Mechanisms of allosteric and mixed mode aromatase inhibitors

Samson A. Souza[1], Abby Held[2], Wenjie Lu[3], Brendan Drouhard[1], Bryant Avila[1], Raul Leyva-Montes[1], Michelle Hu[3], Bill R. Miller III[2], Ho Leung Ng[1*]

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

[2] Department of Chemistry, Truman State University, Kirksville, MO

[3] Department of Chemistry, University of Hawai'i at Mānoa, Honolulu, HI

[*]Corresponding author: hng@ksu.edu

Abstract

Aromatase (Cyp19) catalyzes the last biosynthetic step of estrogens in mammals and is a primary therapeutic target for postmenopausal women with hormone-related breast cancer. However, treatment with aromatase inhibitors is often associated with adverse effects and drug resistance. In this study, we used virtual screening to target a potential cytochrome P450 reductase binding site to discover four novel non-steroidal aromatase inhibitors. The inhibitors have potencies comparable to the noncompetitive tamoxifen metabolite, endoxifen. Our two most potent inhibitors, AR11 and AR13, exhibit mixed-type and competitive-type inhibition. The cytochrome P450 reductase binding site likely acts as a transient binding site. Modeling shows that our inhibitors actually bind better at various sites near the catalytic site. These structures may serve as chemical scaffolds to inhibit aromatase with different adverse effects profiles than clinically used aromatase inhibitors.

Keywords

Aromatase, aromatase inhibitor, P450 inhibitor, allosteric inhibitor

Introduction

Aromatase (Cyp19) catalyzes the three-step transformation of androgens to estrogens in mammals by a mechanism that is well-characterized. The first two steps produce a C-19 gem-diol [1]. 1β-hydrogen abstraction and electron transfer furnish the A-ring aromatization and loss of a formate molecule to complete a single turnover [2]. As such, aromatase inhibition is a common strategy to treat patients with hormone-dependent breast cancer [3]. Third-generation aromatase inhibitors (AIs) are used as primary therapeutics and long-term adjuvants in postmenopausal women with breast cancer.

AIs in clinical use include the non-steroidal AIs (NSAIs), anastrozole and letrozole, and the steroidal AI, exemestane, which interacts covalently at the active site [4,5]. The most potent NSAI, letrozole, binds a distinct site exhibiting both noncompetitive and mixed-mode type Michaelis inhibition [6]. This type of behavior is also observed for the noncompetitive AI endoxifen (K_i = 4 μM), a potent metabolite of tamoxifen, a clinically-used estrogen receptor α (ERα) antagonist [7]. Metabolic N-demethylation of endoxifen produces a competitive AI, norendoxifen (K_i = 35 nM) [8]. This demonstrates two important points. First, small chemical modifications of inhibitors (letrozole to anastrozole, and endoxifen to norendoxifen) can change the inhibitory mechanism in unexpected ways. Second, Cyp19 can be allosterically modulated with high potency. The mechanism of mixed mode inhibition for Cyp19 and other cytochrome P450 (CYP450) enzymes is unclear.

Steroidal AIs interact at the active site, also known as the distal heme site, with high efficacy. However, steroidal analogs frequently exert similar adverse effects. Cyp19 inhibition at an alternative allosteric site provides opportunities for the discovery of novel NSAIs chemically distinct from current therapeutics and with different toxicity profiles.

In this work, we investigate the binding of small molecule inhibitors to the Cyp19 heme distal site, the heme proximal site (the predicted site of CPR binding), and the substrate access channel through a combination of experimental and computational approaches. We first predict and model the interfacial contacts of cytochrome P450 reductase (CPR) with Cyp19. We performed virtual screening against a library of over a million compounds to identify potential inhibitors. We use molecular dynamic simulations to model how they interact at the CPR-Cyp19 interface. We then provide experimental enzyme inhibition data for four new NSAIs we discovered, AR11, AR13, AR19, and AR20. In addition, we use optical absorption spectra to characterize the effects of inhibition on the heme chemical environment. We characterize the inhibition modes of AR11, AR13, AR19, and AR20, and their likeliest binding sites closer to the catalytic site. The cytochrome P450 reductase binding site is likely a transient, low-affinity binding site for multiple ligands.

Materials and Methods

Protein-protein docking. For protein-protein docking, we used the crystal structures of human CYP19 (PDB 4KQ8) with a human-yeast chimeric enzyme [9] (PDB 3FJO) in an open conformation. There is currently no full-length human CPR crystal structure with an exposed FMN-containing face. Human CPR has an intrinsically dynamic hinge region causing the N-and C-terminal domains to adopt a similar open conformation [10,11]. The chimeric protein maintains functionality reducing both cytochrome c and human P450s [9]. The Haddock[12] webserver was used for the interface prediction-driven docking [13] of Cyp19 (PDB 4KQ8) to CPR. Haddock uses the CPORT [13–18] ensemble with multiple interface predictors to predict the interfacial residues to input as restraints to return high scoring clusters of complexes. Residues that are solvent-exposed and predicted to reside at the CYP19-CPR interface were defined as "Active" for Haddock docking, and surrounding solvent-exposed residues were defined as "Passive" (Table S1). The Gibbs free energy and dissociation constant estimates were calculated by the PRODIGY prediction webserver [19,20]. Conserved Cyp19 surfaces were calculated using Bayes' theorem with the ConSurf server [21,22]. Human Cyp19 (PDB 4KQ8) was queried against 150 homologs across different P450 families by setting a 30 % maximum and 10 % minimum sequence identity cutoff.

Virtual screening and docking. Idock [23,24] was used to virtually screen a library of over one million drug-like molecules from the ZINC database [25]. The highest scoring molecules were rescored with the DSX knowledge-based scoring function, which has been shown to be more predictive of binding affinity than docking scoring functions [26]. Compounds were assessed by the interactions made at the proximal heme site and their DSX scores. The top hits were selected for *in vitro* enzyme inhibition screening. Compounds that exhibited moderate to potent anti-aromatase activity were docked against a crystal structure of Cyp19 (PDB 3S79) with a more stringent protocol for MD simulations, involving 10 independent docking runs with Autodock Vina using different random seeds [27]. Autodock Vina exhaustiveness was set to 128, and the number of binding modes was capped at 20 per run. Visual inspection by VMD [28] of clustered docking results was used to predict the best binding mode to be used for MD simulations. The stereochemistry of AR13 was ambiguous from the vendor's (Enamine) molecular description. Both the cis- and trans-cyclopropane forms were used for docking and further modeling.

Molecular dynamics simulations. The AMBER16 [29] molecular dynamics package was used to produce MD simulations. Parameters for the unliganded heme described by Shahrock et al. were used [30]. The FF14SB and GAFF forcefields were used for protein and ligands, respectively. Simulations were carried out with the TIP3P explicit solvent model in an octahedral box truncated 11 Å from the protein surface. Chloride ions were parameterized for charge neutralization of the system. Energy minimizations were achieved with an 8 Å cutoff for the non-bonded energy term for each atom. A subsequent 2 ns heat step to 300 K with the SHAKE algorithm fixed bond lengths involving hydrogen atoms. A 3.5 ns equilibration step preceded the 1 µs production runs. Simulations were run in triplicate. Trajectories were processed using the AMBER cpptraj tools [31]. The binding free energies and decomposition free energies were calculated using MMPBSA.py [32].

Molecular graphics depictions and regression analyses. Standard molecular graphics were visualized with Pymol (The PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC.) Conservation maps were generated with UCSF Chimera [33] software. All regression analyses were performed in GraphPad Prism version 8.0.0 for Windows, GraphPad Software, San Diego, California USA. GraphPad functions used for curve-fitting are reported in the appropriate methods sections.

IC50 assays of top candidates. Compounds were purchased from vendors including Enamine and ChemBridge. We resolved the stereochemistry of compound AR13 using 1D and 2D NMR, which supported a trans configuration at the cyclopropane moiety. In brief, very weak through-space coupling and a ^{3}J value of 3.4 supported protons with an uneclipsed dihedral (Figure S1, S2).

Enzyme activity was measured by monitoring the conversion of exogenous substrate 7-methoxy-4-(trifluoromethyl)coumarin (MFC) to its fluorescent product 7-hydroxy-4-(trifluoromethyl)coumarin (HFC). A Cyp19/MFC high-throughput screening kit (Corning) was used to measure inhibition of aromatase activity for compounds AR11 and AR13. Briefly, 2X NADPH regeneration system (16.25 µM $NADP^{+}$, 825 µM $MgCl_2$, 825 µM glucose-6-phosphate, 0.4 U/mL G6P dehydrogenase) was prewarmed with inhibitors (1:2 serial dilutions) at 37 □C for 10 minutes in black 96-well plates. Reactions were initiated with prewarmed 2X enzyme-substrate mix (15 nM P450 microsomes enriched with oxidoreductase, 50 µM MFC) and incubated at 37 □C for 30 minutes. The 200 µL reaction mixtures were terminated with 75 µL 0.5 M Tris-base (80% ACN). A FluoDia T70 plate reader measured HFC product formation with excitation/emission filters of 400/530 nm. Reactions were performed in duplicate then repeated twice more on separate days.

The reaction conditions were repeated for measuring the inhibitory activity of compounds AR19 and AR20. Modifications are highlighted here. Supersomes containing Cyp19 + CPR (Corning), NADPH regeneration system (Corning), MFC (Chemodex), and ketoconazole (Selleck Chemicals) were used for the reaction mixture. Temperature-controlled incubations of the 96-well plate were carried out with a dry-plate. A Tecan fluorescent plate reader measured the fluorescent product in the circle-read mode at the optimum gain with excitation/emission filters of 405/535 nm. Reactions were measured in duplicate.

Reaction blanks were used for data corrections and these results were normalized to the fluorescence response in the absence of inhibitor. Analyses were performed in GraphPad Prism 8 software, and data were fit to a 4- parameter logistic model on semilog axes.

Soret shifts by absorption spectroscopy. Codon-optimized Cyp19A1 cDNA in the pCW expression vector was a generous gift from the F. Peter Guengerich lab (Vanderbilt University). Preparation of the DNA construct is detailed by Sohl and Guengerich [1]. Recombinant protein was produced, purified, and characterized by methods detailed in the supplementary materials section.

Purified Cyp19 was diluted with 100 mM potassium phosphate buffer (pH 7.4) to 50 µL to a final concentration of 2–3 µM P450. Inhibitors were titrated such that the endpoint would not exceed 3 % ACN. Absorption scans were read with a single-beam Agilent 8453 UV-Vis spectrophotometer after resuspension and a 10-minute incubation period at 25 □C. Buffer A components, endoxifen, AR11, AR13, AR19, and AR20 did not contribute to hyperchromic shifts in the Soret peak region.

AR11 and AR13 kinetics assay. Cyp19 + reductase microsomal preparations (BTI-TN-5B1-4, Corning) were used to measure the conversion of substrate MFC to HFC. Reactions contained 96 µL of NADPH regeneration solution and 4 µL of inhibitor from serial stocks. Mixtures were prewarmed at 37 □C in black 96-well round-bottom plates. Reactions were initiated with a single stream of 100 µL prewarmed enzyme-substrate solution in sequence. Parafilm and aluminum foil were applied to plates before incubation within the linear range at 37 □C for 20 minutes. The reactions were quenched with 75 µL of 0.5 M Tris-base (80% ACN). HFC was measured with a Tecan plate reader in circle-read mode at 52 gain with excitation/emission filters of 405/535 nm. Data was generated from 10 reads and a 40 µs integration time. Reaction mixtures contained 100 mM PPB (pH 7.4), 10 nM P450, 0.325 mM $NADP^+$, 0.825 mM glucose 6-phosphate, 0.825 $MgCl_2$, 0.1 U/mL glucose 6-phosphate dehydrogenase, MFC (9.9, 14.8, 22.2, 33.3, and 50 µM), and various concentrations of inhibitor. Ligand sequestration was minimized by working at the lowest enzyme/microsomal concentration that would return a quantifiable

fluorescence response. At steady-state conditions, the effect on MFC is negligible since the MFC concentration is greater than 1000-fold that of the enzyme. We used an NADPH regeneration system to minimize the effects of NADPH depletion on the fluorescence response. GraphPad Prism 8 software was used for generating regression curves and analyzing data. 5 x 5 Lineweaver-Burk and 5 x 4 Dixon plots were used to diagnose AR11 and AR13 inhibitory modes. Nonlinear regression curves were fit to mixed- or competitive-type Michaelis functions to return kinetic constants.

Results

Predicted Cyp19-CPR interactions. P450s share a conserved fold with low sequence identity between different enzymes. Nearly all the conserved residues are in the buried regions. We used ConSurf, which identifies conserved surfaces [21], to find four major surface-exposed sites that are conserved across different P450 families (Figure 1).

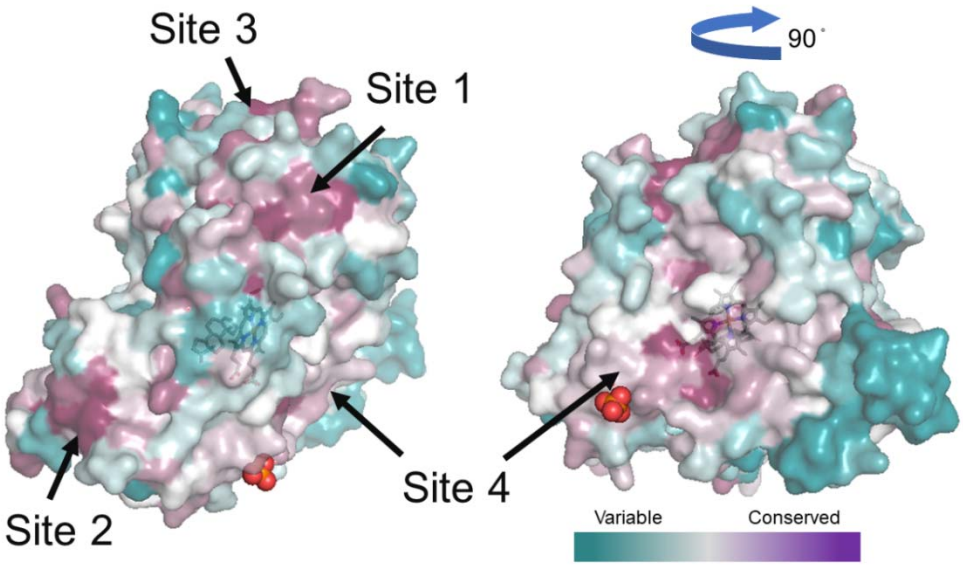


Figure 1. Surface representation of variable and conserved regions of aromatase coded as a cyan to magenta color-gradient. The four conserved regions are labeled as Site 1 – 4. Site 2 is the N-terminus linked to the transmembrane segment. Site 4 corresponds to the proximal heme site.

Only two of the sites were predicted by the protein-protein interface algorithm, CPORT [13], to actively participate in protein-protein interactions (sites 2 and 4 in Figure 1). They correspond to the N-terminal region (αA' and β1- 2) and the proximal heme site. The likeliest CPR binding surface was selected based on three criteria. Firstly, the proximity of the N5 atom of FMN to the iron center of the heme group should be within a distance that is physiologically sound. Secondly, the orientation of the N-terminus of the reductase and Cyp19 should be positioned in the same direction since they are truncations of transmembrane segments. Lastly, only the highest-scoring clusters from docking aromatase with CPR with Haddock [12] were assessed. Only site 4, corresponding to the proximal heme region, fulfilled all three criteria.

The probable binding mode contrasts with that of the crystal structure of *B. megaterium* P450BM3 fusion (PDB 1BVY). The N5 to iron distance in our structure is 14.4 Å, whereas the distance is

22.7 Å in P450BM3. This is within the 14 - 15 Å threshold limit for electron transfer in most physiological processes [34]. The possibility of through-bond tunneling at much longer distances in P450 BM3 was previously refuted due to faster experimental kinetic rates than predicted from theoretical models [35]. Physicochemical descriptions of the modeled and fusion P450-CPR complexes are compared in Table S2. In docking Cyp19 against CPR in a closed conformation (PDB 3QE2), the closest N5 to iron distance was 34.9 Å. The negative electrostatic potential of the FMN domain interacts with the positive potential of the FAD/NADPH domains. In this conformation, $FADH_2$ is in a closer proximity to reduce the FMN cofactor. In the proposed end-on interaction, 3FJO adopts an extended conformation to expose the buried FMN-binding interface for the reduction of Cyp19 at the proximal heme site (Figure 2, S3).

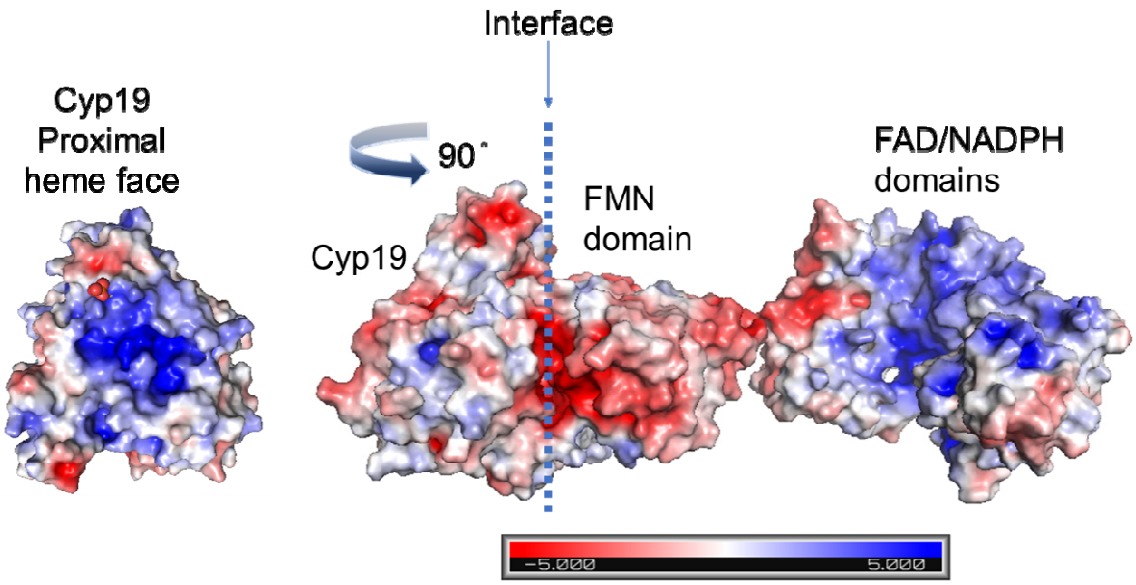


Figure 2. Electrostatic potential surface map of the proposed end-on binding mode of Cyp19 (PDB 4KQ8) in complex with CPR (PDB 3FJO) in an open conformation. Negatively charged potentials (red) to positive potentials (blue) are represented as a color gradient with neutral (gray) regions.

At the Cyp19-CPR interface there are 17 total polar-polar, polar-charged, and charged-charged bond pairs. The most important of these contacts involve sidechains from the Cyp19 residues K108, Y424, K440, and Y441 (Figure 3, Table S2).

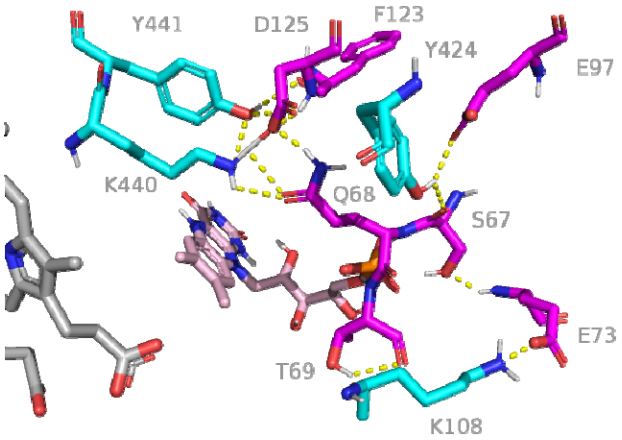


Figure 3. Predicted electrostatic contacts (yellow dotted trace) between Cyp19 (cyan) and the CPR FMN domain (magenta). The FMN cofactor (pink) and heme (gray) are within the 15 Å limit of electron tunneling processes.

IC50 values of top hits. The DSX (knowledge-based scoring function)[26] scores and chemical structures of the top four hit compounds are provided in Figure 4 and Table S3. These compounds showed IC_{50} values < 70 µM. AR13 showed 3-fold increased potency over the control inhibitor, ketoconazole, an antifungal with IC_{50} = 3.08 µM. The IC_{50} value we report for AR19, 66 µM, was fit to a Hill coefficient of 1 with a predicted 14% activity at saturation. The R^2 is suboptimal due to the poor solubility of AR19. AR11 and AR20 exhibited potency comparable to an active tamoxifen metabolite, endoxifen. Dose-response curves and activity data are presented in Figure 5 and Table 1.

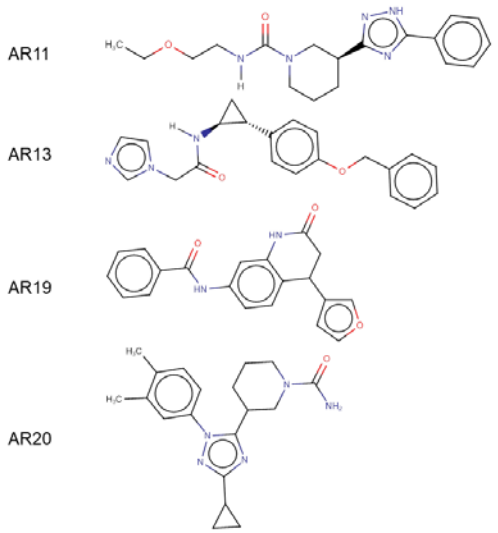


Figure 4. Four most active aromatase inhibitors from virtual screening.

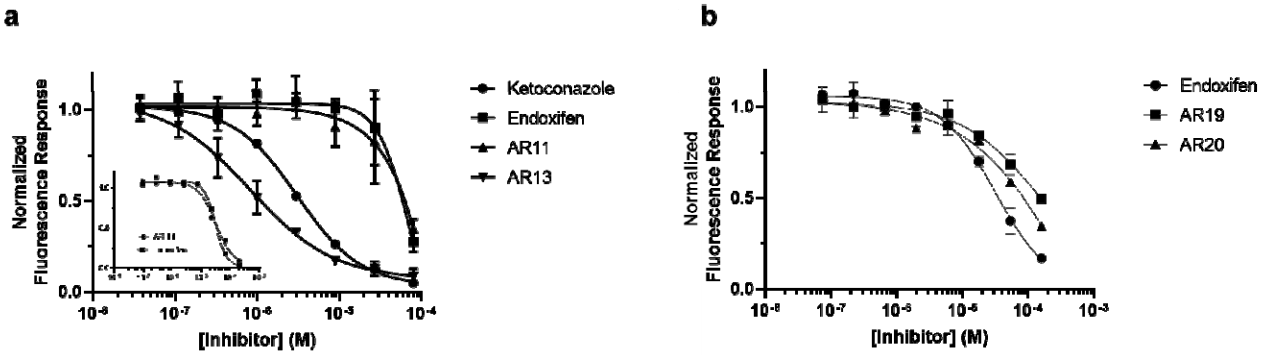


Figure 5. Dose-response curves at 7.5 nM P450 and 25 µM MFC substrate with varying concentrations of inhibitors a) AR11 and AR13, and b) AR19 and AR20. Semi-log curves are fit to a 4-parameter logistic model with data from positive control inhibitors ketoconazole and endoxifen. Reaction mixtures with a 3-fold increase in concentration for AR11 and endoxifen are provided as an inset in (a).

Table 1. Activity data collected from active compounds fit to a 4-parameter logistic 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
AR13	0.82	0.86	0.071	0.978
*AR19**	65.87	n/a	0.144	0.845
AR20	42.83	0.86	0.028	0.992

*3-parameter model reported with a Hill coefficient of 1.

Optical absorption properties of active compounds. P450s exhibit signature Soret peaks that are detectable in the 390 – 460 nm range [36]. Perturbations to the heme environment yield absorption shifts that are conserved. In the absence of substrate, water typically occupies the sixth site of the iron octahedral complex. This is evidenced by a Soret peak typically in the 415 – 417 nm range [36]. Interactions between the heme iron and stronger field ligands will induce absorption shifts at longer wavelengths. The addition of the native substrate, androstenedione (ASD), causes water displacement due to the C-19 protrusion from the ASD backbone [37]. This is indicated by a hypsochromic shift typically

in the 390 – 394 nm range due to iron's adoption of a 5-coordinate high-spin state (5-CHS) [36]. We report that the heme cofactor in recombinant Cyp19 displays a Soret peak shift in 100 mM PPB (pH 7.4) at 416 nm (blue trace) and 395 nm in absence and presence of ASD, respectively. Titration of AR13 induces an 8 nm bathochromic shift to 424 nm in the absence of ASD (Figure 6a). This indicates a direct interaction with the iron by a stronger field ligand than water. Additionally, the enzyme may reversibly adopt the 5-CHS in the presence of AR13 with a 25:1 molar excess of ASD to inhibitor (Figure 6b). Further, the Cyp19-AR13 adduct can be reduced with dithionite to bind CO reversibly. The subsequent addition of AR13 fully recovers the 424 nm peak after 1 hour.

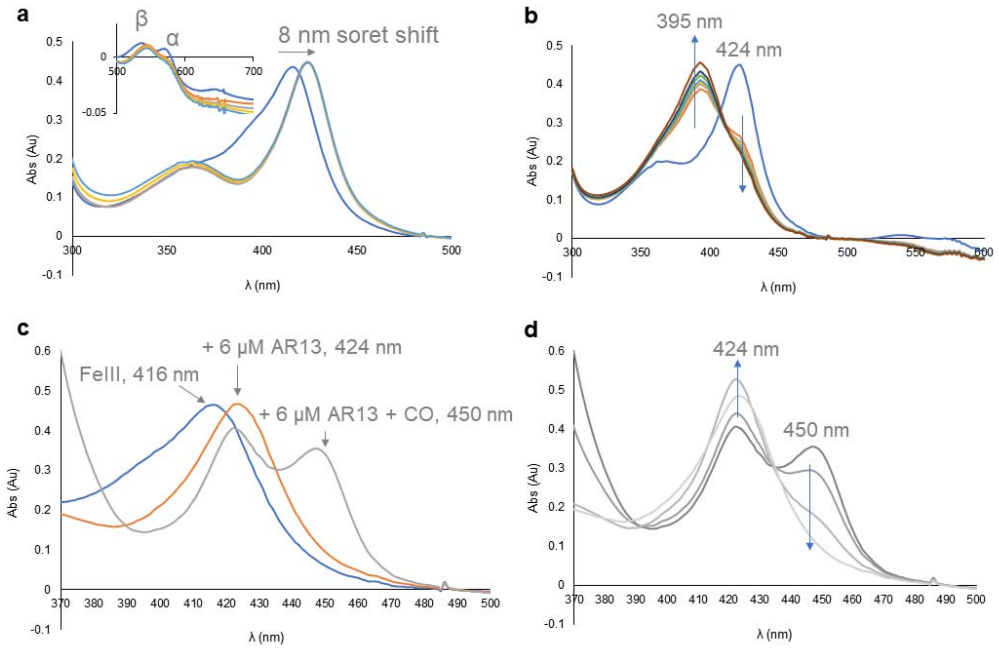


Figure 6. Soret peak shifts in the presence of inhibitor AR13. a) Bathochromic shift of Soret peak maximum from 416 to 424 nm upon titration with AR13 at 6, 12, 60, and 120 μM AR13. Inset shows a 5 nm shift in the β band from 535 to 545 nm and loss of the α band at 570 nm. b) Time-dependent increase of P450 in the high-spin state after the addition of 70 μM ASD to a reaction mixture with 2.5 μM AR13. c) P450 peak is observed after dithionite and CO addition to a reaction mixture of 3 μM P450 and 6 μM AR13. Subsequent addition of 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 (d).

Titration of up to 100 μM of endoxifen did not induce a Soret peak shift from 395 nm in the Cyp19 ASD-bound state (Figure 7a). This indicates that ASD remains in the active site in the presence of a 50-fold molar excess of endoxifen. This behavior is expected of a noncompetitive inhibitor where the K_i is unchanged in the presence of substrate. At a 3-fold (6 μM) molar excess, AR11 prompted the

appearance of a peak shoulder near 416 nm. Figure 7b illustrates the gradual increase of the peak shoulder at 414 nm with an increase in the concentration of inhibitor. This indicates that AR11 causes the enzyme to favor a shift to the 6-coordinate low-spin state and the displacement of ASD in the active site. At 100 µM AR11, the inhibitor-bound 6-coordinate low-spin state is apparent at 416 nm (Figure 7d). The inset in Figure 7d illustrates that AR11 causes Cyp19 to favor the low-spin state in the absence of its native substrate.

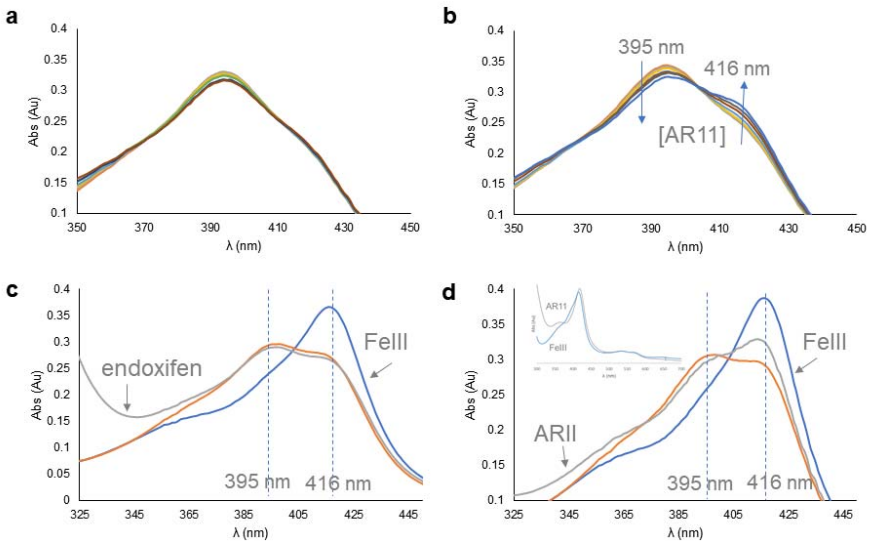


Figure 7. Optical absorption spectra of reaction mixtures in the presence of 3 µM P450, 2 µM ASD, and various concentrations of endoxifen (a and c), and AR11 (b and d). Titration with inhibitor from 6 – 36 µM (a – b), and at 100 µM (c – d). The addition of AR11 favors a transition to the 6-coordinate low spin state indicated by a hyperchromic shift at 416 nm. Blue arrows in panel b indicate the absorption trend as the concentration of AR11 increases. Inset in panel d shows Cyp19 in the absence and presence of 100 µM AR11.

AR19 and AR20 share the same Soret peak trends in the presence and absence of 2 µM ASD. Both compounds induce a gradual shift to the iron low-spin state indicated by simultaneous hyperchromic and red shifting towards 420 nm. This indicates that the population of Cyp19 with iron in the 6-coordinate state increases in the presence of inhibitor. The red shift from 416 nm (blue trace) to 420 nm (red trace) is apparent in Figures 8c and 8d. At 100 µM AR19, 2 µM ASD introduces a faint peak

shoulder at 395 nm. Although there is a drop in the absorption at 420 nm, a peak shoulder is not apparent at this concentration. Higher concentrations of AR19 are required to induce the same Soret peak effects as AR20, suggesting that AR20 is a more potent inhibitor.

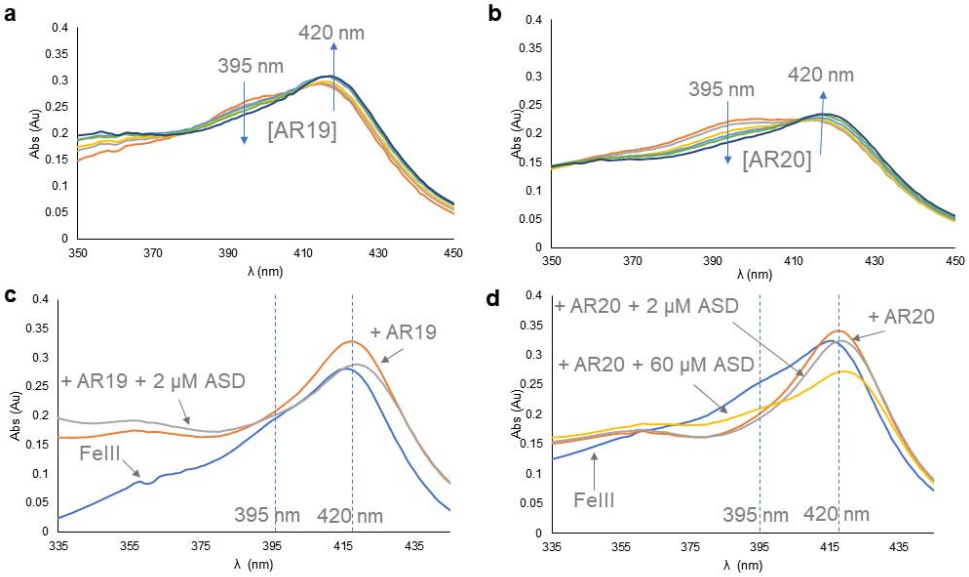


Figure 8. Optical absorption spectra of reaction mixtures in the presence of 2 μM P450, 2 μM androstenedione (ASD), and various concentrations of AR19 (a and c), and AR20 (b and d). a-b) Titration with inhibitor 6 – 36 μM favors a transition to the 6-coordinate low spin state. Blue arrows indicate the absorption trends as the concentration of inhibitor increases. c – d) Addition of 100 μM inhibitor to Cyp19 induces a 4 nm shift from 416 nm to 420 nm. A 395 nm peak shoulder is apparent after the subsequent addition of the native substrate, ASD.

AR11 and AR13 steady-state kinetic assays. Kinetic assays of our two most potent inhibitors were performed under steady-state conditions to determine their Michaelis-Menten and inhibition constants (Figures S4, S5). These assays were used as a diagnostic tool to confirm whether AR11 and AR13 interact at the active site or a distinct site. Inhibition assays were performed at various concentrations of MFC up to 2 X K_m. Nonlinear regression analyses in the absence of inhibitor yielded substrate MFC V_{max} and K_m values of 0.689 pmol HFC/min/pmol P450 and 24.38 μM.

Lineweaver-Burk plots of AR11 yielded functions that intersected the y-axis at different inhibitor concentrations. These corresponded to different apparent V_{max} values, suggesting that both inhibitors do not act competitively. Visual inspection of Dixon-type plots also suggested that AR11 reduces the

V_{max}. In Figure 9, the intersection of regression curves in panel A corresponds to the K_i value. The intersection of the curves in panel C corresponds to a K_i' or inhibitor dissociation constant in the presence of substrate MFC. This value infers non-mutual exclusive binding. As such, AR11 was fit to a mixed-type nonlinear function.

Lineweaver-Burk analyses of MFC conversion at different concentrations of AR13 inferred nonmutual exclusivity between substrate MFC and AR13 since the apparent V_{max} varied with inhibitor concentration. On the contrary, Dixon-type plots suggested that MFC and AR13 are mutually exclusive since the K_i' value was indeterminate (Figure 9d). Nonlinear regression fit to a competitive-type function yields a K_i of 39 nM, 35 – 44 nM for a 95 % confidence interval. This is within an order of magnitude from the working enzyme concentration and supports tight-binding behavior. In a Lineweaver-Burk plot, the $1/V_0$ values of tight-binders begin to converge at high substrate concentrations. Curvature in the AR13 data points is most recognizable at high inhibitor concentrations.

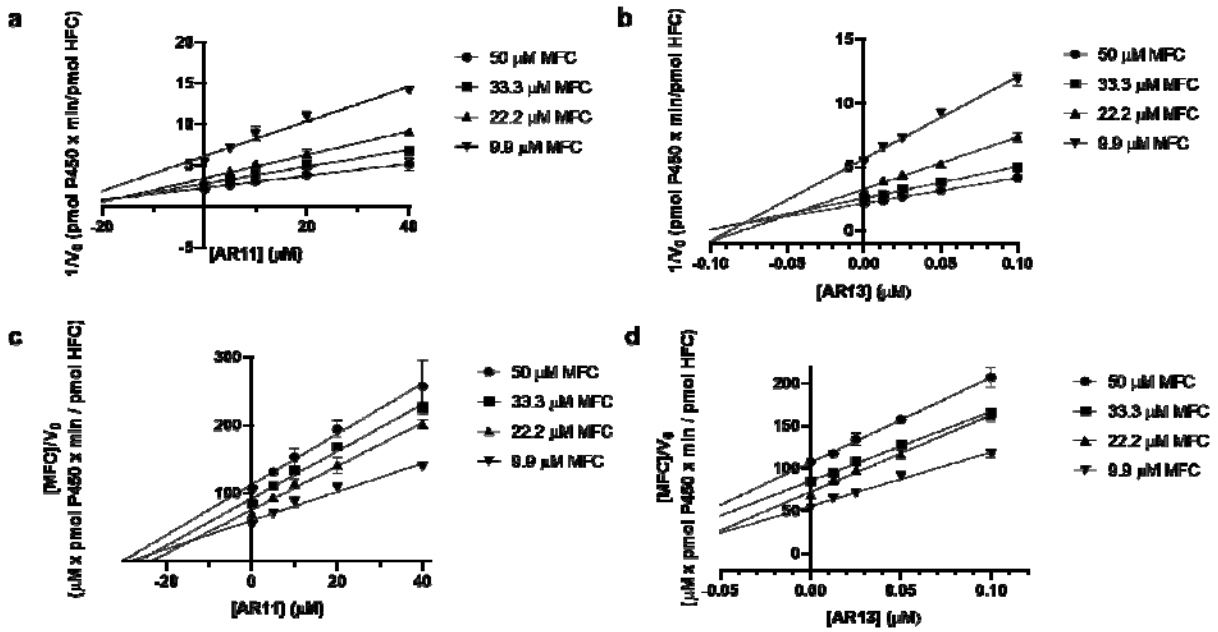


Figure 9. 5 x 4 Dixon-type plots for AR11 (a-b) and AR13 (c-d).

Non-linear regression for AR11 and AR13 fit to their diagnosed mode of inhibition returned K_m values of 24.60 and 22.01 µM at 10 nM P450. Table S4 presents the effects of each inhibitor on these kinetic constants. The apparent K_m increases in the presence of each inhibitor, indicating that more substrate MFC is needed to achieve ½ V_{max}. Meanwhile, the maximum velocity only decreased with

AR11. This indicates that MFC saturation will not attain full enzyme activity in the presence of AR11. Therefore, both substrate and AR11 bind Cyp19 at discrete sites. In contrast, the ability to achieve the V_{max} with a molar excess of MFC over AR13 indicates that both compounds bind the active site.

Table 2. Steady-state kinetic constants in the presence of inhibitors AR11 and AR13 at 10 nM P450.

	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	-	-

MD simulations of compounds docked against Cyp19. Molecular dynamics simulations were performed to observe the residence time of the inhibitors in the proximal heme binding pocket. All our active compounds, except AR11, dissociated from the proximal heme site in less than 1 μs suggesting modest binding strength. AR13 exhibited the weakest interaction at the proximal heme site. It remained in the pocket for less than 100 ns in all replicates.

More rigorous redocking of AR13 enantiomers to Cyp19 with 10 independent runs found that it bound more often near the catalytic active site than the proximal heme pocket (Figure 10). The enantiomers bound the active site with free energies of -36 ± 4 kcal/mol (1R,2S) and -36 ± 2 kcal/mol (1S,2R) as calculated by MM-PBSA. Both enantiomers are projected to form a π-π interaction with F221 in the substrate access channel and a low energy π-π interaction between the terminal azole and the heme's porphyrin system. The exposed (1S, 2R) azole nitrogen directly interacts with the iron at a 3.45 Å distance. Along these lines, Soret shifts support direct interactions of the imidazole with the heme iron, and the kinetic assays support competitive-type behavior.

Similarly, more rigorous redocking of AR11 located it in the substrate access channel more often than in the proximal heme site. AR11 formed π-π interactions with F221 and W224 and a hydrogen bond between the triazole and D309. This suggests that the triazole partially obstructs a space occupied by ASD since E309 is involved in a critical contact with the C3-carbonyl of androgens. Production runs yielded an average free energy of -29 ± 8 kcal/mol. In the proximal heme site, the free energy was -32 ± 4 kcal/mol, despite a destabilizing effect from E357. We report the putative binding modes at both the proximal heme site and substrate access channel in Figure 11 since their binding energies have

overlapping confidence intervals. We include the decomposition scores of the interactions with the greatest contributions to these energies.

Table 3. Docking and simulation metrics of endoxifen and novel Cyp19 inhibitors.

Ligand	*Proximal heme site average dissociation time (ns)*	*Proximal heme site average binding free energy (kcal/mol)*	*Substrate access average binding free energy (kcal/mol)*	*Active site average binding free energy (kcal/mol)*
E-end	< 200 ns	-17 ± 4		
Z- end	< 500 ns	-24 ± 3		
AR11	>1000	-32 ± 4	-29 ± 8	-
AR13 (1R,2S)	<100	-	-	-36 ± 4
AR13 (1S,2R)	-	-	-	-36 ± 2
AR19	-	-	-33 ± 3	-
AR20	800	-21.92 ± 0.04*	-	-

*denotes a value from 1 replicate run

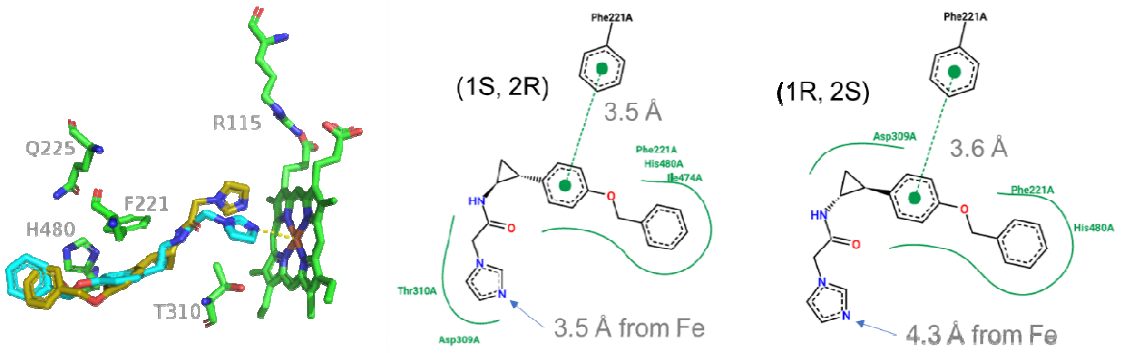
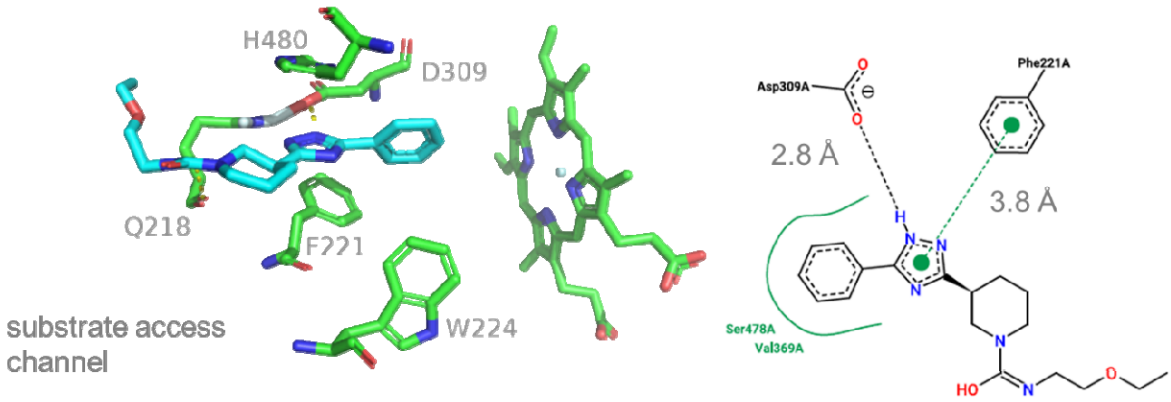


Figure 10. Interaction diagrams of the proposed binding modes of both AR13 enantiomers in the Cyp19 active site. The 1S, 2R (cyan) enantiomer forms a direct interaction with the iron center. The azole of 1R, 2S (yellow) enantiomer forms a π-π interaction with the porphyrin ring system. Both enantiomers are stabilized by F221 in the substrate access channel.

a

AR11 group	Residue- Group	Decomposition score (kcal/mol)
Ph	F221- Ar	-2.00 ± 0.08
Az-NH	Q218- CO	-1.86 ± 0.03
Az	H480- Ar	-1.48 ± 0.05
Az-NH	D309- COO^-	-1.05 ± 0.04
Ph	W224- Ar	-0.226 ± 0.002



b

AR11 group	Residue- Group	Decomposition score (kcal/mol)
Az-NH	F418- CO	-2.090 ± 0.013
carbamide-N	Y361- OH	-1.65 ± 0.02
Az	N421	-1.476 ± 0.003
	E357	0.762 ± 0.002

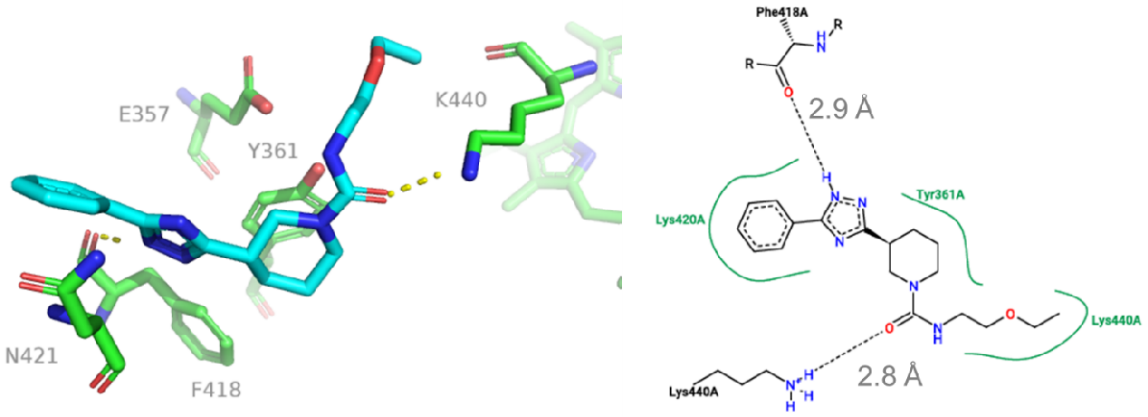
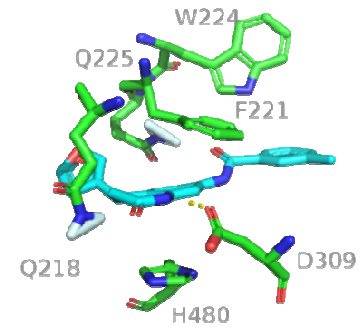


Figure 11. 3-D and 2-D interaction diagrams for AR11 with decomposition scores of the projected protein-ligand interactions at the substrate access channel (a) and the proximal heme site (b). 2-D depictions were generated with PoseView software [38]. Az- azole, Ph- phenyl, Ar- aromatic groups.

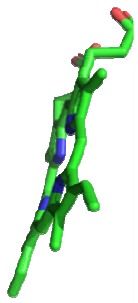
AR19 and AR20 dock to the proximal heme site but dissociate after developing unfavorable interactions with E357, as demonstrated by positive decomposition scores (Figures S6, S7). We also note that control inhibitors, E- and Z-endoxifen, disassociate due to positive free energy contributions from E357. In 1 out of 3 simulations, AR20 re-associates and remains in the site > 1 μs with -22 kcal/mol average binding free energy (Figure S8). We do not report decomposition scores for AR20 because it only transiently remains in a single binding mode throughout the production run. AR19 formed stable interactions in the substrate access channel with F221 and E309. Further, it had a proclivity to migrate closer into the active site for the duration of the production runs (Figure 12a). The average free energy was -33 kcal/mol.

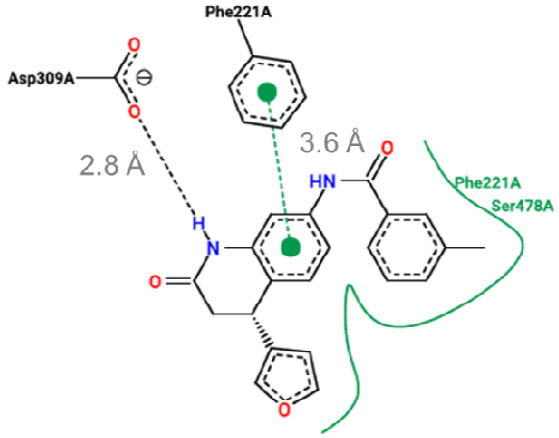
a

AR19 group	Residue- Group	Decomposition score (kcal/mol)
Ph	F221- Ar	-3.33 ± 0.04
amide-NH	Q225- $CONH_2$	-2.02 ± 0.05
Ph	H480- Ar	-1.873 ± 0.010
lactam-CO	Q218- $CONH_2$	-1.471 ± 0.011
Ph	W224- Ar	-0.913 ± 0.009



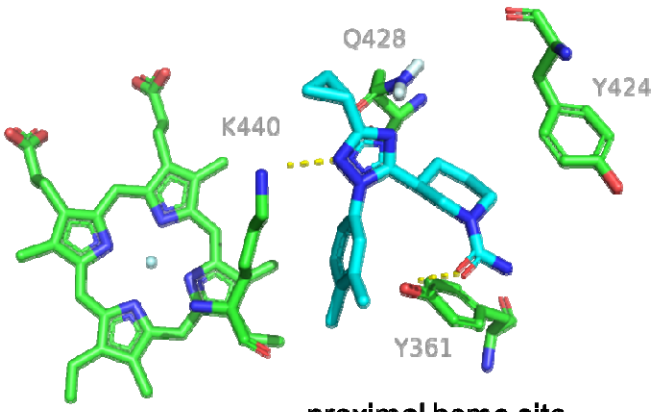
substrate access channel





b

AR20 group	Residue- Group	Decomposition score (kcal/mol)
Az-N	K440- NH^{3+}	-
carbamide-CO	Y361- OH	-



proximal heme site

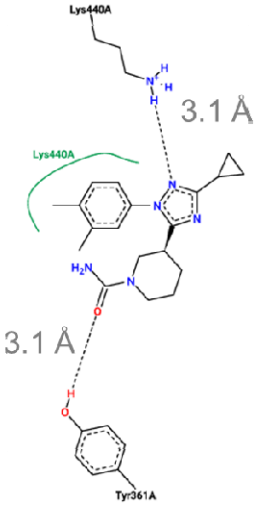


Figure 12. 3-D and 2-D interaction diagrams with decomposition scores of the projected protein-ligand interactions for AR19 (a) and AR20 (b). 2-D depictions were generated with PoseView software [38]. AR20 binding mode represented here is for replicate 2, and >280 ns. Decomposition scores are not available for AR20. Az- azole, Ph- phenyl, Ar- aromatic sidechain.

Discussion

MD simulations support that none of the inhibitors identified bind stably at the Cyp450 reductase binding site. Rather, AR13 binds the enzymatic active site, correlating well with the absorption data we present. The 8 nm Soret peak shift in the presence of AR13 indicates a direct interaction with the heme iron in the active site. Strong red shifts such as from a water-bound heme typify occupation of a strong, nitrogenous sigma-donor ligand. This is likely an interaction with AR13's terminal imidazole moiety. Cyp19's ability to bind ASD and CO in the presence of AR13 demonstrates its capacity to maintain its functionality. Above all, the need for a 28-fold molar excess of ASD to induce a blue shift to 395 nm supports the inhibitor's high potency. Among all our inhibitors, AR13 was the most potent. This correlates with its predicted average binding energy being the greatest over AR11, AR19, AR20, and endoxifen. AR13 has a 40-fold lower IC_{50} value than that of endoxifen and interacts with Cyp19 on the same order of magnitude as norendoxifen (K_i = 35 nM). Furthermore, its chemical structure is remarkably distinct from current third generation NSAIs. To this end, AR13 may be useful as a scaffold to design new NSAIs.

Titration with AR11 against androstenedione induces a peak shoulder between 414 – 416 nm, indicating that AR11 displaces ASD while substrate access to the heme iron is maintained. The average binding free energies at the proximal heme site and substrate access channel have overlapping confidence intervals and the Hill coefficient from a dose-response curve is greater than 1.5. As such, we project that both sites presented in this work are occupied at saturating concentrations of AR11. Likely, AR11 preferentially interacts at the substrate access channel since E357 is projected to be destabilizing. The transition to a 5-coordinate out of-plane position may cause the observed 2.6-fold increase in the K_i' at the proximal heme site. However, there are currently no crystal structure of Cyp19 in the 6-coordinate water-bound state for comparison.

We note that our control inhibitor, E/Z- endoxifen, exhibits similar trends as AR11 in this study. Both dock to the proximal heme site and consistently yield a Hill coefficient close to 2 suggesting multiple binding sites. We note two important differences. Firstly, MD simulations predict that E- and Z- endoxifen do not remain in the proximal heme site, which is consistent with work by Sgrignani et al [39].

Secondly, E/Z-endoxifen does not cause dissociation of ASD from the active site unlike AR11. This is potentially due to AR11 entering a pocket that is partially occupied by ASD.

Titration with AR19 and AR20 causes a 4 nm red shift from a water-bound iron to 420 nm. AR20 is a 1,2,4-triazole that may induce such a shift if there is a direct interaction with the heme iron or formation of a water-bridged ternary complex. P450 ternary complexes have been reported with 1,2,3- and 1,2,4- triazoles [40]. Although this suggests that AR19 and AR20 bind similarly to the active site, small Soret peak shifts may also arise from a change in the iron's chemical environment. Heme perturbation induced by allosteric interactions may cause such Soret peak shifts.

Altogether, we conclude that E/Z- endoxifen, AR11, and AR19 preferentially occupy the substrate access channel, inhibiting Cyp19 on the same order of magnitude. Since the substrate access channel is dynamic, compounds forming weaker interactions such as AR19 may migrate towards the active site to form a direct interaction with iron. E/Z- endoxifen may bind the substrate access channel with high affinity and a second site at saturating levels of inhibitor. We expect that AR11 interacts at the substrate access channel and the proximal heme site at saturating levels. In the substrate access channel, AR11 can bind deeper into the pocket such that it partially occupies space by ASD near the catalytically active E309. AR20 is projected to transiently interact at the proximal heme site.

The proximal heme site is acting as a low-affinity binding site for multiple inhibitors. The proximal heme site is an attractive target to modulate P450 activity because it is a conserved feature among P450s and distinct from the catalytic core. Despite class 2 P450s having less than a 40% sequence identity, all of them are expected to partner with CPR at this site. CYP19 may serve as a model to study its druggability for three reasons. Firstly, a crystal structure with the pentameric PEG bound at the proximal heme site is available and PEG inhibited enzyme activity in a time-dependent manner [41]. Secondly, Cyp19 has a well-defined cavity of 584 $Å^3$, roughly 200 $Å^3$ larger than the active site, that can accommodate a larger library of compounds [41]. Lastly, 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 2 P450 enzymes. This disparity offers a niche to selectively target CYP19 as less than 5% of P450s have a meander loop as lengthy [42]. It is likely that higher affinity inhibitors to the proximal heme site can be identified.

Other P450s that may be allosterically inhibited at the proximal heme site include Cyp3A4 and Cyp1A2. They are the major enzymes in drug metabolism and have well-documented heterotrophic effectors [43–45]. It is possible that these P450s can be modulated at the proximal heme site as the cavities are large enough to accommodate ligand-binding. The Cyp3A4 and Cyp1A2 pocket is 1.1 and 0.7 times

the size of the Cyp19 proximal cavity [41]. More research is needed to investigate alternative ligand binding sites in P450s [46–49].

In a broader context, nearly all exogenous compounds are subject to redox chemistry by P450 enzymes. Azoles and furan-containing xenobiotics, such as those presented here, are often P450 substrates [50]. There are also many important azole-containing drugs that are used to treat fungal infections [51]. However, they also inhibit human P450 isozymes and undergo various metabolic reactions. Off-target interaction with P450s may lead to toxicity or adverse drug interactions.

Acknowledgments

We thank Wesley Yoshida (Univ. Hawaii), Joshua Gurr (Univ. Hawaii), and the Kansas State University Biomolecular NMR Lab for the analytical work on compounds AR11 and AR13. In addition, we acknowledge the MERCURY Consortium for use of Skylight in the MD simulations presented in this work. We thank F. Peter Guengerich (Vanderbilt University) for providing the aromatase expression plasmid and general technical advice. HN was funded by the NSF CAREER Award 1833181, the Victoria and Bradley Geist Foundation, and the Hawaii Community Foundation.

Conflicts of Interest

The authors declare no competing interests.

References

1 C. D. Sohl and F. P. Guengerich, Kinetic Analysis of the Three-step Steroid Aromatase Reaction of Human Cytochrome P450 19A1, *J. Biol. Chem.*, 2010, **285**, 17734–17743.
2 F. K. Yoshimoto and F. P. Guengerich, Mechanism of the Third Oxidative Step in the Conversion of Androgens to Estrogens by Cytochrome P450 19A1 Steroid Aromatase, *J. Am. Chem. Soc.*, 2014, **136**, 15016–15025.
3 P. Fan and V. C. Jordan, New insights into acquired endocrine resistance of breast cancer, *Cancer Drug Resist.*, , DOI:10.20517/cdr.2019.13.
4 S. Maurelli, M. Chiesa, E. Giamello, G. Di Nardo, V. E. V. Ferrero, G. Gilardi and S. Van Doorslaer, Direct spectroscopic evidence for binding of anastrozole to the iron heme of human aromatase. Peering into the mechanism of aromatase inhibition, *Chem. Commun.*, 2011, **47**, 10737.
5 D. Ghosh, J. Lo, D. Morton, D. Valette, J. Xi, J. Griswold, S. Hubbell, C. Egbuta, W. Jiang, J. An and H. M. L. Davies, Novel Aromatase Inhibitors by Structure-Guided Design, *J. Med. Chem.*, 2012, **55**, 8464–8476.
6 C. Egbuta, J. Lo and D. Ghosh, Mechanism of Inhibition of Estrogen Biosynthesis by Azole Fungicides, *Endocrinology*, 2014, **155**, 4622–4628.
7 W. J. Lu, Z. Desta and D. A. Flockhart, Tamoxifen metabolites as active inhibitors of aromatase in the treatment of breast cancer, *Breast Cancer Res. Treat.*, 2012, **131**, 473–481.
8 W. J. Lu, C. Xu, Z. Pei, A. S. Mayhoub, M. Cushman and D. A. Flockhart, 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.*, 2012, **133**, 99–109.
9 L. Aigrain, D. Pompon, S. Moréra and G. Truan, Structure of the open conformation of a functional chimeric NADPH cytochrome P450 reductase, *EMBO Rep.*, 2009, **10**, 742–747.
10 D. Campelo, T. Lautier, P. Urban, F. Esteves, S. Bozonnet, G. Truan and M. Kranendonk, The Hinge Segment of Human NADPH-Cytochrome P450 Reductase in Conformational Switching: The Critical Role of Ionic Strength, *Front. Pharmacol.*, 2017, **8**, 755.
11 D. Hamdane, C. Xia, S.-C. Im, H. Zhang, J.-J. P. Kim and L. Waskell, Structure and Function of an NADPH-Cytochrome P450 Oxidoreductase in an Open Conformation Capable of Reducing Cytochrome P450, *J. Biol. Chem.*, 2009, **284**, 11374–11384.
12 G. C. P. van Zundert, J. P. G. L. M. Rodrigues, M. Trellet, C. Schmitz, P. L. Kastritis, E. Karaca, A. S. J. Melquiond, M. van Dijk, S. J. de Vries and A. M. J. J. Bonvin, The HADDOCK2.2 Web Server: User-Friendly Integrative Modeling of Biomolecular Complexes, *J. Mol. Biol.*, 2016, **428**, 720–725.
13 S. J. de Vries and A. M. J. J. Bonvin, CPORT: A Consensus Interface Predictor and Its Performance in Prediction-Driven Docking with HADDOCK, *PLoS ONE*, 2011, **6**, e17695.
14 S. Liang, C. Zhang, S. Liu and Y. Zhou, Protein binding site predictin using an empirical scoring function, *Nucleic Acids Res.*
15 H. Neuvirth, R. Raz and G. Schreiber, ProMate: A Structure Based Prediction Program to Identify the Location of Protein–Protein Binding Sites, *J. Mol. Biol.*, 2004, **338**, 181–199.
16 A. Porollo and M. Jaroslaw, Prediction-based fingerprints of protein-protein interactions, *Proteins Struct. Funct. Bioinforma.*
17 I. Kufareva, L. Budagyan, E. Raush, M. Totrov and R. Abagyan, PIER: Protein interface recognition for structural proteomics, *Proteins Struct. Funct. Bioinforma.*, 2007, **67**, 400–417.
18 H. Chen and H.-X. Zhou, Prediction of interface residues in protein-protein complexes by a consensus neural network method: Test against NMR data, *Proteins Struct. Funct. Bioinforma.*
19 A. Vangone and A. M. Bonvin, Contacts-based prediction of binding affinity in protein–protein complexes, *eLife*, 2015, **4**, e07454.

20 L. C. Xue, J. P. Rodrigues, P. L. Kastritis, A. M. Bonvin and A. Vangone, PRODIGY: a web server for predicting the binding affinity of protein–protein complexes, *Bioinformatics*, 2016, btw514.
21 H. Ashkenazy, S. Abadi, E. Martz, O. Chay, I. Mayrose, T. Pupko and N. Ben-Tal, ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules, *Nucleic Acids Res.*, 2016, **44**, W344–W350.
22 M. Landau, I. Mayrose, Y. Rosenberg, F. Glaser, E. Martz, T. Pupko and N. Ben-Tal, ConSurf 2005: the projection of evolutionary conservation scores of residues on protein structures, *Nucleic Acids Res.*, 2005, **33**, W299–W302.
23 H. Li, K.-S. Leung, P. J. Ballester and M.-H. Wong, istar: A Web Platform for Large-Scale Protein-Ligand Docking, *PLoS ONE*, 2014, **9**, e85678.
24 H. Li, K.-S. Leung and M.-H. Wong, in *2012 IEEE Symposium on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB)*, IEEE, San Diego, CA, USA, 2012, pp. 77–84.
25 J. J. Irwin, T. Sterling, M. M. Mysinger, E. S. Bolstad and R. G. Coleman, ZINC: A Free Tool to Discover Chemistry for Biology, *J. Chem. Inf. Model.*, 2012, **52**, 1757–1768.
26 G. Neudert and G. Klebe, DSX: A knowledge-based scoring function for the assessment of protein–ligand complexes, *J. Chem. Inf. Model.*, 2011, **51**, 2731–2745.
27 O. Trott and A. J. Olson, AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading, *J. Comput. Chem.*, 2009, NA-NA.
28 W. Humphrey, S. Schulten and A. Dalke, VMD: Visual Molecular Dynamics, *J. Mol. Graph.*, 1996, **14**, 33–38.
29 D. A. Case, R. M. Betz, D.S. Cerutti, T. Cheatham, T. Darden, R. E. Duke, T.J. Giese, H. Gohlke, A. W. Götz, N. Homeyer, S. Izadi, P. A. Janowski, J. Kaus, A. Kovalenko, Tai-Sung Lee, S. LeGrand, P. Li, C. Lin, T. Luchko, R. Luo, B. Madej, D. Mermelstein, K. Merz, G. Monard, H. Nguyen, H. Nguyen, I. Omelyan, A. Onufriev, D. R. Roe, A. E. Roitberg, C. Sagui, C. Simmerling, W. M. Botello-Smith, J. M. Swails, R.C. Walker, J. Wang, R.M. Wolf, Xiongwu Wu, L. Xiao and P. A. Kollman, Amber 16, University of California, San Francisco., , DOI:10.13140/RG.2.2.27958.70729.
30 K. Shahrokh, A. Orendt, G. S. Yost and T. E. Cheatham, Quantum mechanically derived AMBER-compatible heme parameters for various states of the cytochrome P450 catalytic cycle, *J. Comput. Chem.*, 2012, **33**, 119–133.
31 D. R. Roe and T. E. Cheatham, PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data, *J. Chem. Theory Comput.*, 2013, **9**, 3084–3095.
32 B. R. Miller, T. D. McGee, J. M. Swails, N. Homeyer, H. Gohlke and A. E. Roitberg, MMPBSA.py: An Efficient Program for End-State Free Energy Calculations, *J. Chem. Theory Comput.*, 2012, **8**, 3314–3321.
33 E. F. Pettersen, T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng and T. E. Ferrin, UCSF Chimera- A visualization system for exploratory research and analysis, *J. Comput. Chem.*, 2004, **25**, 1605–1612.
34 C. C. Page, C. C. Moser, X. Chen and P. L. Dutton, Natural engineering principles of electron tunnelling in biological oxidation–reduction, *Nature*, 1999, **402**, 47–52.
35 A. W. Munro, D. G. Leys, K. J. McLean, K. R. Marshall, T. W. B. Ost, S. Daff, C. S. Miles, S. K. Chapman, D. A. Lysek, C. C. Moser, C. C. Page and P. L. Dutton, P450 BM3: the very model of a modern flavocytochrome, *Trends Biochem. Sci.*, 2002, **27**, 250–257.
36 A. Luthra, I. G. Denisov and S. G. Sligar, Spectroscopic features of cytochrome P450 reaction intermediates, *Arch. Biochem. Biophys.*, 2011, **507**, 26–35.
37 N. Kagawa, H. Hori, M. R. Waterman and S. Yoshioka, Characterization of stable human aromatase expressed in E. coli, *Steroids*, 2004, **69**, 235–243.
38 K. Stierand and M. Rarey, Drawing the PDB: Protein-ligand complexes in two dimensions, *ACS Med. Chem. Lett.*, 2010, **1**, 540–545.

39 J. Sgrignani, M. Bon, G. Colombo and A. Magistrato, 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.*, 2014, **54**, 2856–2868.
40 K. P. Conner, P. Vennam, C. M. Woods, M. D. Krzyaniak, M. K. Bowman and W. M. Atkins, 1,2,3-triazole–heme interactions in cytochrome P450: Functionally competent triazole–water–heme complexes, *Biochemistry*, 2012, **51**, 6441–6457.
41 D. Ghosh, C. Egbuta and J. Lo, Testosterone complex and non-steroidal ligands of human aromatase, *J. Steroid Biochem. Mol. Biol.*, 2018, **181**, 11–19.
42 D. Sirim, M. Widmann, F. Wagner and J. Pleiss, Prediction and analysis of the modular structure of cytochrome P450 monooxygenases, *BMC Struct. Biol.*, 2010, **10**, 34.
43 V. Polic and K. Auclair, Allosteric Activation of Cytochrome P450 3A4 via Progesterone Bioconjugation, *Bioconjug. Chem.*, 2017, **28**, 885–889.
44 C. M. Woods, C. Fernandez, K. L. Kunze and W. M. Atkins, Allosteric Activation of Cytochrome P450 3A4 by α-Naphthoflavone: Branch Point Regulation Revealed by Isotope Dilution Analysis, *Biochemistry*, 2011, **50**, 10041–10051.
45 Y. Satomi and H. Nishino, Inhibition of the enzyme activity of cytochrome P450 1A1, 1A2 and 3A4 by fucoxanthin, a marine carotenoid, *Oncol. Lett.*, 2013, **6**, 860–864.
46 K. W. Anderson, N. Mast, J. W. Hudgens, J. B. Lin, I. V. Turko and I. A. Pikuleva, Mapping of the Allosteric Site in Cholesterol Hydroxylase CYP46A1 for Efavirenz, a Drug That Stimulates Enzyme Activity, *J. Biol. Chem.*, 2016, **291**, 11876–11886.
47 I. G. Denisov, Y. V. Grinkova, P. Nandigrami, M. Shekhar, E. Tajkhorshid and S. G. Sligar, Allosteric Interactions in Human Cytochrome P450 CYP3A4: The Role of Phenylalanine 213, *Biochemistry*, 2019, **58**, 1411–1422.
48 H. Du, J. Li, Y. Cai, H. Zhang, G. Liu, Y. Tang and W. Li, Computational Investigation of Ligand Binding to the Peripheral Site in CYP3A4: Conformational Dynamics and Inhibitor Discovery, *J. Chem. Inf. Model.*, 2017, **57**, 616–626.
49 M. Redhair, J. C. Hackett, R. D. Pelletier and W. M. Atkins, Dynamics and Location of the Allosteric Midazolam Site in Cytochrome P4503A4 in Lipid Nanodiscs, *Biochemistry*, 2020, **59**, 766–779.
50 L. A. Peterson, Reactive Metabolites in the Biotransformation of Molecules Containing a Furan Ring, *Chem. Res. Toxicol.*, 2013, **26**, 6–25.
51 R. J. M. Brüggemann, J. C. Alffenaar, N. M. A. Blijlevens, E. M. Billaud, J. G. W. Kosterink, P. E. Verweij and D. M. Burger, Clinical Relevance of the Pharmacokinetic Interactions of Azole Antifungal Drugs with Other Coadministered Agents, *Clin. Infect. Dis.*, 2009, **48**, 1441–1458.