have the capacity to interrupt enzyme-coupling, attenuate electron transfer, or allosterically destabilize the heme altogether. Competitive P450 inhibitors that interact with the active site iron, often incur unwanted effects due to off-target binding of other P450s. Allosteric regulation of Cyp19 provides a framework to rationally design P450 drugs with diverse chemical profiles, thereby expanding our library of candidate molecules of therapeutic value.

Chapter 5 - Aromatase inhibition by novel inhibitors

In this chapter, we aim to uncover inhibitors that regulate aromatase function by a non-competitive mechanism. We take a standard approach by first screening a library of millions of compounds *in silico*, then measuring the level of activity *in vitro*. We then characterize the mode of inhibition in three ways. The first is through measuring Cyp19 absorption shifts since P450s exhibit signature soret shifts when the chemical environment of the heme is altered. The second means is by describing enzyme kinetic behavior at various concentrations of substrate and inhibitor. Finally, we plan to illustrate allosteric binding by providing a cocrystal structure. Currently, there are no Cyp19 structures with a drug-like molecule bound at an allosteric site.

In this chapter, I define non-competitive binding as any event that occurs at an allosteric site- or a locale distant from the active site such that ligand and substrate do not demonstrate mutual exclusivity. Noncompetitive will be defined as a mixed-type of inhibition where the inhibitor has an equal affinity for the enzyme in the absence and presence of the substrate (as described by Lehninger).

Introduction

In 2011, the first *in vitro* study that demonstrated anti-aromatase activity by principal tamoxifen metabolites was published by Dr. Wenjie Lu et al. from the lab of late Dr. David Flockhart at Indiana University School of Medicine.[54] This would spur interest in the complex role of selective estrogen receptor modulators (SERMs). Up to this point, tamoxifen was administered by breast cancer patients as an ERα antagonist. In this study, the racemates N-desmethyl tamoxifen (NDMT), and 4-hydroxy N-desmethyl tamoxifen (endoxifen) exhibited

noncompetitive behavior with K_i values of 15.9 and 4.0 μM, respectively.[54] The following year, the Flockhart group recounted ongoing work in which N,N-didesmethyl-4-hydroxytamoxifen (norendoxifen) was in fact the most potent metabolite of tamoxifen with a K_i of 35 nM.[76] More importantly, it exhibited competitive-type kinetics, illustrating that simple modifications to ligand structure can change not just the binding mode, but the mechanism by which it modulates enzyme or receptor function.

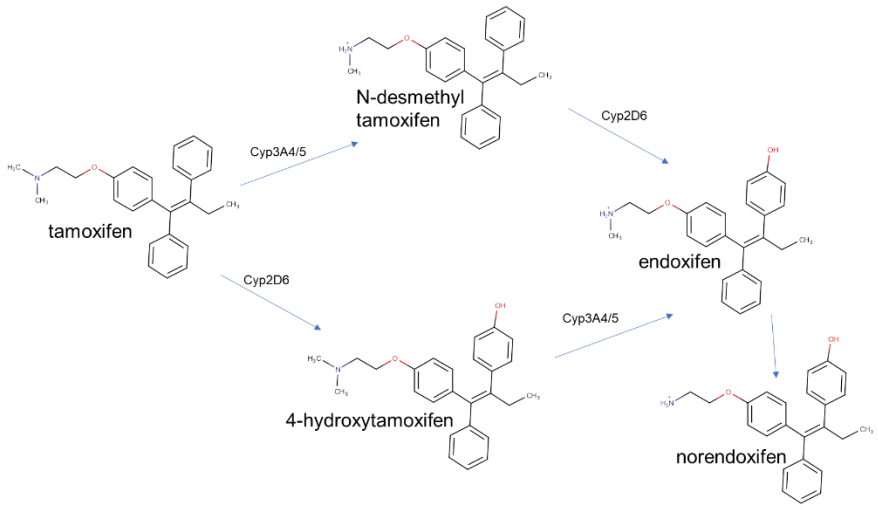


Figure 5.1 Tamoxifen and Cyp19 active metabolites.
P450s that catalyze the redox reactions include Cyp3A4/5 in N-demethylation and Cyp2D6 in 4' hydroxylations.

After Dr. Lu joined our team as a postdoctoral researcher, the Flockhart group screened norendoxifen analogs with substitutions at the ethyl group and para position of the phenyl groups.[77] Meanwhile, we sought to expand on the work by docking endoxifen against Cyp19, then screening the putative binding site for top hit compounds. Endoxifen scored well as a binder at the proximal heme site. We also noted that this site is the putative interface for Cyp19-CPR coupling. Further, it had a high druggability score, and it was in a region that was relatively well-conserved as a P450

enzyme. Moving forward, we explored this pocket as a potential target to modulate aromatase function. Protein-protein interfaces are inherently difficult regions to target, which made this pursuit even more challenging yet rewarding independent of the outcome.

Materials and Methods

In silico screening for allosteric inhibitors

Idock and FragP were used for the virtual screening, and to select the top hit compounds from a library of over millions of small molecules. Candidates selected for activity-screening were top-hits for binding in the putative pocket for endoxifen. Drug scores, according to the DSX[78] scoring function, are provided in Appendix D.

Determining the IC_{50} of top hit inhibitors

A CYP19/MFC HTS (high-throughput screening) Kit (Corning) was used to measure the anti-aromatase activity of candidate inhibitors AR11- AR13, AR11-2, and RN1. The activity was monitored in the conversion of substrate 7-methoxy-4-(trifluoromethyl)coumarin (MFC) to its fluorescent metabolite 7-hydroxy-4-(trifluoromethyl)coumarin. A 2X NADPH regeneration system was used for 1:2 serial dilutions of inhibitor in black 96-well plates (Corning), and prewarmed for 10 minutes in an incubator at 37 ˚C. Reactions were initiated by the addition of a 2X enzyme/substrate solution for a final 200 µL reaction volume. Reaction components included 7.5 nM P450 microsomes enriched with oxidoreductase, 25 µM MFC, 8.125 µM $NADP^+$, 412.5 µM $MgCl_2$, 412.5 µM glucose-6-phosphate (G6P), 0.2 U/mL G6P dehydrogenase (G6PDH). Reactions were incubated for 30 minutes at 37 ˚C, then terminated with 75 µL 0.5 M Tris-base (80 % ACN). A FluoDia T70 plate reader measured HFC product with ex/em filters, 400/530 nm from

duplicate reactions. Reads from two cycles were averaged to complete a data set for each inhibitor. Three assays were completed on separate days to yield three data sets for analysis.

Reaction conditions for measuring anti-aromatase activity for AR15-AR20 were replicated as detailed in the prior section. Modifications are highlighted here. Reaction components CYP19 + P450 reductase supersomes (Corning), NADPH regeneration system (Corning), MFC (Chemodex), ketoconazole (Selleck Chemicals) were purchased separately, and buffer and stop solution were made "in-house." A dry-plate was used for temperature-controlled incubations of the 96-well plate. A Tecan fluorescent plate reader measured product formation at an optimum gain with ex/em filters, 405/535 nm, by a circle-read mode. Reactions were measured in duplicate and averaged for analysis.

Time-dependence measurements of AR13

CYP19 turnover of 7-methoxy-4-(trifuoromethyl)coumarin (MFC)was monitored by a time-course experiment at different concentration of AR13 (0, 0.03, 0.3, and 3 μM). In short, a 100 μL NADPH regeneration system and inhibitor were preincubated in a temperature-controlled incubator at 37˚C for 10 minutes per reaction in a black round-bottom 96-well plate (Corning). 7.5 nM P450 and 25 μM MFC were incubated separately in a water bath at 37˚C. Each reaction was initiated with 100 μL enzyme/substrate solution. Reactions were quenched with 75 μL of 0.5 M Tris-base at various time points. At t=0, enzyme solution was added after the Tris-base stop solution. A FluoDia T70 plate reader measured substrate conversion to HFC at ex/em 400/530 nm filters. Progress curves were generated in Microsoft Excel. Regression of progress curves was analyzed in a time frame that achieved linearity in the absence of inhibitor.

CPR control assay

Cytochrome P450 reductase (CPR) microsomal preparations were purchased from corning to measure potential interactions with lead compounds AR11 and AR13. Reductase activity was measured by monitoring the absorption of substrate cytochrome c (cyt c) at 550 nm. Reactions were preincubated at room temperature for 10 minutes in 1 mL disposable cuvettes, then used to zero the spectrophotometer. Preincubation mixture contained inhibitor (0, 4, 16, and 64 μM) with 0.075 mg/mL CPR, 0.5 mg/mL cyt c, and 100 mM potassium phosphate buffer (pH 7.4). After measuring a steady baseline for 30-seconds on kinetic mode, a response was initiated with NADPH (6 and 20 μM),[79] and the cuvette was inverted to facilitate mixing. Initial velocity (V_0) was recorded, and reactions were repeated twice. Diphenyleneiodonium was used as a control to test the CPR inhibitory potential.

Determining the mode of inhibition

Microsomes prepared from insect cells (BTI-TN-5B1-4) overexpressing Cyp19 and CPR (Corning) were used to fluorometrically determine the inhibitory mode of compounds AR11 and AR13. In a 96-well plate format, 96 μL NADPH regeneration solution ($NADP^+$, G-6P, $MgCl_2$, G-6P dehydrogenase) and 4 μL of serial diluted inhibitor-stocks (4 μL ACN for non-inhibitor and reaction-blanks) were pipetted in each well and prewarmed for 10 minutes. Reactions were initiated sequentially with a quick stream of 100 μL preincubated enzyme/substrate solution to ensure homogenous mixing across the samples. Plates were layered with parafilm and foil to avoid evaporation and photodegradation of the fluorometric substrate. After a 20-minute incubation period on a dry-plate at 37°C, reactions were quenched with 75 μL 100 mM Tris-base (80 % ACN) in-sequence. Conversion to HFC-product was measured with a Tecan plate reader with ex/em,

405/535 nm filters in circle read mode, 10 reads, 40 μs integration time at 52 gain. Each 200 μL reaction contained 100 mM PPB (pH 7.4), 10 nM P450, 0.325 mM $NADP^+$, 0.825 mM G-6P, 0.825 $MgCl_2$, 0.1 U/mL G-6PDH, substrate MFC (9.9, 14.8, 22.2, 33.3, and 50 μM), and inhibitor.

Data analysis of fluorescence-based assays

Regression analyses were performed in GraphPad Prism version 8.0.0 for Windows, GraphPad Software, San Diego, California USA. Dose-response curves were fit to a four-parameter logistic model. Accordingly, IC_{50} values reflect experimental data without overfitting to a hill slope of 1. Michaelis-Menten kinetic constants generated from non-linear regression curves were fit to mixed-mode and competitive inhibition types, and Lineweaver-Burk and Dixon-type plots were assessed to determine inhibition type.

Selection of mutants and Cyp19 constructs for crystallization

To characterize the protein-ligand interactions, recombinant mutants were selected that we predict would increase the likelihood of crystallization. Our wildtype construct is ranked 4/5 and 5/10 in crystallization classes by the Expert Pool method and the Random Forest Classifier (RF). The lower number indicates a protein optimal for crystallization, while a higher number indicates a lesser promising outcome.[80] One of the metrics included in the RF score is the surface entropy contribution of solvent-exposed residues. Amino acid residues, lysine, glutamine, and glutamate, have side chains that contribute a more negative value to the entropy term (more positive free energy). As such, they prevent protein-protein interactions necessary to form an ordered crystal lattice. Therefore, by mutating clusters at the surface to alanine, the surface entropy is reduced, ultimately improving the likelihood of crystal formation. However, introducing clusters of apolar

residues at the surface has its own implications that need to be weighted. From the top three clusters, the triple mutant E269A/E270A/K271A (denoted E269A in this work) had the highest score of 7.48. The second and third clusters were double mutants with the scores 5.4, and 5.3, respectively (App. B). In this case, higher scores represent an increased likelihood of growing crystals. The triple mutant was selected to observe functionality and subsequent crystallization trials. E181A was selected because it was reported to consistently express at levels 15-fold greater than that of the wildtype. This was thought to be the result of tightening the loop between αD/E that ultimately increases protein stability such that less is packaged in inclusion bodies during induction.[19] Single mutants V80S, G156A, L240S, and A419S, were selected for crystallization because they are surface exposed, and predicted to either improve rigidity or stability. Further, they do not interrupt secondary structure (at the terminus of helices or at looped regions), and to our knowledge, no prior work had been reported.

Preparation of CYP19A1 DNA constructs for expression in E. coli

CYP19A1 cDNA in pCW expression vector was a generous gift from the Guengerich Lab. The DNA construct includes an N-terminal truncation of the P450 transmembrane segment and a C-terminal His_6 tag to maintain solubility during induction and aid in purification. The DNA construct was codon-optimized for expression in *E. coli*. Details on the cloning strategy are provided in the experimental procedures section of Sohl, C and Guengerich, P.[15]

Site-directed mutants were made by PCR amplification with Q5 Site-Directed Mutagenesis Kit (NEB) according to the manufacturers' protocol. Codons were selected based on their highest frequency in *E. coli* expression and those that curtailed self-dimerization, heterodimers, and hairpin formation. Q5 2X master mix, *CYP19A1* codon-optimized template, and primers were

resuspended in type 1 water to 25 µL final volume. Primers and cycling parameters were optimized to facilitate the successful amplification of the targeted mutant. The general cycling conditions are provided in Table 5.1. Modification of these conditions for each mutant is appended. Following DNA amplification, 1 µL of PCR product, 2X KLD reaction buffer, 10X KLD enzyme mix, and type 1 water were incubated at room temperature for 5 minutes. After that, 5 µL of the reaction mixture was used for a standard transformation protocol by heat shock with E. cloni 10G competent cells (Lucigen). Overgrowth cultures were plated on carbenicillin-selective LB-agar media. Transformants were picked, and DNA was extracted with a Monarch plasmid miniprep kit (NEB) for sequencing at Genewiz (South Plainfield, NJ).

Table 5.1 PCR cycling parameters for Q5 site-directed mutagenesis.

Stage	*Step*	*Temperature (°C)*	*Time*
1	Initial denaturation	98	30 s
2 (25 cycles)	Denaturation	98	10 s
	Annealing	T_a	25 s
	Extension	72	t_{ext}
3	Final extension	72	2 min
4	Hold	4	∞

His_6-MBP-TEV-CYP19 LIC cloning strategy

To increase expression levels, *CYP19A1* was cloned into an empty pET vector with an N-terminal His_6-tagged MBP fusion by a LIC method. Briefly, primers were designed to amplify CYP19 with LIC fusion tags at the 5' ends. 0.75 µg of pET-MBP plasmid was digested with 10 U high-fidelity SspI (NEB) in CutSmart buffer at 37 °C for 15 min. The reaction was terminated by heat denaturation at 65 °C for 20 min, and the linearized vector was gel purified. The vector and

insert were subjected to T4 DNA polymerase treatment separately in 50 μL reaction volumes. Mixtures contained 1 μg of DNA in NEB 2.1 buffer with 1 U of enzyme. Synthesis of complementary overhangs was facilitated by 2.6 mM dCTP, and dGTP in the reactions mixtures of DNA insert and vector, respectively. Reactions were incubated for 30 min at 22 ˚C and terminated by heat denaturation for 20 min at 75 ˚C. Mixtures were temperature-controlled in a TC-3000G thermocycler (Techne). The vector and insert were annealed at a 1:2 and a 1:3 molar ratio. The equation: 0.65 x # base pairs = ng/pmol was used to determine the appropriate amount of DNA in ng quantities (approximately 42 ng : 20 ng and 42 ng : 30 ng, respectively). Reactions were incubated at 25 ˚C for 5 min. The annealing reaction was supplemented with 1 μL 25 mM EDTA, followed by an additional 5-min incubation at room temperature. E. cloni 10G competent cells were transformed with 3 μL of the reaction mixture for cloning.

Colony PCR with orientation-specific primers were used to confirm transformants that successfully acquired vector with CYP19 DNA. In short, each selected colony was suspended in 20 μL of type 1 water. One microliter of colony suspension was mixed with 1X colony PCR master mix (Lucigen), and MBP forward/ CYP19 reverse primers. Cycling parameters were followed as recommended by the manufacturer. PCR products were quantified on 1.7 % agarose gel supplemented with GelGreen nucleic acid stain (Biotium) for visualizing band size and intensity with an LED transilluminator (blue light with orange filter). Colonies with the correct band size were miniprepped, and plasmids were sent to Genewiz for sequencing. All primers and sequencing results are appended.

Production and purification of CYP19 and its mutants

E. cloni 10G (Lucigen) competent cells were used for cloning and expression cultures. Pre-cultures were grown for 12 -16 hours at 37 °C, 250 rpm in LB lennox broth, and 10 µg/mL carbenicillin. Expression cultures were inoculated at a 1:50 ratio in TB by volume and grown at 37 °C, 250 rpm. At OD_{600} 0.6 – 0.8, cultures were incubated at 25 °C, 150 rpm, for 1 hour before induction with 1 mM iPTG and 1 mM δ-aminolevulinic acid. Cultures were harvested and washed thrice with 200 mM NaCl, 100 mM sodium phosphate buffer (pH 7.4) at 4 times the cell pellet volume for 20 minutes, 6000 rpm (5,524 x g).

Cell pellets were resuspended in lysis buffer (buffer A with 15 mM imidazole, 1 % tween 20, 1 mM DTT, 0.1mM EDTA, and 0.1 mM PMSF) prior to sonication. Lysates were ultracentrifuged at 40,000 rpm (164,700 x g) for 20 minutes at speed. The supernatant was passed through Ni-NTA resin and washed with 10 column volumes buffer A + 0.1 % tween 20 + 15 mM imidazole. The column was washed with 5 column volumes of buffer + 50 mM imidazole for further polishing. His_6-tagged recombinant protein was eluted with 250 mM imidazole. Red fractions were pooled, and buffer exchanged into 20 mM Tris-HCl, 50 mM NaCl, 1mM DTT (pH 7.4) with 500 µL-volume 30,000 NWCO spin columns (Amicon) at 4 °C. The concentrate was allowed to sit at room temperature, and then it was loaded onto a 1 mL prepacked DEAE FF/ Q XL HiTrap column (GE Healthcare). The flow-through was collected, then injected onto a Superdex75 10/300 GL column with FPLC (AKTA) system, and recombinant protein was isocratically eluted with buffer A at 0.45 mL/min flowrate. Alternatively, Superdex200 Increase 10/300 GL columns were used at 0.7 ml/min to achieve more resolved peaks 55 – 110 kDa-although co-elution of contaminating proteins within this range required further purification for crystallization assays. Colored fractions with monomeric and dimeric aromatase were pooled and

concentrated in 0.5 mL Amicon spin columns. DTT was added to concentrated samples to 1 mM, and protein was stored at -76 °C in PCR tubes, 5 - 100 µL aliquots.

Conditions for the highest expression levels of MBP-tagged CYP19 are described here. Protein was expressed in Rosetta 2 cells for 24 hrs, 25 °C, at 150 rpm. Preculture conditions and lysate preparation are described in the aromatase and mutant methods section. The recombinant protein was eluted from a Ni-NTA column at 70 mM imidazole. Further polishing involved elution from a 1 mL prepacked MBP column (GE Healthsciences) in Buffer A with 10 µM maltose. Eluate was concentrated with 500 µL-volume 100,000 NWCO filters (Amicon) at 4 °C.

Characterization of functional CYP19 and its mutants

CYP19 size was quantified by SDS-PAGE and blot analysis. Protein purity was confirmed by SDS-PAGE with detection by coomassie-based Acquastain (Bulldog Bio)- capable of detecting nanogram quantities of protein after 1 hour of soaking. For blot analyses, the protein was loaded into precast SDS-PAGE gels and transferred to a FluoroTrans PVDF membrane in Rapid Transfer (VWR) buffer for 25 min at 80 V to facilitate the efficient transfer of CYP19. Blots were blocked in Rapid Block (VWR) buffer, then incubated overnight at 4 ̊C with Ni^{2+}-conjugated horseradish peroxidase (HRP) for detection of His_6-tagged protein. Mouse anti-MBP conjugated HRP was used to detect MBP-tagged protein at 0.1 µg/mL. Membranes were washed 3 times for 10 minutes with TBST buffer. 3,3'-diaminobenzidine tablet (DAB, Sigma) was reconstituted in peroxide/TBS solution for colorimetric detection.

CD Spectra of wild type, E181A, and E269A mutants were observed from 190 – 260 nm, at 50 nm/min with a Jasco J-815 Circular Dichroism (CD) Spectropolarimeter. The average from

5 acquisitions was smoothed, and mdeg was converted to mean residue ellipticity (MRE) by the equation

MRE = mdeg/(n x C x l),

where n = number of peptide bonds, C = concentration (M), and l = cell path length (cm).

Percent helicity was estimated by the 222 nm method[81] in which $[\theta]_{222}$= -36,000 deg cm^2 $dmol^{-1}$ for 100 % helicity, and fraction helicity = $[\theta]_{222}$ – 3000/(-39,000)

The binding affinity of androstenedione and the quantity of active P450 of wild type and mutants were used to measure protein functionality. In brief, CYP19 was diluted in 100 mM PPB (7.4) to 3 μM P450 in a quartz UltraMicro 50-μL cuvette (Agilent). Androstenedione (1.0 – 46.0 μM) was titrated into the cuvette and gently resuspended before each absorption scan. Assays to determine ASD binding affinity for mutants Y361W, K440A, and Y441V differed in that ~ 0.4 mg/mL protein was titrated with 20 – 120 μM ASD. Spectra were corrected for lamp drift at 490 nm and subtracted from the enzyme in the absence of substrate. The change in peak to trough absorption between 350 and 500 nm from the resultant difference spectra was graphed as a function of substrate concentration. Data were fit to a single-site binding isotherm in GraphPad Prism 8.0 software.

The concentration of P450 reported here was quantified after 1 freeze-thaw to quantify active P450 after storage at -76 ̊C. The method used is detailed by Sohl and Guengerich[79] with modifications noted here. The protein solution was diluted with Buffer A to 50 μL in an UltraMicro quartz cuvette (Agilent). A stream of nitrogen was used to displace air in the cuvette. $Na_2S_2O_4$ grains were picked with a p20-200 micropipette tip, then added to the protein solution. The cuvette lip was parafilmed, then placed in a single beam UV-Vis 8453 spectrophotometer (Agilent) for an absorption scan of the reduced enzyme. A 3-mL luer-lock syringe with septum

was loaded with 0.5 mL – 1 mL of CO gas. The septum was replaced with a syringe needle, and CO was slowly bubbled (1 s^{-1}) into the protein solution. Absorption scans were taken until the peak maximum at 450 reached a plateau. A change in absorption from 450 to 490 nm of the CO difference spectrum was reported to quantify the concentration of active CYP19. P450 content was calculated by the equation

$$\Delta Abs_{450\text{-}490} = 91{,}000\ M^{-1}cm^{-1} \times l \times C,$$

where l = cuvette path length (cm), and C = concentration of P450 (M).

Inhibitor binding affinity for CYP19 mutants

Mutant primers were designed to introduce amino acid substitutions at the predicted allosteric binding pocket. Site-directed mutagenesis was used to amplify and transform *E. coli* by the Q5 protocol detailed earlier in this chapter. Materials and methods for expression, purification, and characterization were repeated for these mutants highlighted earlier. Table APPDX presents primers and PCR cycling parameters for codon substitution. The proposed interaction each mutant disrupts is presented in Table 5.2.

Table 5.2 Proposed protein-inhibitor interactions loss with corresponding mutants.

Mutant	*AR11*	*AR13*	*AR19*	*AR20*
Y361W	Binding (steric)	Binding (steric), H-bond with Tyr OH	Binding (steric)	π – π with phenyl and triazole- may strengthen
K440A	H-bond with Lys NH_3^+	H-bond with Lys NH_3^+	-	-
Y441V	π – π with phenyl	-	Binding (complementarity)	-

Results

IC_{50} of top-hit compounds from virtual screening

Inhibitor candidates AR11, 12, and 13 were initially screened for anti-aromatase activity with a high-throughput screening kit from Corning. AR11 returned an IC_{50} value of 31.07 μM, comparable to positive control endoxifen with a value of 30.69 μM. AR13 IC_{50} was in the submicromolar range, exhibiting a nearly 3-fold increased level of potency than positive control ketoconazole. Analogs of AR11 and 13, AR11-2, and RN1 respectively did not inhibit CYP19 to any appreciable extent, although RN1 exhibited weak to modest activity at 87 μM, albeit the same order of magnitude as endoxifen and AR11. AR19 and AR20 yielded IC_{50} values comparable to endoxifen, although AR19 was only capable of attenuating activity to 14%. Compounds without reported IC_{50} values in Table 5.3 did not exhibit anti-aromatase activity in a range to warrant further study. Dose-response curves for all inhibitors are provided in Appendix B, while curves for the most potent inhibitors are compiled and presented here (Table 5.3). The ratio of residual values of a 3-parameter model to a 4-parameter model (Table B.4) for IC_{50} values reported here are greater than 0.990, with the exception of endoxifen (0.973).

Table 5.3 CYP19 responsiveness to inhibitor candidates fit to a four-parameter model.

Inhibitor	*IC_{50} (μM)*	*Hill slope*	*Fractional activity at saturation*	*Goodness of fit (R^2)*
ketoconazole	3.08	-1.13	0.035	0.999
endoxifen	30.69	-2.26	0.021	0.997
AR11	31.07	-1.51	0.051	0.986
AR12	-	-	-	-
AR13	0.82	-0.86	0.071	0.978
AR11-2	-	-	-	0.803
RN1	86.62	-1.32	-0.059	0.988
AR15	-	-	-	0.569
AR16	-	-	-	0.956
AR17	-	-	-	0.908
AR18	-	-	-	0.866
AR19*	65.87	-1	0.144	0.846
AR20	42.83	-0.86	0.028	0.992

*Extrapolated and fit to a 3-p model

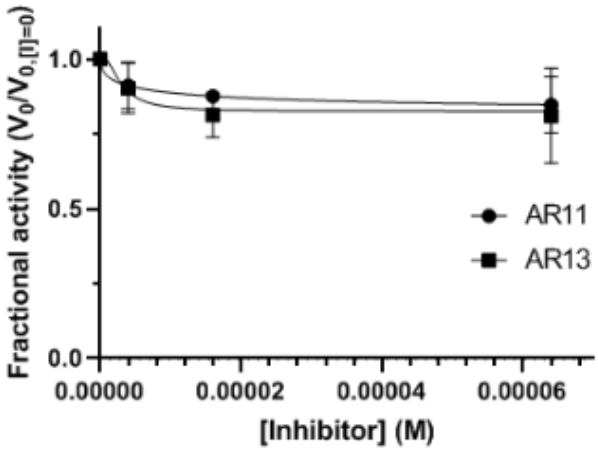


Figure 5.2 CPR fractional velocity at various AR11 and AR13 concentration. Initial velocities were measured at 0.075 μg/mL enzyme, and 6 μM NADPH.

At the approximate K_m measured for NADPH (data not shown), cytochrome P450 reductase activity was reduced to 85 % and 81 % at 64 μM inhibitor concentration for AR11 and AR13, respectively. The reduced activity plateaus when the molar ratio for inhibitor was at the same order of magnitude as NADPH. At 20 μM, the substrate concentration at 75 % the measured V_{max}, there was no appreciable loss of enzyme activity at an inhibitor concentration up to 64 μM.

The data collected is inadequate to return reliable K_i values from a Dixon plot. A linear fit of the data yielded R^2 values between 0.12 and 0.98, albeit the data clearly demonstrates AR11 and AR13 proclivity to selectively inhibit aromatase.

AR13 has an IC_{50} that is nearly 4-fold more potent than positive control ketoconazole, and within two orders of magnitude from the CYP19 concentration. As such, AR13 was assayed for time-dependent behavior that may indicate a tight-binding or irreversible inhibitor. Given the reaction conditions for determining the IC_{50}, the linear range was within the first 10 minutes of the reaction. Figure 5.3 illustrates that in the presence of the inhibitor, reaction progress curves do not deviate from linearity.

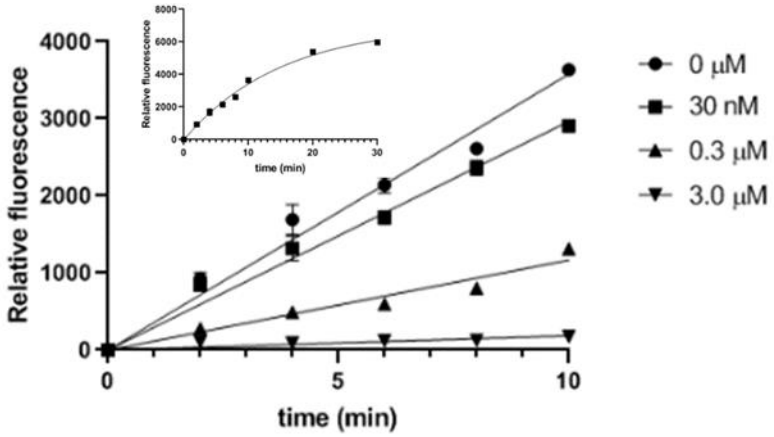


Figure 5.3 Progress curves of 25 μM MFC conversion to fluorescent product at 7.5 nM Cyp19 in the absence and presence of AR13.

Uninhibited reactions were linear within the first 10 minutes. Linearity was maintained at 0.03, 0.3, and 3 μM AR13.

Characterization of functional recombinant CYP19 and its mutants

Most strategies for recombinant expression of aromatase require induction of *E. coli* DH5α cells at longer incubation times (up to 48 hours).[19,82,83] A codon-optimized construct with a C-terminal His_6 tag yielded the highest expression levels in E. cloni 10G cells (Lucigen) over

the Rosetta 2 expression cell line. Figure 5.4 presents blots with a probe that binds His-tagged protein for detection by DAB substrate. Cyp19 is present at 55 kDa, with histidine-rich contaminants at 25 kDa, and 15 kDa (not shown in the figure).

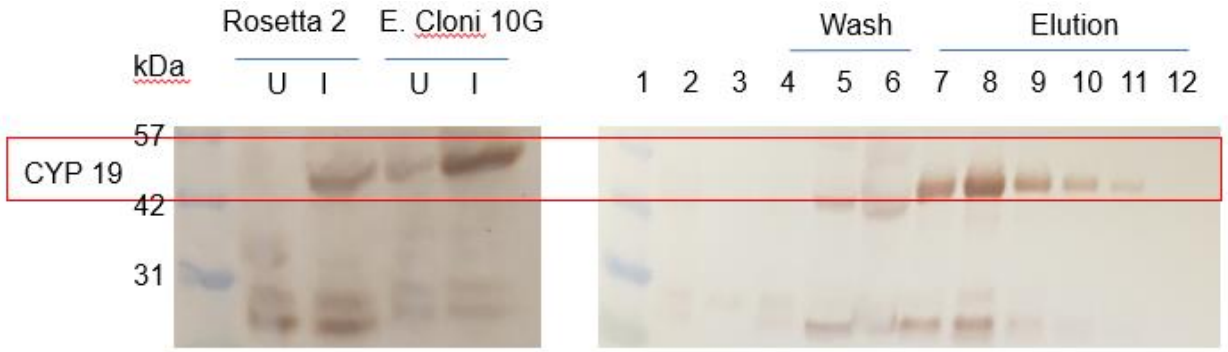


Figure 5.4 Blots with Ni-HRP probe and DAB substrate detection after a wet transfer from SDS-PAGE gels.

Cyp19 is boxed in red at 55 kDa. Smaller, histidine-rich contaminants are detected at 25 kDa. Uninduced and induced cultures from each cell line indicate higher expression levels after a 24 hr induction time in E. cloni 10 G cells relative to the Rosetta 2 cell line (left blot). Sequential 500 µL-fractions from a nickel purification include sample (lane 2), flow-through (3), wash (4 – 6), and elution (7 – 12) (right blot).

The most productive reaction conditions included induction at 1 mM iPTG for 48 hours, 27 – 28 ˚C. In contrast, Rosetta 2 cells produced higher expression levels of the N-terminal MBP-tagged construct after 24 hours (data not shown). A representative gel filtration chromatogram after nickel and DEAE purification is presented here (Table 5.4). The chromatogram from a modified purification procedure for recombinant protein used in binding studies is appended (Ni-purified protein onto a SEC200 Increase gel filtration column).

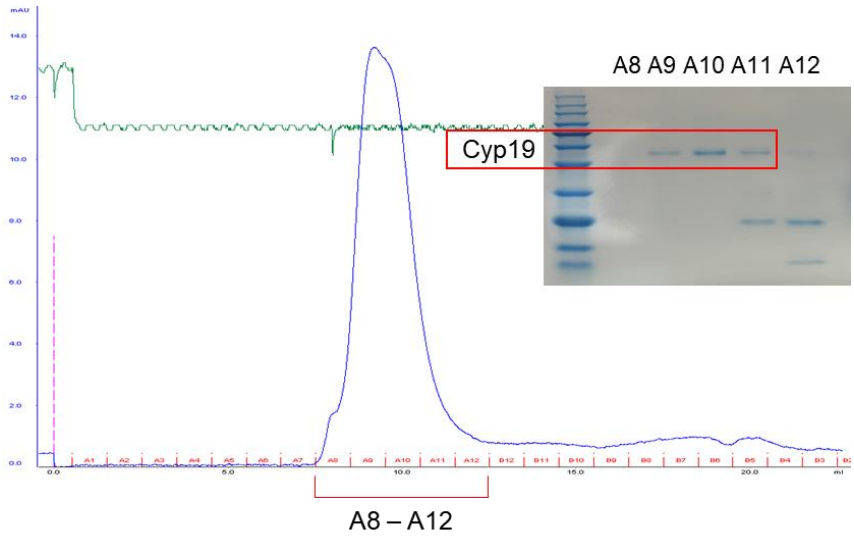


Figure 5.5 Gel filtration chromatogram of Cyp19 after Ni- (70 mM wash) and DEAE-purification.

The sample was injected onto a Superdex 75 column at 0.5 mL/min. UV trace indicates protein absorption at 280 nm with constant column pressure (green trace). SDS-PAGE of fractions A8 – A12 reveals the coelution of Cyp19 with a 25 kDa protein in fraction A11. Cyp19 is boxed in red.

Fractions A9 – A10 were pooled and concentrated in 30 – 50 kDa cut-off filters prior to storage in Buffer A at -76 ̊C. Figure 5.6 is an SDS-PAGE gel that shows the sample purity after each purification step. The most intense band greater than 55 kDa is a Cyp19 dimer at 110 kDa that can be detected in blots (not shown).

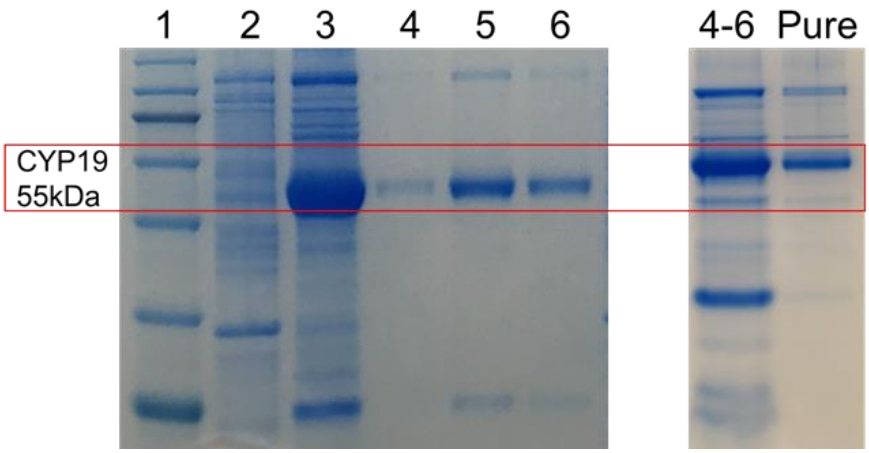


Figure 5.6 SDS-PAGE gels of Cyp19 after expression in E. cloni 10G competent cells after a 48 hr induction period.

Samples loaded include the pellet (lane 2), Ni-purified (3), and 1-mL DEAE flow-through fractions (4-6). The second gel includes a pooled concentrate of samples loaded in lanes 4 – 6, and purified enzyme after gel filtration with a dimer at 110 kDa.

The amount of active enzyme was quantified by measuring the sample absorption at 450 nm from a CO-difference spectrum. Figure 5.7 presents the absolute and difference spectra of the wildtype and E269A triple mutant. The insets also show between 35 – 45 % helical content. Cyp19 is 52 % helical according to the crystal structure and prior CD work.[82] The disparity is likely due to contributions from inactive partially unfolded protein, sample inhomogeneity that disrupts secondary structure, and protein contaminants.

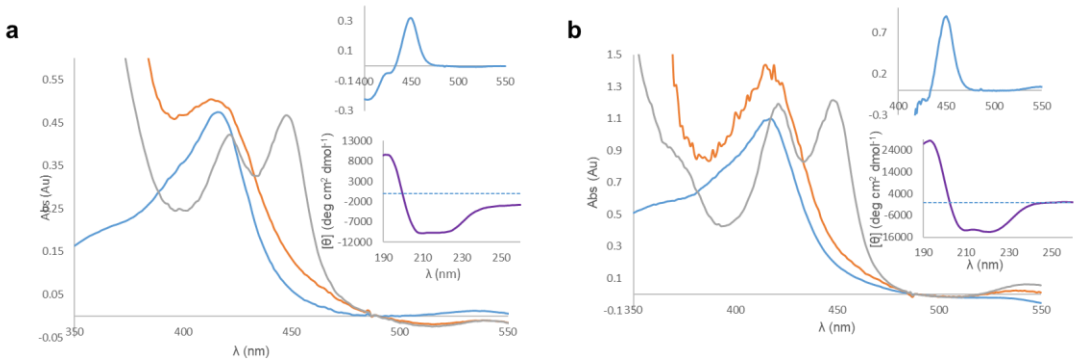


Figure 5.7 Absolute absorption spectra of oxidized (blue), reduced (orange), and CO-bound (grey) Cyp19 for a) wildtype and b) E269A mutant with a peak maximum at 450 nm.

Insets provide CO difference spectra (top) and CD spectra (bottom). CD spectra of 0.8 mg/mL wildtype, and 0.5 mg/mL E269A triple mutant have 35 % and 46 % helicity, respectively.

The binding affinity for endogenous substrate androstenedione was measured optically by its ability to displace heme-bound water- ultimately increasing the soret peak absorption at 394 nm while decreasing the peak absorption at 416 nm. A hyperbolic fit of the change in absorption of the difference spectra's peak to trough as a function of concentration androstenedione yields the binding isotherm. Wildtype and E269A had dissociation constants of 1.9 and 1.3 µM.

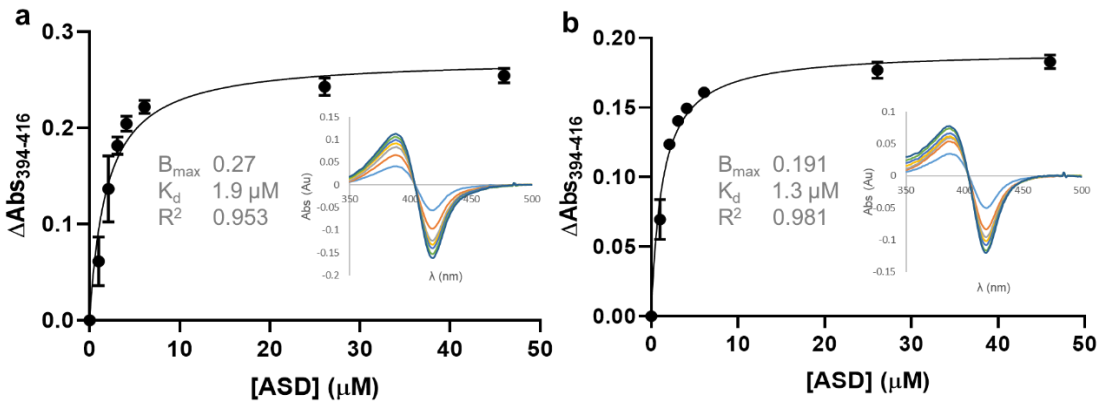


Figure 5.8 Cyp19 binding isotherms for single-site specific binding at 3 μM P450 and concentration ASD at 1.0, 2.3, 3.1, 4.1, 6.1, 26.1, and 46.0 μM.

a) Wildtype enzyme and b) E269A triple mutant have respective ASD K_d values of 1.9 and 1.3 μM. FeIII-ASD difference spectra are provided as insets.

Table 5.4 reports the average yield of active P450 recombinantly produced after one freeze-thaw cycle for Cyp19 and its mutants. Although values are not reported here, time-dependent shifts in the soret peak after the addition of a saturating amount of ASD and CO indicated that MBP-tagged Cyp19 is produced and purified in a functional state (Figure 5.9). Binding isotherms for mutants not shown here are provided in Appendix B.

Table 5.4 Quantification of recombinant aromatase.

Mutant	*mg protein/L culture*	*nmol P450/L culture*	*nmol P450/mg protein*	*K_d ASD (μM)*	*B_{max} (ΔAbs *)*	*B_{max}/K_d*
Wildtype	0.2	1.5	~7.65	1.9	0.27	0.14
MBP-Cyp19	-	-	-	-	-	-
V80S				-	-	-
G156A	0.1 (P450)			-	-	-
E181A	0.1	-	-	-	-	-
L240S				-	-	-
EEK(269-271)AAA	0.1	2.4	~23.06	1.3	0.19	0.15
Y361W	-		~ 0.58	12.68	0.09	7.2×10^{-3}
A419S		-	-	-	-	-
K440A	-	**	**	103.6 †	0.07 †	6.7×10^{-4} †
Y441V	-		~ 0.42	11.26	0.11	9.8×10^{-3}

*B_{max} values are reported as the maximum change in absorption units between peak and trough of resultant difference spectra. Peak and trough wavelengths for respective mutants are reported in the materials and methods section.
** unmeasurable quantity
† extrapolated value

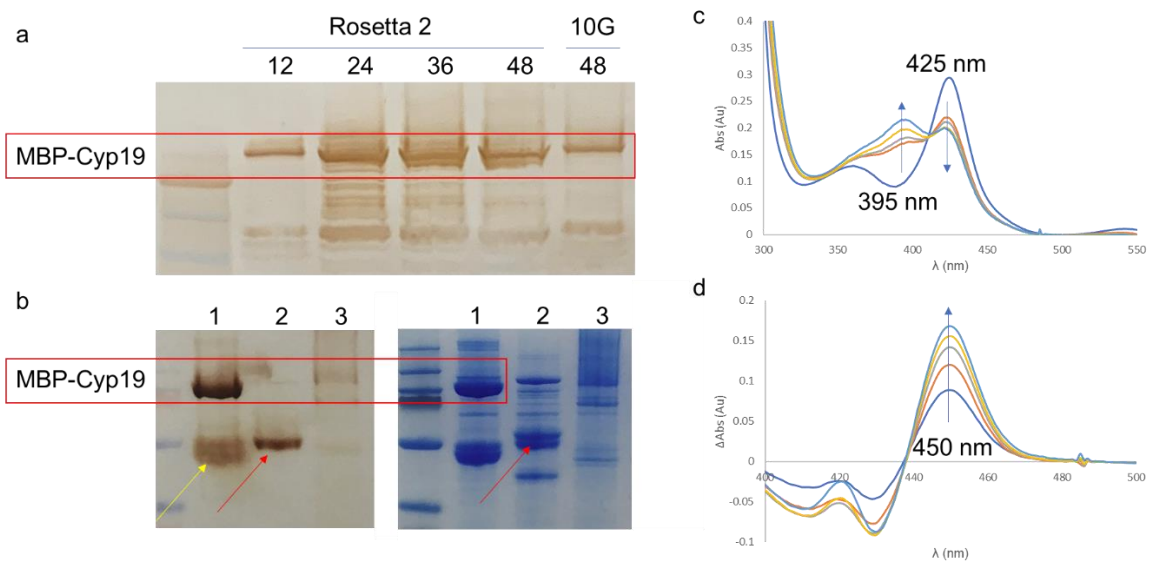


Figure 5.9 Characterization of MBP-Cyp19 fusion protein by western blot, SDS-PAGE, and optical absorption.

a) Western blot of clarified lysate after various induction times, 12 – 48 hours. Mouse anti-MBP IgG1 antibody detected MBP-tagged protein and degraded product. b) Blot and SDS-PAGE of purified MBP-Cyp19 (lane 1), Cyp19 (2), and MBP-Cyp19 after TEV cleavage (3). Lane 3 shows protein aggregation of the sample after TEV protease treatment. His-probe was used for the detection of recombinant protein on the blot. c) Absolute absorption spectra of purified MBP-Cyp19 after the addition of 40 µM ASD indicates a time-dependent increase of the high-spin state at 395 nm for up to 1 hr. Absorption scans were taken at 0, 5, 30, and 60 minutes. d) CO-difference spectra reached a plateau at 450 nm after 8 minutes. Scans depicted here were taken at 0, 2, 4, 6, and 8 min. MBP-Cyp19 bands at 97 kDa are boxed in red, Cyp19 band at 55 kDa (red), and MBP band ~42 kDa (yellow) are indicated by arrows in panel b. Blue arrows indicate the direction of hyperchromic stretching with time in panels c and d.

Optical absorption properties of CYP19 in the presence of novel inhibitors

Soret shifts in P450 enzymes arise from changes in the heme-iron environment, including coordination number, spin state, and ligand field strength. A covalent interaction at the iron 6-coordinate site causes a bathochromic shift from a 5- coordinate state. Water or a hydroxy ligand typically occupies this site in substrate-free CYP19, causing a shift from ~393 nm to ~414 nm. Compounds that bind iron with a higher affinity are known as type 2 inhibitors. The size of this shift is dictated by the chemical specie that occupies the sixth site, arising from the nature of the ligand's orbital overlap with higher energy d orbitals.

AR13 titration in 3.5 µM P450 from 6 µM to 12 µM caused a red shift from 413 nm to 424 nM. Saturation was achieved at 6 µM, with no increase in absorption after the first titration event. Addition of ASD to the reaction mixture- up to 150 µM -did not recover the 5-coordinate high-spin state. A time-dependent increase of P450 in the high-spin state was attained when the inhibitor to the ASD molar ratio was decreased two-fold. Figure 5.10 presents a bathochromic shift of the soret peak upon titration with AR13 and the reversible recovery of the 5-coordinate state at a 28-fold molar excess of the endogenous substrate to the inhibitor.

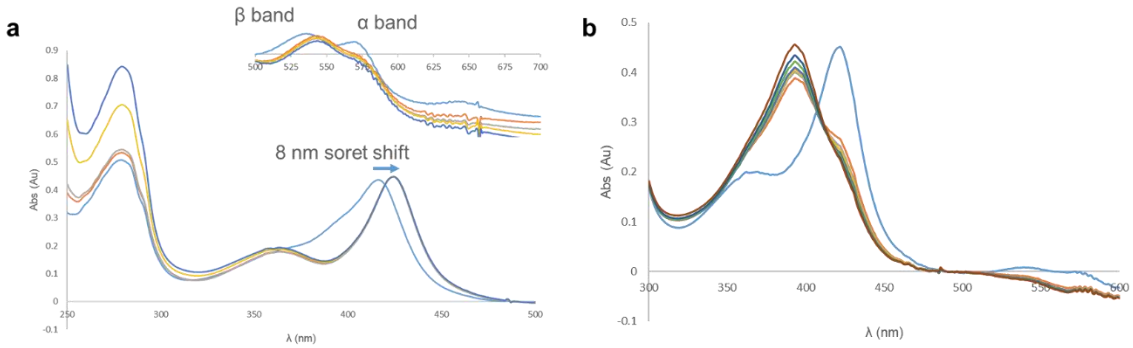


Figure 5.10 Soret peak shifts reversibly in the presence of AR13 and androstenedione.
A bathochromic shift is observed upon titration with AR13, while a 5-coordinate complex can be achieved at a 1 to 25 molar ratio of AR13 to ASD. a) Titration of AR13 at 6, 12, 60, and 120 µM with the inset that shows a 5 nm shift in the β band from 535 to 545 nm, and loss of the α band at 570 nm. b) Time-dependent increase of P450 in the high-spin state after the addition of 70 µM ASD to a reaction mixture of 2.5 µ AR13.

At 3 µM and 44.7 µM ASD, CO-bound iron was achieved after bubbling 1 mL of CO into a dithionite-reduced reaction mixture. The addition of AR13 at saturating conditions (6 µM) caused a loss of the 448 nm peak. Subsequent reduction and CO-bubbling of the reaction mixture did not result in the reappearance of the 448 nm peak. A fresh reaction of 3.5 µM P450 with 6 µM AR13 caused a soret peak shift from 416 nm to 424 nm (Figure 5.11). The addition of dithionite and CO yielded an A_{423}/A_{447} peak absorbance ratio of 1.14. The ratio increased to 1.49 after the

immediate addition of AR13 for a final concentration of 12 µM. Thirty minutes in aerobic conditions increased this ratio to 2.86. Complete oxidation of dithionite in solution was evident by a loss in absorption at 314 nm after 60 minutes. The peak ratio increased to 4.01, with complete loss of the peak shoulder at 447 nm.

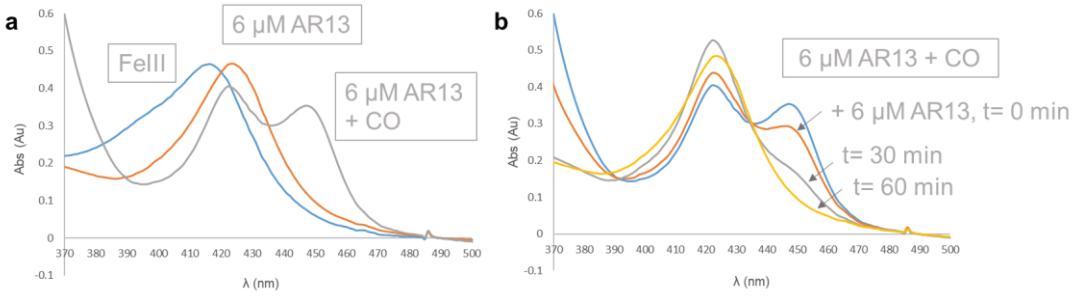


Figure 5.11 Soret shifts in the presence of AR13 and CO indicate reversible competition for heme coordination in active P450.

a) A P450 peak is observed after dithionite and CO addition to a reaction mixture of 3 µM P450 and 6 µM AR13. Subsequent addition of b) 6 µM AR13 results in the recovery of the 424 nm soret peak and complete loss of CO-bound enzyme in a time-dependent manner.

At 2 µM ASD, the soret shift maximum was at 394 nm. Titration of AR11 to 36 µM resulted in the appearance of a shoulder peak near 414 nm. A shift to the 6-coordinate water-bound state was evident at 100 µM, where the shoulder peak became the new peak maximum. Endoxifen was previously reported to display non-competitive kinetics,[54] and used as a control. Enzyme absorption properties were unchanged in the presence of endoxifen at concentrations 50-fold higher than that of ASD.

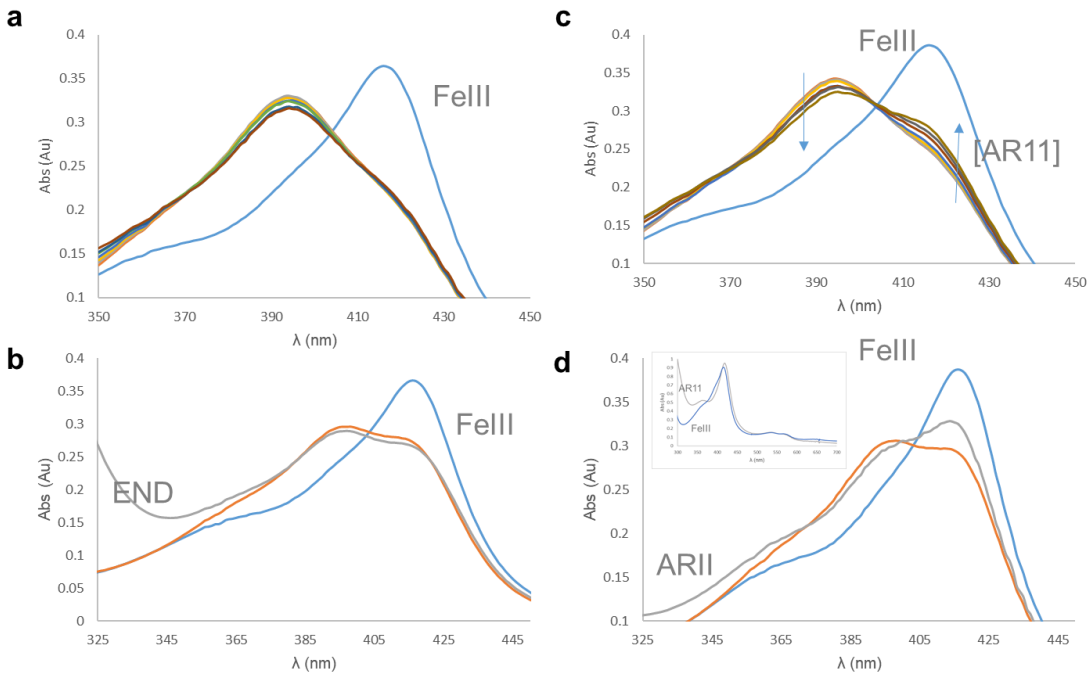


Figure 5.12 Optical absorption spectra of reaction mixtures in the presence of 3 µM P450, 2 µM ASD, and various concentrations of endoxifen a-b and AR11 c-d.

Titration with inhibitor 6 – 36 µM (a and c), and 100 µM (b and d) favor a transition to the 6-coordinate low-spin state for AR11. Arrows indicate the peak absorption trend as the concentration of AR11 increases. Inset in panel d shows CYP19 in the absence and presence of 100 µM AR11.

At 2 µM ASD, and 3 – 6 µM P450, a shoulder peak indicated an equilibrium between the low- and high-spin state of CYP19. The subsequent addition of AR19 or AR20 favored a transition of the heme iron to the low spin state. This behavior mimics that of AR11, with the exception that AR19 induces a weaker hyperchromic stretch. In the absence of substrate, 100 µM AR19 and 20 does not cause a soret peak shift, indicating that water is not displaced upon inhibitor binding at saturating conditions.

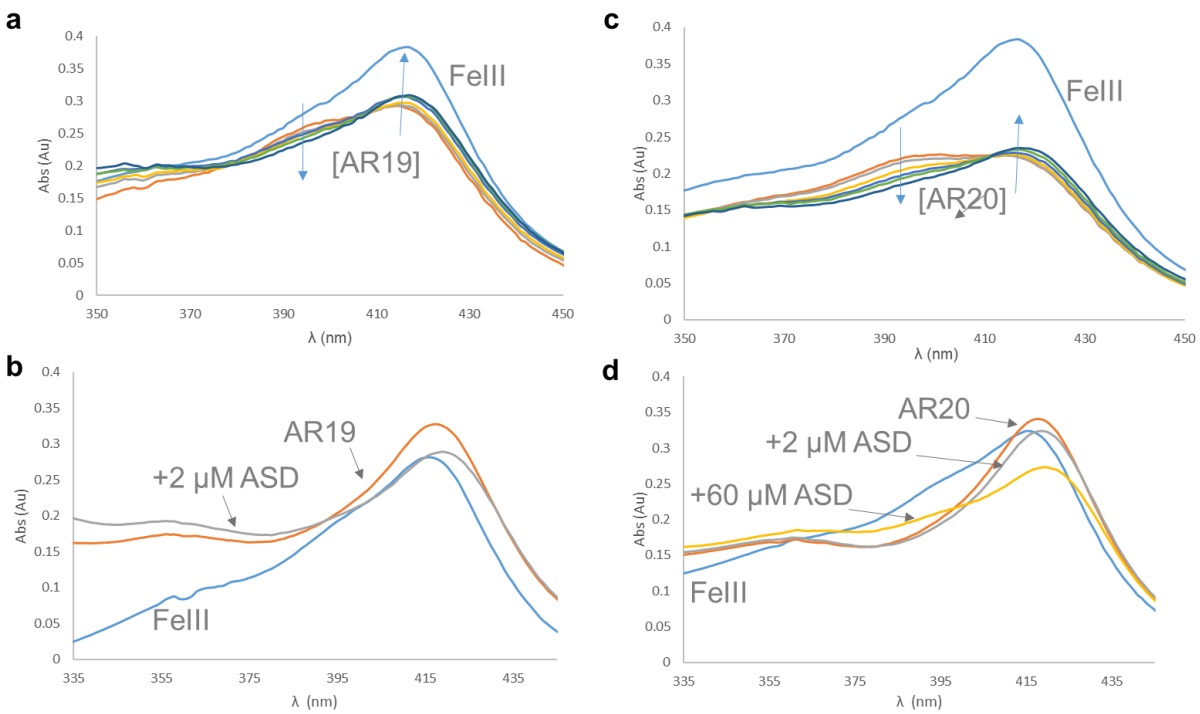


Figure 5.13 Optical absorption spectra of reaction mixtures in the presence of 2 μM P450, 2 μM androstenedione (ASD), and various concentrations of AR19 a-b, and AR20 c-d.

Titration with inhibitor from 6 – 36 μM (a and c), and 100 μM (b and d) favor a transition to the water-bound, 6-coordinate low-spin state. Blue arrows indicate the absorption trend as the concentration of inhibitor increases.

Fluorescence-based reversible inhibition assays

Fluorescence-based assays were performed such that all reactions were within the linear range at 10 nM P450 with the highest MFC concentration at 50 μM due to the inherent solubility limit at 2 % acetonitrile. Nonlinear regression analysis yielded V_{max}, and K_m values 0.689 pmol HFC/min/pmol P450, and 24.38 μM (Figure B.16). A range of inhibitor concentrations was selected, such that no inhibitor would return initial velocities less than 0.2 pmol/min/pmol P450 (~ 1/3 V_{max}) at 50 μM MFC. This increased R^2 values, and consequently the reliability of 5 x 5 Lineweaver-Burk plots, and 4 x 5 Dixon-type plots used to diagnose the inhibition types.

Lineweaver-Burk plots for AR11 and AR13 resulted in curves that intersected the Y-axis at different Y values. These correspond to different apparent V_{max} ($V_{max,app}$). The intersection of curves at different concentrations of MFC from a Dixon plot $1/V_0$ by [I] corresponds to the inhibitor binding affinity in the absence of substrate. The intersection of curves from a plot of $[MFC]/V_0$ by [AR11] corresponds to AR11 binding affinity in the presence of MFC substrate (K_i'). The curves were near parallel for AR13, inferring an α value that approaches infinite (Figure 5.15). As such nonlinear regression for AR11 was fit to a mixed-type inhibition model to return values of K_i and K_i', while AR13 was fit to a competitive-type model to return a K_i value.

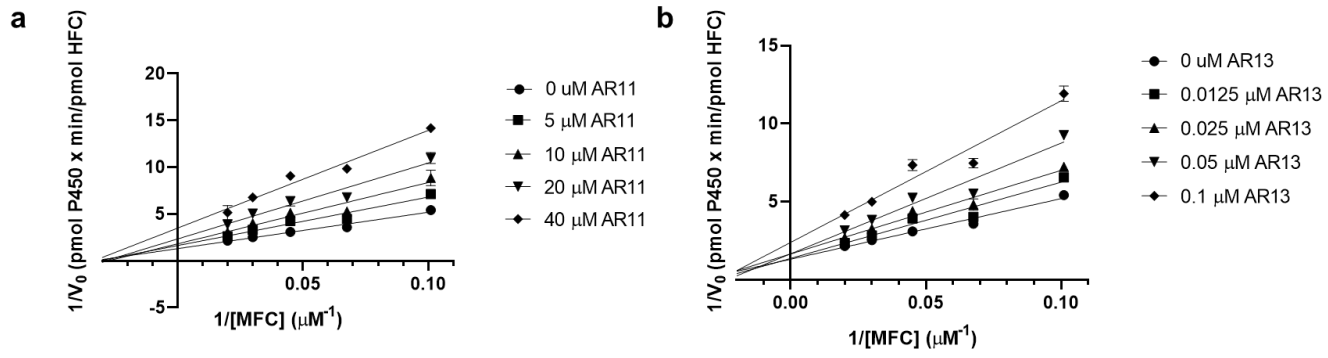


Figure 5.14 Line-Weaver Burke plots for reactions performed in triplicate at 10 nM P450 and various concentrations of inhibitors a) AR11, and b) AR13.

Y intercepts represent reciprocal $V_{max,app}$, and are a crude diagnostic for determining the presence of a competitive inhibitor.

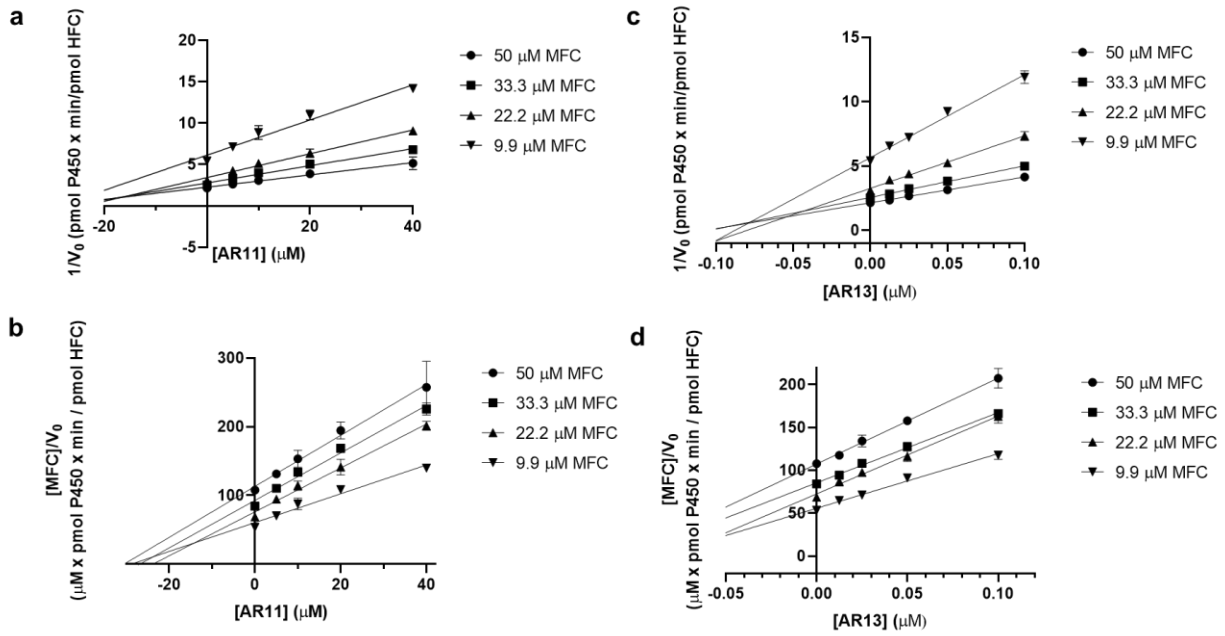


Figure 5.15 Dixon-type plots for a-b) AR11 and c-d) AR13, where the intersection for curves in panels a and c represent the Ki, and the intersection of curves for panels b and d represent the Ki'.

In panel d, curves for AR13 are near parallel and demonstrate that substrate MFC, and inhibitor AR13 bind at a mutually exclusive site.

Kinetic constants we report are returned from a nonlinear regression analysis with AR11 and AR13 fit to their respective inhibitory model (Table 5.5, curves are appended). The apparent Michaelis constant increased in the presence of each inhibitor, while the apparent maximum velocity decreased in the presence of AR11 (Table 5.6).

Table 5.5 Steady-state kinetic constants in the absence and presence of AR11 and AR13.

	V_{max} (pmol HFC/min/pmol P450)	K_m (μM)	K_i (μM)	α	K_i'(μM)
No Inhibitor	0.689	24.38	-	-	-
AR11	0.678	24.60	14.920	2.567	38.300
AR13	0.654	22.01	0.039	-	-

Table 5.6 Observed V_{max} and K_m in the presence of various inhibitor concentrations.

	$V_{max,app}$ (pmol HFC/min/pmol P450)	$K_{m,app}$ (μM)	R^2
[AR11] (μM)			
0	0.678	24.60	0.988
5	0.600	29.05	0.980
10	0.538	32.58	0.953
20	0.446	37.82	0.973
40	0.332	44.29	0.931
[AR13] (nM)			
0	0.654	22.01	0.989
0.0125		29.06	0.981
0.025		36.12	0.984
0.05		50.23	0.978
0.1		78.45	0.960

Discussion

AR13 causes CYP19 absorption trends typical of a type II aromatase inhibitor

Soret peak shifts of CYP 19 in the presence of AR13 indicates the behavior of a type II inhibitor. In the presence of dithionite and CO, an equilibrium exists between CO-bound and inhibitor-bound enzyme. The inhibitor-bound enzyme is favored by increasing AR13 concentration indicated by the loss of a peak at 450 nm. Mechanistically, AR13 acts competitively for direct interaction with the heme iron center. An 8 nm shift in the soret peak from a water-bound iron typifies the direct interaction of a nitrogenous sigma-donor ligand. AR13 has a terminal imidazole moiety common in current AI azoles. Additionally, the enzyme-inhibitor complex does not switch the enzyme to an inactive P420 form, indicating that this is a reversible interaction.

AR11, 19, and 20 exhibit soret shift trends that remove androstenedione from the active site

Androstenedione has reported K_d values that range in the submicromolar to the micromolar range (table of reported K_d values with sources from Sohl and Guen 0.13 ± 0.07 μM, and Kagawa 0.86 ± 0.04 μM). In the presence of ASD, the iron of the heme center adopts a 5-coordinate high spin state recognized by a soret peak at 394 nm. AR11 titration of a solution with enzyme-bound ASD causes the appearance of a shoulder peak near 414 nm, indicating a change to a 6-coordinate water-bound iron. As such, AR11 may act as a competitive inhibitor in which the 6-coordinate site is accessible to solvent, or it acts by allostery such that the energy landscape of the active site is altered to favor dissociation of the endogenous substrate. AR19 and 20 do not interact with the heme group or cause water displacement upon binding. The addition of androstenedione causes a shift in the equilibrium from the water-bound state to the 5-coordinate state. From this, we understand that androgens and our inhibitors are mutually exclusive.

AR11 exhibits the kinetic behavior of a mixed inhibitor, while AR13 demonstrates competitive inhibition by a tight-binder

Lineweaver-Burk plots were initially evaluated to diagnose the mode of inhibition for AR11 and AR13. The intersection of linear regression lines at each concentration inhibitor did not have the same value and therefore appeared to have different V_{max} values. This would indicate non-mutual exclusivity, thereby acting as mixed-type inhibitors and consequently binding at an allosteric site. Data for AR13 did not correlate with the absorption data that would indicate Fe interaction with a stronger field ligand than water. When nonlinear regression curves were fit to a mixed type inhibitory mode, the K_i was 46 – 85 nM at a 95 % confidence interval. At 10 nM P450,

the estimated K_i was within an order of magnitude from the concentration of enzyme. As such, trends from a Lineweaver-Burk plot would resemble that of a mixed inhibitor rather than a competitive one. For this reason, the Y values begin to converge as x approaches zero. The curvature is recognizable at high substrate concentrations for curves at high inhibitor concentrations. Therefore, a nonlinear regression analysis was fit to a competitive-type mechanism to return a K_i of 39 nM, 35 – 44 nM 95 % confidence interval. This value should be taken as an approximation since steady-state conditions are not maintained for K_i / [enzme] less than 1000.

When linear transformations are performed, extra caution should always be taken as to which variables are weighted heavier. For example, in a Lineweaver-Burk analysis, velocities at lower substrate concentrations weight heavier on the outcome of the plot. Therefore, Dixon plots were also used during these analyses since they offer greater reliability in determining inhibition mode.[84] Figure 5.15 panel d demonstrates that the linear functions do not intersect. Consequently, there is no K_i' value because the α coefficient ($K_i' = \alpha K_i$) approaches infinite. The high α value indicates a preference to bind the substrate-free form of Cyp19, or in this case, only the substrate free form because the substrate and inhibitor are mutually exclusive. Altogether, AR13 behaves as a reversible semi tight-binding inhibitor. AR13 does not exhibit time-dependent behavior. Therefore, either the predicted K_i is orders of magnitude too low, or the reaction conditions were not optimal for peak activity. The latter is likely to be the case since this level of potency is also observed in the IC_{50} values when a high-throughput kit was used during the initial screening. Docking of both trans-AR13 enantiomers in the active site reveals that the molecule spans the length of the active site, and it makes a pi interaction with F221. Further, direct interaction with the iron contributes to its potency. The stereoisomer in Figure 5.16 panel b favors a binding mode

by which the sp_2 hybridized nitrogen is closer to interact with iron, although direct interaction would require at least a 2.5 Å distance.

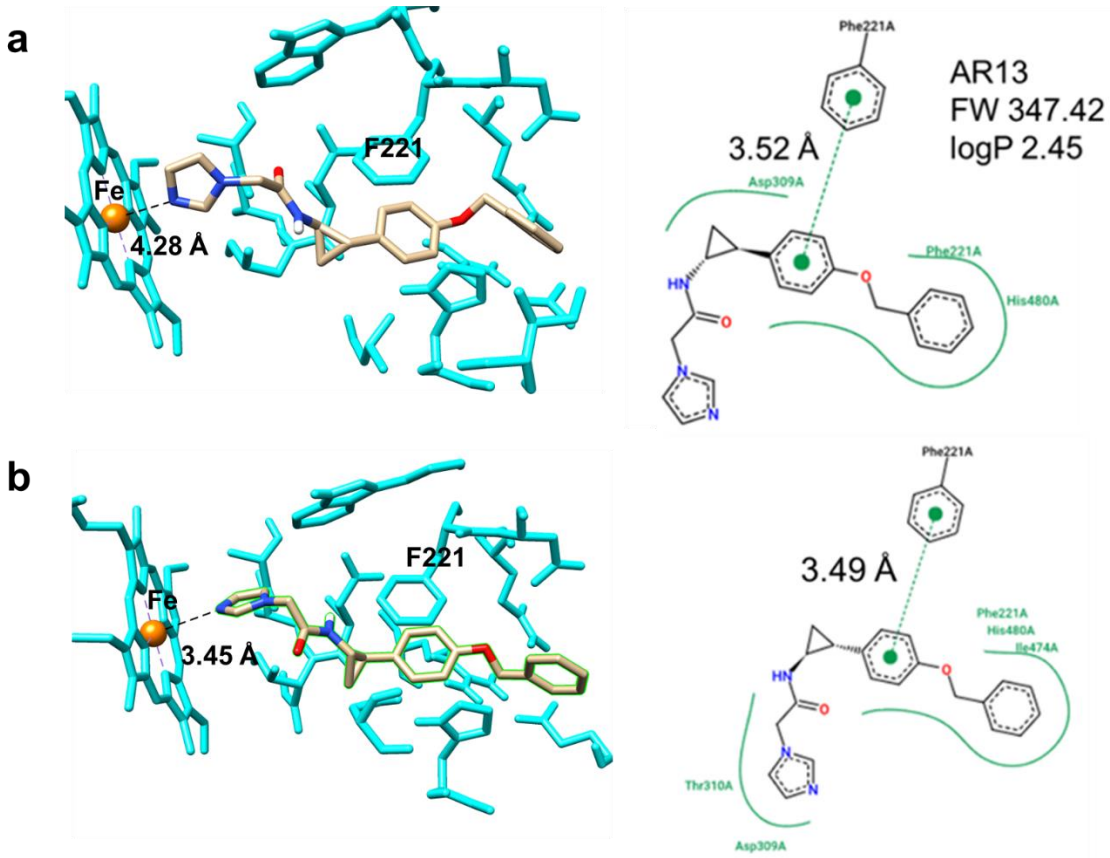


Figure 5.16 Enantiomers of trans-AR13 docked in Cyp19 active site.

Shape complementarity and a critical pi interaction with F221 contribute to AR13 potency. The stereoisomer in (a) does not allow nitrogen to approach iron at a distance for direct metal-ligand interaction. The distance of this structure was modeled 4.3 Å away, while the molecule in (b) was close enough for an end-on interaction at 3.5 Å.

Nonlinear regression analyses for AR11 correlated well with a mixed inhibitor. Dixon plots revealed intersecting lines in panels a and b of Figure 5.15. These values correspond to the K_i values of the enzyme in the substrate-bound and unbound form. The α value is 2.57, indicating a preference for the substrate-free enzyme. I should note here that in Figure 5.15 panel b, the

absence of the trace representing [AR11] = 9.9 μM, would yield a graph that demonstrates no intersections or value for K_i'. Should this be the case that AR11 acts competitively, the binding mode would be one that induces a change in the heme environment since the soret peak exhibits a 2 – 3 nm red shift. This type of shift is observed in 1,2,3-triazoles and some 1,2,4-triazoles that form water-bridged ternary complexes.[85] In these cases, water is more basic and therefore acts as a stronger field ligand that induces only slightly red-shifted soret, β, and α peaks. AR11 is a 3, 5-disubstituted 1,2,4-triazole where direct heme coordination is not spatially possible. Structurally, this molecule is exciting in that either of the three nitrogen can act as a proton donor since the ring substitutions occur at the two carbons.

Mutants Y361W, K440A, and Y441V were unstable relative to the wildtype enzyme. Nickel-purified protein solutions loss the characteristic deep red color of heme-containing proteins, and the enzyme had a diminished capacity to bind CO- a hallmark to measure active cytochrome P450 content. Y361W may have sterically agitated helix K, while Y441V may have interrupted the network of pi-pi interactions in the proximal heme pocket. Perhaps, the most noticeable effect with the K440A is the abolishment of an electrostatic interaction with the G431 CO backbone of the K"-L loop. It contains C437 thiolate that binds the heme iron. Further R435 protrudes from the backbone, forming a network of charged-polar interactions, one of these being a carboxylate tail of the heme group. In this work, the K440A mutant perturbed Fe-enzyme coordination such that no appreciable amount of P450 was measurable. Therefore, this mutant was not suitable for binding studies.

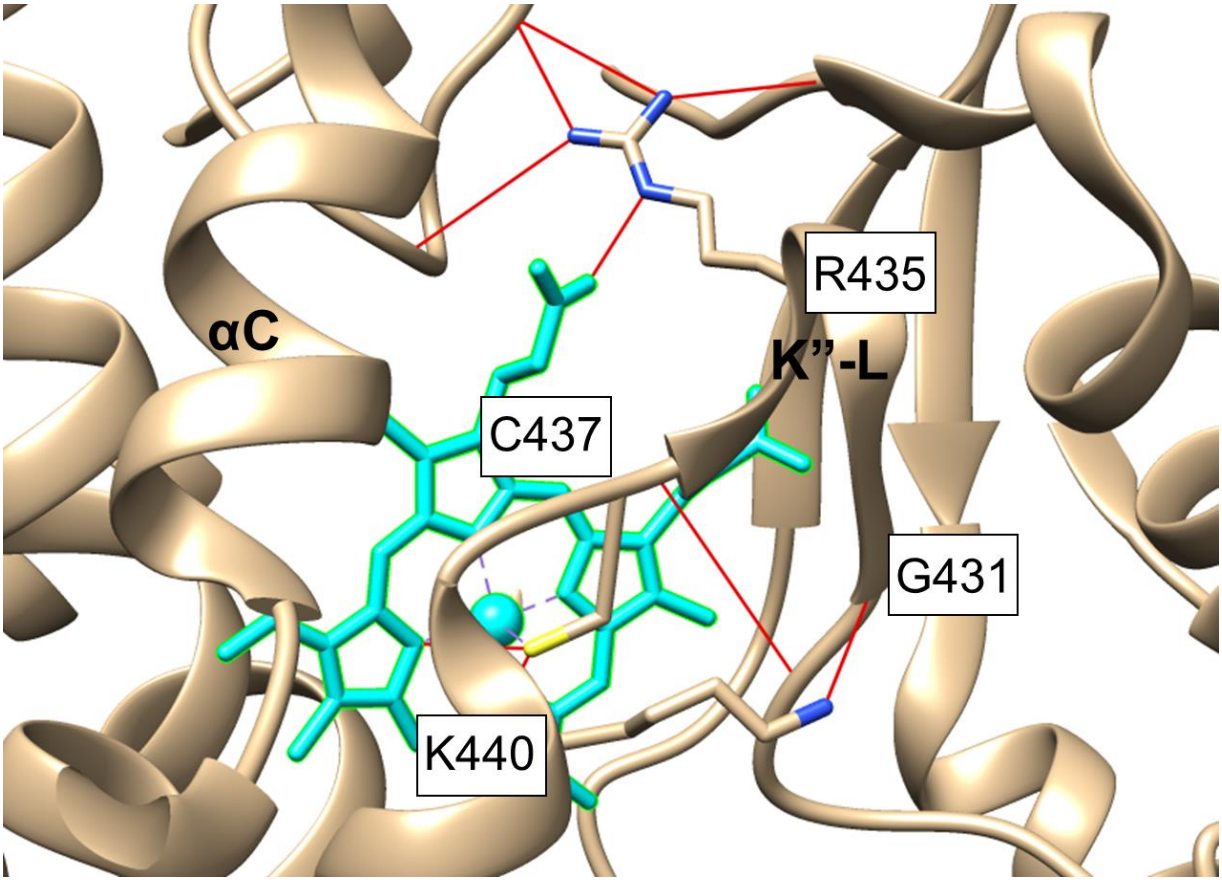


Figure 5.17 Ribbon representation of the proximal heme site with intramolecular interactions at the K"-L loop that contributes to heme stability.

Hydrogen bonding is highlighted as a red trace. Loss of the charged-polar interaction between K440 ε-amine and G431 CO destabilizes the K”-L loop such that hydrogen bonds between the backbone of the loop, and R435-COO of a heme carboxylate are perturbed.

AR19 and AR20 are novel lead compounds for Cyp19 regulation

AR19 and AR20 exhibited anti-aromatase activity with IC_{50} values on the same order of magnitude as endoxifen and AR11. However, the fractional activity at higher concentrations was extrapolated to predict the activity at saturation. AR20 is a trisubstituted 1,2,4- triazole, that would likely bind the heme iron if any two of the substituents were absent since small, monosubstituted azoles appear to bind hemoproteins indiscriminately. Whether AR20 binds competitively or

allosterically, this illustrates that active trisubstituted 1,2,4-triazoles have the potential to uniquely interact with Cyp19 creating a novel subclass of non-steroidal AIs. Additionally, as a triazole, it is metabolically stable with the opportunity to form pi interactions and the ability to act as a proton acceptor at multiple sites. Further, its carbon skeleton has a distinct framework from other active 1,2,4- triazoles in that most are mono and disubstituted. For example, current NSAIs in clinical use, letrozole, and anastrozole are monosubstituted. This trend is also observed in 1,2,3- triazoles. This drug class has gained considerable attention for its click chemistry and high yield.[85] Current 1,2,3- triazole sulfonamides show IC_{50} values in the submicromolar range and have their cytotoxicities characterized.[86,87] Further, letrozole has been modified to a 1,2,3-triazole, and it was found that the corresponding 1,2,5- analog was inactive- inferring that the nitrogen at the 3 and 4 positions was essential for activity. Likely, because they shared a similar binding mode- which brings me to my point. As a trisubstituted 1,2,4- triazole, AR20 is likely to exhibit a distinct binding mode, and as such, could serve as a scaffold to build other drug-like compounds.

AR19 is the only non-azole inhibitor in this work. It is a furan-containing lactam with an IC_{50} analogous to AR20. Yet this value is somewhat misleading since AR19 does not inhibit the enzyme in its entirety. Instead, it reduces and maintains 14 % activity, assuming a hill slope of 1. Docked in the allosteric site, AR19 had only one identifiable hydrogen bond with N428. Despite this, it complemented the subpocket well enough for Van der Waals interactions to make larger contributions. AR19 is a lead compound that can serve as a template for non-azole Cyp19 therapeutics. The surface representation in Figure 5.18 shows the spatial feasibility for chemical modifications at the 1,3 disubstituted phenyl group.

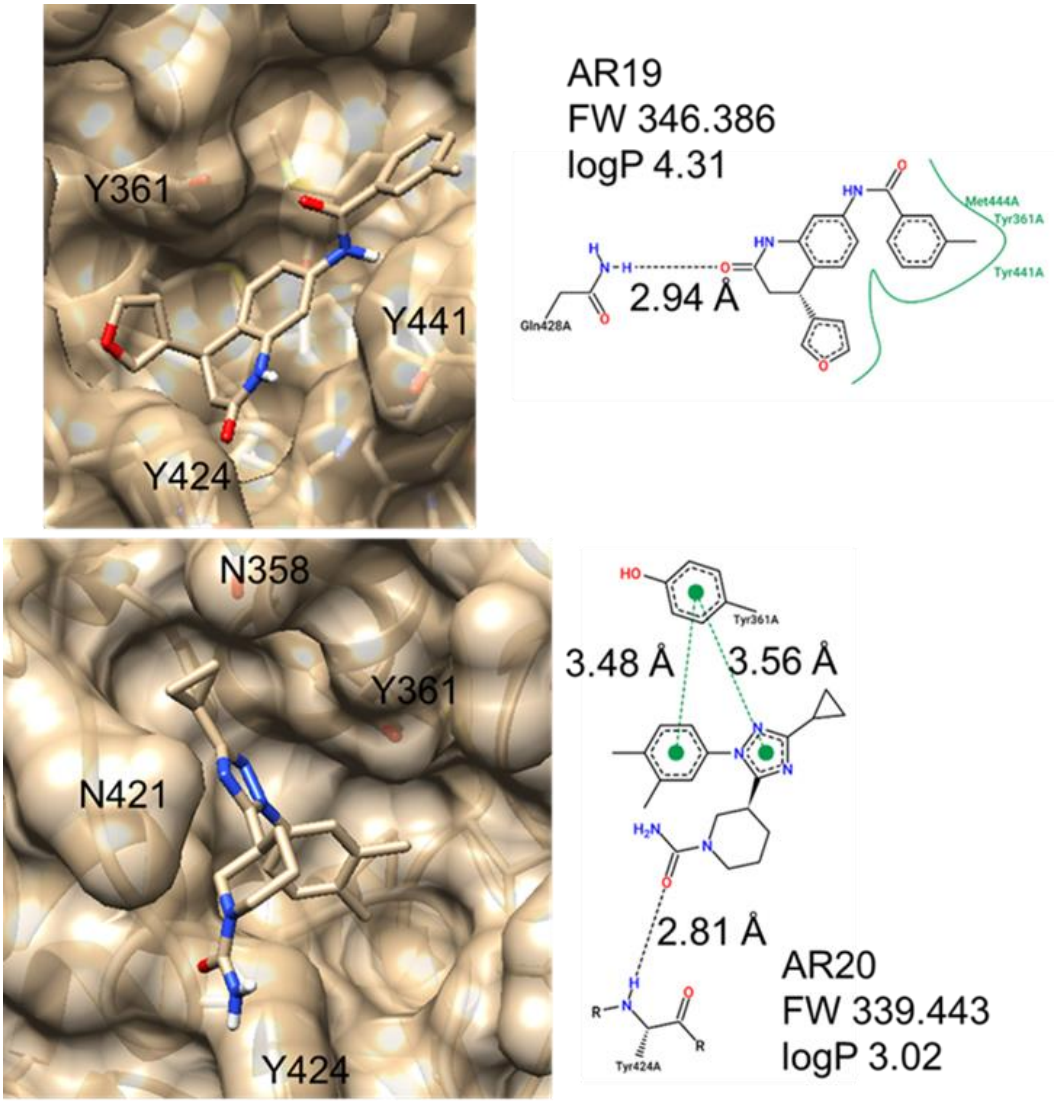


Figure 5.18 AR19 and AR20 docked in the Cyp19 active site.

Surface representation maps show that both inhibitors occupy sub pockets with room for chemical modifications.

Opportunities and ongoing work with AR11, AR13, and AR20

Although we originally set out to find allosteric inhibitors, I include AR13 as part of this discussion since its carbon skeleton is geometrically distinct from current third-generation NSAIs, anastrozole and letrozole. Currently, we understand that AR11 and AR13 interact reversibly at Cyp19 and potency on the same order of magnitude or better than endoxifen. However, in working

with microsomes, there are variables inherent to the system that must be addressed. Although the microsomes are enriched with active P450 and reductase, endogenous enzymes may contribute to metabolic products that may be active or inactive against Cyp19. For this reason, control microsomes are made available by the manufacturer. In the unlikely event that this is the case, work to quantify and chemically determine these potential products is ongoing. Further, the variable that may contribute to variances in the IC_{50}, and kinetic constants are substrate and inhibitor sequestration in the microsomes. Although reactions in the presence and absence of control microsomes can be performed, there is no definitive way to ensure that the amount of free substrate in solution is available. To this end, kinetic constants should always be interpreted with caution in working with these systems. Additionally, as an enzyme-coupled reaction, there is less control over the system- the reductase is assumed to exist at saturating levels. Altogether, the convenience of using microsomes to study P450 activity has a suite of inherent costs that must be addressed.

Moving forward, we need to measure the level of anticancer activity in the MCF7-aro[88] cell line. This cell line is ERα dependent and transfected to overexpress aromatase. Following this, in the absence of a crystal structure, two avenues need to be explored to justify the need for studying these lead compounds. The first is to assay potential activity against AI-resistant cell lines. The second is to observe any cytotoxic effects in control cell lines such as HELA and HEK293 cells. Accordingly, drug-drug interactions need to be evaluated since azoles are notorious for off-target effects. Antifungal triazoles are often administered by immunocompromised individuals, requiring careful surveillance.[89] Drugs such as itraconazole and fluconazole are known to act as substrate and/or inhibitors of metabolic enzymes, drug transporters, and nuclear receptors.[89]

Limitations and opportunities for Cyp19 crystallization with allosteric inhibitors

There are crystal structures of human aromatase from both an *E. coli* recombinant expression system and human placenta in the presence and absence of active site substrates. Therefore, we know that Cyp19 is a crystallizable protein target. However, the enzyme's inability to maintain a stable heme in the absence of a ligand in the active site is an inherent drawback to attaining protein crystals with allosteric inhibitors. We were unable to grow crystals by conditions currently published. We expect that interaction at the allosteric site will interrupt the lattice if a soaking method was employed. Therefore, in our current study, we performed sparse-matrix screens in the presence of inhibitors AR11 and AR13.

We designed and expressed mutants predicted to increase the likelihood of attaining crystals- by increasing enzyme stability or reducing surface entropy. However, performing initial crystal screens was made difficult by our inability to attain milligram quantities of pure protein. This result was due to low expression levels, and a purification process exacerbated by the loss of substantial amounts of protein from successive polishing routines. After each purification step, the progressive loss of Cyp19 was evident by a diminished band intensity in SDS-PAGE gels and the loss of the hemoprotein's characteristic red color. We simplified the purification process and increased expression levels by producing an MBP fusion protein. However, tag cleavage formed insoluble Cyp19 aggregates since MBP was only transiently stabilizing. The introduction of 200 mM NaCl, at a saturating concentration of ASD substrate, and a cocktail of EDTA/E64/PMSF protease inhibitors did not hinder unfolding and subsequent aggregate formation. In working with P450s, renaturation to a functional form after loss of the heme is not a viable procedure. Moving forward, we only analyzed TEV protease activity at time points from 30 min – 6 days in a 30 ˚C

water bath. Therefore, other reaction conditions- such as decreasing the temperature- should be observed. Overcoming this bottleneck will provide convincing evidence that an MBP fusion can be used for the overexpression of functional Cyp19 for further study.

We acknowledge the Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco, with support from NIH P41-GM103311 for backing the UCSF Chimera software used in generating many of the images in this work.

Chapter 6 - Future directions

In this chapter, I briefly highlight the achievements and any opportunities to expand our understanding of each project. I close with remarks on the direction that the drug discovery field appears to be progressing.

BioA

In the *B. subtilis* BioA project, we provided unequivocal evidence that lysine donates its ε-amine to PLP in the first half-reaction of transamination. We proposed a mechanism in which a water molecule acts as the catalytic base, and that substrate turnover is likely due to proper alignment of the ε-carbon substituents.

GPER

When we undertook this project, we set out to find a DNA construct that would express well in a recombinant system. The goal was to eventually produce milligram quantities of functional protein for structural studies. After exploring *E. coli*, and [briefly] yeast expression systems to no avail, we turned our attention to cell-free expression. Concurrently, our lab made strides after two years of optimizing GFP overexpression. Therefore, we tailored the system with a few adjustments to the production of GPER. This demonstrated that for the first time, GPER could be produced recombinantly in a functional state. Moving forward, we have not performed the reaction on a preparative scale. In doing so, GPER can be isotopically labeled for binding assays and NMR studies since the synthesis is performed in an open system. Along these lines, cytotoxic effects and endogenous host responses are evaded, which makes CFPS an attractive system for recombinant protein production.

Aromatase

In this project, we identified 4 novel compounds and an AR13 analog- RN1 (current lysine demethylase 1 inhibitor- cytotoxic in ovarian cancer cell lines) that exhibit antiaromtase activity with mild to moderate responses that compare to current lead compound and active metabolite endoxifen. AR13 exhibited competitive behavior with a K_i nearly equivalent to that of the most potent tamoxifen metabolite, norendoxifen. AR11 demonstrated moderate activity by a mixed-mode. AR19 and AR20 showed mild to moderate inhibitory activities with spectral properties that mimic that of AR11. By characterizing how AR11 and AR20 interact with Cyp19, we can better predict how substitutions at the azole ring will affect binding mode, and potencies. The need for new potent AIs is evidenced by successes in the sequential use of steroidal and NSAIs in patients who have incurred AI resistance. Moreover, elucidating protein-ligand interactions of allosteric inhibitors may uncover a new class of AIs that are therapeutic to those with AI resistance on current regimens.

Onward

With the integration of machine learning in drug discovery, uncharacterized proteins can be scored on their likeliness to be druggable.[90] Further, the interactome can be probed to find 'hot spots' at an interface to modulate protein-protein interactions.[91,92] While computational approaches are invaluable in defining the druggable proteome, strides in chemical biology have redefined this space. The onset of protease-targeting chimeras (PROTACs) has provided greater opportunities for therapy to individuals afflicted by illness caused by the dysregulation of proteins that were once considered undruggable. These include proteins where treatments are no longer

effective due to evading mechanisms,[93] and those where treatment is best by protein removal altogether such as tau degradation for treating Alzheimer's Disease.[94] PROTACs are heterobifunctional molecules that interact with target proteins for the recruitment of the E3 ubiquitin ligase complex. In turn, the target protein is ubiquitinated for degradation by the ubiquitin proteosome pathway.[93] Antagonism not by ligand occupancy, but rather degradation ultimately expands what we understand to be druggable.

References

1. 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 settings. *Adv. Drug Deliv. Rev.* **23**, 3–25 (1997).
2. The International Transporter Consortium. Membrane transporters in drug development. *Nat. Rev. Drug Discov.* **9**, 215–236 (2010).
3. Shultz, M. D. Two Decades under the Influence of the Rule of Five and the Changing Properties of Approved Oral Drugs: Miniperspective. *J. Med. Chem.* **62**, 1701–1714 (2019).
4. Du Vigneaud, V., Melville, D. B., Gyorgy, P. & Rose, C. S. On the identity of Vitamin H with biotin. *Science* **92**, 62–63 (1940).
5. Lin, S. & Cronan, J. E. Closing in on complete pathways of biotin biosynthesis. *Mol. Biosyst.* **7**, 1811 (2011).
6. Knowles, J. R. The mechanism of biotin-dependent enzymes. *Annu. Rev. Biochem.* 29 (1989).
7. Zempleni, J., Hassan, Y. I. & Wijeratne, S. S. Biotin and biotinidase deficiency. *Expert Rev. Endocrinol. Metab.* **3**, 715–724 (2008).
8. Yu, J. *et al.* MMAR_2770, a new enzyme involved in biotin biosynthesis, is essential for the growth of Mycobacterium marinum in macrophages and zebrafish. *Microbes Infect.* **13**, 33–41 (2011).
9. Gomez, J. E. & McKinney, J. D. M. tuberculosis persistence, latency, and drug tolerance. *Tuberculosis* **84**, 29–44 (2004).
10. Gray, K. M., Keer, J., Williams, H. D. & Smeulders, M. J. Mutants of Mycobacterium smegmatis impaired in stationary-phase survival. *Microbiology* **146**, 2209–2217 (2000).

11. Mårtensson, U. E. A. *et al.* Deletion of the G protein-coupled receptor 30 impairs glucose tolerance, reduces bone growth, increases blood pressure, and eliminates estradiol-stimulated insulin release in female mice. *Endocrinology* **150**, 687–698 (2009).
12. Prossnitz, E. R. & Barton, M. The G-protein-coupled estrogen receptor GPER in health and disease. *Nat. Rev. Endocrinol.* **7**, 715–726 (2011).
13. Girgert, R., Emons, G. & Gründker, C. Inactivation of GPR30 reduces growth of triple-negative breast cancer cells: possible application in targeted therapy. *Breast Cancer Res. Treat.* **134**, 199–205 (2012).
14. Marjon, N. A., Hu, C., Hathaway, H. J. & Prossnitz, E. R. G Protein–Coupled Estrogen Receptor Regulates Mammary Tumorigenesis and Metastasis. *Mol. Cancer Res.* **12**, 1644–1654 (2014).
15. Sohl, C. D. & Guengerich, F. P. Kinetic Analysis of the Three-step Steroid Aromatase Reaction of Human Cytochrome P450 19A1. *J. Biol. Chem.* **285**, 17734–17743 (2010).
16. Yoshimoto, F. K. & Guengerich, F. P. Mechanism of the Third Oxidative Step in the Conversion of Androgens to Estrogens by Cytochrome P450 19A1 Steroid Aromatase. *J. Am. Chem. Soc.* **136**, 15016–15025 (2014).
17. Cocconi, G. First generation aromatase inhibitors- aminoglutethimide and testololactone. *Breast Cancer Res. Treat.* **30**, 57–80 (1994).
18. Ghosh, D., Jiang, W., Lo, J. & Egbuta, C. Higher order organization of human placental aromatase. *Steroids* **76**, 753–758 (2011).
19. Lo, J. *et al.* Structural Basis for the Functional Roles of Critical Residues in Human Cytochrome P450 Aromatase. *Biochemistry* **52**, 5821–5829 (2013).

20. Ghosh, D., Egbuta, C. & Lo, J. Testosterone complex and non-steroidal ligands of human aromatase. *J. Steroid Biochem. Mol. Biol.* **181**, 11–19 (2018).
21. Agarwal, V., Lin, S., Lukk, T., Nair, S. K. & Cronan, J. E. Structure of the enzyme-acyl carrier protein (ACP) substrate gatekeeper complex required for biotin synthesis. *Proc. Natl. Acad. Sci.* **109**, 17406–17411 (2012).
22. Dey, S., Lane, J. M., Lee, R. E., Rubin, E. J. & Sacchettini, J. C. Structural Characterization of the *Mycobacterium tuberculosis* Biotin Biosynthesis Enzymes 7,8-Diaminopelargonic Acid Synthase and Dethiobiotin Synthetase,. *Biochemistry* **49**, 6746–6760 (2010).
23. Manandhar, M. & Cronan, J. E. A Canonical Biotin Synthesis Enzyme, 8-Amino-7-Oxononanoate Synthase (BioF), Utilizes Different Acyl Chain Donors in Bacillus subtilis and Escherichia coli. *Appl. Environ. Microbiol.* **84**, e02084-17 (2017).
24. Eisenberg, M. A. & Stoner, G. L. Biosynthesis of 7, 8-Diaminopelargonic Acid, a Biotin Intermediate, from 7-Keto-8- Aminopelargonic Acid and S-Adenosyl-L- M4ethionine. *J. Bacteriol.* 6 (1971).
25. Kabsch, W. XDS. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 125–132 (2010).
26. McCoy, A. J. *et al.* Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674 (2007).
27. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 486–501 (2010).
28. Murshudov, G. N., Vagin, A. A. & Dodson, E. J. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D Biol. Crystallogr.* **53**, 240–255 (1997).

29. Murshudov, G. N. *et al.* REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr. D Biol. Crystallogr.* **67**, 355–367 (2011).
30. Winn, M. D. *et al.* Overview of the CCP4 suite and current developments. *Acta Crystallogr. D Biol. Crystallogr.* **67**, 235–242 (2011).
31. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 486–501 (2010).
32. Potterton, E., Briggs, P., Turkenburg, M. & Dodson, E. A graphical user interface to the CCP4 program suite. *Acta Crystallogr. D Biol. Crystallogr.* **59**, 1131–1137 (2003).
33. Lebedev, A. A. *et al.* JLigand: a graphical tool for the CCP4 template-restraint library. *Acta Crystallogr. D Biol. Crystallogr.* **68**, 431–440 (2012).
34. Pettersen, E. F. *et al.* UCSF Chimera- A visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).
35. Hayashi, H. Pyridoxal Enzymes: Mechanistic Diversity and Uniformity. *J. Biochem. (Tokyo)* **118**, 463–473 (1995).
36. Ginguay, A., Cynober, L., Curis, E. & Nicolis, I. Ornithine aminotransferase, an important glutamate-metabolizing enzyme at the crossroads of multiple metabolic pathways. *Biology* **6**, 18 (2017).
37. Mani Tripathi, S. & Ramachandran, R. Direct Evidence for a Glutamate Switch Necessary for Substrate Recognition: Crystal Structures of Lysine ε-Aminotransferase (Rv3290c) from Mycobacterium tuberculosis H37Rv. *J. Mol. Biol.* **362**, 877–886 (2006).
38. Dunathan, H. C. Conformation and reaction specificity in pyridoxal phosphate enzymes. *Proc. Natl. Acad. Sci.* **55**, 712–716 (1966).

39. Kumar, R. & Thompson, E. B. Transactivation functions of the N-terminal domains of nuclear hormone receptors: Protein folding and coactivator interactions. *Mol. Endocrinol.* **17**, 1–10 (2003).

40. Pietras, R. J. & Szego, C. M. Specific binding sites for oestrogen at the outer surfaces of isolated endometrial cells. *Nature* **265**, 69–72 (1977).

41. Filardo, E. J., Quinn, J. A. & Frackelton, A. R. Estrogen action via the G protein-coupled receptor, GPR30: Stimulation of adenylyl cyclase and cAMP-mediated attenuation of the epidermal growth factor receptor-to-MAPK signaling axis. 15.

42. Bologa, C. G. *et al.* Virtual and biomolecular screening converge on a selective agonist for GPR30. *Nat. Chem. Biol.* **2**, 207–212 (2006).

43. Revankar, C. M., Cimino, D. F., Sklar, L. A., Arterburn, J. B. & Prossnitz, E. R. A Transmembrane Intracellular Estrogen Receptor Mediates Rapid Cell Signaling. **307**, 7 (2005).

44. Drew, D. *et al.* GFP-based optimization scheme for the overexpression and purification of eukaryotic membrane proteins in Saccharomyces cerevisiae. *Nat. Protoc.* **3**, 784–798 (2008).

45. Kim, T.-W. *et al.* Simple procedures for the construction of a robust and cost-effective cell-free protein synthesis system. *J. Biotechnol.* **126**, 554–561 (2006).

46. Oliver, R. C. *et al.* Tuning micelle dimensions and properties with binary surfactant mixtures. *Langmuir* **30**, 13353–13361 (2014).

47. Minde, D. P., Maurice, M. M. & Rüdiger, S. G. D. Determining biophysical protein stability in lysates by a Fast Proteolysis Assay, FASTpp. *PLoS ONE* **7**, e46147 (2012).

48. Na, Y.-R. & Park, C. Investigating protein unfolding kinetics by pulse proteolysis. *Protein Sci.* **18**, 268–276 (2009).

49. Viegas, A., Manso, J., Nobrega, F. L. & Cabrita, E. J. Saturation-transfer difference (STD) NMR: A simple and fast method for ligand screening and characterization of protein binding. *J. Chem. Educ.* **88**, 990–994 (2011).

50. Sirim, D., Widmann, M., Wagner, F. & Pleiss, J. Prediction and analysis of the modular structure of cytochrome P450 monooxygenases. *BMC Struct. Biol.* **10**, 34 (2010).

51. Hong, Y., Li, H., Yuan, Y.-C. & Chen, S. Molecular characterization of aromatase. *Ann. N. Y. Acad. Sci.* **1155**, 112–120 (2009).

52. Hong, Y., Li, H., Yuan, Y.-C. & Chen, S. Sequence–function correlation of aromatase and its interaction with reductase. *J. Steroid Biochem. Mol. Biol.* **118**, 203–206 (2010).

53. Connick, J. P., Reed, J. R. & Backes, W. L. Characterization of interactions among CYP1A2, CYP2B4, and NADPH-cytochrome P450 reductase: Identification of specific protein complexes. *Drug Metab. Dispos.* **46**, 197–203 (2018).

54. Lu, W. J., Desta, Z. & Flockhart, D. A. Tamoxifen metabolites as active inhibitors of aromatase in the treatment of breast cancer. *Breast Cancer Res. Treat.* **131**, 473–481 (2012).

55. Fauman, E. B., Rai, B. K. & Huang, E. S. Structure-based druggability assessment—identifying suitable targets for small molecule therapeutics. *Curr. Opin. Chem. Biol.* **15**, 463–468 (2011).

56. Halgren, T. A. Identifying and Characterizing Binding Sites and Assessing Druggability. *J. Chem. Inf. Model.* **49**, 377–389 (2009).

57. Volkamer, A., Kuhn, D., Rippmann, F. & Rarey, M. DoGSiteScorer: a web server for automatic binding site prediction, analysis and druggability assessment. *Bioinformatics* **28**, 2074–2075 (2012).

58. Kozakov, D. *et al.* The ClusPro web server for protein–protein docking. *Nat. Protoc.* **12**, 255–278 (2017).

59. Kozakov, D. *et al.* How good is automated protein docking? *Proteins Struct. Funct. Bioinforma.* **81**, 2159–2166 (2013).

60. Vajda, S. *et al.* New additions to the ClusPro server motivated by CAPRI: Development of the ClusPro Server. *Proteins Struct. Funct. Bioinforma.* **85**, 435–444 (2017).

61. de Vries, S. J. & Bonvin, A. M. J. J. CPORT: A Consensus Interface Predictor and Its Performance in Prediction-Driven Docking with HADDOCK. *PLoS ONE* **6**, e17695 (2011).

62. Janin, J. *et al.* CAPRI: A Critical Assessment of PRedicted Interactions. *Proteins Struct. Funct. Genet.* **52**, 2–9 (2003).

63. de Vries, S. J., van Dijk, M. & Bonvin, A. M. J. J. The HADDOCK web server for data-driven biomolecular docking. *Nat. Protoc.* **5**, 883–897 (2010).

64. Campelo, D. *et al.* The Hinge Segment of Human NADPH-Cytochrome P450 Reductase in Conformational Switching: The Critical Role of Ionic Strength. *Front. Pharmacol.* **8**, 755 (2017).

65. Hamdane, D. *et al.* Structure and Function of an NADPH-Cytochrome P450 Oxidoreductase in an Open Conformation Capable of Reducing Cytochrome P450. *J. Biol. Chem.* **284**, 11374–11384 (2009).

66. Aigrain, L., Pompon, D., Moréra, S. & Truan, G. Structure of the open conformation of a functional chimeric NADPH cytochrome P450 reductase. *EMBO Rep.* **10**, 742–747 (2009).

67. Kastritis, P. L., Rodrigues, J. P. G. L. M., Folkers, G. E., Boelens, R. & Bonvin, A. M. J. J. Proteins feel more than they see: Fine-tuning of binding affinity by properties of the non-interacting surface. *J. Mol. Biol.* **426**, 2632–2652 (2014).

68. Vangone, A. & Bonvin, A. M. Contacts-based prediction of binding affinity in protein–protein complexes. *eLife* **4**, e07454 (2015).

69. Xue, L. C., Rodrigues, J. P., Kastritis, P. L., Bonvin, A. M. & Vangone, A. PRODIGY: a web server for predicting the binding affinity of protein–protein complexes. *Bioinformatics* btw514 (2016) doi:10.1093/bioinformatics/btw514.

70. *The PyMOL Molecular Graphics System, Version 2.1.0 Schrodinger, LLC.*

71. Nebert, D. W., Wikvall, K. & Miller, W. L. Human cytochromes P450 in health and disease. *Philos. Trans. R. Soc. B Biol. Sci.* **368**, 20120431 (2013).

72. Munro, A. W. *et al.* P450 BM3: the very model of a modern flavocytochrome. *Trends Biochem. Sci.* **27**, 250–257 (2002).

73. van Zundert, G. C. P. *et al.* The HADDOCK2.2 Web Server: User-Friendly Integrative Modeling of Biomolecular Complexes. *J. Mol. Biol.* **428**, 720–725 (2016).

74. Shimada, T., Mernaugh, R. L. & Guengerich, F. P. Interactions of mammalian cytochrome P450, NADPH-cytochrome P450 reductase, and cytochrome b5 enzymes. *Arch. Biochem. Biophys.* **435**, 207–216 (2005).

75. Sgrignani, J., Bon, M., Colombo, G. & Magistrato, A. Computational approaches elucidate the allosteric mechanism of human aromatase inhibition: A novel possible route to small-molecule regulation of CYP450s activities? *J. Chem. Inf. Model.* **54**, 2856–2868 (2014).

76. Lu, W. J. *et al.* The tamoxifen metabolite norendoxifen is a potent and selective inhibitor of aromatase (CYP19) and a potential lead compound for novel therapeutic agents. *Breast Cancer Res. Treat.* **133**, 99–109 (2012).

77. Lv, W., Liu, J., Skaar, T. C., Flockhart, D. A. & Cushman, M. Design and synthesis of norendoxifen analogues with dual aromatase inhibitory and estrogen receptor modulatory activities. *J. Med. Chem.* **58**, 2623–2648 (2015).

78. Neudert, G. & Klebe, G. DSX: A knowledge-based scoring function for the assessment of protein–ligand complexes. *J. Chem. Inf. Model.* **51**, 2731–2745 (2011).

79. Guengerich, F. P., Martin, M. V., Sohl, C. D. & Cheng, Q. Measurement of cytochrome P450 and NADPH–cytochrome P450 reductase. *Nat. Protoc.* **4**, 1245–1251 (2009).

80. Jahandideh, S., Jaroszewski, L. & Godzik, A. Improving the chances of successful protein structure determination with a random forest classifier. *Acta Crystallogr. D Biol. Crystallogr.* **70**, 627–635 (2014).

81. Wei, Y., Thyparambil, A. A. & Latour, R. A. Protein helical structure determination using CD spectroscopy for solutions with strong background absorbance from 190 to 230nm. *Biochim. Biophys. Acta BBA - Proteins Proteomics* **1844**, 2331–2337 (2014).

82. Baravalle, R. *et al.* Impact of R264C and R264H polymorphisms in human aromatase function. *J. Steroid Biochem. Mol. Biol.* **167**, 23–32 (2017).

83. Kagawa, N., Hori, H., Waterman, M. R. & Yoshioka, S. Characterization of stable human aromatase expressed in E. coli. *Steroids* **69**, 235–243 (2004).

84. Kolhatkar, V. & Polli, J. E. Reliability of inhibition models to correctly identify type of inhibition. *Pharm. Res.* **27**, 2433–2445 (2010).

85. Conner, K. P. *et al.* 1,2,3-triazole–heme interactions in cytochrome P450: Functionally competent triazole–water–heme complexes. *Biochemistry* **51**, 6441–6457 (2012).

86. Prachayasittikul, V. *et al.* Discovery of novel 1,2,3-triazole derivatives as anticancer agents using QSAR and in silico structural modification. *SpringerPlus* **4**, 571 (2015).

87. Xu, X., Zhang, L. & Xie, X. Somatostatin receptor type 2 contributes to the self-renewal of murine embryonic stem cells. *Acta Pharmacol. Sin.* **35**, 1023–1030 (2014).

88. Lui, K., Tamura, T., Mori, T., Zhou, D. & Chen, S. MCF-7aro/ERE, a novel cell line for rapid screening of aromatase inhibitors, ERα ligands and ERRα ligands. *Biochem. Pharmacol.* **76**, 208–215 (2008).

89. Nivoix, Y. *et al.* The enzymatic basis of drug-drug interactions with systemic triazole antifungals. *Clin. Pharmacokinet.* **47**, 779–792 (2008).

90. Han, L. Y. *et al.* Support vector machines approach for predicting druggable proteins: recent progress in its exploration and investigation of its usefulness. *Drug Discov. Today* **12**, 304–313 (2007).

91. Bai, F., Morcos, F., Cheng, R. R., Jiang, H. & Onuchic, J. N. Elucidating the druggable interface of protein−protein interactions using fragment docking and coevolutionary analysis. *Proc. Natl. Acad. Sci.* **113**, E8051–E8058 (2016).

92. Morrow, J. K. & Zhang, S. Computational prediction of hot spot residues. 23 (2013).

93. Kargbo, R. B. Treatment of cancer and Alzheimer's Disease by PROTAC degradation of EGFR. *ACS Med. Chem. Lett.* **10**, 1098–1099 (2019).

94. Kargbo, R. B. Treatment of Alzheimer's by PROTAC-Tau Protein Degradation. *ACS Med. Chem. Lett.* **10**, 699–700 (2019).

Appendix A - Copyright permissions

Appendix A acknowledges illustrations reprinted for use in this work.

Figure 1.1 and Figure 3.1 Reused with permission by Springer Nature and Copyright Clearance Center. Springer Nature, Nature Reviews Endocrinology. The G-protein-coupled estrogen receptor GPER in health and disease, Eric R. Prossnitz, Matthias Barton. 2011. The Springer Nature license agreement is provided in the following pages.

Figure 3.1 Springer Nature license no. 4777470390196 applied for use prior to March 23, 2020.

SPRINGER NATURE LICENSE
TERMS AND CONDITIONS

Mar 26, 2020

This Agreement between Samson A Souza ("You") and Springer Nature ("Springer Nature") consists of your license details and the terms and conditions provided by Springer Nature and Copyright Clearance Center.

License Number	4794960822228
License date	Mar 23, 2020
Licensed Content Publisher	Springer Nature
Licensed Content Publication	Nature Reviews Endocrinology
Licensed Content Title	The G-protein-coupled estrogen receptor GPER in health and disease
Licensed Content Author	Eric R. Prossnitz et al
Licensed Content Date	Aug 16, 2011
Type of Use	Thesis/Dissertation
Requestor type	academic/university or research institute
Format	electronic
Portion	figures/tables/illustrations

Number of figures/tables /illustrations	2
High-res required	no
Will you be translating?	no
Circulation/distribution	100 - 199
Author of this Springer Nature content	no
Title	Protein-ligand interactions of druggable protein targets
Institution name	Kansas State University
Expected presentation date	Mar 2020
Portions	Images for Figures 2 and 3.
Requestor Location	Mr. Samson Souza 1941 College Heights Rd #5 MANHATTAN, KS 66502 United States Attn: Mr. Samson Souza
Total	0.00 USD

Terms and Conditions

Springer Nature Customer Service Centre GmbH
Terms and Conditions

This agreement sets out the terms and conditions of the licence (the **Licence**) between you and **Springer Nature Customer Service Centre GmbH** (the **Licensor**). By clicking

'accept' and completing the transaction for the material (**Licensed Material**), you also confirm your acceptance of these terms and conditions.

1. Grant of License

1. 1. The Licensor grants you a personal, non-exclusive, non-transferable, world-wide licence to reproduce the Licensed Material for the purpose specified in your order only. Licences are granted for the specific use requested in the order and for no other use, subject to the conditions below.

1. 2. The Licensor warrants that it has, to the best of its knowledge, the rights to license reuse of the Licensed Material. However, you should ensure that the material you are requesting is original to the Licensor and does not carry the copyright of another entity (as credited in the published version).

1. 3. If the credit line on any part of the material you have requested indicates that it was reprinted or adapted with permission from another source, then you should also seek permission from that source to reuse the material.

2. Scope of Licence

2. 1. You may only use the Licensed Content in the manner and to the extent permitted by these Ts&Cs and any applicable laws.

2. 2. A separate licence may be required for any additional use of the Licensed Material, e.g. where a licence has been purchased for print only use, separate permission must be obtained for electronic re-use. Similarly, a licence is only valid in the language selected and does not apply for editions in other languages unless additional translation rights have been granted separately in the licence. Any content owned by third parties are expressly excluded from the licence.

2. 3. Similarly, rights for additional components such as custom editions and derivatives require additional permission and may be subject to an additional fee. Please apply to Journalpermissions@springernature.com/bookpermissions@springernature.com for these rights.

2. 4. Where permission has been granted **free of charge** for material in print, permission may also be granted for any electronic version of that work, provided that the material is incidental to your work as a whole and that the electronic version is essentially equivalent to, or substitutes for, the print version.

2. 5. An alternative scope of licence may apply to signatories of the STM Permissions Guidelines, as amended from time to time.

3. Duration of Licence

3. 1. A licence for is valid from the date of purchase ('Licence Date') at the end of the relevant period in the below table:

Scope of Licence	Duration of Licence
Post on a website	12 months
Presentations	12 months
Books and journals	Lifetime of the edition in the language purchased

4. Acknowledgement

4. 1. The Licensor's permission must be acknowledged next to the Licenced Material in print. In electronic form, this acknowledgement must be visible at the same time as the figures/tables/illustrations or abstract, and must be hyperlinked to the journal/book's homepage. Our required acknowledgement format is in the Appendix below.

5. Restrictions on use

5. 1. Use of the Licensed Material may be permitted for incidental promotional use and minor editing privileges e.g. minor adaptations of single figures, changes of format, colour and/or style where the adaptation is credited as set out in Appendix 1 below. Any other changes including but not limited to, cropping, adapting, omitting material that affect the meaning, intention or moral rights of the author are strictly prohibited.

5. 2. You must not use any Licensed Material as part of any design or trademark.

5. 3. Licensed Material may be used in Open Access Publications (OAP) before publication by Springer Nature, but any Licensed Material must be removed from OAP sites prior to final publication.

6. Ownership of Rights

6. 1. Licensed Material remains the property of either Licensor or the relevant third party and any rights not explicitly granted herein are expressly reserved.

7. Warranty

IN NO EVENT SHALL LICENSOR BE LIABLE TO YOU OR ANY OTHER PARTY OR ANY OTHER PERSON OR FOR ANY SPECIAL, CONSEQUENTIAL, INCIDENTAL OR INDIRECT DAMAGES, HOWEVER CAUSED, ARISING OUT OF OR IN CONNECTION WITH THE DOWNLOADING, VIEWING OR USE OF THE MATERIALS REGARDLESS OF THE FORM OF ACTION, WHETHER FOR BREACH

OF CONTRACT, BREACH OF WARRANTY, TORT, NEGLIGENCE, INFRINGEMENT OR OTHERWISE (INCLUDING, WITHOUT LIMITATION, DAMAGES BASED ON LOSS OF PROFITS, DATA, FILES, USE, BUSINESS OPPORTUNITY OR CLAIMS OF THIRD PARTIES), AND
WHETHER OR NOT THE PARTY HAS BEEN ADVISED OF THE POSSIBILITY OF SUCH DAMAGES. THIS LIMITATION SHALL APPLY NOTWITHSTANDING ANY FAILURE OF ESSENTIAL PURPOSE OF ANY LIMITED REMEDY PROVIDED HEREIN.

8. Limitations

8. 1. ***BOOKS ONLY:*** Where **'reuse in a dissertation/thesis'** has been selected the following terms apply: Print rights of the final author's accepted manuscript (for clarity, NOT the published version) for up to 100 copies, electronic rights for use only on a personal website or institutional repository as defined by the Sherpa guideline (www.sherpa.ac.uk/romeo/).

9. Termination and Cancellation

9. 1. Licences will expire after the period shown in Clause 3 (above).

9. 2. Licensee reserves the right to terminate the Licence in the event that payment is not received in full or if there has been a breach of this agreement by you.

Appendix 1 — Acknowledgements:

For Journal Content:
Reprinted by permission from [**the Licensor**]: [**Journal Publisher** (e.g. Nature/Springer/Palgrave)] [**JOURNAL NAME**] [**REFERENCE CITATION** (Article name, Author(s) Name), [**COPYRIGHT**] (year of publication)

For **Advance Online Publication papers:**
Reprinted by permission from [**the Licensor**]: [**Journal Publisher** (e.g. Nature/Springer/Palgrave)] [**JOURNAL NAME**] [**REFERENCE CITATION** (Article name, Author(s) Name), [**COPYRIGHT**] (year of publication), advance online publication, day month year (doi: 10.1038/sj.[JOURNAL ACRONYM].)

For Adaptations/Translations:
Adapted/Translated by permission from [**the Licensor**]: [**Journal Publisher** (e.g. Nature/Springer/Palgrave)] [**JOURNAL NAME**] [**REFERENCE CITATION** (Article name, Author(s) Name), [**COPYRIGHT**] (year of publication)

Note: For any republication from the British Journal of Cancer, the following credit line style applies:

Reprinted/adapted/translated by permission from [**the Licensor**]: on behalf of Cancer Research UK: : [**Journal Publisher** (e.g. Nature/Springer/Palgrave)] [**JOURNAL NAME**] [**REFERENCE CITATION** (Article name, Author(s) Name), **[COPYRIGHT]** (year of publication)

For **Advance Online Publication** papers:
Reprinted by permission from The [**the Licensor**]: on behalf of Cancer Research UK: [**Journal Publisher** (e.g. Nature/Springer/Palgrave)] [**JOURNAL NAME**] [**REFERENCE CITATION** (Article name, Author(s) Name), [**COPYRIGHT**] (year of publication), advance online publication, day month year (doi: 10.1038/sj. [JOURNAL ACRONYM])

For Book content:
Reprinted/adapted by permission from [**the Licensor**]: [**Book Publisher** (e.g. Palgrave Macmillan, Springer etc) [**Book Title**] by [**Book author**(s)] [**COPYRIGHT**] (year of publication)

Other Conditions:

Version 1.2

Questions? customercare@copyright.com or +1-855-239-3415 (toll free in the US) or +1-978-646-2777.

Appendix B - Assay development and supporting results by chapter

Chapter 3

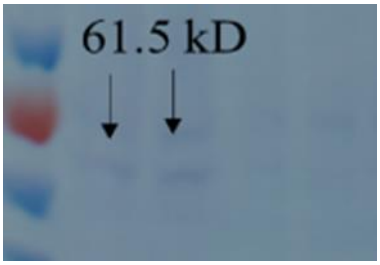


Figure B.1 Blot of GPER-GFP fusion protein with Ni-HRP to probe the His_{10} tag.

Bands were detected at approximately 61.5 kDa (black arrows). From left to right, wells were loaded with prestained laddder, uninduced, induced (+DMSO), uninduced- one freeze thaw, induced (+DMSO)- one freeze-thaw.

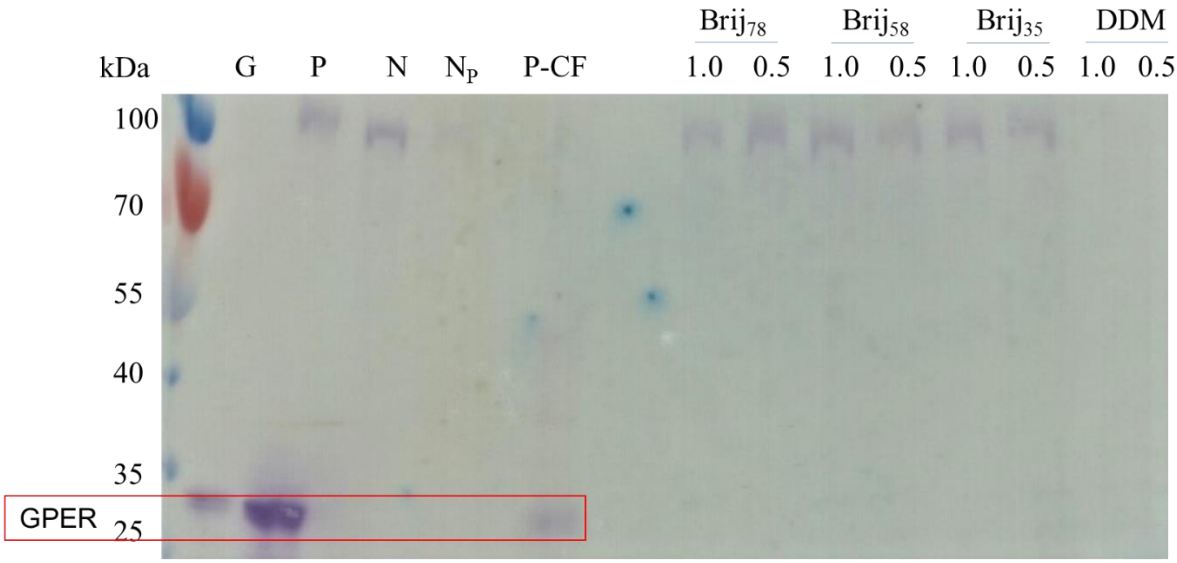


Figure B.2 Blot with His-probe with GPER at apparent mass 27 kDa.

A DCF method did not produce detectable levels of GPER, while a P-CF method showed a band for GPER when it was solubilized with SDS and LMPG detergents. G (GFP sample), P (GPER solubilized with SDS in denaturing conditions), N (supernatant of the cell-free reaction in the absence of DNA), N_P (solubilized pellet of the cell-free reaction in the absence of DNA), P-CF (GPER solubilized with LMPG). The remaining wells are samples from cell-free reactions by a D-CF method with the detergent and concentration noted.

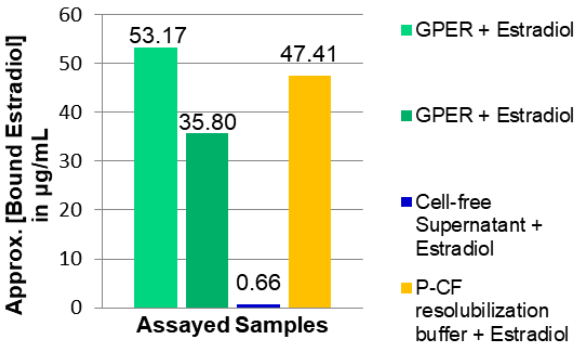


Figure B.3 GPER-bound estradiol as determined by Estradiol EIA Kit.

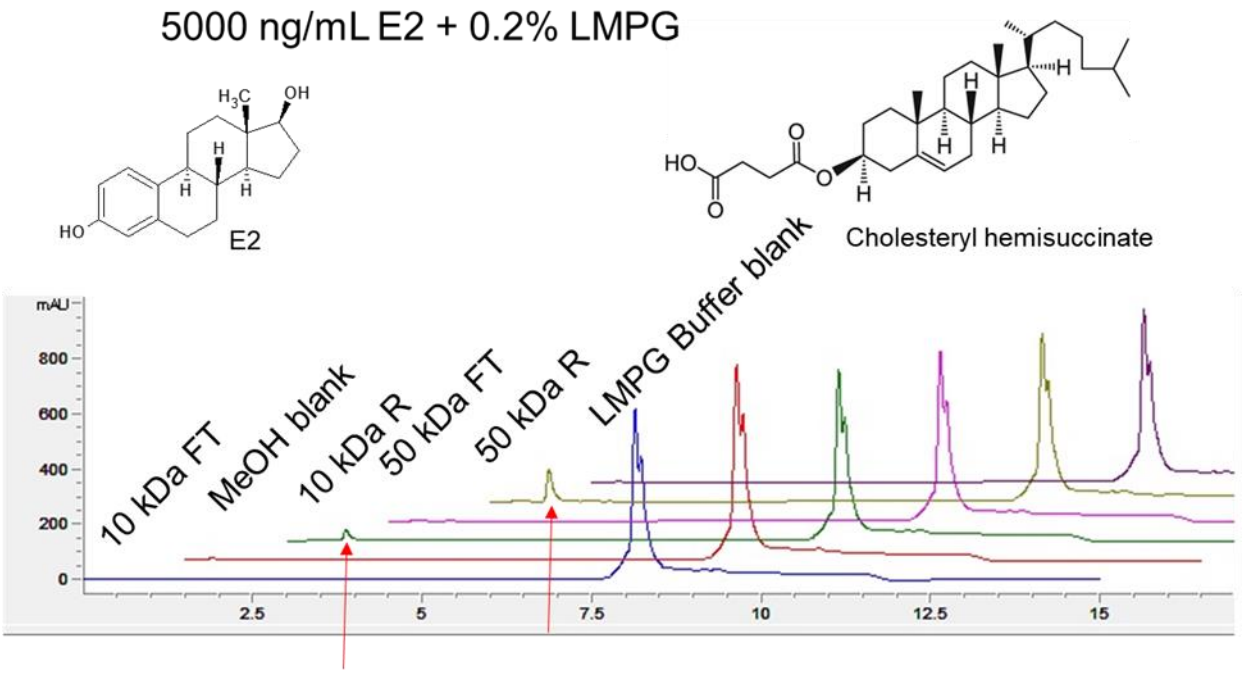


Figure B.4 HPLC chromatogram of control samples with a 10 µL auto injection on column.

The red arrows indicate that estradiol was present in the retentate (R) of 10 and 50 kDa MWCO spin columns. The flow-through (FT) had a substantially lower signal for estradiol at 0.83 min LMPG has an approximate 44 kDa micellar size.[46]

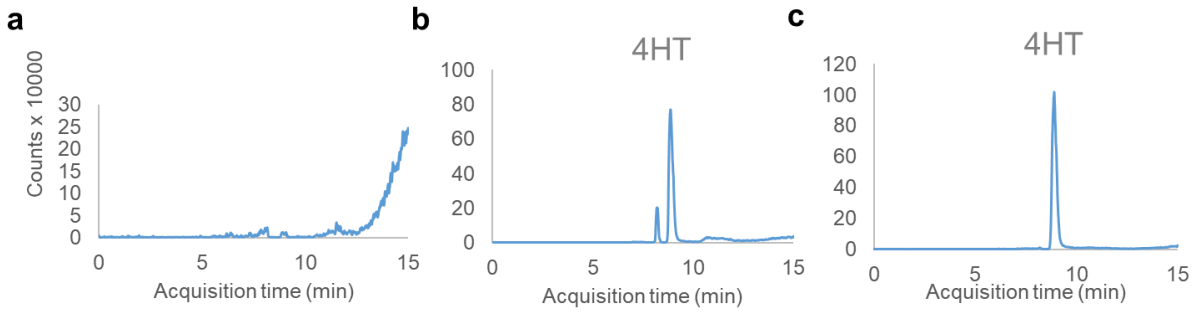


Figure B.5 Extracted ion chromatogram at m/z 372 shows tamoxifen (4HT) in HEPES buffer eluting from the LC column at 8.8 min.

A negative control (a), positive control – buffer + 4HT (b), and GPER + 4HT were loaded onto the column with an autosampler.

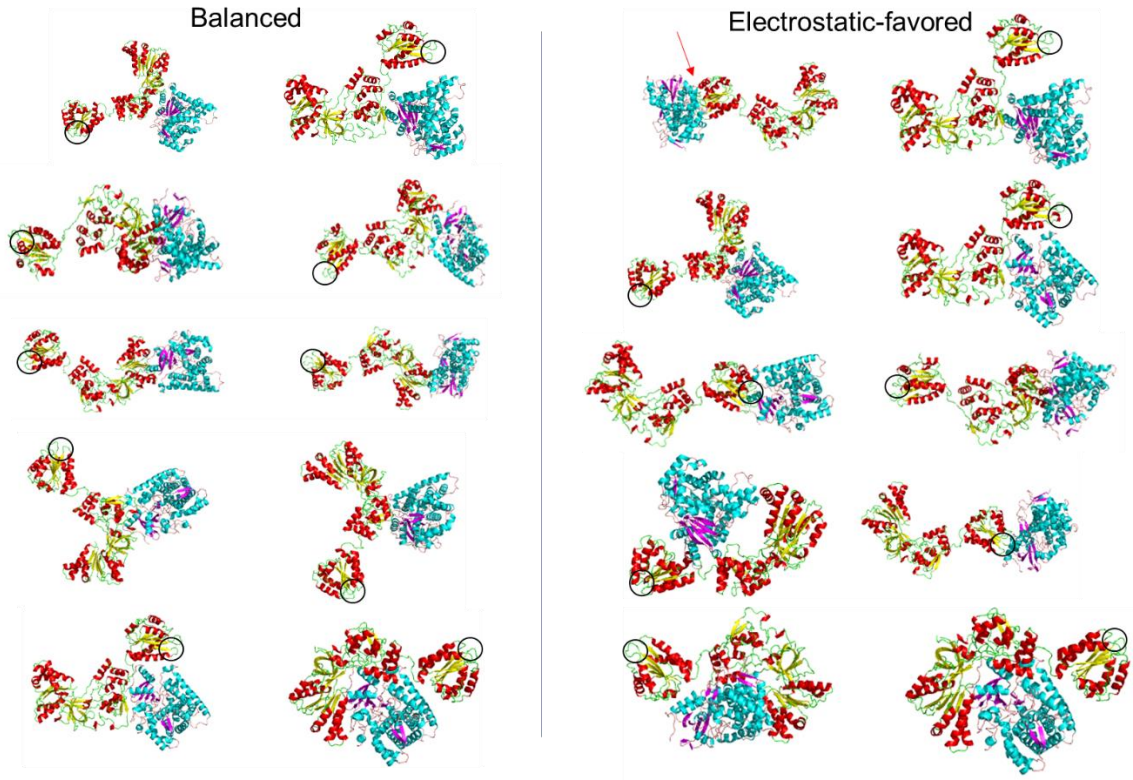


Figure B.6 Top hit *ab initio* docking models as predicted by ClusPro for balanced (left), and electrostatic-favored options.

Sites of the flavin-donor cofactor in the FMN domain of CPR are circled in unproductive enzyme coupling. Red arrows indicate a productive binding mode between FMN and heme cofactors. The balanced and electrostatic-favored models yield 0, and 1 hit respectively.

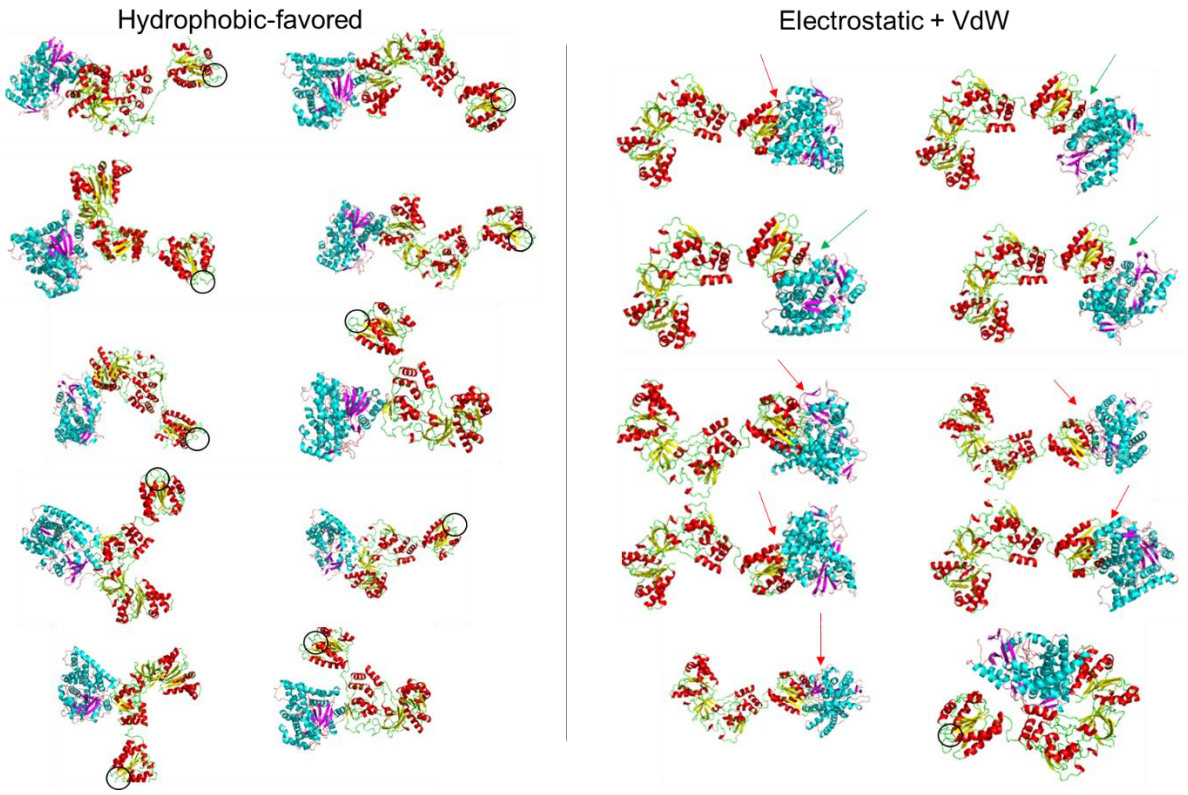


Figure B.7 Top hit *ab initio* docking models as predicted by ClusPro for hydrophobic-favored (left), and electrostatic + Van der Waals weighted options.

Sites of the flavin-donor cofactor in the FMN domain of CPR are circled in unproductive enzyme coupling. A red arrow indicates a productive binding mode between FMN and heme cofactors, and a green arrow indicates FMN domain interaction with the proximal heme site, but with improper FMN cofactor orientation. The hydrophobic-favored and electrostatic + Van der Waals models yield 0, and 7 properly oriented cofactors at the protein-protein interface, respectively. Overall, 9 out of 10 binding modes from the electrostatic + Van der Waals model resulted in the FMN domain interacting with the proximal heme site.

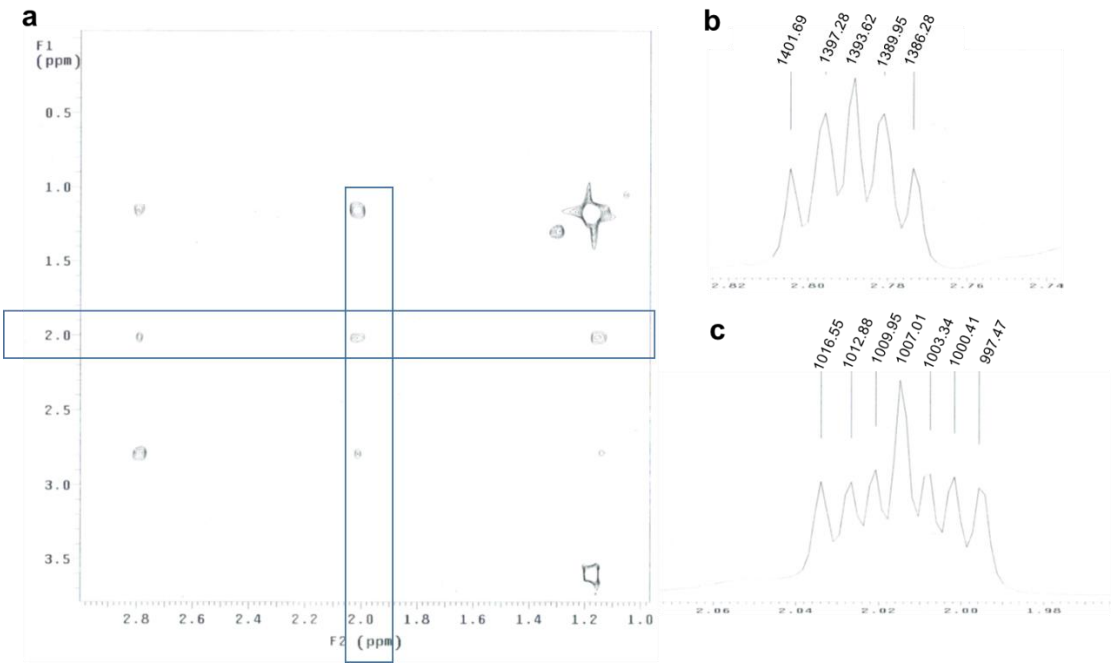


Figure B.8 Proton peaks at 1.15, 2.01, and 2.79 at the cyclopropyl group of AR13 in MeOD, and representative ^{3}J values indicate geminal hydrogens, and two methine groups with trans hydrogen.

a) ^{1}H 2-D TOCSY indicate that crosspeaks at 1.15, 2.01, and 2.79 are in the same spin system of the cycloproyl moiety. b) Peak splitting in hertz at δ 2.79 yield ^{3}J values 3.70, 3.70, and 7.97. c) ^{3}J values at δ 2.01 are 3.4, 6.5, and 9.6. Coupling constants for the proton at each chemical shift indicate one cis, and two trans neighboring protons. In $CDCl_3$ solvent with ^{1}H decoupling at δ 2.79, ^{3}J values 6.52, and 9.78 Hz remain.

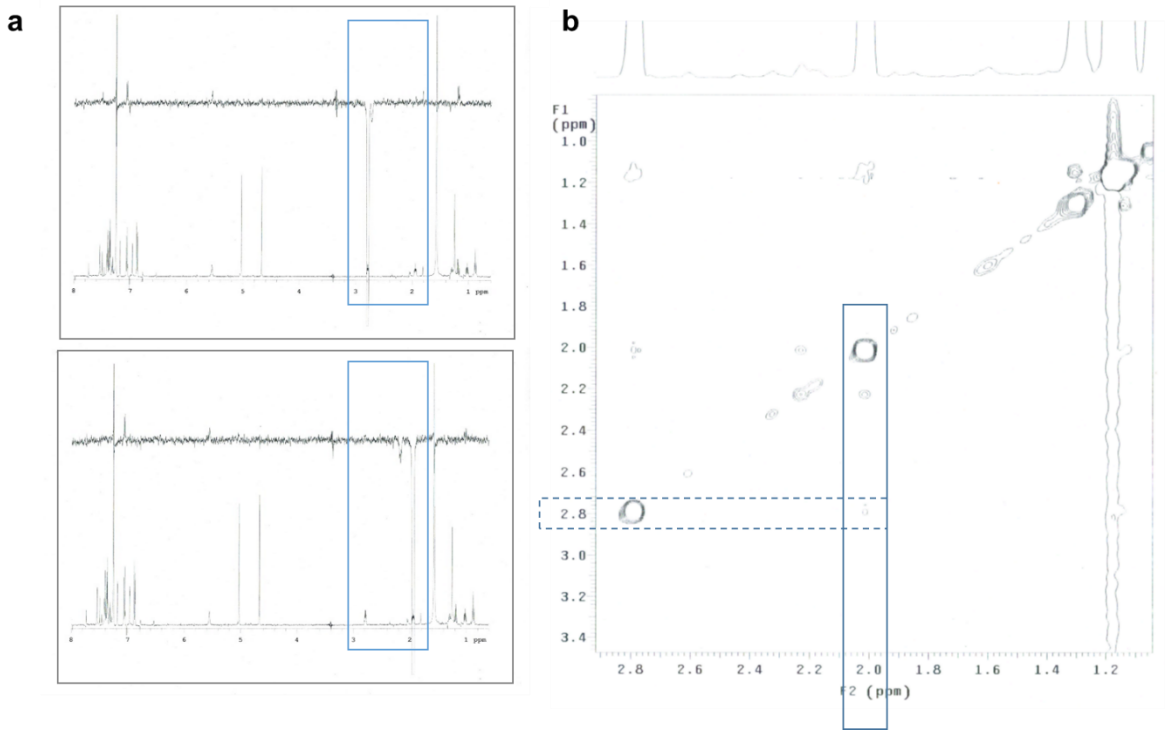


Figure B.9 NOE indicate weak through-space coupling between protons at δ 2.01, and 2.79.

a) 1D NOE return no noticeable coupling between protons boxed in blue. Top frame indicates irradiation at δ 2.79, with bottom frame at δ 2.01. b) 2D NOESY has very weak crosspeaks between protons at δ 2.01 and 2.79.

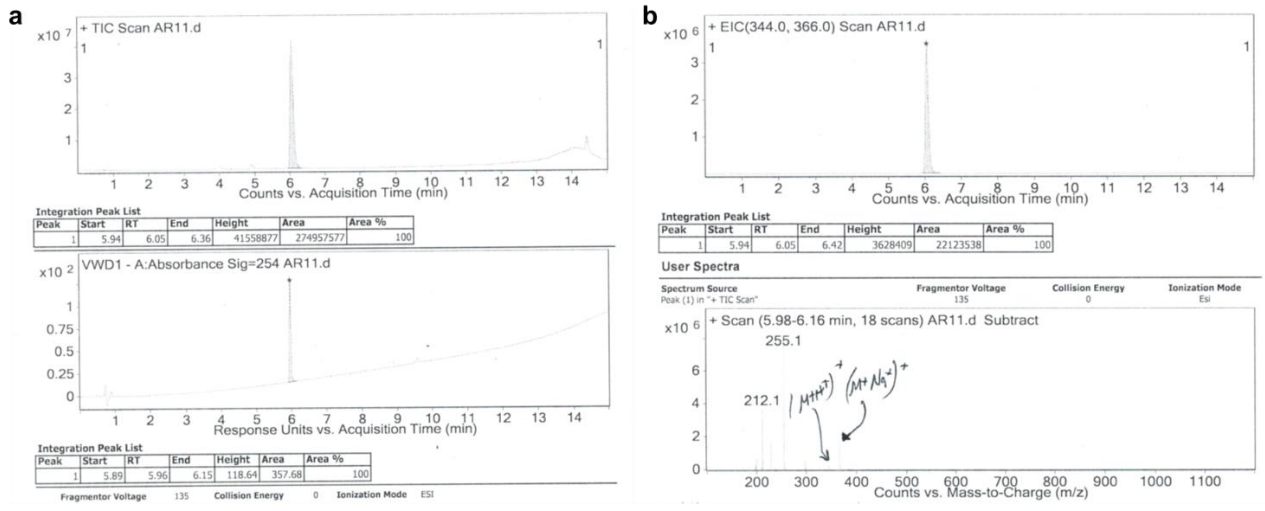
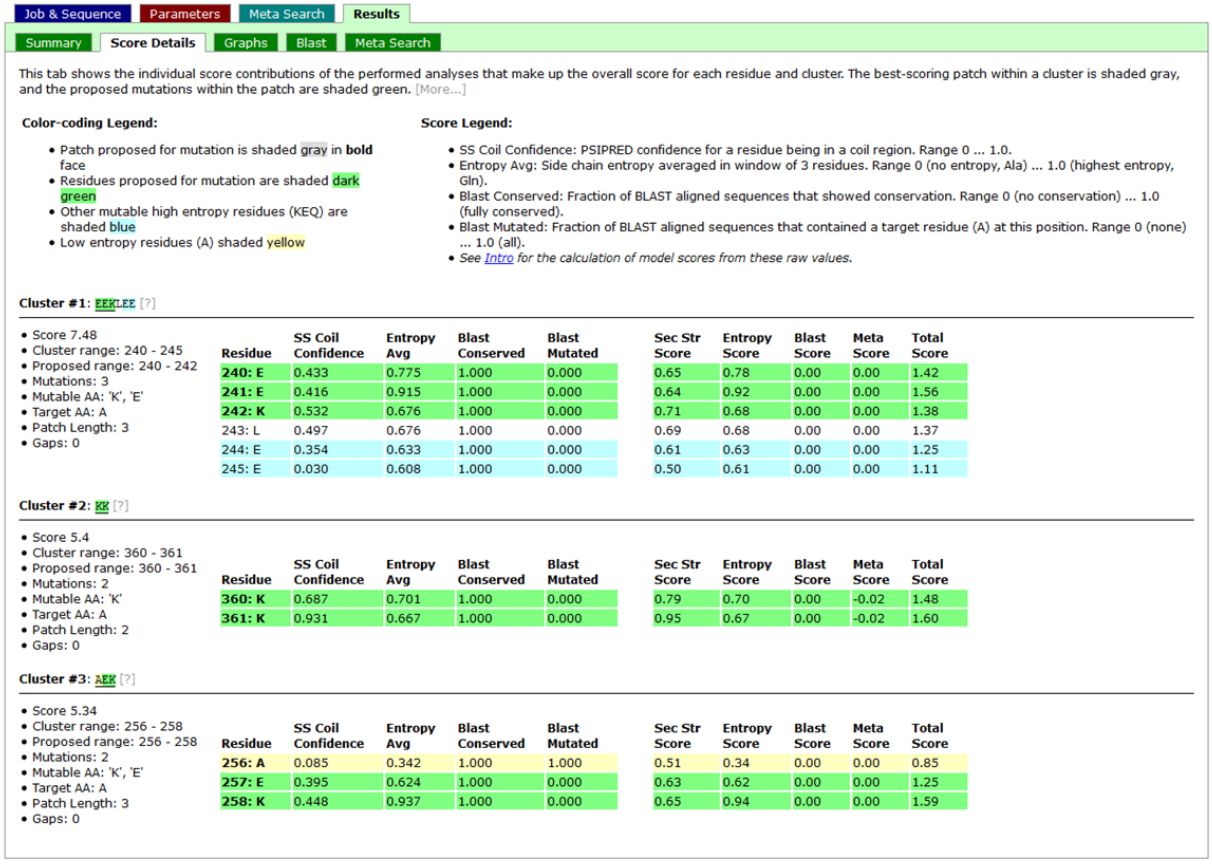


Figure B.10 ESI-MS of AR11 agrees with expected mass 344.

a) TIC scan, and absorbance yield a single peak at 6 minutes. b) EIC scan at 344.0 yields a peak at 6 minutes, and a scan between 6.0 and 6.2 minutes retention time returns m/z of $M+H^+$, and $M+Na^{+-}$ at 345, and 367 respectively.



Job & Sequence | Parameters | Meta Search | Results

Summary | Score Details | Graphs | Blast | Meta Search

This tab shows the individual score contributions of the performed analyses that make up the overall score for each residue and cluster. The best-scoring patch within a cluster is shaded gray, and the proposed mutations within the patch are shaded green. [More...]

Color-coding Legend:

- Patch proposed for mutation is shaded gray in **bold** face
- Residues proposed for mutation are shaded dark green
- Other mutable high entropy residues (KEQ) are shaded blue
- Low entropy residues (A) shaded yellow

Score Legend:

- SS Coil Confidence: PSIPRED confidence for a residue being in a coil region. Range 0 ... 1.0.
- Entropy Avg: Side chain entropy averaged in window of 3 residues. Range 0 (no entropy, Ala) ... 1.0 (highest entropy, Gln).
- Blast Conserved: Fraction of BLAST aligned sequences that showed conservation. Range 0 (no conservation) ... 1.0 (fully conserved).
- Blast Mutated: Fraction of BLAST aligned sequences that contained a target residue (A) at this position. Range 0 (none) ... 1.0 (all).
- *See Intro for the calculation of model scores from these raw values.*

Cluster #1: EEKLEE [?]

- Score 7.48
- Cluster range: 240 - 245
- Proposed range: 240 - 242
- Mutations: 3
- Mutable AA: 'K', 'E'
- Target AA: A
- Patch Length: 3
- Gaps: 0

Residue	SS Coil Confidence	Entropy Avg	Blast Conserved	Blast Mutated	Sec Str Score	Entropy Score	Blast Score	Meta Score	Total Score
240: E	0.433	0.775	1.000	0.000	0.65	0.78	0.00	0.00	1.42
241: E	0.416	0.915	1.000	0.000	0.64	0.92	0.00	0.00	1.56
242: K	0.532	0.676	1.000	0.000	0.71	0.68	0.00	0.00	1.38
243: L	0.497	0.676	1.000	0.000	0.69	0.68	0.00	0.00	1.37
244: E	0.354	0.633	1.000	0.000	0.61	0.63	0.00	0.00	1.25
245: E	0.030	0.608	1.000	0.000	0.50	0.61	0.00	0.00	1.11

Cluster #2: KR [?]

- Score 5.4
- Cluster range: 360 - 361
- Proposed range: 360 - 361
- Mutations: 2
- Mutable AA: 'K'
- Target AA: A
- Patch Length: 2
- Gaps: 0

Residue	SS Coil Confidence	Entropy Avg	Blast Conserved	Blast Mutated	Sec Str Score	Entropy Score	Blast Score	Meta Score	Total Score
360: K	0.687	0.701	1.000	0.000	0.79	0.70	0.00	-0.02	1.48
361: K	0.931	0.667	1.000	0.000	0.95	0.67	0.00	-0.02	1.60

Cluster #3: AEK [?]

- Score 5.34
- Cluster range: 256 - 258
- Proposed range: 256 - 258
- Mutations: 2
- Mutable AA: 'K', 'E'
- Target AA: A
- Patch Length: 3
- Gaps: 0

Residue	SS Coil Confidence	Entropy Avg	Blast Conserved	Blast Mutated	Sec Str Score	Entropy Score	Blast Score	Meta Score	Total Score
256: A	0.085	0.342	1.000	1.000	0.51	0.34	0.00	0.00	0.85
257: E	0.395	0.624	1.000	0.000	0.63	0.62	0.00	0.00	1.25
258: K	0.448	0.937	1.000	0.000	0.65	0.94	0.00	0.00	1.59

Figure B.11 Top scoring clusters that predict mutable residues for reducing the surface entropy of Cyp19 for crystallization.

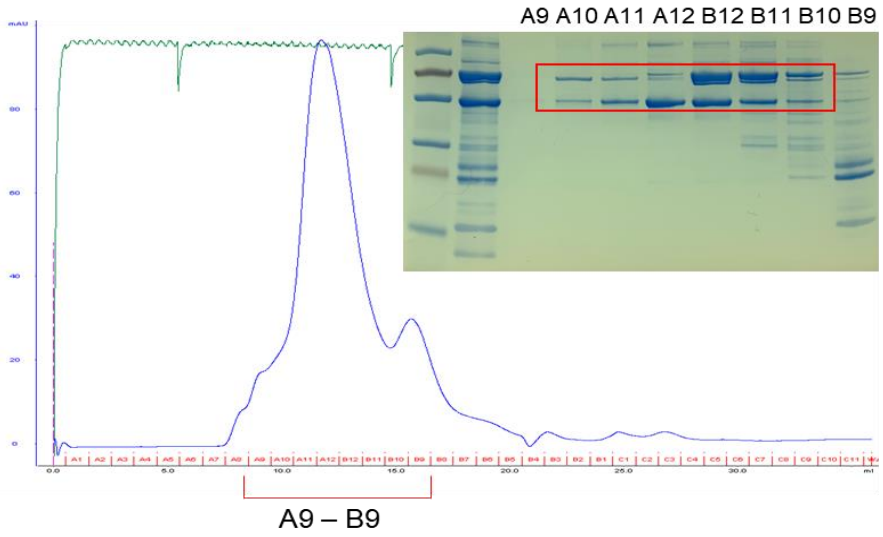


Figure B.12 Gel filtration chromatogram of Cyp19 after Ni-purification (50 mM imidazole wash).

Samples were injected onto a Superdex 200 Increase column at 0.7 mL/min. Cyp19 coelutes with a 70 kDa protein in fractions A10 – B12. Proteins with mass less than 55 kDa coelute in fractions B11 - B9. Cyp19 and 70 kDa protein are boxed in red. In purifying mutants Y361W, K440A, and Y441V, fractions A10 – B12 were pooled.

Table B.1 Primers and PCR conditions of CYP19 mutants.

Mutant	*Purpose*	*SDM Insertion Site (bp)*	*Q5 SDM Primers*
E181A[a]	Eliminate intermolecular interaction with K440 of proximal heme site; increase expression of solubilized monomer	454 – 456 GAG > GCC	E181AF- AGTCACCAATGCCTCCGGTTATGTTGACGTTC E181AR- TCTTCCAGACGGTCCAGG T_a 64 ˚C, t_{ext} 3 min
E269A[a] *E270A* *K271A* *(triple mutant)*	Improve crystallizability	718 – 726 GAAGAAAAA > GCAGCCGCC	E269AF- CGCCCTCGAAGAGTGCATGGAC E269AR- GCTGCGGTGCTGATACGACGACG T_a 62 ˚C, t_{ext} 2 min 40 s, 0.25 ng template
V80S[b]	Improve Stability	151 – 153 GTT > TCT	V80SF- CTACAACCGTTCTTACGGCGAATTC V80SR- TAGTTGCACGCAGAACCA T_a 62 ˚C, t_{ext} 3 min
G156A[b]	Improve Stability	379 – 381 GGT > GCG	G156AF- GTCTGGTCCGGCGCTGGTTCGTA G156AR- AGCGCTTTCATAAAGAACG

			T_a 60 ˚C, t_{ext} 3 min
L240S[a]	Improve Stability	631 – 633 CTG > TCT	L240SF- AATTTCCTGGTCTTACAAAAAGTACGAAAAGTC L240SR- TTGAAGAAAATGTCCGGC T_a 56 ˚C, t_{ext} 3 min
A419S[a]	Improve Stability	1168 – 1170 GCG > TCT	A419SF- CGAGAATTTTTCTAAGAATGTTCCGTACC A419SR- AGGGTGAATTCGTTCGGT T_a 61 ˚C, t_{ext} 3 min 30 s
Y361W[a]	Interrupt inhibitor binding	994 – 996 TAC > TGG	Y361WF- AAACTTCATCTGGGAGTCTATGCGTTAC Y361WR- TCCATAACTTTCAGTTTCTG T_a 59 ˚C, t_{ext} 3 min 30 s
K440A[a]	Interrupt inhibitor binding	1231 - 1233	K440AF- CGCGGGTGCATACATCGCAATGGTTATG K440AR- CAACCACGCGGACCGAAG T_a 61 ˚C, t_{ext} 4 min 30 s, 5 % DMSO, + 1.25 nmol dNTPs, 18 ng template
Y441V[a]	Interrupt inhibitor binding	1234 - 1236	Y441VF- CGCGGGTAAAGTTATCGCAATGGTTATGATG Y441VR- CAACCACGCGGACCGAAG T_a 67 ˚C, t_{ext} 3 min 15 s

a- Sequencing primer G19seqF (341 – 360 bp) – 5'-CCACCCGTCCGTTCTTTATG-3'

b- Sequencing primer CYP19bR seq (855- 882 bp) – 5'-CGCGATGAGGAACAGCATGAAGAACAG-3'

Table B.2 Primers for LIC cloning MBP fusion protein.

Primer	*Oligonucleotide*
LIC cloning- CYP19 forward primer	5'- TACTTCCAATCCAATGCAATGGCCAAAAAAACCTCTTCTAAAGGTC -3'
LIC cloning- CYP19 reverse primer	5'- TTATCCACTTCCAATGTTATTATTCCAGACAACGGTCAGAATTACG- 3'
Colony PCR- forward	5'- GATGAAGCCCTGAAAGACGCGCAG-3'
Colony PCR- reverse	5'-CGCGATGAGGAACAGCATGAAGAACAG-3' (CYP19bR seq)
Sequencing- MBP forward primer	5'- CCGCAGATGTCCGCTTTCT-3'
Sequencing- T7 terminal Reverse primer	5'-GCTAGTTATTGCTCAGCGG-3'

Table B.3 Sequencing results of CYP19 mutants.

Mutant	*Region of codon substitution*
MBP-CYP19	GCCGCCAGCGGTCGTCAGACTGTCGATGAAGCCCTGAAAGACGCGCAGACTAA TGGGATCGAGGAAAACCTGTACTTCCAATCCAATGCAATGGCCAAAAAAACCT CTTCTAAAGGTC (LIC forward) GACGGCGCTCGAATTCGGATCCGTTATCCACTTCCAATGTTATTATTCCAGACA ACGGTCAGAATTACG (LIC reverse)
V80S	CGCATGAATTCGCCGTAAGAACGGTTGTAGTAGTTGCACGCAGAACCAATACC (reverse complement)
G156A	GTAACCATACGAACCAGCGCCGGACCAGACAGCGCTTTCATAAAGAACGGAC G (reverse complement)
E181A	AAAACCCACCTGGACCGTCTGGAAGAAGTCACCAATGCCTCCGGTTATGTTGA CGTTCTGAC
L240S	AGCCGGACATTTTCTTCAAAATTTCCTGGTCTTACAAAAAGTACGAAAAGTCTG TTAAAGACCTGAAGG
EEK269-271AAA	AAGCGTCGTCGTATCAGCACCGCAGCCGCCCTCGAAGAGTGCATGGACTTTGC GACCGAG
Y361W	TCAAAATTGACGACATCCAGAAACTGAAAGTTATGGAAAACTTCATCTGGGAG TCTATGCGT
A419S	TCTTCCCGAAACCGAACGAATTCACCCTCGAGAATTTTTCTAAGAATGTTCCGT ACCGTTACTT
K440A	GCCGTTCGGCTTCGGTCCGCGTGGTTGCGCGGGTGCATACATCGCAATGGTTAT GATGAA
Y441V	GCTTCGGTCCGCGTGGTTGGCGGGTAAAGTTATCGCAATGGTTATGATGAAGG

** codon substitutions are highlighted in yellow, MBP LIC primers are highlighted in teal*

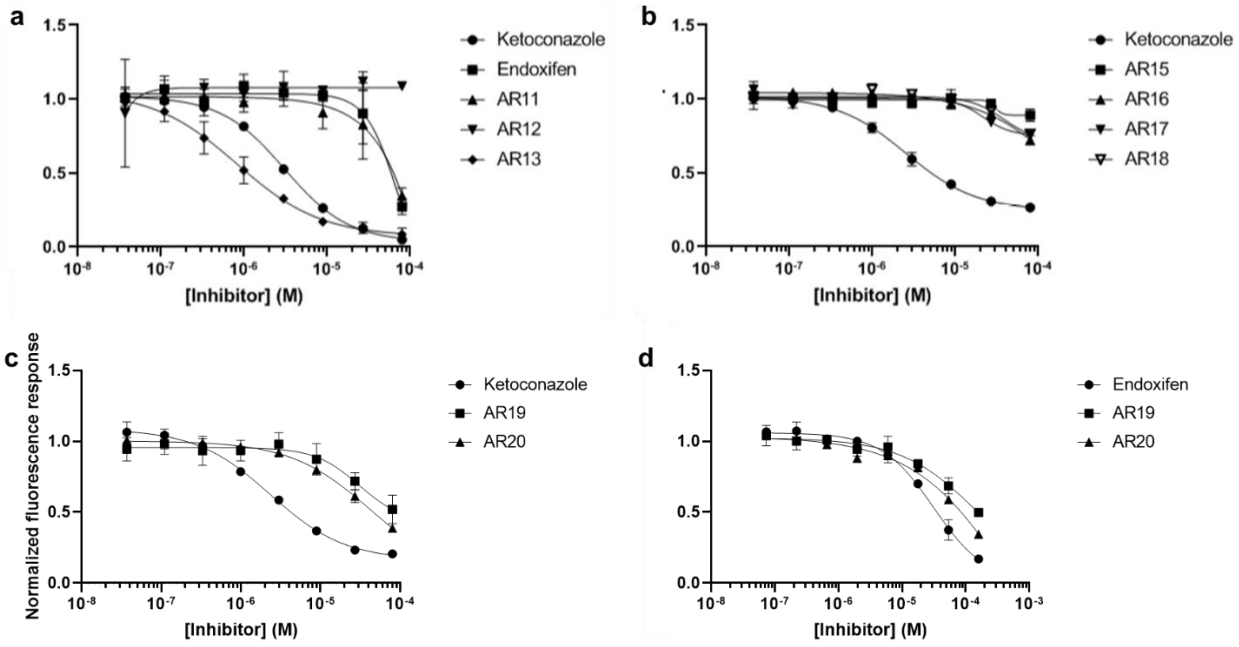


Figure B.13 Dose response curves fit to a 4-paramter logistic model at 7.5 nM P450 at various concentration a) AR11-13, b) AR15-18, and c-d) AR19-20 relative to control inhibitors ketoconazole and endoxifen.

Table B.4 Data collected from active inhibitors according to 4-parameter (4-p) and 3-parameter (3-p) logistic models.

Inhibitor	*IC_{50} (μM) 4-p, 3-p*	*Hill slope*	*Fractional activity at saturation 4-p, 3-p*	*Goodness of fit (R^2) 4-p, 3-p*	*$R^2_{3\text{-}p}$ / $R^2_{4\text{-}p}$*
ketoconazole	3.08, 3.18	-1.13	0.035, 0.008	0.999, 0.999	0.999
endoxifen	30.69, 49.80	-2.26	0.021, -0.350	0.997, 0.970	0.973
AR11	31.07, 46.56	-1.51	0.051, -0.162	0.986, 0.980	0.993
AR13	0.82, 0.87	-0.86	0.071, 0.093	0.978, 0.977	0.999
RN1	86.62, 157.6	-1.32	-0.059, -0.483	0.988, 0.986	0.998
AR19*	30.55 , 65.87	-1.56	0.424, 0.144	0.846, 0.845	0.999
AR20	42.83, 30.04	-0.86	0.028, 0.163	0.992, 0.992	0.999

A 3-p logistic model fits the dose-response function to a hill value of 1. *3-p model was reported.

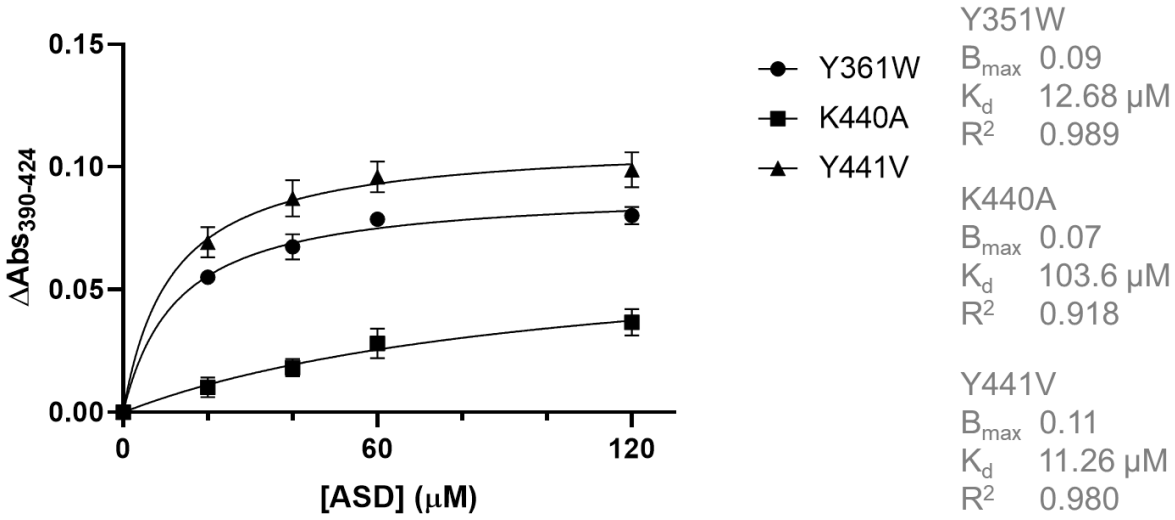


Figure B.14 Binding isotherm fit to a single-site specific binding model for Cyp19 mutants Y361W, K440A, and Y441V at 0.4 mg/mL total protein.

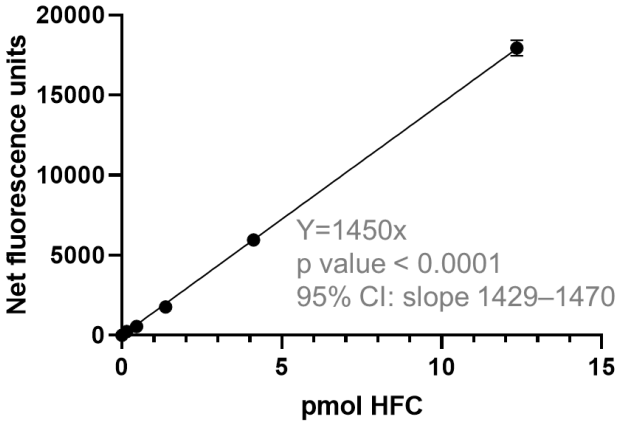


Figure B.15 Standard curve for fluorescence units conversion to pmol HFC product in fluorescence-based reversible inhibition assays at 52 gain.

Concentration at 0.15, 0.46, 1.37, 4.12, and 12.35 pmol HFC returned values within a measurable range, [HFC] yielding RFU > 50,000 were out of range at 52 gain.

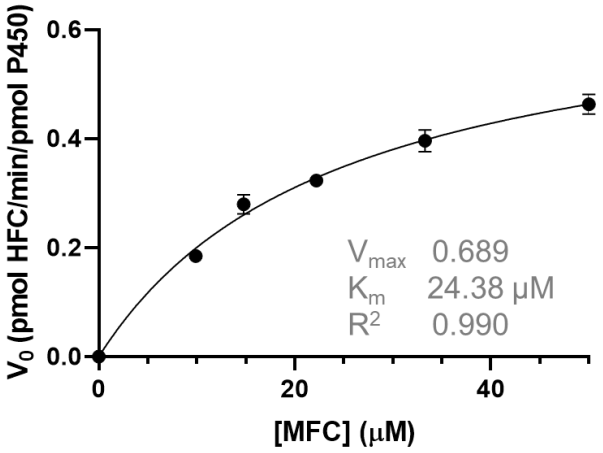


Figure B.16 Non-linear regression at 10 nM P450, and [MFC] at 0, 9.9, 14.8, 22.2, 33.3, and 50 μM.

V_{max}, and K_m are reported from non-linear regression analysis assuming steady-state kinetic behavior.

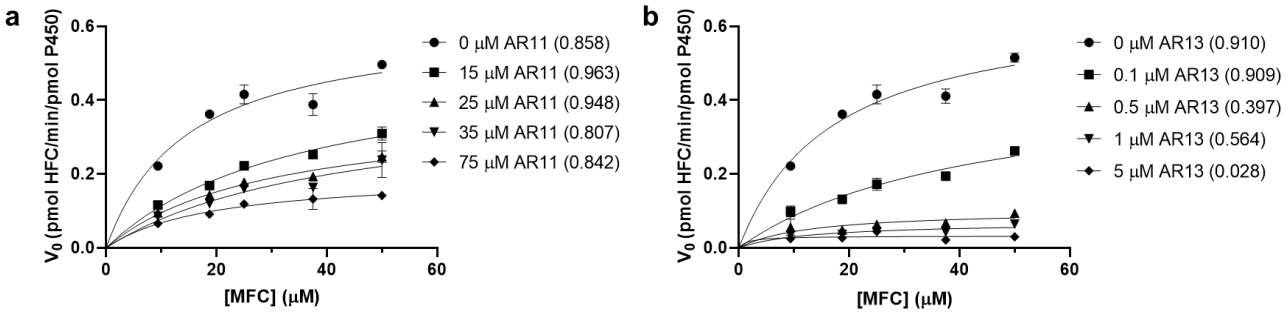


Figure B.17 Non-linear regression at a range of inhibitor concentration less than 10 x IC_{50} for a) AR11 and b) AR13.

Parenthesized R^2 values are lower at higher inhibitor concentration.

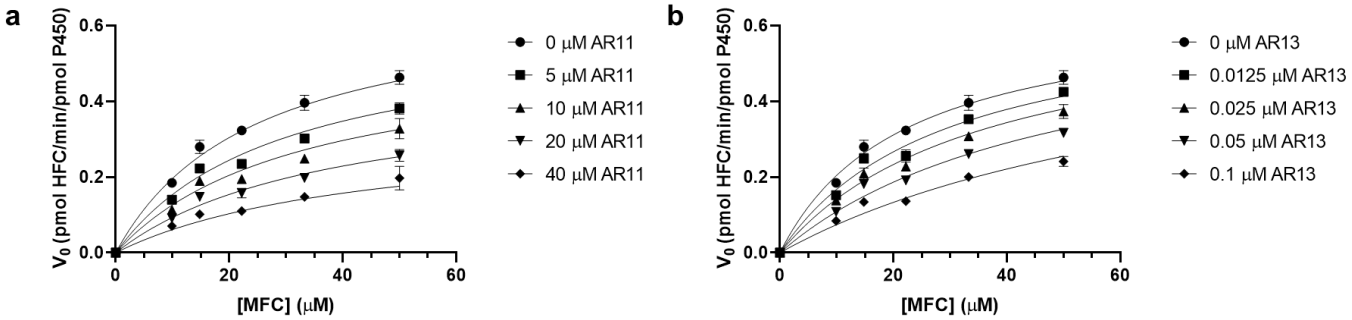
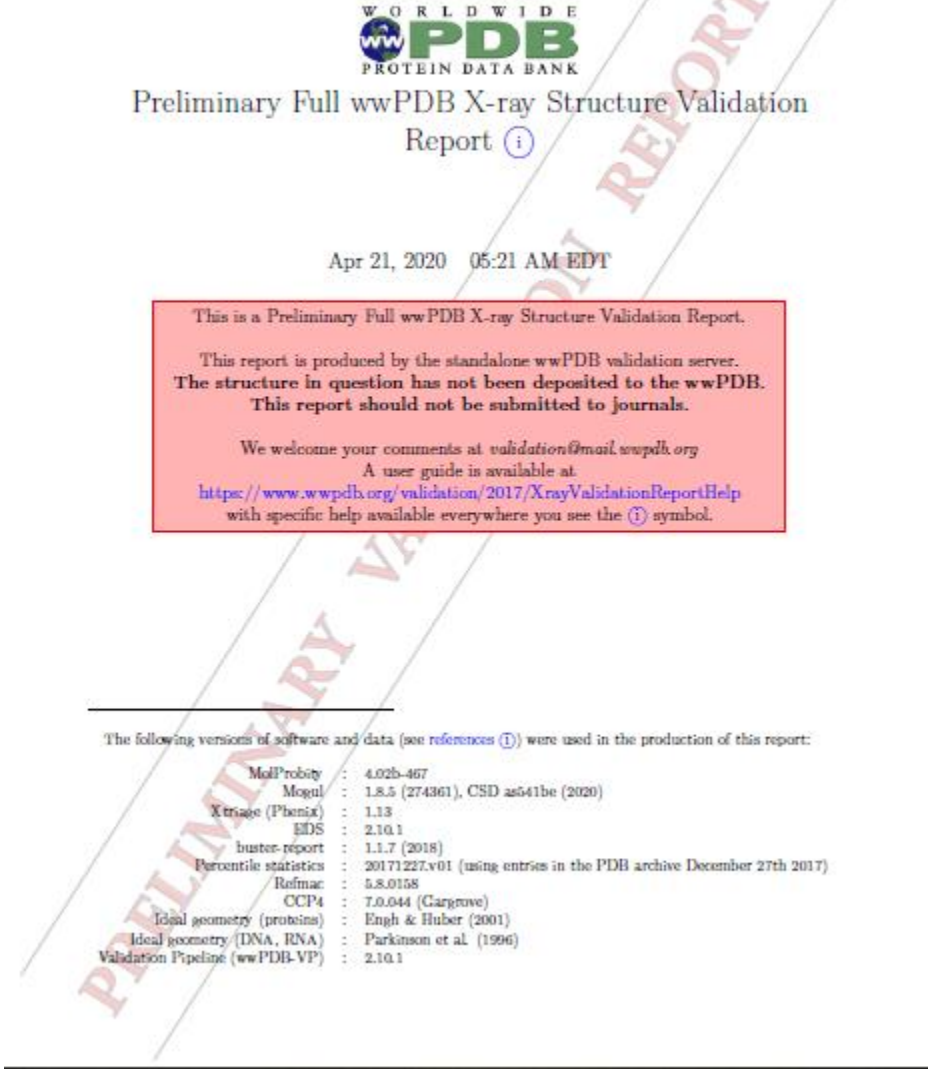


Figure B.18 Non-linear regression of inhibition assays performed in triplicate at 10 nM P450, 0, 9.9, 14.8, 22.2, 33.3, and 50 uM MFC.

Curves for AR11 and AR13 are fit to a mixed-type and competitive-type inhibition models respectively.

Appendix C - BioA preliminary full wwPDB validation report



WORLDWIDE PDB
PROTEIN DATA BANK

Preliminary Full wwPDB X-ray Structure Validation Report (i)

Apr 21, 2020 05:21 AM EDT

This is a Preliminary Full wwPDB X-ray Structure Validation Report.

This report is produced by the standalone wwPDB validation server.
The structure in question has not been deposited to the wwPDB.
This report should not be submitted to journals.

We welcome your comments at *validation@mail.wwpdb.org*
A user guide is available at
https://www.wwpdb.org/validation/2017/XrayValidationReportHelp
with specific help available everywhere you see the (i) symbol.

The following versions of software and data (see references (i)) were used in the production of this report:

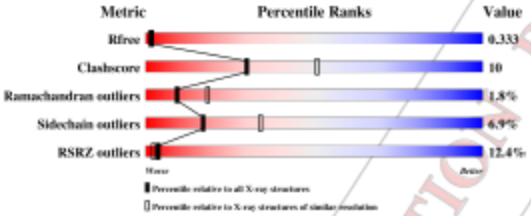
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.10.1
buster-report	:	1.1.7 (2018)
Percentile statistics	:	20171227.v01 (using entries in the PDB archive December 27th 2017)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.10.1

1 Overall quality at a glance

The following experimental techniques were used to determine the structure:
X-RAY DIFFRACTION

The reported resolution of this entry is 2.59 Å.

Percentile scores (ranging between 0-100) for global validation metrics of the entry are shown in the following graphic. The table shows the number of entries on which the scores are based.



Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	111664	2767 (2.60-2.60)
Clashscore	122126	3110 (2.60-2.60)
Ramachandran outliers	120053	3062 (2.60-2.60)
Sidechain outliers	120020	3062 (2.60-2.60)
RSRZ outliers	108989	2706 (2.60-2.60)

The table below summarises the geometric issues observed across the polymeric chains and their fit to the electron density. The red, orange, yellow and green segments on the lower bar indicate the fraction of residues that contain outliers for >=3, 2, 1 and 0 types of geometric quality criteria respectively. A grey segment represents the fraction of residues that are not modelled. The numeric value for each fraction is indicated below the corresponding segment, with a dot representing fractions <=5% The upper red bar (where present) indicates the fraction of residues that have poor fit to the electron density. The numeric value is given above the bar.

Mol	Chain	Length	Quality of chain
1	A	420	13% 76% 21%
2	B	416	14% 74% 23%

2 Entry composition

There are 4 unique types of molecules in this entry. The entry contains 13204 atoms, of which 6586 are hydrogens and 0 are deuteriums.

In the tables below, the ZeroOcc column contains the number of atoms modelled with zero occupancy, the AltConf column contains the number of residues with at least one atom in alternate conformation and the Trace column contains the number of residues modelled with at most 2 atoms.

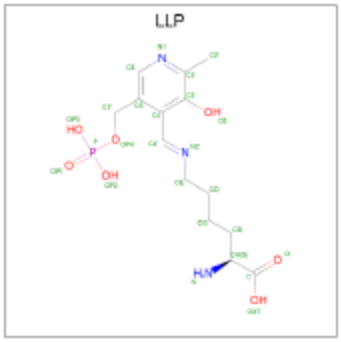
- Molecule 1 is a protein.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
			Total	C	H	N	O	S			
1	A	420	6595	2108	3290	548	627	22	0	0	0

- Molecule 2 is a protein.

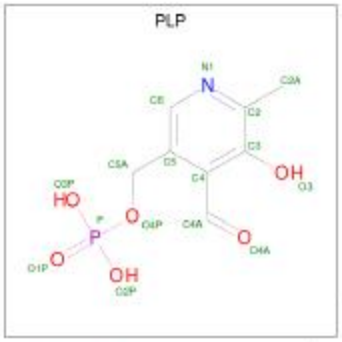
Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
			Total	C	H	N	O	S			
2	B	416	6540	2090	3268	542	618	22	0	0	0

- Molecule 3 is (2S)-2-amino-6-[[3-hydroxy-2-methyl-5-(phosphonooxymethyl)pyridin-4-yl]methylideneamino]hexanoic acid (three-letter code: LLP) (formula: $C_{14}H_{22}N_3O_7P$).



Mol	Chain	Residues	Atoms						ZeroOcc	AltConf
			Total	C	H	N	O	P		
3	C	1	46	14	21	3	7	1	0	0

- Molecule 4 is PYRIDOXAL-5'-PHOSPHATE (three-letter code: PLP) (formula: $C_8H_{10}NO_6P$).



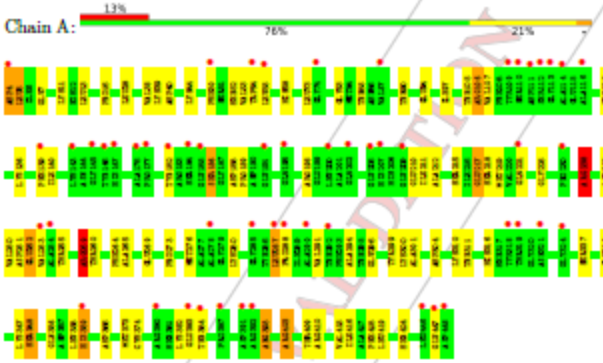
Mol	Chain	Residues	Atoms						ZeroOcc	AltConf
			Total	C	H	N	O	P		
4	D	1	23	8	7	1	6	1	0	0

3 Residue-property plots (i)

These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of the various outlier classes displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and electron density. Residues are color-coded according to the number of geometric quality criteria for which they contain at least one outlier: green = 0, yellow = 1, orange = 2 and red = 3 or more. A red dot above a residue indicates a poor fit to the electron density (RSRZ > 2). Stretches of 2 or more consecutive residues without any outlier are shown as a green connector. Residues present in the sample, but not in the model, are shown in grey.

- Molecule 1:

Chain A:



- Molecule 2:

Chain B:

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1	Depositor
Cell constants a, b, c, α, β, γ	58.30Å 60.54Å 62.72Å 96.15° 106.04° 99.23°	Depositor
Resolution (Å)	59.51 – 2.59 38.90 – 2.59	Depositor EDS
% Data completeness (in resolution range)	91.8 (59.51-2.59) 91.8 (38.90-2.59)	Depositor EDS
R_{merge}	0.11	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.35 (at 2.58Å)	Xtriage
Refinement program	REFMAC 5.8.0155	Depositor
R, R_{free}	0.232 , 0.340 0.244 , 0.333	Depositor DCC
R_{free} test set	1157 reflections (5.02%)	wwPDB-VP
Wilson B-factor (Å^2)	48.3	Xtriage
Anisotropy	0.033	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.42 , 42.6	EDS
L-test for twinning[2]	$< \|L\| > = 0.49$, $< L^2 > = 0.32$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.91	EDS
Total number of atoms	13204	wwPDB-VP
Average B, all atoms (Å^2)	46.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 6.05% of the height of the origin peak. No significant pseudotranslation is detected.*

[1] Intensities estimated from amplitudes.
[2] Theoretical values of $< |L| >$, $< L^2 >$ for acentric reflections are 0.5, 0.333 respectively for untwinned datasets, and 0.375, 0.2 for perfectly twinned datasets.

5 Model quality

5.1 Standard geometry

Bond lengths and bond angles in the following residue types are not validated in this section: LLP, PLP

The Z score for a bond length (or angle) is the number of standard deviations the observed value is removed from the expected value. A bond length (or angle) with $|Z| > 5$ is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Chain	Bond lengths		Bond angles	
		RMSZ	#|Z| >5	RMSZ	#|Z| >5
1	A	0.81	11/3373 (0.3%)	0.89	8/4550 (0.2%)
2	B	0.79	11/3339 (0.3%)	0.92	12/4504 (0.3%)
All	All	0.80	22/6712 (0.3%)	0.90	20/9054 (0.2%)

Chiral center outliers are detected by calculating the chiral volume of a chiral center and verifying if the center is modelled as a planar moiety or with the opposite hand.A planarity outlier is detected by checking planarity of atoms in a peptide group, atoms in a mainchain group or atoms of a sidechain that are expected to be planar.

Mol	Chain	#Chirality outliers	#Planarity outliers
1	A	0	1

All (22) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	5	LEU	C-N	-12.19	1.06	1.34
1	A	393	ARG	C-N	9.63	1.56	1.34
1	A	210	GLU	C-N	-8.24	1.15	1.34
2	B	412	LEU	C-N	-8.20	1.18	1.33
1	A	4	ASP	C-N	-7.72	1.16	1.34
2	B	106	ARG	CZ-NH2	7.21	1.42	1.33
1	A	106	ARG	CZ-NH2	7.20	1.42	1.33
1	A	415	VAL	C-N	7.20	1.50	1.34
1	A	106	ARG	CZ-NH1	6.95	1.42	1.33
2	B	106	ARG	CZ-NH1	6.90	1.42	1.33
1	A	358	LEU	C-N	6.86	1.49	1.34
2	B	238	ARG	CZ-NH2	6.81	1.41	1.33
2	B	410	ARG	CZ-NH2	6.76	1.41	1.33
2	B	183	ARG	CZ-NH1	6.26	1.41	1.33
1	A	259	ARG	CZ-NH2	6.03	1.40	1.33
2	B	410	ARG	CZ-NH1	5.90	1.40	1.33

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(A)	Ideal(A)
2	B	183	ARG	CZ-NH2	5.86	1.40	1.33
2	B	368	ARG	C-N	-5.74	1.20	1.34
2	B	238	ARG	CZ-NH1	5.58	1.40	1.33
2	B	244	TYR	C-N	-5.36	1.21	1.34
1	A	252	GLU	CD-OE1	-5.32	1.19	1.25
1	A	259	ARG	CZ-NH1	5.07	1.39	1.33

All (20) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	B	337	SER	CB-CA-C	-10.25	90.62	110.10
1	A	393	ARG	NE-CZ-NH1	8.13	124.37	120.30
2	B	238	ARG	NE-CZ-NH1	-7.55	116.52	120.30
2	B	443	VAL	O-C-N	-7.52	110.67	122.70
2	B	445	SER	O-C-N	6.84	133.64	122.70
2	B	338	GLU	N-CA-CB	6.80	122.84	110.60
1	A	287	LEU	C-N-CD	-6.51	106.27	120.60
1	A	259	ARG	NE-CZ-NH1	-6.37	117.12	120.30
2	B	367	ILE	O-C-N	-6.09	112.95	122.70
2	B	338	GLU	C-N-CA	-5.97	106.78	121.70
2	B	445	SER	CA-C-N	-5.89	104.24	117.20
1	A	393	ARG	NE-CZ-NH2	-5.86	117.37	120.30
1	A	238	ARG	NE-CZ-NH1	5.80	123.20	120.30
1	A	403	ARG	NE-CZ-NH1	5.58	123.09	120.30
2	B	338	GLU	O-C-N	5.49	131.49	122.70
1	A	416	ILE	O-C-N	5.38	131.31	122.70
2	B	244	TYR	O-C-N	5.19	131.01	122.70
1	A	393	ARG	CG-CD-NE	5.13	122.56	111.80
2	B	367	ILE	CA-C-N	5.04	128.30	117.20
2	B	338	GLU	N-CA-C	-5.01	97.48	111.00

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	4	ASP	Mainchain

5.2 Too-close contacts

In the following table, the Non-H and H(model) columns list the number of non-hydrogen atoms and hydrogen atoms in the chain respectively. The H(added) column lists the number of hydrogen

atoms added and optimized by MolProbity. The Clashes column lists the number of clashes within the asymmetric unit, whereas Symm-Clashes lists symmetry related clashes.

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	A	3305	3290	3288	74	5
2	B	3272	3268	3268	72	5
3	C	25	21	18	7	0
4	D	16	7	7	1	0
All	All	6618	6586	6581	133	5

The all-atom clashscore is defined as the number of clashes found per 1000 atoms (including hydrogen atoms). The all-atom clashscore for this structure is 10.

All (133) close contacts within the same asymmetric unit are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:259:ARG:HG3	1:A:373:MET:HG3	1.44	0.99
2:B:206:GLU:HG3	2:B:207:HIS:CD2	1.99	0.98
1:A:260:THR:HG21	1:A:265:ALA:HB2	1.46	0.97
1:A:13:LEU:O	2:B:106:ARG:NH2	1.97	0.96
1:A:218:SER:OG	1:A:252:GLU:OE2	1.85	0.94
2:B:206:GLU:HG3	2:B:207:HIS:NE2	1.84	0.93
1:A:410:ARG:NE	3:C:1:LLP:OXT	2.05	0.90
2:B:206:GLU:CG	2:B:207:HIS:CD2	2.55	0.88
2:B:107:VAL:HG13	2:B:291:VAL:HG13	1.60	0.81
1:A:139:PHE:CE2	1:A:212:ALA:HB3	2.17	0.80
2:B:57:VAL:HG13	2:B:422:LEU:HB2	1.66	0.78
2:B:186:SER:O	2:B:188:ASP:N	2.17	0.78
2:B:97:ILE:HD11	2:B:107:VAL:HG23	1.66	0.75
2:B:260:THR:HG22	2:B:268:HIS:NE2	2.02	0.75
1:A:259:ARG:HG3	1:A:373:MET:CG	2.18	0.74
1:A:136:LYS:HD3	1:A:211:ILE:O	1.88	0.72
2:B:78:HIS:CE1	2:B:318:TYR:HA	2.24	0.72
1:A:55:LEU:HD22	2:B:318:TYR:CZ	2.26	0.70
1:A:105:THR:O	1:A:106:ARG:HD2	1.91	0.68
1:A:218:SER:CB	1:A:252:GLU:OE2	2.41	0.68
1:A:90:THR:HB	2:B:5:LEU:HD11	1.76	0.67
1:A:217:GLU:OE1	1:A:221:GLN:OE1	2.13	0.66
1:A:54:TRP:CD1	1:A:419:LEU:HD22	2.31	0.65
2:B:107:VAL:HG13	2:B:291:VAL:CG1	2.27	0.65
2:B:445:SER:OG	2:B:446:LEU:HD12	1.97	0.64
2:B:17:PHE:O	2:B:410:ARG:NH2	2.30	0.63
1:A:106:ARG:NH1	1:A:296:GLU:OE1	2.28	0.63

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:52:SER:O	1:A:55:LEU:HD12	1.99	0.62
2:B:258:GLY:HA3	2:B:422:LEU:HD21	1.82	0.61
1:A:217:GLU:CD	1:A:221:GLN:OE1	2.39	0.61
1:A:259:ARG:CG	1:A:373:MET:HG3	2.26	0.61
2:B:260:THR:CG2	2:B:268:HIS:NE2	2.64	0.60
1:A:347:LYS:NZ	1:A:424:SER:O	2.22	0.59
2:B:206:GLU:HG3	2:B:207:HIS:CE1	2.37	0.59
1:A:418:PHE:C	1:A:419:LEU:HD12	2.23	0.58
2:B:445:SER:OG	2:B:446:LEU:N	2.36	0.58
2:B:133:LYS:HB3	2:B:135:GLU:OE2	2.03	0.58
2:B:206:GLU:HG2	2:B:207:HIS:CD2	2.37	0.58
2:B:361:LEU:O	2:B:382:LYS:NZ	2.37	0.57
2:B:5:LEU:HG	2:B:29:ILE:HD13	1.86	0.57
2:B:112:SER:OG	2:B:115:GLU:HG3	2.05	0.57
1:A:105:THR:C	1:A:106:ARG:HD2	2.24	0.56
2:B:265:ALA:O	2:B:268:HIS:CD2	2.57	0.56
2:B:5:LEU:HD23	2:B:41:ILE:HD13	1.87	0.56
1:A:259:ARG:CG	1:A:373:MET:CG	2.84	0.56
1:A:128:TRP:CE3	1:A:131:ILE:HD11	2.40	0.55
2:B:259:ARG:H	2:B:422:LEU:HD21	1.70	0.55
2:B:265:ALA:O	2:B:268:HIS:HD2	1.90	0.54
1:A:55:LEU:HD22	2:B:318:TYR:CE2	2.43	0.54
2:B:183:ARG:NH2	2:B:385:LYS:O	2.40	0.54
1:A:106:ARG:HG3	1:A:299:TYR:HB2	1.89	0.53
2:B:186:SER:C	2:B:188:ASP:H	2.08	0.53
1:A:13:LEU:C	2:B:106:ARG:HH21	2.03	0.52
3:C:1:LLP:C5'	3:C:1:LLP:NZ	2.73	0.52
1:A:107:VAL:HG13	1:A:291:VAL:CG1	2.40	0.51
2:B:312:PHE:CE2	2:B:314:HIS:HB3	2.46	0.51
1:A:252:GLU:OE1	1:A:264:PHE:HB3	2.10	0.51
1:A:344:VAL:O	1:A:348:SER:OG	2.27	0.51
1:A:251:ASP:OD1	3:C:1:LLP:H2'2	2.11	0.50
2:B:16:PRO:O	2:B:18:THR:HG23	2.11	0.50
2:B:331:ASN:OD1	2:B:335:PHE:HE2	1.95	0.50
2:B:409:THR:OG1	2:B:417:ALA:O	2.29	0.50
1:A:97:ILE:HD12	1:A:105:THR:HA	1.94	0.49
1:A:139:PHE:CE2	1:A:212:ALA:CB	2.95	0.49
1:A:85:THR:CG2	2:B:29:ILE:HG12	2.42	0.49
2:B:82:LEU:HD21	3:C:1:LLP:HB2	1.95	0.49
2:B:258:GLY:HA3	2:B:422:LEU:CD2	2.43	0.49
1:A:393:ARG:NH1	2:B:306:TYR:CG	2.81	0.49

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
2:B:185:GLU:OE1	2:B:185:GLU:N	2.42	0.48
1:A:11:LYS:C	2:B:106:ARG:NH1	2.67	0.48
1:A:186:SER:C	1:A:188:ASP:H	2.17	0.48
1:A:40:ASP:OD1	1:A:40:ASP:C	2.52	0.48
2:B:54:TRP:N	2:B:54:TRP:CD1	2.82	0.48
1:A:260:THR:HG21	1:A:265:ALA:CB	2.33	0.47
2:B:16:PRO:O	2:B:18:THR:CG2	2.63	0.47
1:A:128:TRP:CE2	1:A:136:LYS:HD2	2.49	0.47
1:A:343:GLN:O	1:A:347:LYS:HG3	2.14	0.47
1:A:58:HIS:HE1	1:A:340:ILE:HD13	1.80	0.47
1:A:217:GLU:HB2	1:A:221:GLN:NE2	2.29	0.47
1:A:83:GLY:H	2:B:18:THR:HG21	1.79	0.47
1:A:7:GLU:HG3	1:A:11:LYS:HE3	1.96	0.47
1:A:253:VAL:HG11	3:C:1:LLP:C6	2.44	0.47
1:A:250:VAL:HG21	1:A:273:PRO:HB3	1.97	0.47
1:A:287:LEU:HD22	2:B:287:LEU:HD22	1.96	0.46
1:A:186:SER:HB2	1:A:188:ASP:H	1.79	0.46
2:B:78:HIS:HE1	2:B:318:TYR:HA	1.77	0.46
1:A:255:THR:HG22	1:A:373:MET:SD	2.56	0.46
2:B:65:LEU:O	2:B:66:ASP:C	2.54	0.46
2:B:181:VAL:HA	2:B:184:SER:HB3	1.97	0.45
1:A:13:LEU:N	2:B:106:ARG:NH2	2.64	0.45
2:B:396:TYR:O	2:B:399:SER:OG	2.28	0.45
1:A:218:SER:HB2	1:A:250:VAL:CG1	2.47	0.45
1:A:359:HIS:O	1:A:382:LYS:NZ	2.49	0.45
2:B:133:LYS:CB	2:B:135:GLU:OE2	2.65	0.45
2:B:174:TYR:OH	2:B:210:GLU:OE1	2.32	0.44
2:B:259:ARG:N	2:B:422:LEU:HD21	2.31	0.44
1:A:73:LEU:HD22	2:B:66:ASP:OD1	2.18	0.44
1:A:252:GLU:HG3	1:A:276:MET:SD	2.57	0.44
1:A:55:LEU:HD23	1:A:280:LYS:HG2	1.98	0.44
1:A:16:PRO:HG3	2:B:319:THR:HG21	1.99	0.44
1:A:140:ILE:HD11	1:A:211:ILE:HD13	1.99	0.44
1:A:54:TRP:N	1:A:54:TRP:CD1	2.86	0.44
2:B:361:LEU:HD23	2:B:361:LEU:HA	1.85	0.44
2:B:206:GLU:CG	2:B:207:HIS:NE2	2.66	0.43
2:B:57:VAL:CG1	2:B:422:LEU:HB2	2.44	0.43
1:A:250:VAL:CG2	1:A:273:PRO:HB3	2.48	0.43
2:B:6:ILE:HA	2:B:29:ILE:HD11	2.01	0.43
2:B:338:GLU:O	2:B:339:ASN:C	2.54	0.43
1:A:73:LEU:HD11	2:B:69:ILE:CG2	2.48	0.43

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:28:LEU:HD22	1:A:403:ARG:HD3	2.00	0.42
1:A:374:CYS:HB2	1:A:418:PHE:HB3	2.02	0.42
1:A:107:VAL:HG13	1:A:291:VAL:HG13	2.01	0.42
1:A:106:ARG:HD3	1:A:296:GLU:OE1	2.19	0.42
1:A:39:LYS:HA	1:A:44:LYS:O	2.20	0.42
2:B:50:PHE:O	2:B:51:SER:C	2.58	0.42
3:C:1:LLP:NZ	3:C:1:LLP:H5'2	2.33	0.42
2:B:352:HIS:HA	2:B:369:GLN:OE1	2.19	0.42
1:A:316:HIS:HB3	4:D:1:PLP:O3P	2.19	0.42
1:A:55:LEU:HD21	2:B:317:SER:OG	2.20	0.42
2:B:394:ILE:O	2:B:398:VAL:HG23	2.20	0.42
1:A:238:ARG:O	1:A:238:ARG:HD3	2.19	0.41
1:A:128:TRP:CD2	1:A:136:LYS:HD2	2.55	0.41
1:A:215:SER:HA	1:A:249:ILE:HB	2.02	0.41
1:A:253:VAL:HG11	3:C:1:LLP:H6	2.02	0.41
2:B:124:ALA:HB1	2:B:213:ALA:HB2	2.03	0.41
1:A:106:ARG:HG2	1:A:294:ALA:O	2.20	0.41
1:A:217:GLU:OE1	1:A:221:GLN:NE2	2.53	0.41
2:B:142:MET:O	2:B:143:LYS:C	2.57	0.41
2:B:106:ARG:HG3	2:B:299:TYR:CG	2.56	0.41
2:B:93:ALA:O	2:B:97:ILE:HG12	2.21	0.41
2:B:58:HIS:HA	2:B:423:ALA:HB2	2.03	0.41
1:A:182:TYR:CD1	1:A:366:ASP:HB2	2.56	0.40
2:B:216:ILE:HD11	2:B:229:MET:HE1	2.03	0.40

All (5) symmetry-related close contacts are listed below. The label for Atom-2 includes the symmetry operator and encoded unit-cell translations to be applied.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:383:GLU:OE2	2:B:186:SER:O[1_444]	0.99	1.21
1:A:383:GLU:CD	2:B:186:SER:O[1_444]	1.79	0.41
1:A:300:LYS:HZ2	2:B:239:GLU:OE1[1_455]	1.42	0.18
1:A:383:GLU:OE2	2:B:186:SER:C[1_444]	2.06	0.14
1:A:300:LYS:NZ	2:B:239:GLU:OE1[1_455]	2.11	0.09

5.3 Torsion angles

5.3.1 Protein backbone

In the following table, the Percentiles column shows the percent Ramachandran outliers of the chain as a percentile score with respect to all X-ray entries followed by that with respect to entries of similar resolution.

The Analysed column shows the number of residues for which the backbone conformation was analysed, and the total number of residues.

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
1	A	416/420 (99%)	377 (91%)	31 (8%)	8 (2%)	9	17
2	B	412/416 (99%)	369 (90%)	36 (9%)	7 (2%)	10	20
All	All	828/836 (99%)	746 (90%)	67 (8%)	15 (2%)	9	18

All (15) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
2	B	187	GLY
1	A	225	GLY
1	A	337	SER
2	B	43	GLY
2	B	179	PRO
1	A	288	PRO
1	A	301	ALA
2	B	51	SER
2	B	280	LYS
2	B	424	SER
1	A	189	PRO
1	A	219	MET
1	A	259	ARG
2	B	76	ILE
1	A	53	VAL

5.3.2 Protein sidechains

In the following table, the Percentiles column shows the percent sidechain outliers of the chain as a percentile score with respect to all X-ray entries followed by that with respect to entries of similar resolution.

The Analysed column shows the number of residues for which the sidechain conformation was analysed, and the total number of residues.

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
1	A	355/355 (100%)	334 (94%)	21 (6%)	21	43
2	B	351/351 (100%)	323 (92%)	28 (8%)	13	26
All	All	706/706 (100%)	657 (93%)	49 (7%)	17	34

All (49) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	5	LEU
1	A	38	VAL
1	A	50	PHE
1	A	94	GLU
1	A	129	LYS
1	A	186	SER
1	A	198	ARG
1	A	217	GLU
1	A	238	ARG
1	A	245	ASP
1	A	248	MET
1	A	269	GLU
1	A	304	ASP
1	A	310	LYS
1	A	311	THR
1	A	348	SER
1	A	356	GLN
1	A	359	HIS
1	A	384	THR
1	A	409	THR
1	A	447	GLU
2	B	4	ASP
2	B	18	THR
2	B	34	THR
2	B	40	ASP
2	B	50	PHE
2	B	51	SER
2	B	54	TRP
2	B	58	HIS
2	B	61	ARG
2	B	102	LYS
2	B	143	LYS
2	B	191	GLU
2	B	215	SER
2	B	231	GLU

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
2	B	234	LEU
2	B	263	MET
2	B	292	THR
2	B	332	LEU
2	B	346	GLU
2	B	359	HIS
2	B	379	VAL
2	B	409	THR
2	B	412	LEU
2	B	425	THR
2	B	427	GLU
2	B	442	GLU
2	B	443	VAL
2	B	444	THR

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (2) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	58	HIS
1	A	137	GLN

5.3.3 RNA

There are no RNA molecules in this entry.

5.4 Non-standard residues in protein, DNA, RNA chains

There are no non-standard protein/DNA/RNA residues in this entry.

5.5 Carbohydrates

There are no carbohydrates in this entry.

5.6 Ligand geometry

2 ligands are modelled in this entry.

In the following table, the Counts columns list the number of bonds (or angles) for which Mogul statistics could be retrieved, the number of bonds (or angles) that are observed in the model and

the number of bonds (or angles) that are defined in the Chemical Component Dictionary. The Link column lists molecule types, if any, to which the group is linked. The Z score for a bond length (or angle) is the number of standard deviations the observed value is removed from the expected value. A bond length (or angle) with $|Z| > 2$ is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	$\#\|Z\| > 2$	Counts	RMSZ	$\#\|Z\| > 2$
4	PLP	D	1	-	16,16,16	2.86	3 (18%)	20,23,23	1.55	3 (15%)
3	LLP	C	1	-	21,25,25	2.85	5 (23%)	26,34,34	1.44	3 (11%)

In the following table, the Chirals column lists the number of chiral outliers, the number of chiral centers analysed, the number of these observed in the model and the number defined in the Chemical Component Dictionary. Similar counts are reported in the Torsion and Rings columns. '-' means no outliers of that kind were identified.

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
4	PLP	D	1	-	-	1/8/8/8	0/1/1/1
3	LLP	C	1	-	-	3/15/19/19	0/1/1/1

All (8) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
4	D	1	PLP	C3-C2'	8.38	1.49	1.40
3	C	1	LLP	C3-C2'	8.36	1.49	1.40
3	C	1	LLP	C4'-NZ	5.73	1.46	1.27
4	D	1	PLP	C4-C5	5.44	1.48	1.42
3	C	1	LLP	C4-C5	5.34	1.48	1.42
3	C	1	LLP	C4-C3	5.09	1.48	1.40
4	D	1	PLP	C4-C3	5.02	1.48	1.40
3	C	1	LLP	C4-C4'	2.75	1.51	1.46

All (6) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
4	D	1	PLP	C4-C3-C2	-4.10	117.65	120.19
3	C	1	LLP	C4-C3-C2	-3.93	117.75	120.19
3	C	1	LLP	C4-C4'-NZ	-3.18	109.72	124.31
4	D	1	PLP	O4A-C4A-C4	-2.72	118.98	124.91
4	D	1	PLP	C6-N1-C2	2.62	124.02	119.17
3	C	1	LLP	C6-N1-C2	2.58	123.94	119.17

There are no chirality outliers.

All (4) torsion outliers are listed below:

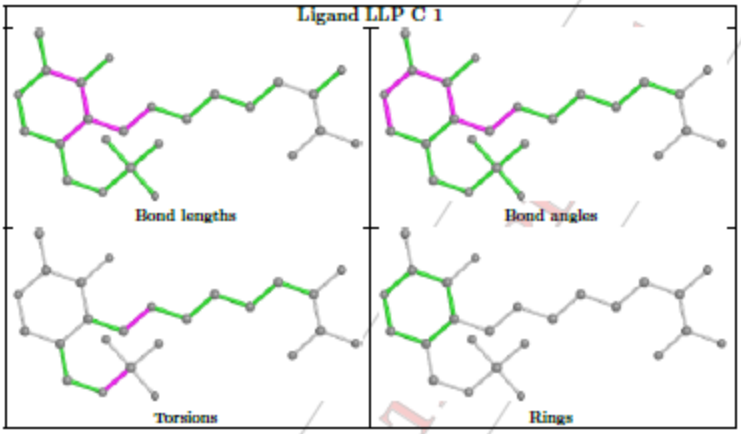
Mol	Chain	Res	Type	Atoms
3	C	1	LLP	C4-C4'-NZ-CE
3	C	1	LLP	C5'-OP4-P-OP1
3	C	1	LLP	C5'-OP4-P-OP2
4	D	1	PLP	C4-C5-C5A-O4P

There are no ring outliers.

2 monomers are involved in 8 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
4	D	1	PLP	1	0
3	C	1	LLP	7	0

The following is a two-dimensional graphical depiction of Mogul quality analysis of bond lengths, bond angles, torsion angles, and ring geometry for all instances of the Ligand of Interest. In addition, ligands with molecular weight > 250 and outliers as shown on the validation Tables will also be included. For torsion angles, if less then 5% of the Mogul distribution of torsion angles is within 10 degrees of the torsion angle in question, then that torsion angle is considered an outlier. Any bond that is central to one or more torsion angles identified as an outlier by Mogul will be highlighted in the graph. For rings, the root-mean-square deviation (RMSD) between the ring in question and similar rings identified by Mogul is calculated over all ring torsion angles. If the average RMSD is greater than 60 degrees and the minimal RMSD between the ring in question and any Mogul-identified rings is also greater than 60 degrees, then that ring is considered an outlier. The outliers are highlighted in purple. The color gray indicates Mogul did not find sufficient equivalents in the CSD to analyse the geometry.



5.7 Other polymers

There are no such residues in this entry.

5.8 Polymer linkage issues

The following chains have linkage breaks:

Mol	Chain	Number of breaks
1	A	4
2	B	2

All chain breaks are listed below:

Model	Chain	Residue-1	Atom-1	Residue-2	Atom-2	Distance (Å)
1	B	146:TYR	C	174:TYR	N	15.06
1	A	147:HIS	C	173:SER	N	14.77
1	B	412:LEU	C	413:GLY	N	1.18
1	A	4:ASP	C	5:LEU	N	1.16
1	A	210:GLU	C	211:ILE	N	1.15
1	A	5:LEU	C	6:ILE	N	1.06

6 Fit of model and data

6.1 Protein, DNA and RNA chains

In the following table, the column labelled '#RSRZ> 2' contains the number (and percentage) of RSRZ outliers, followed by percent RSRZ outliers for the chain as percentile scores relative to all X-ray entries and entries of similar resolution. The OWAB column contains the minimum, median, 95th percentile and maximum values of the occupancy-weighted average B-factor per residue. The column labelled 'Q< 0.9' lists the number of (and percentage) of residues with an average occupancy less than 0.9.

Mol	Chain	Analysed	<RSRZ>	#RSRZ>2		OWAB(Å^2)	Q<0.9
1	A	420/420 (100%)	0.85	55 (13%)	3 2	19, 42, 68, 85	0
2	B	416/416 (100%)	0.91	49 (11%)	4 2	19, 46, 72, 89	0
All	All	836/836 (100%)	0.88	104 (12%)	4 2	19, 44, 71, 89	0

All (104) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	B	384	THR	5.9
1	A	184	SER	5.7
2	B	391	ASP	5.1
1	A	185	GLU	4.9
2	B	333	ALA	4.7
2	B	185	GLU	4.3
1	A	74	GLY	4.0
1	A	110	SER	3.9
2	B	195	GLN	3.8
1	A	448	ASP	3.7
1	A	4	ASP	3.7
1	A	392	ARG	3.5
2	B	145	GLY	3.5
1	A	112	SER	3.4
2	B	388	TYR	3.3
1	A	318	TYR	3.3
2	B	224	SER	3.2
1	A	146	TYR	3.2
1	A	446	LEU	3.2
1	A	145	GLY	3.2
1	A	113	GLY	3.1
2	B	344	VAL	3.1
1	A	139	PHE	3.0
2	B	401	LYS	2.9

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
2	B	111	ASP	2.9
2	B	445	SER	2.9
2	B	186	SER	2.9
2	B	318	TYR	2.8
1	A	177	PRO	2.8
1	A	209	GLU	2.8
2	B	112	SER	2.8
1	A	191	GLU	2.8
2	B	390	ALA	2.7
1	A	387	PRO	2.7
2	B	404	GLU	2.7
1	A	202	GLN	2.7
1	A	114	ALA	2.7
2	B	113	GLY	2.6
2	B	446	LEU	2.6
1	A	111	ASP	2.6
1	A	391	ASP	2.6
1	A	200	LEU	2.6
2	B	43	GLY	2.6
1	A	55	LEU	2.5
1	A	87	VAL	2.5
2	B	405	LEU	2.5
2	B	117	MET	2.5
1	A	143	LYS	2.5
2	B	183	ARG	2.4
2	B	443	VAL	2.4
1	A	116	ALA	2.4
1	A	447	GLU	2.4
2	B	343	GLN	2.4
1	A	383	GLU	2.4
2	B	146	TYR	2.4
1	A	50	PHE	2.4
2	B	212	ALA	2.4
1	A	287	LEU	2.3
1	A	54	TRP	2.3
1	A	207	HIS	2.3
1	A	380	ARG	2.3
2	B	356	GLN	2.3
2	B	355	LEU	2.3
2	B	144	ASN	2.3
2	B	184	SER	2.3
2	B	199	GLU	2.3

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	A	253	VAL	2.3
1	A	278	ALA	2.2
2	B	357	ASP	2.2
1	A	319	THR	2.2
2	B	441	HIS	2.2
2	B	360	ALA	2.2
2	B	316	HIS	2.2
2	B	141	ALA	2.2
1	A	285	GLY	2.2
2	B	387	PRO	2.2
1	A	221	GLN	2.2
1	A	254	ALA	2.2
1	A	324	GLY	2.2
1	A	176	ALA	2.2
2	B	309	LEU	2.2
1	A	359	HIS	2.1
2	B	134	PRO	2.1
1	A	206	GLU	2.1
2	B	175	LYS	2.1
2	B	398	VAL	2.1
1	A	384	THR	2.1
2	B	277	ALA	2.1
1	A	321	ASN	2.1
1	A	109	TYR	2.1
1	A	230	PRO	2.0
1	A	288	PRO	2.0
2	B	315	GLY	2.0
1	A	147	HIS	2.0
1	A	292	THR	2.0
2	B	307	GLU	2.0
1	A	195	GLN	2.0
2	B	424	SER	2.0
2	B	203	LEU	2.0
2	B	442	GLU	2.0
1	A	290	ALA	2.0
2	B	54	TRP	2.0
1	A	182	TYR	2.0
1	A	343	GLN	2.0

6.2 Non-standard residues in protein, DNA, RNA chains

There are no non-standard protein/DNA/RNA residues in this entry.

6.3 Carbohydrates (i)

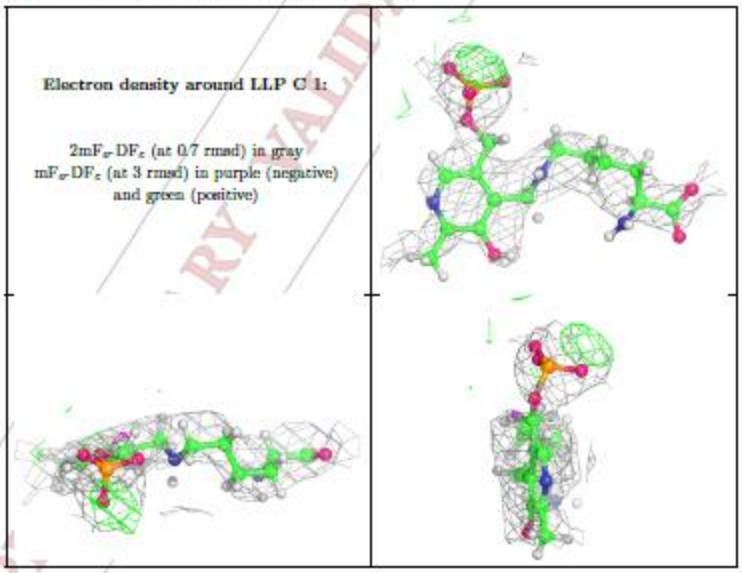
There are no carbohydrates in this entry.

6.4 Ligands (i)

In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. The B-factors column lists the minimum, median, 95th percentile and maximum values of B factors of atoms in the group. The column labelled 'Q< 0.9' lists the number of atoms with occupancy less than 0.9.

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(Å^2)	Q<0.9
3	LLP	C	1	25/?	0.76	0.39	61,71,76,77	0
4	PLP	D	1	16/?	0.93	0.29	49,76,84,94	0

The following is a graphical depiction of the model fit to experimental electron density of all instances of the Ligand of Interest. In addition, ligands with molecular weight > 250 and outliers as shown on the geometry validation Tables will also be included. Each fit is shown from different orientation to approximate a three-dimensional view.



6.5 Other polymers

There are no such residues in this entry.

Appendix D - Physicochemical properties of chemical compounds

Compound	*I.d.*	*DSX score*	*FW (neutral)*	*logD (7.4)*	k_m	k_i	λ *(nm)/*ε *($M^{-1}cm^{-1}$)*
E1			270.37	4.31		10 uM (CYP19)	
E2			272.38	3.74	3 nM (GPER)		
Androstenedione			286.41	3.93	15-30 nM (CYP19)		238/14,550
G1 agonist			412.3				
MFC							203/-, 334/6,790
HFC							203/-, 341/1,891 (391/-, 341 peak shoulder)
Tamoxifen							
4-OH tamoxifen							
Endoxifen						4.0 uM (CYP 19)	243/22,850 282/14,440
Norendoxifen						70 nM (CYP19)	
Norgestrel			312.45	3.96			
AR11		-136	343.43	2.58			245/20, 010
AR11-2			355.44	2.59			
AR12		-135	474.64	4.63			
AR13		-136	347.42	2.45			(234/26,090 far UV peak shoulder) 282/4,240
AR13-2 (RN1)			381.52		20 mM (water)		
AR15	Z1396098874 (enamine)	-116	329.40				
AR16	Z654473400 (enamine)	-115	347.42				
AR17	58801956 (chembridge)	-118	335	2.21			
AR18	64459905 (chembridge)	-118	357	3.28			
AR19	10939549 (chembridge)	-122	346	3.04			
AR20	32312044 (chembridge)	-117	339	1.75			238/9,500

Higher ranking drugs have DSX values that are more negative. Extinction coefficients were experimentally determined.

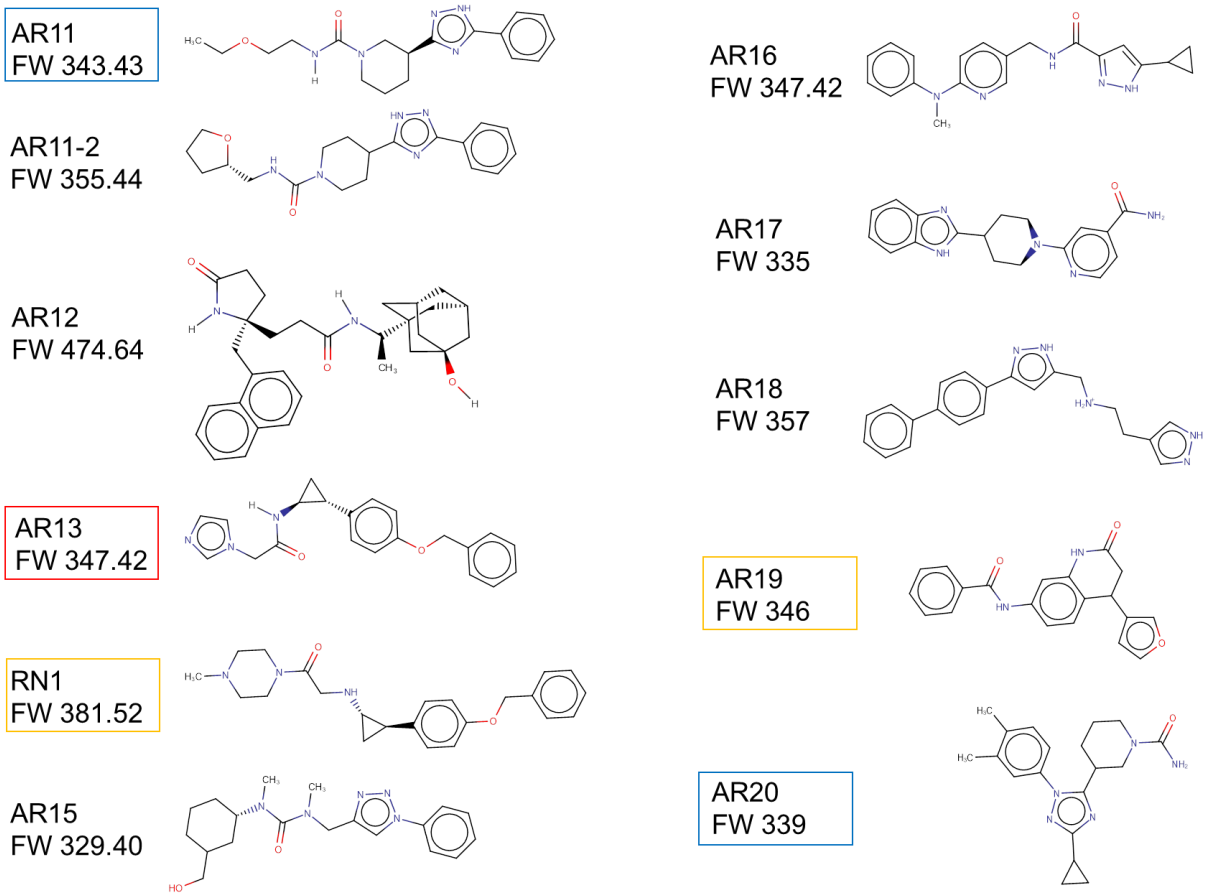


Figure D.1 Compounds screened for antiaromatase activity.

Compounds are color-coded according to the level of inhibition. IC_{50} < 1 μM (red), < 45 μM (blue), and < 90 μM (yellow).